The Effect of the Prepaid Health Care Act on the Demand for Health Insurance, Demand for Medical Services and Labor Force Utilization in Hawai‘i

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI‘I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ECONOMICS

May 2005

By

Abdul Jabbar

Dissertation Committee:

Gerard Russo, Chairperson

James Moncur

PingSun Leung

Sang-Hyop Lee

Lawrence Nitz
ACKNOWLEDGMENTS
I would like to express my sincere gratitude and appreciation to all members of the committee for their advice, patience and support throughout my dissertation. I am deeply indebted to my dissertation committee chairman, Dr. Gerard Russo, for stimulating my interest in the topic, for his detailed comments and for continued financial and moral support. He always read the manuscript very carefully and gave me clear theoretical ideas and brightened the research path with meaning beyond what I could see. This dissertation has benefited greatly from his comments for which I am most grateful. I would like to thank Dr. Sang-Hyop Lee who always patiently looked through the econometric and empirical modeling sections, giving detailed comments freely and long lists of literature. My sincere thanks also go to Dr. Pingsun Leung who raised theoretical issues and gave intelligent comments throughout my dissertation period. Dr. James Moncur deserves special mention for his valuable comments and crucial role in this dissertation. I always felt indebted to Dr. Lawrence Nitz for his detailed and valuable comments. All the committee members, particularly the committee chair, generously contributed their time and energy to this dissertation.

Discussions with Dr. Thamana Lekprichakul and other fellows have been helpful during the course of writing this dissertation. I would like to thank Mr. David Petherbridge and Mr. Tim Rupright who voluntarily helped edit the English of the manuscript. I am forever grateful to them. Though this dissertation benefited from the contributions of all the committee members and the reviewers, I am solely responsible for any errors while expressing the hope that they will not be numerous nor too troublesome.
ABSTRACT

Hawai‘i established a unique health policy under the Prepaid Health Care Act (PHCA) of 1974 which requires, with some exceptions, private-sector employers to offer full-time workers (i.e., more than 20 hours per week) health insurance benefits. A number of concerns were raised that the law would increase coverage, reduce money wages, reduce the price responsiveness of demand for health insurance, increase the demand for health care and alter labor force utilization in Hawai‘i. The last three concerns are investigated in this study using three different data sets: the Current Population Survey (CPS), the National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEPS). Concerns about increases in coverage and reductions in money wages are not examined in this study, but are thoroughly investigated in other studies.

The first essay uses data from CPS to estimate the demand for health insurance by workers in Hawai‘i with a view to understanding how mandated coverage alters the price responsiveness of demand for employer-sponsored health insurance (ESI). Variations in marginal tax rates are used to identify the price-elasticity of demand for health insurance with respect to after-tax premiums in Hawai‘i, the United States overall and several states including California, Florida, Michigan and Nevada. The estimated price elasticity of demand for health insurance of full-time workers in Hawai‘i is significantly less than the estimate for the United States overall, -0.13 and -0.34 respectively. The implication is that PHCA has caused the demand for coverage to be more price inelastic than otherwise would have occurred. By comparison, the estimated price elasticities of demand for health insurance among full-time workers in the states of California, Florida, Michigan...
and Nevada are -0.44, -0.58, -0.39, and -0.55, respectively, implying an even greater disparity in price responsiveness between more and less regulated health insurance markets. The results, in part, explain why employer-sponsored insurance coverage in Hawai‘i has remained stable under the pressure of rising premiums while a comparable erosion of such coverage has occurred in the United States overall.

The second essay examines the effect of health insurance on the demand for physician visits in Hawai‘i using MEPS and NHIS data from 1996-2002. It also examines the effects of age and health status on the demand for health care. The results of the study show that the demand for physician visits is slightly higher in Hawai‘i (i.e., 3.51) than in the United States (i.e., 3.42) because of higher coverage and because Hawai‘i has a larger proportion of people 65 and older than the nationwide. Higher demand for health care in the future will put upward pressure on the costs of health care in Hawai‘i.

Health insurance is a very significant determinant of the demand for physician visits. People who are insured make more frequent visits to physicians than those who are uninsured because of lower out-of-pocket prices. The number of physician visits responds negatively to changes in the amount paid out-of-pocket. A reduction in the level of cost sharing increases the demand for physician visits. Three alternative models are used to estimate the pure price response: the Poisson model, the Negative Binomial model and the Multinomial logit model. These three models suggest that price elasticities for health care are in the -0.1 to -0.2 range which are slightly lower than the United States. These values are consistent with those in the lower range of the non-experimental literature,
which vary from -0.1 to -2.1. Price elasticity estimates reported by the RAND experimental study are -0.2 which are very close to the estimates found in this study. The magnitude of the price elasticity of physician visits has very important implications for the generosity of health insurance. If price elasticity is higher, then insurance coverage should be less generous. The results also show that increased family income positively affects physician visits. Depending on the insurance plan, higher income people have 15 to 30 percent more physician visits than low income people. The main reason for the rise in health care costs is not increased life expectancy but the invention of expensive medical technology, expensive surgical procedures and blockbuster pills.

The third essay examines the impact of PHCA on labor force utilization and labor market sorting in Hawai‘i using empirical methods. PHCA requires private-sector employers to provide health insurance to full-time workers. However, certain classes of workers are exempt from this regulation including part-time employees (i.e., less than 20 hours per week), self-employed workers, family workers, public-sector workers and employees working under a collective bargaining contract. I hypothesize that PHCA will cause employment to shift from the exempt class to the regulated class of workers and that among regulated employees, the utilization of labor will rise. Using four decades of data from CPS 1963-2004, I produce direct estimates (weighted tabulations) and model-based estimates (multinomial logit regression) of the distribution of the labor force by hours employed and across sectors. These estimates are produced for Hawai‘i before and after the implementation of the 1974 regulation. Using workers of the other 49 states as a control group, we implement a difference-in-difference approach to sweep out economy-
wide social and structural changes. The results indicate a modest shift in the labor force with more workers in the regulated class and greater utilization of labor among full-time workers than would have otherwise occurred.
Table of Contents

Abstract..................................................................................................................ii
Introduction...........................................................................................................1

Essay One

Employee’s Demand for Employer-Sponsored Health Insurance in Hawai‘i

Abstract...............................................................................................................6
1.1. Introduction......................................................................................................8
1.2. Literature Review............................................................................................18
1.3. Conceptual and Theoretical Framework.......................................................26
1.4. Tax Incentives and Worker’s Decision to Purchase Health Insurance..........33
1.5. Empirical Model Specification......................................................................34
1.6. Data and Methodology....................................................................................48
1.7. Estimates of the Demand for Health Insurance in Hawai‘i...........................52
1.8. Compassion of Results with Other Major Studies........................................69
1.9. Policy Implications.........................................................................................72
1.10. Conclusion.....................................................................................................73
1.11 Bibliography....................................................................................................77
1.12 Appendix........................................................................................................81

Essay Two

Demand for Physician Visits in Hawai‘i

Abstract..................................................................................................................91
2.1. Introduction.....................................................................................................93
2.2. Literature Review............................................................................................97
2.3. How Health Insurance Affects Demand for Health Care............................102
2.4. Conceptual Framework..................................................................................105
2.5. Determinants of the Demand for Physician Visits .................. 107
2.6. Model Specification and methodology ................................ 111
2.7. Data ............................................................................. 120
2.8. Empirical Results ............................................................ 122
2.9. Further Research ............................................................. 125
2.10. Conclusion ..................................................................... 126
2.11. Bibliography ................................................................... 128
2.12 Appendix ........................................................................ 133

Essay Three

The Effect of the Prepaid Health Care Act on Labor Force Utilization and Labor Sorting in Hawai‘i

Abstract ............................................................................... 139
3.1. Introduction ..................................................................... 140
3.2. Literature Review ........................................................... 142
3.3. Method and Methodology .................................................. 146
3.4 Data ................................................................................. 148
3.5. Empirical Results ............................................................. 151
3.6. Comparison with Other Studies ......................................... 163
3.7. Further Research ............................................................. 164
3.8. Conclusion ...................................................................... 165
3.9 Bibliography ..................................................................... 167
3.10. Appendix ....................................................................... 170
Introduction

Since the Prepaid Health Care Act of 1974 has been implemented, a number of concerns have been raised. These concerns include the propositions that it will alter the price responsiveness of demand for health insurance, the demand for physician visits, and cause labor sorting and labor force utilization in Hawai‘i. These concerns are investigated in this study using three data sets: the Current Population Survey (CPS), the National Health Interview Survey (NHIS) and the Medical Expenditure Panel Survey (MEPS). The estimated price-elasticity of demand for health insurance by full-time workers in Hawai‘i is significantly less than the estimate for the United States overall, -0.13 and -0.34 respectively. This implies the intuitive result that the Prepaid Health Care Act has caused the demand for coverage to be more price inelastic than otherwise would occur. The National Health Interview Survey and the Medical Expenditure Panel Survey data for 1996 to 2002 confirm that the demand for physician visits is relatively higher in Hawai‘i than in the United States. The results also indicate a modest shift in the labor force with more workers in the regulated class and greater utilization of labor among full-time workers than would have otherwise occurred.

Hawai‘i established a unique health policy under the Prepaid Health Care Act (PHCA) which was passed on September 2, 1974 and became effective on January 1, 1975. It was suspended by Supreme Court order on October 5, 1981. PHCA was ineffective for fifteen months until on January 14, 1983 when congress granted a special exemption from
ERISA that allowed PHCA to resume.\(^1\) PHCA was passed the same year as the federal Employee Retirement Income Security Act (ERISA) which superceded all state laws related to employee benefits.

PHCA requires all employers to provide health insurance benefits for full-time workers who work 20 hours or more a week for four consecutive weeks. The purpose of the act was to expand coverage and increase access to health care in Hawai‘i. The underlying rationale was that increased coverage will reduce the role of price as a barrier to a more equitable distribution of health care services. Therefore, it will affect health care utilization in Hawai‘i. It was also expected that this law would cause labor sorting from the exempt sector to the non-exempt sector and impact the number of hours worked. The effects of PHCA on these issues are examined in this study. PHCA has also increased the coverage rate and reduced money wage rates in Hawai‘i. These issues are not investigated in this study but are thoroughly examined in other studies.

The thesis is organized into three essays. Essay one investigates worker’s price elasticity of demand for employer-sponsored health insurance in Hawai‘i and finds it is significantly lower than in other selected states and the overall United States. The states were selected based upon the industrial structure and state income tax codes. Two states, Nevada and Florida, have a similar industrial structure to Hawai‘i regarding travel and service industries. The other two selected states, Michigan and California, have completely different industrial structures from Hawai‘i.

\(^1\) 29 U.S.C. 1144 §514 (b) 5(A).
Price and income elasticities are estimated for these five states: California, Florida, Hawai‘i, Michigan, Nevada, and the overall United States. Estimation of the demand for ESI also provides information about other factors that affect the demand. These factors can be altered by policy formation, particularly the price of health insurance. The effect of many health policies critically depends upon these elasticities. It has also analyzed the effect of different policy proposals on the demand, such as the effect of tax subsidies and public health programs. It has also examined whether tax subsidies are an effective tool to encourage workers to purchase coverage.

Public policy is also encouraging competition in the health care market to reduce and control the cost of health care. The increased role of the market in the health care sector has increased the importance of demand analysis. The demand for health care increases as society ages and as people have more and better coverage. The second essay estimates the demand for physician visits in Hawai‘i. It also examines the effect of health insurance, age and health status on the demand for physician visits. The Medical Expenditure Panel Survey (MEPS) and National Health Interview Survey (NHIS) data from 1996 to 2002 are used.

The results of the study show that the average per capita demand for physician visits is slightly higher in Hawai‘i (i.e., 3.51) than in the United States (i.e., 3.42) because of higher coverage and because Hawai‘i has the largest proportion of people 65 and older. Health insurance is a very significant determinant of the demand for physician visits.
Covered people make more frequent visits to physicians than uninsured people because of lower out-of-pocket costs. The number of physician visits responds negatively to changes in the amount paid out-of-pocket. A reduction in the level of cost sharing increases the demand for physician visits. Three alternative models are used to estimate the pure price response: the Poisson model, the Negative Binomial model, and the Multinomial logit model. These three models suggest that price elasticities for physician visits are in the -0.1 to -0.2 range. The RAND experimental study reported that price elasticity is -0.2. These values are consistent with those in the lower range of the non-experimental literature which varies from -0.1 to -2.1. The magnitude of the price elasticity of physician visits is affected by the generosity of health insurance. If price elasticity is higher, then insurance coverage should be less generous. The results also show that increased family income positively affects physician visits. Depending on the insurance plan, higher income people have 15 to 30 percent more physician visits than low income people. The main reason for the rise in health care cost is not increased life expectancy but the invention of expensive medical technology, expensive surgical procedures and blockbuster pills.

Many economists and public policy makers thought that the Prepaid Health Care Act would split labor force utilization and cause labor sorting from the exempt to the non-exempt sectors in Hawai‘i. The third essay examines the impact of PHCA on labor force utilization and labor market sorting in Hawai‘i using empirical methods. It is hypothesized that PHCA will cause employment to shift from the exempt class to the regulated class of workers and that among regulated employees the utilization of labor
will rise. Four decades of CPS data, 1963 to 2004, are used to calculate direct estimates (weighted tabulations) and model-based estimates (multinomial logit regressions). These estimates are produced for Hawai‘i before and after the implementation of the 1974 regulation. Using workers in the other 49 states as a control group, we implement a difference-in-difference approach to sweep out economy-wide social and structural changes. The results indicate a modest shift in the labor force with more workers in the regulated class and greater utilization of labor among full-time workers than would have otherwise occurred.

The maximum likelihood estimation method is used to estimate all regression results in this study. Program routines of this method are included in many statistical packages like SAS, SPSS and STATA. All the data analysis in this study is done by STATA SE 8.2. It is a popular general purpose statistical package developed by STATA Corporation and performs many standard econometric methods of data analysis. It also provides a powerful programming language that enables data analysts to write their own programs for complex models in a relatively simple way.

The estimates of this study provide up-to-date information about the demand for health insurance, physician visits, labor sorting and labor force utilization in Hawai‘i. This information will help policy makers better understand the attitudes of employers, improve the accuracy of their economic projections, and enable them to design more appropriate policies to reduce the cost of health care and the number of uninsured in Hawai‘i. A
reduction in the uninsured population as well as a reduction of the cost of health care will
provide more health care to more people and improve the quality of life in Hawai‘i.
Employee’s Demand for Employer-Sponsored Health Insurance in Hawai‘i

Abstract
This essay uses data from the Current Population Survey (CPS) to estimate the demand for health insurance by private workers in Hawai‘i with a view to understanding how mandated insurance coverage alters the price responsiveness of demand. Hawai‘i established a unique health policy under the Prepaid Health Care Act (PHCA) of 1974 which requires, with some exceptions, private-sector employers to offer full-time workers (i.e., more than 20 hours per week) health insurance and requires these employees to accept the employer-sponsored insurance (ESI). Because the other 49 states and the District of Columbia are explicitly prohibited from implementing an equivalent mandate under the federal Employee Retirement and Income Security Act (ERISA) of 1974, comparisons between Hawai‘i and the remainder of the United States are natural ways of assessing the impact of this Hawai‘i-specific regulation. Variations in marginal tax rates are used to identify the price-elasticity of demand for health insurance with respect to after-tax premiums in Hawai‘i, the United States overall, and several comparative states including California, Florida, Michigan and Nevada.

The estimated price-elasticity of demand for health insurance by private workers in Hawai‘i is significantly less than the estimate for the United States overall, -0.13 and -0.34 respectively, implying the intuitive result that PHCA has caused the demand for coverage to be more price inelastic than otherwise would have occurred. The estimated
price-elasticities of demand for health insurance among private workers in the comparative states of California, Florida, Michigan and Nevada are -0.44, -0.58, -0.39, and -0.55, respectively implying an even greater disparity in price responsiveness between more and less regulated health insurance markets. The results, in part, explain why in Hawai‘i, employer-sponsored insurance coverage has remained stable under the pressure of rising premiums while a comparable erosion of such coverage has occurred in the United States overall.
1.1. Introduction
The main purpose of this essay is to examine the effects of the Prepaid Health Care Act (PHCA) of 1974 on the price responsiveness of a private worker’s demand for ESI in Hawai‘i which is the only state in the United States that regulates its health insurance system through the PHCA. This regulation requires that an employer must provide health insurance benefits to full-time employees and pay for at least half of the premium. In the early 1990s, many other states also tried to follow the Hawaiian model of the PHCA. These states introduced reforms to encourage firms to offer health insurance to their workers. The effects of these reforms critically depend upon the price elasticity of demand for ESI. For example, if the demand is inelastic with respect to its price, then the tax subsidy will not have much effect on health insurance coverage. On the other hand, if the demand is elastic with respect to its price, then the tax subsidy will have a much larger effect on health insurance coverage. This essay also estimates the effect of other factors that determine the demand for ESI in Hawai‘i. These estimates provide valuable information about the factors affecting the demand for health insurance that may be of interest to planners and decision makers who formulate health policy for Hawai‘i.

Employer-Sponsored health Insurance (ESI) has been a principal source of health insurance in Hawai‘i and rest of the United States, since 1940. Presently, 70% of the non-elderly population of Hawai‘i is covered directly or indirectly by ESI which is significantly higher than the nationwide estimate of 64% (Employee Benefit Research Institute, 2000). The United States Census Bureau reported that 175.3 million workers and their dependents received health insurance through employers in the year 2000. In
1997, around 76% of the employed population purchased health insurance through their employers. During the same year, 33% of unemployed and 45% of those who were not in the labor force were covered by ESI (Current Population Reports, December 2002). Nationwide, in 1998 almost 90% of the non-elderly population who had private coverage were covered through their employer (Gruber, 2002).

Employer-Sponsored health Insurance is regulated under the PHCA in Hawai‘i. Policy-makers in Hawai‘i want to know the effect of this regulation on ESI in order to design more precise and accurate health coverage policies. In spite of the need, these state-level estimates were not available due to data limitations. However, sufficient data is now available to get reliable state-level estimates. This essay estimates the demand for ESI and examines the effect of the PHCA on the price responsiveness of ESI in Hawai‘i. It is found that the state-level estimates are significantly different from state to state. Therefore, using nationwide estimates to formulate state-level policy does not provide precise and accurate policy forecasts.

Workers buy health insurance through their employer because it is less costly than if they were to buy it privately and directly from service providers. Two main reasons account for this: the employer’s premium contribution and tax subsidies. Employers pay for a significant fraction of the total insurance premium which reduces the cost of insurance to employees by an equal amount. On average, employers pay around 75.6% of the overall health insurance premium, and workers contribute only 24.4% (EBRI, 2000). Employers pay the entire premium for 43% of their workers (Zedlewski, 1992). Phelps (1986)
reported that over 80% of the total premium is paid by employers. Gruber and McKnight (2002) reported that the share of the employer’s contribution is declining over time, but it is still over 80% of the total premium. The employer’s premium contribution is estimated to be 5.6% of private industry’s total compensation in 2001. This is the largest non-wage benefit in most workers’ compensation packages (Department of Labor, 2001).

Federal and State governments do not collect income or payroll taxes from employees on the employer’s contributions to the premium, but employers can still deduct that amount from their cost. This tax subsidy makes health insurance cheaper for the workers. The tax subsidy is a more effective incentive for high-wage workers while the employer’s contribution is more effective for low-wage workers. The government gives this tax incentive to encourage workers to purchase health insurance. Raising the number of people who are insured will likely increase their access to health care and improve the quality of their lives. These two benefits provide strong incentives to workers to purchase health insurance through their workplace.

ESI would probably still be provided even if the tax subsidy and employer’s contribution did not exist because the worker group has low administrative costs, low probability of adverse selection and more predictable health risks. If a group assembled itself for the purpose of buying health insurance, such a group would largely attract only sickly individuals who expect high medical costs (selection bias). The insurance company would have to pay higher-than-expected benefits to the individuals of that group. The insurance company might also have a high probability of adverse selection and a low
predictability of health risks for that group. Therefore, the health insurance policy for that group will be sold at a higher premium. Individual policies also cost more than group policies for the same level of coverage because of the filtering of high-risk patient costs and higher administrative costs. In the US, the loading fee for group health insurance is about 15% to 20% of the premium while for non-group health insurance it is about 80%. The huge difference in the loading fee substantially affects the premiums for health insurance.

A number of concerns have been raised about the tax treatment of the employer’s contributions for employee health insurance. The first and most important concern that has received the most attention lately is that it is one of the most costly federal subsidies to workers. The Federal government foregoes over $100 billion in revenue annually by excluding employer’s premium contributions from income and payroll taxes, (Sheils and Hogan, 1999). Second, the exclusion of the employer’s contributions from the employee’s federal and state income taxes provides a significant subsidy for the purchase of health insurance. This may encourage workers to purchase more insurance than they would if they were using taxable income to purchase it. This overinsurance leads to a lower out-of-pocket price of medical services which stimulates the utilization of health care services. It also creates upward pressure on the cost of health care (Feldstein and Friedman, 1977, page 176). Third, and more importantly, since most of the tax subsidy is given to workers only, it distorts the relative price of health insurance in the market that creates an efficiency loss in the health care market. Fourth, some empirical studies found that this subsidy has a very small impact on the number of uninsured.
These concerns have triggered extensive debate among health economists and policy makers to eliminate or cap the dollar value of health insurance benefits that can be excluded from taxation. In May 1993, the government of Quebec, Canada, eliminated the tax subsidy of health and dental benefits from an employee’s provincial taxable income. Some politicians are also recommending that the Government of Hawai'i consider this option to improve its budgetary position. Some economists favor this point of view because it will also solve the problem of “overinsurance” and reduce resource inefficiency because the consumers share a greater degree of cost to purchase health insurance. It will also reduce pressure on medical expenditures and ultimately reduce medical expenditure inflation because over insurance is often viewed as contributing to high and rising medical expenditures nationwide.

On the other hand, some health policy makers are arguing in favor of extending the amount of this tax subsidy. They argue that the rising cost of health care has made it impossible for many low-income workers to purchase health insurance. As a result, they argue the government should either provide comprehensive coverage to poor workers or at least substantially subsidize the purchase of health insurance. Therefore, they recommend that the government extend the tax subsidy further.

The effectiveness of this tax subsidy as a policy instrument depends critically upon reliable estimates of the price and income elasticities of the demand for health insurance. The price and income elasticities of ESI are of considerable practical importance to the
public health policy-makers because these are the basis for calculations of the effect of a
tax subsidy on employer-provided health insurance. There are no such estimates available
for Hawai‘i. Yet published estimates for the US as a whole vary widely. Estimates of the
price elasticity of demand for health insurance for the US range from -3 by Woodbury
and Hamermesh (1992) to -0.16 by Holmer (1984). This wide range of elasticity
estimates does not allow for quantification of the effect of a tax subsidy.

If the price elasticity is small, then the after-tax price has very little impact on the
probability of insurance coverage but has a large impact on federal and/or state revenues.
On the other hand, if the price elasticity is large, then it has a large impact on the
probability of insurance coverage but small impact on federal and/or state revenues.

There is no previous study that has estimated the price and income elasticities of demand
for employer-provided health insurance in Hawai‘i. The policy-makers have to rely on
national estimates for creating health policy in Hawai‘i. The national estimates vary
widely, so the recent debate on this subsidy has further increased the need for these
estimates by health policy makers to accurately formulate future health policy for
Hawai‘i.

Employer-Sponsored health Insurance is the most important type of health insurance in
Hawai‘i. There are two types of health insurance providers in Hawai‘i, public and private.
Public health insurance is provided by state and federal nonprofit organizations and is
mainly provided through two public health care programs, Medicare and Medicaid.
Medicare is a universal and mandatory health insurance program for the elderly
introduced in 1965. It is the largest government health insurance program. Medicaid is mainly for low-income people. There are several other government health insurance programs such as military health care systems, the Veterans Administration system, mental health programs, and maternal and child health programs.

The Current Population Survey (CPS) 2003 estimates that 32% of the population in Hawai‘i is covered by public health insurance programs. However, most of the population is covered by private health insurance. Commercial insurance companies provide private health insurance, most of which are profit-oriented. Private health insurance is purchased either through the workplace or directly from the insurer. Approximately 66% of the total population of Hawai‘i has purchased health insurance through employers. Only 8% of the total population has purchased health insurance directly from an insurer. About 15% of the total population of Hawai‘i has dual coverage. Approximately 9.6% of the Hawaiian population has no health insurance. Therefore, employer-sponsored health insurance is the primary source of coverage in Hawai‘i. Figure 1.1 summarizes the above-mentioned descriptive statistics of Hawai‘i’s population by insurance status. The total population of Hawai‘i is at the bottom and the insured population by employer-provided insurance is at the top.
During the last four decades, the cost of health care has increased much faster than per capita income growth and the general rate of inflation. Various government administrations have pursued a range of innovative health care cost-controlling efforts in the past, but did not succeed in any significant way. Presently, some health economists feel that health care costs are getting out of control since they are growing at a rate of 8% per year. That is much higher than the 2% growth rate in per capita income (Navarro, 1988). In 2003, the growth rate in health care costs slowed down from 8% to 6% per annum, but it is still more than three times the rate of overall price inflation. Currently, 15.2% of gross national product (GNP) is spent on health care as compared to 14% in 1994, 9.2% in the 1980s, 6% in the 1960s and 4% in the 1920s (DOL, various reports). Large employers and the federal government bear most of this rising cost of health care. Currently, public policy is encouraging competition in the health insurance market in order to stop the escalation of health care costs, which includes the cost of physicians’ services, hospitals, drugs, and the like.
The health insurance market is highly regulated in the United States. Therefore, market forces cannot work perfectly. However, the government cannot regulate every decision as it is physically impossible for the government to perfectly monitor employers and employees. As a result, market forces play a significant role regardless of the degree of government intervention. Demand for health insurance plays an important role in determining the price and quantity of health insurance. This chapter only analyzes the demand for employment-based health insurance because the price for public insurance is not available. Furthermore, non-group insurance covers only a small fraction of the population, so employment-based health insurance is the most important type of health insurance in Hawai'i. It is financed on a premium basis that is paid to the service provider by the employer. The employer contributes at least 50% of the worker’s premium and employees can only be required to contribute a maximum of 1.5% of their wage. For low-income workers, the employee’s contribution comes to a very small amount, so the employer pays the entire premium.

Beginning in the 1930s and 1940s, employers voluntarily provided health insurance for their workers and their dependents. In 1974, it became law in Hawai'i that all employers must offer health insurance benefits to their full-time workers. Part-time workers who usually work less than 20 hours a week, or earn less than $542 per month (86.67 multiplied by $6.25 minimum wage per hour) are exempt from the law. Government employees, self-employed workers, and employees covered under collective bargaining are also exempted. It was thought that the mandated employment-based health insurance would significantly increase the insured population in Hawai'i without having a
damaging impact on business. Employers offer voluntary health insurance in other states because their contribution to the premium is deductible from federal taxes and also improves the health of their workers. Healthier workers are more productive, absent less from their jobs and work more hours.

The remainder of the chapter has six sections which are organized as follows. The first section summarizes the previous literature on ESI. The second section presents a simple theoretical framework and derives the demand for health insurance that underlies the entire chapter. The next section specifies the empirical model, discusses the expected results and the estimation procedure. It also provides the definitions of variables used in the analysis. The fourth section contains the main empirical results of this chapter. In the next to last section of the chapter, the results of this study are compared with other related studies. In the last section, broader conclusions are drawn from the empirical results.
1.2. Literature Review
This section reviews historical economic literature that is related to the evolution of the estimation of demand for employer-provided health insurance. In 1738, Daniel Bernoulli was the first to propose a theory of the demand for health insurance (Sommer, 1954). He derived the demand for health insurance of a representative consumer by maximizing the expected utility function subject to his budget constraint. He assumed that consumers are risk averse, implying that the marginal utility of income or wealth is diminishing. This behavior can be represented by a concave utility function in income or wealth. He assumed a specific concave utility function to derive the demand for health insurance. Bernoulli proved that a voluntary purchase of health insurance would make the consumer better off. He showed that the utility level achieved, after paying the insurance premium, is higher than the expected utility level from remaining uninsured. Since no one was able to satisfactorily measure utility for more than two centuries, not much progress was made in this field during that time.

In 1944, von Neumann-Morgenstern (vNM) developed a practical method for measuring utility as a function of income or wealth. At that time it became possible (with this utility function) to look at the various shapes of utility functions and predict how individuals with differently shaped utility function will respond to the opportunity to purchase health insurance. This invention boosted interest in the theory of the demand for health insurance.
Four years later, Friedman and Savage (1948) refined the theory of the demand for health insurance by introducing the gamble theoretic format to the insurance problem. They proved that risk averse individuals prefer to pay a small certain loss (the insurance premium) to reduce a large, but actuarially equivalent, uncertain future financial loss if they get sick. Obviously, people differ in the amount they are willing to reduce or eliminate financial risk. Thus, a worker’s attitude toward risk strongly affects the demand for health insurance. By purchasing insurance, the consumer is choosing certainty over uncertainty.

Arrow (1963) agreed with Friedman and Savage and proved that the risk-averse consumer would be better off buying health insurance if an actuarially fair premium is offered to the consumer who maximizes his expected utility subject to his budget constraint. Later on, Mossin (1968), Smith (1968), Pratt (1964) and many others also studied the theory of demand for health insurance. They all came to the same conclusion: that risk-averse individuals demand health insurance to reduce uncertainty. Consumer preferences regarding risk play an important role in the demand for health insurance. All of these studies show that the premium, an individual’s income, and institutional features play an important role in the determination of the demand for health insurance.

Nyman (2003) introduced a new theory of demand for health insurance. He claimed that the consumer’s demand for health insurance is for an income transfer if he gets ill, not for certainty as claimed by the conventional theory of the demand for health insurance. He used a Bernoulli-type utility function which is diminishing in a certain income. While the
conventional theory of demand for insurance is derived from vNM type utility function which combines consumer preferences for income and risk in one utility function. He claimed that the decision to purchase health insurance depended upon the shape of the consumer's Bernoulli utility function for a certain income. The consumer's preferences regarding risk were irrelevant in the determination of the demand for health insurance. He claimed that the voluntary purchase of health insurance makes the consumer better off. The question then arises, "Why do so many people not buy health insurance?"

The central idea of the conventional theory is that people demand health insurance in order to avoid the risk of large financial losses. There is always a risk of illness or injury which causes a high level of spending on medical services. To convert the possibilities of large expenses into smaller and certain payments, risk-averse consumers seek and obtain insurance that covers some of the costs of medical care. Their motivation to purchase health insurance is to gain access to health care services that would otherwise be unaffordable. On the other hand, the new theory claims that individuals demand health insurance for income transfer. The new theory of health insurance also claims that the consumers' preferences regarding risk are irrelevant in determining the demand for health insurance.

Pauly (1968) studied the theory of demand for health insurance with special emphasis on the moral hazard problem. He argued that health insurance lowers the price of health care to consumers. The low price of health care provides incentive for the consumer to use more health care than if it were at full cost. The health insurance does not affect the cost
of health care to society. However, this inefficiency due to moral hazard will lead to a welfare loss to the society.

Theories of the demand for insurance emerged in economic literature in the late 1960's. Since then a number of empirical studies have been conducted to estimate the demand for health insurance using different data sets. The first study to estimate the demand for health insurance was done by Fuchs and Kramer (1972). They estimated the price and income elasticities of demand for health insurance. Their estimates showed that the income effect is positive. Later on, Feldstein (1973) estimated the demand for health insurance and examined the effect of tax subsidy on the demand for health insurance. He concluded that as the price of health care increases the demand for health insurance should also increase. He used the fact that demand for health care is inelastic so that any increase in the price of health care services would also increase the total spending for these services. It will increase the risk of large financial losses which in turn will cause an increase in the demand for health insurance. He also concluded that as marginal tax rates decrease, the demand for health insurance will decrease.

Phelps (1973, 1976) comprehensively studied the demand for health insurance. He derived the demand for health insurance for a risk averse person by maximizing a von Neumann-Morgenstern-type utility function subject to his budget constraint. He also estimated the demand for health insurance using the 1963 survey data on families’ health insurance coverage. He estimated the price elasticity of demand to be equal to -0.67 and income elasticity to be in the range between +0.12 and +0.19. Since there was no data on
the price of health insurance or exogenous sources of variation in insurance prices such as loading fees or the families' marginal tax rates, he used the employment-group size as a health insurance price proxy. He assumed that administrative costs decline as the number of enrollees increase. Even though a strong relationship is found in the empirical work between the loading fee and the employment-group size, it is still not a proper measure for the price. Presently, workers' marginal tax rates are available in the CPS data and more accurate estimates of the demand for health insurance can be produced. This study has also calculated more accurate estimates of price and income elasticities of demand for health insurance. Another objection of Phelps's study is that he did not look at the effect of public programs like Medicare and Medicaid on the demand for health insurance since he used data from before these programs existed. This study has also remedied that problem.

Feldstein and Friedman (1977) conducted another comprehensive experimental study for the demand for health insurance. They used a theoretical-choice simulation model for the estimation of the aggregate demand for health insurance. This demand is measured by the average effective coinsurance rate instead of the aggregate premiums. Presently, this aggregate premium data is available, so this study will use actual data to estimate the demand for health insurance.

Gruber (1994) analyzed the effect of mandatory benefits in employer-provided health insurance on coverage and the labor market. Different state regulations mandate that group health insurance plans must include certain benefits. These mandated benefits raise
the cost to employers for providing health insurance to its workers. Gruber used data for employees in small firms from May CPS supplements for the years 1979, 1983 and 1988. Gruber found two important results. First, the state-mandated health insurance benefits are not an important cause of the low rate of health insurance coverage among small-firm workers. Second, a significant amount of the cost of these mandated employer-provided health insurance benefits are shifted to the targeted group in the form of a reduction in their wages.

Woodbury and Hamermesh (1992) estimated the demand for health insurance using panel data for 1477 institutions of higher learning in the US. They found that the demand for health insurance in these institutions is very responsive to the changes in real income and the after-tax price of the insurance. They used three alternative definitions of the after-tax price of health insurance. The first definition assumes that workers receive an actuarially fair return on the payroll taxes, \( P = (1 - \tau_f - \tau_s) \). The second definition assumes that the payroll tax is fully borne by employers. \( P = (1 - \tau_f - \tau_s - \tau_{ss} - \tau_{me}) \). The third definition assumes that the worker bears the entire burden of both payroll taxes.

\[
P = \left[ \frac{1 - \tau_f - \tau_s - \tau_{ss} - \tau_{mc}}{1 + \tau_{ss} + \tau_{mc}} \right].
\]

They concluded that the elasticities are robust with respect to the varying definitions of the after-tax price.

Gruber and Krueger (1992) found that higher insurance costs have a negative but insignificant effect on employment. They also suggested that a “substantial portion of the cost to employers is shifted to employees in the form of lower wages”. Krueger (1993)
said that the long-run cost of mandated health insurance is likely to be shifted to labor in the form of lower wages with a smaller reduction in the number of jobs. His analysis was based on the consensus values of elasticities of labor supply and demand. Gruber and Hanratty (1993) found opposite results than the theoretical prediction from Canadian data. Employment and wages both increased with the National Health Insurance (NHI) in Canada.

Gruber and Poterba (1996) estimated the self-employed workers' demand for health insurance. They developed a new framework for measuring the net tax subsidy to purchase health insurance. They also introduced a detailed tax code in their after-tax price of health insurance. They combined data from the TAXSIM program with survey data from the National Medical Expenditures Survey for estimation and analysis. They estimated after-tax price elasticity of health insurance to be about -1.8 which is much larger than previous studies. Therefore, the estimated effects of tax subsidies by previous studies are heavily underestimated. They proposed that the effect of ending the exclusion of employer-provided health insurance on the demand for health insurance would be much larger than predicted by previous studies. They have also shown a concern that this tax subsidy will lead to overinsurance. This overinsurance might have contributed to the rise in health care cost. They suggested that ending this tax subsidy would reduce the overinsurance problem and the cost of health care.
Estimates of price and income elasticities of the demand for health insurance in the United States are summarized in table 1.1.

**Table 1.1:*
**Price and Income elasticities of demand for health insurance in the United States**

<table>
<thead>
<tr>
<th>Study</th>
<th>Elasticity Definition</th>
<th>Source of price variation</th>
<th>Price elasticity</th>
<th>Income elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phelps (1973)</td>
<td>Share of Premium expenditure</td>
<td>Employment-group size</td>
<td>-0.67%</td>
<td>0.15%</td>
</tr>
<tr>
<td>Taylor and Wilensky (1983)</td>
<td>Premium expenditure</td>
<td>Marginal tax rates</td>
<td>-0.21%</td>
<td>0.02%</td>
</tr>
<tr>
<td>Woodbury (1983)</td>
<td>Premium expenditure</td>
<td>Marginal tax rates</td>
<td>-1.7% to -3.5%</td>
<td></td>
</tr>
<tr>
<td>Holmer (1984)</td>
<td>Premium expenditure</td>
<td>Marginal tax rates</td>
<td>-0.16%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Farley and Wilensky (1985)</td>
<td>Premium expenditure</td>
<td>Marginal tax rates</td>
<td>-0.41%</td>
<td>0.01%</td>
</tr>
<tr>
<td>Marquis and Phelps (1987)</td>
<td>Premium expenditure</td>
<td></td>
<td>-0.6%</td>
<td></td>
</tr>
<tr>
<td>Manning and Marquis (1989)</td>
<td>Insurance enrolment</td>
<td>Premiums</td>
<td>-0.54%</td>
<td>0.07%</td>
</tr>
<tr>
<td>Short and Taylor (1989)</td>
<td>Insurance enrolment</td>
<td>Employee out-of-pocket premium</td>
<td>-0.32%</td>
<td>0.13%</td>
</tr>
<tr>
<td>Feldman et al. (1989)</td>
<td>Insurance enrolment</td>
<td>Employee out-of-pocket premium</td>
<td>-0.31%</td>
<td>-----</td>
</tr>
<tr>
<td>Woodbury and Hamermesh (1992)</td>
<td>Share of premium expenditures</td>
<td>Marginal tax rates</td>
<td>-2 % to -3%</td>
<td></td>
</tr>
<tr>
<td>Gruber and Poterba (1994) (Self-employed)</td>
<td>After-tax price</td>
<td>Marginal tax rates</td>
<td>-1.8 %</td>
<td></td>
</tr>
</tbody>
</table>

In summary, previous studies of the U.S. demand for health insurance report a wide range of price elasticities, from -0.16 to -3, which is much too wide to provide reliable guidance in assessing the effectiveness of a tax subsidy policy in Hawaiʻi. Therefore, there is a need for reliable estimates for Hawaii to formulate more appropriate tax subsidy policies.
1.3. Conceptual and Theoretical Framework
This section describes conceptual and theoretical background of the empirical
specification. Demand for ESI is different from the demand for most goods and services.
The ESI is not purchased as a final consumption good but as a means of paying for the
future stochastic purchases of health services. As usual, the demand for ESI represents
the amount of insurance coverage that a worker is willing to buy at different prices
(premiums) for health insurance, with their given level of income, tastes, prices of other
goods and services, and institutional features. Additional insurance coverage will be
purchased if the insurance premium declines and vice versa. The marginal benefit of
additional coverage declines for the worker as he purchases more comprehensive
coverage. Full coverage of all medical expenses would be demanded only when the
loading fee is zero or policies are sold at actuarially fair premium.

It is a challenging problem to measure the price of health insurance. The most commonly
used measure for its price is an average premium. This includes an actuarially fair
premium and the loading fee. The fair premium is equal to the expected benefits the
individual gets from the health insurance. The loading fee includes claims processing,
administrative and other costs, and normal profits of the insurer for risk-bearing. Some
health economists disagree with the previous definition and consider that only the loading
fee is the proper price of health insurance because an insurance policy returns money to
the insured person as a benefit of the policy. So the price of insurance is the amount that
exceeds the expected benefits of the policy, which is the loading fee. If the premium is
just equal to the expected benefits then the insurance itself is free. Insurance companies
that work for profit and bear risks cannot offer free insurance at equilibrium. It must
charge the consumer an insurance premium that at least covers the expected benefits it will pay out plus a loading fee. Therefore, a positive loading fee will be charged in equilibrium. The average loading fee in the US for group health insurance is about 15% to 20%, while for non-group policies, it is about 80%. In general, the higher the price of the health insurance, the lower is the demand for health insurance and vice versa. Very few studies were able to use this variable because of the non-availability of data.

There are various models that explain the behavior of individual workers in purchasing health insurance. In most of these models, workers are expected utility maximizers. The non-purchasing choice can result in consuming either public health care or private health care with full cost paid out-of-pocket. The expected utility models predict the demand for health insurance very well. A general model is explained in the following section and a more specific model is then defined. The demand for health insurance is derived from maximizing the expected utility function of a representative worker subject to his budget constraint. The worker follows a von Neumann-Morgenstern-type concave expected utility function $EU(Y)$ that is increasing in income and exhibits diminishing marginal utility from income, i.e. $\frac{\partial EU(Y)}{\partial Y} > 0$ and $\frac{\partial^2 EU(Y)}{\partial Y^2} < 0$. The diminishing expected marginal utility property implies that the worker is risk averse. In other words, workers prefer to have any level of income with certainty over a gamble providing the same level of income on average. Therefore, the worker is willing to pay a premium for health insurance policies that will exceed the actuarially fair premium. The gap between the premium that the worker is willing to pay and the actuarially fair premium is called a risk
premium. The risk premium allows insurers to cover expenses above the medical payout including a claims-processing fee, other administrative expenses, and profit.

For simplicity, it is assumed that the demand for medical care is completely price inelastic. Therefore there is no moral hazard problem. A worker without health insurance has to spend $M$ amount of money on medical care if he becomes ill. It is assumed that medical treatment will completely eliminate the non-financial loss of illness in the form of health deterioration. An uninsured worker with original income, $Y^0$, spent $M$ amount on medical care and has disposable income, $Y = Y^0 - M$, left to spend on other goods and services. For simplicity, the price of medical care is normalized to one, so $M$ represents the number of units of health care consumed. It is also assumed that the payoff, $I$, is equal to $M$, the exogenously determined loss. If the exogenously determined probability of becoming ill is $\pi$ then actuarially fair premium, $R = \pi I$.

The worker’s decision to purchase health insurance for an unexpected illness depends on the expected utility with or without health insurance. A worker will purchase health insurance if his expected utility with health insurance is higher than the expected utility without health insurance. If a worker chooses to remain uninsured then his expected utility is given in the following equation (1.1):

$$EU_u = \pi U(Y^0 - M) + (1 - \pi)U(Y^0) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots
$$

He will face a small possibility, $\pi$, of a large loss, $M$, in the event that the illness occurs or the large possibility, $(1 - \pi)$, that the illness will not occur. On the other hand if the worker chooses to buy insurance, $I = M$, then he has to pay a premium, $R$. His disposable
income becomes \( Y = Y^0 - R - M + I \). The expected utility of the insured worker is given in the following equation (1.2):

\[
EU_i = \pi U(Y^0 - \pi M - M + I) + (1- \pi)U(Y^0 - \pi M) \ldots (1.2)
\]

Using the assumption that the payoff, \( I \), is equal to \( M \), the exogenously determined loss.

\[
EU_i = \pi U(Y^0 - \pi M) + (1- \pi)U(Y^0 - \pi M) \ldots (1.3)
\]

Simplifying the above equation (1.3) produces the following equation (1.4).

\[
EU_i = U(Y^0 - \pi M) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1.4)
\]

The insured worker faces a small loss in the form of the insurance premium, \( \pi M \). The worker chooses between being uninsured and having an uncertain outcome with an expected utility given by equation (1.1) and being insured and having a certain outcome with certain utility given by equation (1.4). If his utility after paying the premium is higher than the expected utility without insurance, he will purchase insurance. If the worker chooses to become insured, he prefers a certain level of utility over an uncertain level. Thus the demand for insurance can be interpreted as a demand for certainty.

The decision to purchase health insurance can also be explained by the utility curve, depicted in figure 1.2, used by Friedman and Savage (1948). Suppose the consumer is initially endowed with \( Y^0 \) income. The point \((Y^0-M)\) denotes the income left for other goods and services after becoming ill and spending \( M \) amount on medical care. The utility level enjoyed by the consumer if he remains healthy is \( U(Y^0) \) and if he become ill \( U(Y^0-M) \). Without insurance, expected income is \( Y_u = \pi(Y^0-M) + (1- \pi)(Y^0)=Y^0-\pi M \), and expected utility is \( U_u = \pi U(Y^0-M) + (1- \pi)U(Y^0) \). The expected utility \((Y_u, U_u)\) lies on the straight line between the points \([Y^0, U(Y^0)]\) and \([(Y^0-M), U(Y^0-M)]\). The straight line
represents expected utility of the uninsured worker with a constant loss, \( M \), but different probabilities of illness occurrence, \( 0 < \pi < 1 \). As the probability that the loss, \( M \), will occur increases, the expected utility value moves down to the left on the straight line, closer to the point represented by \( U(Y^0 - M) \) on the curve. With insurance, the income and utility both are certain and represented by \( Y_i = [Y^0 - \pi M] \) and \( U_i = U(Y^0 - \pi M) \), respectively.

The expected income with or without insurance is the same, \( Y_u = [Y^0 - \pi M] = Y_i \), but the utility of being insured is higher than the utility of being uninsured, \( U_i > U_u \). This utility gained by avoiding risk is the key to the demand for health insurance. As is clear from the figure, the individual prefers to purchase health insurance relative to self-insuring. Therefore, the health insurance will be purchased. The amount of utility gained from insurance is equal to the vertical distance between \( U_i \) and \( U_u \). The greater the concavity of the utility curve, the greater the risk aversion and the greater the utility gain from insurance. This gain can be measured in monetary terms, representing a willingness to pay for health insurance. Health insurance is never sold at its actuarially fair value because there is a loading fee. If the consumer is charged the actuarially fair premium plus an additional very small amount, \( \varepsilon \), as a loading fee, the consumer would still be willing to purchase insurance because

\[
U(Y^0 - \pi M - \varepsilon) > \pi U(Y^0 - M) + (1 - \pi)U(Y^0) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1.5).
\]

\( U_i \) (the left-hand side of inequality (1.5)) is still greater than \( U_u \) (the right side of the inequality (1.5)). If \( \varepsilon = E \) the premium would be such that the utility with insurance, \( U_i \), would just be equal to the expected utility without insurance \( U_u \). Therefore, \( E \) is the
largest amount that the consumer would be willing to pay in excess of the actuarially fair premium. In insurance literature, E is called the risk premium. It is a horizontal distance between the expected utility line and the utility curve. The risk premium, E, and the actuarially fair premium, πM, change as the probability of the event’s occurring changes.

![Figure 1.2](image)

The insurer charges a premium that covers the expected spending on medical care (expected payoff) and the loading fee. As long as the insurer sets a premium that is less than the sum of the fair premium (expected payoff) and the risk premium, the insurance will be purchased by the risk-averse consumer.

The relationship between the price of insurance and the quantity demanded can be explained by figure 1.3. The price of insurance is measured on the vertical axis and the probability of an event’s occurrence, π, on the horizontal axis. The curved line starts at “0” probability of the event’s occurring and ends at a probability of “1”. The area under
the curve is taken from figure 1.2. It shows the distance between the actual utility curve and the expected utility line. It also shows the amount above the fair premium the worker is willing to pay for the insurance. The price of insurance is represented by line AA. It is increasing because as the probability of an event occurring increases the cost of processing claims increases. The worker will purchase insurance for the events that fall between $\pi_1$ and $\pi_2$. Between these two points, the price a worker has to pay for insurance is less than his expected benefits. Therefore, the insurance will be purchased for those events. If the price of the insurance increases, the price line shifts upward and fewer events will be insured. If price rises to line BB, the worker will not buy insurance for any event and prefers to remain uninsured. Thus, as the price of insurance rises, the worker will be less likely to get insurance for certain events. This inverse relationship between the price of insurance and the quantity of insurance demanded is called the demand schedule for insurance.

If the magnitude of the loss, $M$, increases, the willingness to pay curve will also shift upward as shown by the dotted curve. This will cause a shift in the demand for health
insurance curve. Less than 100% insurance will be purchased unless it is offered at the actuarially fair premium. There are two situations, the area below $\pi_1$ and above $\pi_2$, in which the price of insurance will exceed the amount that the worker is willing to pay. These workers will be better off not purchasing the insurance. If these workers are required by law, as in Hawai‘i, to purchase the insurance at that price it will make them worse off. This explains why we observe some workers purchasing insurance while others do not in a free market system.

1.4. Tax Incentives and Workers’ decision to Purchase Health Insurance
This section investigates how favorable tax treatment affects a worker’s discrete choice of whether or not to buy health insurance. The tax subsidy to workers for the purchase of health insurance lowers their price which provides them an incentive to purchase more. Without the tax subsidy, the price of health insurance in terms of foregone income would be one. Since the employer’s contribution to health insurance is exempted from a worker’s income and payroll taxes, this reduces the price of health insurance to less than one. This tax exemption affects the after-tax price of ESI in several ways. First, the employer’s contribution is excluded from personal federal and state taxable income, and from payroll taxes. There are two payroll taxes: the Medicare hospital insurance federal payroll tax of 2.9% and the Old Age Security and Disability Insurance (OASDI) federal payroll tax of 12.4%. Both of these payroll taxes are split 50-50 between employer and employee. Second, most of the employees who receive employer-provided health insurance pay their part of the premium from their pre-tax income. Third, taxpayers who file itemized deductions are allowed to deduct expenditures on medical care and directly purchased health insurance in excess of 7.5% of adjusted gross income (AGI) from
federal taxable income. Therefore, the after-tax cost for the non-itemizer is $1 on medical care and directly purchased health insurance. However, for the itemizer, the after-tax cost of such spending is \((1-\alpha \tau_f)\). Where, \(\tau_f\) is the federal marginal tax rate and \(\alpha = 1\) if the marginal dollar of spending exceeds the AGI floor (7.5%) and zero otherwise. Medical expenses are not deductible from state taxable income. A very small proportion of workers file itemized deductions. All three of these tax subsidies are incorporated in the empirical model specified in the following section.

The Tax Reform Act of 1986 made the most sweeping changes in the United States tax code since 1941. The objective of these reforms was to design a tax system which would be simple, fair, revenue neutral and reduce the tax burden on individuals. The individual rate schedule, which had 14 marginal tax rates ranging from 11% to 50% were reduced to five rates, 15%, 28%, 31%, 36% and 39% for different income brackets. The corporate rate was reduced from 46% to 34%. In order to keep these reforms revenue neutral, the government eliminated investment tax credits, slightly decelerated depreciation allowances, etc. The 1986 reforms transfer roughly US$24 billion per year in tax burden from households to corporations. The tax-free status of employer-provided health insurance and many other fringe benefits provided by employers were not touched in these reforms. Therefore, these reforms are irrelevant to the tax-free status of employer-provided health insurance.

1.5. Empirical Model Specification
This section describes the empirical specification of the demand for employer-provided health insurance. Empirical studies of the demand for employer-provided health
insurance have used various model specifications. The empirical model specification is very important for policy recommendations. If the model specification does not correspond to the underlying behavioral structure that drives the demand for health insurance, policy recommendations based on the estimation may not have the desired effects. Some previous specifications are briefly summarized in the following section.

The most important and widely defined variable in these studies is the price of health insurance. Different empirical studies have used different proxies for the price of health insurance due to the non-availability of proper data. Phelps (1973, 1976), the first comprehensive study in this area, used the employment-group size as a proxy for the price of health insurance. He argued that the group policies are sold at substantially lower prices than non-group policies since larger employers can negotiate a lower premium due to lower administrative costs and greater risk pooling. Even though a strong negative relationship between premiums and employment-group size had been observed in the family survey of 1963, a proper variable for the price of health insurance remains to be found. Leibowitz and Chernew (1992) and Marquis and Long (1993) used regional differences as a proxy for the price variation of health insurance. These differences may also explain some variation in the demand for health insurance, but it is still not a good exogenous source of variation in insurance price.

Holmer (1984) used the relative price of insurance as a ratio of the premium to the price of other goods. For simplicity, he normalized the price of other goods to be equal to one. If the employer pays the full premium which is tax deductible, the after-tax price of
health insurance could be written as one minus the employee’s marginal tax rate. On the other hand, if the employee has to pay the entire premium, which does not qualify for any tax subsidy, the price of the health insurance is one. In reality, both employer and employee contribute to the premiums of employer-provided health insurance. Therefore, the after-tax expenditures on health insurance should be a weighted average of the two given in the following equation:

\[ p \cdot h = (1-t)e + (1)(h-e) \quad \ldots \ldots \ldots \quad (1.6) \]

where \( p \) denotes the after-tax price of employer-provided health insurance; \( h \), the total premium; \( e \), the employer’s contribution; and \( t \), the buyer’s marginal tax rate\(^1\). Equation (1.6) can be rearranged and solved for the after-tax price of health insurance, \( p \), which is given in Equation (1.7)

\[ p = 1-te/h \quad \ldots \ldots \ldots \quad (1.7) \]

Taylor and Wilensky (1983) assumed that the ratio of the employer’s contribution to the total premium is constant across all insurance plans and employers. This implies that the price of health insurance varies inversely with the marginal tax rate. Therefore, they used \( (1-t) \) as a regressor in their regression model. Many other empirical studies, such as Woodbury (1983) and Sloan and Adamache (1986), have also used the marginal tax rate as a regressor. Some other empirical studies have used the natural log of \( (1-t) \) as the price of health insurance in their regression for the demand for health insurance. Farley and

\(^1\) An average \( t \approx t_{f} + t_{s} + t_{e} + \frac{t}{2} \)

where, \( t_{f} \) denotes federal marginal income tax on earned income, \( t_{s} \) the state income tax rate, \( t_{e} \) employer and employee payroll tax rate.
Wilensky (1983) used \((1-p)\) rather than \((1-t)\) as the price variable in their regression for the demand for health insurance\(^2\).

Gruber and Poterba (1994) criticized the formulation of the after-tax price of health insurance used in previous studies. They constructed a new, more appropriate formula to calculate the after-tax price of health insurance which overcomes the shortcomings of previous formulas. A closer examination of the Gruber and Poterba (1994) formula indicates that it is very close to Holmer (1984) in spirit. They argued that federal and state taxes substantially subsidize the employer-provided health insurance. They also introduced a loading fee and a detailed tax code in the after-tax price of health insurance. For simplicity, they assumed perfect competition in the labor market which implies that the employer’s contribution to the premium ultimately comes out of what would otherwise have been money wages for labor. They also assumed that the employer’s contribution is excluded from taxation while the employee’s contribution is only partially excluded with IRS section 125 plans. As a result of these tax rules, the after-tax price of employer-provided health insurance to the employee is given by the following equation:

\[
P_{HI} = \left[ \frac{1 - \tau_f - \tau_s - \tau_{ss}}{1 + \tau_{ss}} \right] \left( \frac{E + \delta \cdot G}{E + G} \right) + \left( 1 - \delta \right) \cdot G (1 + \lambda) \ldots \ldots \ldots (1.8)
\]

Where, \(p\) denotes the after-tax price of employer-provided health insurance; \(\tau_f\), the federal marginal income tax rate on earned income; \(\tau_s\), a net-of-federal-tax state income tax rate; \(\tau_{ss}\), federal payroll tax rate for the OASDI program; \(E\), the employer’s contribution; \(G\),

\(^2\) In other words, they used \(te/h\) as a price of health insurance.
the employee’s contribution; \( \delta \), the fraction of the employee’s premium that can be paid for on a pre-tax basis (0.20); \( \lambda \), the loading fee\(^3\).

The US tax code has another federal payroll tax of 2.9% called the Medicare Health Insurance Tax. This tax is also split 50-50 between the employer and the employee. The contribution by each is denoted by, \( \tau_{mc} \) which comes to a 1.45% tax levied on the employee’s wage income. This was either merged with the social security tax or ignored by Gruber and Poterba (1996). This study will modify the formula for the after-tax price of health insurance by explicitly including this federal payroll tax rate in the calculation. The modified formula for the after-tax price of health insurance is given in the following equation (1.9):

\[
P_{HI} = \left( \frac{1 - \tau_f - \tau_s - \tau_{ss} - \tau_{mc}}{1 + \tau_{ss} + \tau_{mc}} \right) \times \left( \frac{E + \delta * G}{E + G} \right) \times \left( \frac{(1 - \delta) * G}{E + G} \right) \times (1 + \lambda) \quad (1.9)
\]

Where, \( \tau_{mc} \) is the federal payroll tax rate for the Medicare HI program and the rest of the symbols are as defined in the equation (1.8).

Absolute money prices, as such, do not determine the demand for health insurance. Rather, it is the relative prices that determine this demand. They defined the after-tax relative price of employer-provided insurance as a ratio of the after-tax cost of employer-provided insurance to the after-tax cost of self-insurance given in equation (1.10).

Individuals are allowed to deduct total medical expenditures from taxable income if these expenditures exceed 7.5% of their AGI.

---

\(^3\) Includes claims processing, marketing, overhead and administrative costs, and normal profits of insurance company, etc. In other words, it is a charge above expected insurance benefits.
\[
P_{relative} = \left( \frac{1 - \tau_f - \tau_s - \tau_{ss} - \tau_{mc}}{1 + \tau_{ss} + \tau_{mc}} \right) \left( \frac{E + \delta \cdot G}{E + G} \right) + \frac{(1 - \delta) \cdot G}{E + G} \right) \cdot (1 + \lambda) \]

Where, \( \alpha = 1 \) if the total medical care expenditures exceed the 7.5% of AGI and zero otherwise. The after-tax cost of the marginal dollar spent on medical care is \((1 - \alpha \tau_c)\) if the individual itemizes the tax return. For a non-itemizer, this deduction is irrelevant because the after-tax cost of such spending is $1. Keeping other things constant, as the after-tax price of health insurance increases, the demand for health insurance will decrease and vice versa.

Several variations of the formula for after-tax prices are used by different studies depending on the availability of data. Farley and Wilensky (1985) and Farley and Monheit (1985) used equation (1.7) in their models rather than including the marginal tax rate as a separate variable. Taylor and Wilensky (1983) and Enthoven (1993) included the marginal tax rate as a regressor. This study will follow Gruber and Poterba (1996) for the tax treatment. The definition of the after-tax price of health insurance given in equation (1.10) could not be applied directly because information on loading fees and the employee’s contribution to premiums is not available. Given the data limitations, this study will apply the following definition for the after-tax price of health insurance,

\[
P = \left[ \frac{1 - \tau_f - \tau_s - \tau_{ss} - \tau_{mc}}{1 + \tau_{ss} + \tau_{mc}} \right].
\]

This definition assumes that the worker bears the entire tax burden of both payroll taxes because the employer’s contributions to health insurance premiums ultimately come out of what would otherwise have been the money wage for workers. So higher medical costs do not harm employers or owners but do reduce money
wages for workers. Lower costs benefit workers, not employers because they add to take-home pay, not profits. The employers voluntarily offer health insurance in other states because it improves workers’ health and attracts better workers. Improved worker health means higher productivity and reduced absenteeism.

An individual’s income is another main factor that affects the demand for health insurance. According to Arrow (1974) risk-aversion decreases as income increases which implies that an individual’s income would have a negative effect on the demand for insurance. On the other hand, Pauly (1978) said that health insurance becomes more valuable as income rises since the size of the loss that health insurance covers rises with income. If medical care is a normal good then the effect of income on the demand for insurance may be positive, even if risk-aversion decreases as income increases.

The study will follow a Gruber and Poterba (1996)-type probit regression model to estimate the demand for employment-based health insurance in Hawai’i. The model relates the demand for employer-provided insurance to the after-tax price of insurance, the income of the individual, and other institutional features. The model will also control for individual and job characteristics which are explained later on in the text.

The probability of purchasing insurance is a function of a number of regressors. The primary regressors are the after-tax price of employer-provided health insurance and the earner’s income. The after-tax price of health insurance is calculated by the formula developed by Gruber and Poterba (1996) given in equation (1.10). The earner’s annual
income is used as a proxy for permanent income since a complete measure of wealth is not available in CPS data sets. Income includes total annual income from all sources. Another variable that affects the demand for health insurance is the entitlement to a public health insurance program (Medicare or Medicaid). These programs are viewed as substitutes for private health insurance. These might have a crowding-out effect on private health insurance. Therefore, it will negatively affect the demand for employment-based health insurance.

The regression also includes age and age-squared as regressors because age captures the risk of illness and expected medical expenditures. The risk of illness is high for new-born children and elderly people and low for young and middle aged people. This assumption is made based on the empirical fact that health expenditures decrease as a child grows and then start increasing as people age. The demand for insurance is also affected by personal characteristics such as education, gender, marital status, and race. These variables are also included as regressors. Demand is also influenced by job characteristics: employment status (full-time or part-time employment), industry, union membership and firm size. This study has also included the regional variables Honolulu and non-Honolulu as a regressor. The residents of Honolulu and non-Honolulu might have different evaluations of health insurance and might behave differently in seeking health insurance. The sign of Honolulu is expected to be positive based on the economic model of medical services. Most medical services are only available in Honolulu. Therefore, travel costs are less for residents of Honolulu. Self-declared health status is also included as a regressor. The number of childless families will vary in the future as
the age distribution of households changes. This means that a reduction in children per
adult in the future will be reflected primarily in a different distribution of children among
those families having at least one child. The variable of childless families is included in
the regression. The consumer’s preferences regarding risk also play an important role in
the determination of the demand for health insurance. A globally risk-averse worker will
purchase health insurance while the risk-loving worker will not. At present no national
survey contains direct information on consumer’s preferences about health care and
health insurance. Therefore, since this variable cannot be included in the regression this
specification treats consumer’s preferences as given. The study also attempted to control
for variation in time and industrial structure by including eight year dummies and thriteen
industry dummies in the regression.

Formally, the empirical model discussed above is described in the following equation:

\[ Y_i^* = X_i \beta + \varepsilon_i \ldots \ldots (1.11) \]

An employee’s demand for health insurance is denoted by \( Y_i^* \) which linearly depends on
the explanatory variables, such as the after-tax price of health insurance, income of the
employee and socio-demographic characteristics denoted by \( X_i \) which is a \( K \times 1 \) row
vector. Its components and the expected signs of the corresponding coefficients are given
in Table 1.1. The employee’s demand for employment-based health insurance \( Y_i^* \) is
unobservable and continuous. Consider, \( Y_i^* \) represents a difference in expected utilities
with and without health insurance = EU(with health insurance) – EU(without health
insurance). Then when \( Y_i^* \geq 0 \) the employee chooses to buy health insurance, whereas
when \( Y_i^* < 0 \) the employee does not. We do not observe the net benefit of the purchase of
health insurance. What is being observed is a dummy variable $Y_i$ which is equal to one if an employee is covered by employment-based health insurance and zero otherwise. The observed variable, $Y_i$, assumes the form $Y_i = 1$ if $Y_i^* \geq 0$ and $0$ if $Y_i^* < 0$. Where $\beta$ is a vector of unknown coefficients to be estimated. A plausible assumption for the disturbance term, $\epsilon_i$, is that it follows the normal distribution, since $Y_i^*$ is a continuous variable measuring utility difference. The probability of $Y_i^*$ follows a cumulative distribution function $F(X_i\beta)$ as in Goldberger (1964). However, due to only two observed values of $Y_i$, only two residuals exist for any single $X_i$ value, $\epsilon_i = -X_i\beta$ if $Y_i = 0$ and $\epsilon_i = 1 - X_i\beta$ if $Y_i = 1$. Therefore, observed values of $Y_i$ do not follow a normal distribution but rather a Bernoulli (Binomial (0,1)) distribution. Thus, the expectation of the random term is $E(\epsilon_i) = (1 - X_i\beta) F(X_i\beta) - [1 - F(X_i\beta)]X_i\beta = F(X_i\beta) - X_i\beta \neq 0$. Similarly, the variance of $\epsilon_i$ is not constant (Heteroskedasticity problem). An individual's decision to purchase health insurance is a random component. The probability of purchasing this health insurance is given in the following equation (1.12):

$$E(Y_i) = \text{prob}(Y_i = 1) = \text{prob}(Y_i^* \geq 0) = \text{prob}(\epsilon_i \geq -X_i\beta) = F(X_i\beta) = 1 - F(-X_i\beta). \quad (1.12)$$

where $F$ is the cumulative distribution function for $\epsilon_i$. The observed values of $Y_i$ are considered as realizations of a binomial process with probabilities given by equation (1.12) that depend on $X_i$. Hence the likelihood function that follows from these probabilities is

$$L = \prod_{i \in \mathcal{Y}_0} [1 - F(X_i\beta)] \prod_{i \in \mathcal{Y}_1} F(X_i\beta) \quad \prod_{i \in \mathcal{Y}_1} \quad (1.13)$$

The functional form of $F(.)$ in equation (1.13) depends on the assumptions made about the probability structure of the error term, $\epsilon_i$. If the cumulative distribution function of $\epsilon_i$
follows standard normal distribution then equation (1.13) becomes a probit regression model. \( F(x, \beta) = \int_{-\infty}^{x, \beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \) \(^{1.14}\)

On the other hand, if the cumulative distribution function of \( \varepsilon \) follows a logistic distribution then equation (1.13) becomes a logit regression model.

\[
F(x, \beta) = \frac{\exp(x, \beta)}{[1 + \exp(x, \beta)]} = \frac{1}{[1 + \exp(-x, \beta)]} \] \(^{1.15}\)

Obviously, the latent variable \( Y_j* \) follows the distribution of \( \varepsilon \). The cumulative normal distribution and the logistic distribution are very close to each other, except at the tails. Both of these distributions have the familiar bell shape of symmetric distributions. The probit regression transforms probabilities of purchasing health insurance into \( Z \) scores from the cumulative standard normal distribution (1.14) rather than into logged odds from the logistic distribution (1.15). Despite this difference, probit and logistic regressions give essentially similar results.

The estimates of \( \beta \) from the two models are not directly comparable. Since the logistic distribution has a known variation of \( \pi^2/3 \) while the standard normal distribution has a variation of \( 1 \), the estimates of \( \beta \) obtained from the logit model have to be multiplied by \( 3^{1/2}/\pi \) to be comparable to the estimates obtained from the probit model. Both models are non-linear and non-additive. In both models, independent variables have a greater effect on the probabilities in the middle of the cumulative distribution curve than near the extremes. Both of these models are estimated by the maximum likelihood estimation method. The maximum likelihood estimation in linear regression leads to weighted least squares, while this is not the case in logistic or probit regressions. In these models, the
log-likelihood is maximized numerically using an iterative algorithm (Collett, 1991). The log-likelihood for a sample of independent observations is given in the following equation (1.16):

\[
\log L = \sum_{i=1}^{n} \{(1 - Y_i) \log(1 - F(X_i, \beta)) + Y_i \log(F(X_i, \beta))\} \quad (1.16)
\]

where \(F(X_i, \beta)\) can be specified as a linear function of \(X_i \beta\) the linear probability model.

The dependent variable is the probability of choosing health insurance. The coefficients of the probit model show a linear and additive change in Z-score units due to a one-unit change in the independent variable. The probabilities can be calculated by using a computer or Z-table.

The decision to include or exclude a variable in the model is made based on the economic theory and the Bayesian Information Criterion (BIC) developed by Raftery (1995), \(\text{BIC} = Z^2 - \ln(n)\). Where \(Z\) is the Z-statistic of the coefficient in question and \(n\) is a sample size. The BIC value refers to the difference in model information with and without the variable in question. If the BIC value for a variable equals or falls below 0, the data provide little support for including the variable in the model. Therefore, the variable is not included in the model. On the other hand, if the BIC for a coefficient is above 0, then the variable should be included in the regression model. Raftery specifies a rule of thumb to evaluate the “grades of the evidence” for the inclusion of a variable. He defines a BIC difference of 0-2 as weak, 2-6 as positive, 6-10 as strong, and greater than 10 as very strong. A few variables are included in the regression because of economic theory even though their BIC ranking was weak.
A regular regression assumes additivity and linearity which means that the effect of an independent variable on the dependent variable stays the same regardless of the levels of other independent variables. The probit and logit models violate the additivity and linearity assumptions between probabilities and independent variables. In other words, if the value of one independent variable is sufficiently high to push the probability of the dependent variable to near 0 or 1, then the effects of other variables cannot have much influence. Thus, the floor and ceiling make the influence of all the independent variables inherently non-additive and interactive. But in the logit model the effects of the independent variables on the logged odds are linear and additive — each X variable has the same effect on the logged odds regardless of its level or the level of other X variables. Similarly, in the probit model the effect of the independent variables on the Z-scores is linear and additive. But the effect of a unit change in the independent variable X on the predicted probability is nonlinear and non-additive — each X variable has a different effect on the probability depending on its level and the level of other independent variables. The effect of X on the probability would be smaller near the 0 or 1 than near the middle because the function bends slowly and smoothly as the values approaches to the floor or ceiling.

The error term also violates the assumption of homoscedasticity or equal variance for all X values. The residuals are relatively small near the floor of 0 and the ceiling of 1 but large near the middle values of X. As a result, the variance of the errors is not constant for all values of X. Therefore, the least squares criterion of estimation does not give
efficient estimates. This is why probit and logit regression models are used. They give more efficient estimates in this case.

The probit or logit regression model has a number of properties. First, the probit or logit models do not have a boundary of 0 or 1. In the logit model, the odds eliminate the upper boundary of probabilities, and the logged odds eliminate the lower bound of probabilities as well. While in the probit model, as probabilities vary between 0 and 1, the corresponding Z-scores vary between negative and positive infinity. Thus, the probit or logit vary from negative infinity to positive infinity. Second, the probit and logit are symmetric around the midpoint probability of 0.5. Third, there is a nonlinear relationship between logged odds and probabilities. That is, the same change in the probability translates into a greater change in the logged odds near the floor or ceiling than near the midpoint. Similarly, the same change in probability results in a bigger change in Z scores as the probabilities approach 0 or 1. Conversely, a one-unit change in the logit translates into a greater change in probabilities near the midpoint than near the extremes. Similarly, one-unit change in the Z scores produces a smaller change in the probabilities near the floor of 0 and near the ceiling of 1 than in the middle. The probit or logit regression relies on maximum likelihood procedures to obtain the coefficient estimates. The maximum likelihood estimation aims to find those coefficients that have the greatest likelihood of producing the observed data. This means maximizing the log likelihood function.

Although probabilities vary between 0 and 1, the logit or the logged odds of the probabilities have no such limits — they vary from negative to positive infinity. There
are a number of S-shaped curves that differ in how rapidly or slowly the tails approach 0 and 1. The logit transformation used in a logistic regression has the advantage of relative simplicity, while the probit transformation appears often in published literature. The cumulative standard normal curve resembles the logistic curve, only with Z-scores instead of logged odds along the horizontal axis. The probit curve approaches the floor and ceiling slightly faster than the logit curve, but the differences are small.

1.6. Data and Methodology

This section describes data sources and descriptive statistics. The data used in this study is not collected especially for the estimation of this model, but extracted from several different sources, primarily the Current Population Survey (CPS) March supplement. The data on the marginal state tax rates is calculated by taxism version 5.0 (a tax calculation program written by the National Bureau of Economic Research (NBER)). The data used in this study is seasonally unadjusted annual data for 1995 – 2003 for each of the five states included and the overall United States. About 3% of the data was excluded because the data about their income tax was incomplete. The sample is limited to labor force participants who earned positive income in the previous year. This restriction is required to calculate the after tax price of health insurance. A more comprehensive specification of the after-tax price of health insurance can be used with improved data.

The CPS is one of the oldest, largest and most well-recognized nationally representative labor surveys in the United States. The Census Bureau has been conducting the CPS on a monthly basis since the 1940s to provide up-to-date information on the labor force and demographic characteristics of the US population. Initially, the CPS was designed to
collect up-to-date facts about the number of Americans who are employed, unemployed or not in the labor force. Over the years, the survey has been expanded in many areas such as earnings and health insurance. In addition to employment facts, in 1988 the CPS March Supplement started collecting information on whether each individual had a health insurance plan provided through an employer or union at any time during the previous year. It also asks whether the employer paid all, part or none of the plan’s premium and who else in the family was covered by the plan. The question about self-declared health status was added in 1995. Since self-declared health status is a major factor in deciding whether to buy health insurance, this study uses data from 1995 to 2003. The present study uses 4308 observations. The eliminated observations do not include sufficient information to construct one or more variables called for by the model.

Currently, the CPS interviews around 99,000 housing units monthly which are selected scientifically. The sample is selected in stages. In the first stage of sampling, the U.S is divided into 792 primary sampling units (PSUs). Every PSU falls within the boundary of a state and is comprised of a metropolitan area, a large county, or a group of smaller counties. The PSUs are then grouped into strata. The strata are constructed so that they are as homogeneous as possible with respect to labor force and other social and economic characteristics. One PSU is randomly selected per stratum. In the second stage of sampling, a sample of four housing units within the sample PSUs is drawn. A selected housing unit is interviewed for 4 consecutive months and then dropped from the sample for the next 8 months and then brought back in the following 4 months and then dropped out from the sample. The sample size of the CPS varies from year to year but has been
significantly increased since 2002, to improve state-level estimates. The CPS sample is based on the civilian non-institutional population of the United States. A person 15 years old or over residing in a scientifically selected household is interviewed to collect information about the whole household.

The sample includes only those individuals who earn a positive amount of income because social security tax rates are only available for those individuals that are required to calculate the after-tax price of health insurance. It has also excluded employed individuals whose insurance information is missing from the data. A family is considered as employed if either the head or spouse is employed. A family is insured if both the family head and spouse reported being insured at any point during the survey year.

Hawaii is composed of six major islands (Hawaii's big island, Kauai, Lanai, Maui, Molokai and Oahu) but the public use CPS data file only allows for the analysis of Honolulu and non-Honolulu counties. The Honolulu county consists of all the residents of Oahu, while non-Honolulu county includes the residents of the other five Islands. The March Supplement Final Weights are used to correct the sampling bias and the probability of being included in the sample.

Table 1.2 presents the list of variables used in the estimation along with their weighted means, standard deviations, and minimum and maximum values. The first column of the table reports the variable names used in the regression. The weighted means of the corresponding variables are reported in the second column of the table. The means of the
dummy variables equal the proportion of cases with a value of 1, and can be interpreted as a probability. The standard deviation and minimum and maximum values are reported in the subsequent columns respectively. The projected effects of the corresponding variables are shown in the last column of the table 1.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th>Exp Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insured</td>
<td>0.92</td>
<td>0.27</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>After-tax-price</td>
<td>0.61</td>
<td>0.06</td>
<td>0.40</td>
<td>0.75</td>
<td>-</td>
</tr>
<tr>
<td>Log of personal income</td>
<td>10.23</td>
<td>0.76</td>
<td>0</td>
<td>13.11</td>
<td>+</td>
</tr>
<tr>
<td>Age</td>
<td>39.09</td>
<td>12.75</td>
<td>15</td>
<td>81</td>
<td>-</td>
</tr>
<tr>
<td>Age squared</td>
<td>1690.32</td>
<td>1075.22</td>
<td>225</td>
<td>6561</td>
<td>+</td>
</tr>
<tr>
<td>College or higher education</td>
<td>0.62</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Male</td>
<td>0.63</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.44</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Non-white race</td>
<td>0.71</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Full-time worker</td>
<td>0.77</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Worker having children</td>
<td>0.43</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Working living in Honolulu County</td>
<td>0.78</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Union member</td>
<td>0.06</td>
<td>0.24</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Firm size 1 - 9</td>
<td>0.12</td>
<td>0.33</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Firm size 10 - 499</td>
<td>0.34</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Firm size 500 and over</td>
<td>0.53</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Eligible for public insurance</td>
<td>0.02</td>
<td>0.13</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>White collar worker</td>
<td>0.47</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Excelent health</td>
<td>0.29</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Very good health</td>
<td>0.37</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Good health</td>
<td>0.29</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Poor health</td>
<td>0.06</td>
<td>0.23</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

The data show that the average age of wage earners in Hawai‘i is approximately 39.8 years, which is slightly higher than the average age of the national labor force in the United States, which is 39.4 years. The current average income of an earner in Hawai‘i is $28000 which is also slightly higher than the national average during a similar time period. The percentage of college graduate earner in Hawai‘i is 62% which is much

51
higher than national average 56%. The percentage of male earners in Hawai‘i is almost 65% which is also much higher than the national average 53%. The percentage of married earners in Hawai‘i is 45% which is far below the national average of 59%. The percentage of full-time earners in Hawai‘i is 76% which is also much less than the national average of 94%. The percentage of earners living in a household having at least one child under the age of 18 is 45% in Hawai‘i, which is less than the national average of 50%. The percentage of earners who are union members is 5% which is much more than the national average of 3.1%. More earners in Hawai‘i either work in small businesses or large firms as compared to national average.

1.7. Estimates of the Demand for Health Insurance in Hawai‘i
This section discusses the results of regression equation (1.14). All the regressions are run on the earner’s characteristics. Regression equation (1.14) is estimated using probit and logit non-linear estimation methods. The results of both the estimation techniques are very similar. Only probit estimates are reported in tables 1.3 and 1.4 since the estimation methods do not make significant differences in the coefficient estimates and most of the published literature reports probit results. In general, the logistic regression coefficients exceed the corresponding probit coefficients by a small fraction, $\lambda$ (Cramer, 2003). The choice of which model to use may be a matter of computational ease of use and personal preference. Theoretically, it is difficult to justify the selection of one model over the other.

The key variables of the model are the after-tax price of health insurance and the worker’s income. The model also controls a detailed set of demographic, job and
individual characteristics: age, education, gender, marital status, race, full or part-time worker, union membership, worker living in a family with children. It also includes other variables such as: self-declared health status, occupation, industry and firm size. The effect of a unit change in any independent variables on the predicted probability would be smaller near the floor or ceiling than near the middle. This is because of the fact that the distribution function is less curved in the middle than near the tails. As values get closer and closer to 0 or 1, the relationship requires a larger and larger change in the independent variable to have the same impact as a smaller change in the independent variable at the middle of the curve. All the coefficients are statistically significant at the 5% level, except for the following dummies: male, race, children, Honolulu and white-collar workers. It is disappointing to find that the regional dummy is not significant even at the 20% level of significance. The regression estimates are weighted by the CPS March supplement’s final weights.

The reference group of the regression given in table 1.3 is an ineligible unmarried white female blue-collar non-union member who works part-time in a large firm, has high school or less education, and lives outside Oahu. Her predicted probability of being covered is equal to 0.096 (the intercept). The predicted probability is computed at the means values of the independent variables. The predicted probability comes from a probit regression model and the formula is \( Y = \Phi (XB) \). Where \( \Phi \) is a normal distribution function determined by the unknown parameters \( \beta \) and the independent variables of \( X \).
The likelihood function varies between 0 and 1 so correspondingly the log likelihood function will vary from negative infinity to zero. The natural log of 1 is equal to 0 and the natural log of 0 is undefined, but as the probability approaches 0, the natural log becomes an increasingly large negative number. The closer the likelihood value to 1, then the closer the log likelihood value is to 0, and the more likely it is that the parameters could produce the observed data. The more distant the negative value from zero, the less likely the parameters could produce the observed data. In other words, the larger the negative value of the log likelihood, the poorer the model. The log likelihood value reflects the likelihood that the data would be observed given the parameter estimates. The model is perfect if log likelihood equals 0. The larger its value (i.e. the closer the negative value to zero), the better the parameters do in producing the observed data.

The log likelihood estimate of the full model is -1028. One way to interpret the size of the log likelihood estimate is to compare the model value to the initial or baseline value assuming all the b coefficients equal zero except the intercept. The baseline log likelihood comes from including only a constant term in the model. The greater the difference between the baseline log likelihood and the model log likelihood, the better the model coefficients do in producing the observed sample values. This difference can be used for goodness of fit as well as for significance tests. Multiplying the difference by -2 gives a chi-square value with degrees of freedom equal to the number of independent variables (not including the constant, but including squared and interaction terms). The model chi-square value is equal to 206 which decisively reject the null hypothesis. This means that the explanatory variables explain the model significantly.
Table 1.3 shows a baseline log likelihood of $\ln l_0 = -1131$ and a full model log likelihood (reached after five iterations) of $\ln l_1 = -1028$. If a model makes no contribution toward explaining the variation in the dependent variable, the difference between these two log likelihood values is very close to zero. On the other extreme, if the model completely explains the variance in the dependent variable, then $\ln l_1 = 0$ and the difference between these two log likelihood value is maximum. If the model is less than perfect, then $\ln l_1 < 0$ and the difference is moderate. The log likelihood ratio multiplied by $-2$ denoted by $-2\text{LLR}$ follows the chi-square distribution with degrees of freedom equal to the number of independent variables in the model. Therefore, multiplying the difference of $-102.8$ by $-2$ gives the calculated chi-square value of 206 with 37 degrees of freedom. This can be compared with tabulated chi-square values. The chi-square test shows that the regression is highly significant.
Table 1.3: Probit regression estimates for the demand for health insurance in Hawaii

<table>
<thead>
<tr>
<th>Variables</th>
<th>coefficients</th>
<th>dy/dx</th>
<th>Std.Err</th>
<th>Z</th>
<th>P&gt;Z</th>
<th>95 % CI</th>
<th>X bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>After-tax-price</td>
<td>-1.9433</td>
<td>-0.2111</td>
<td>0.7522</td>
<td>-2.58</td>
<td>0.01</td>
<td>-3.42</td>
<td>-0.47</td>
</tr>
<tr>
<td>Log income</td>
<td>0.1888</td>
<td>0.0205</td>
<td>0.0587</td>
<td>3.22</td>
<td>0.00</td>
<td>0.07</td>
<td>0.30</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0359</td>
<td>-0.0039</td>
<td>0.0159</td>
<td>-2.26</td>
<td>0.02</td>
<td>-0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>Age square</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.0002</td>
<td>2.81</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>College or higher</td>
<td>0.1188</td>
<td>0.0132</td>
<td>0.0748</td>
<td>1.59</td>
<td>0.11</td>
<td>-0.03</td>
<td>0.27</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0704</td>
<td>-0.0075</td>
<td>0.0783</td>
<td>-0.9</td>
<td>0.37</td>
<td>-0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.4031</td>
<td>0.0425</td>
<td>0.0853</td>
<td>4.73</td>
<td>0.00</td>
<td>0.24</td>
<td>0.57</td>
</tr>
<tr>
<td>Non-white race</td>
<td>0.0076</td>
<td>0.0008</td>
<td>0.0795</td>
<td>0.1</td>
<td>0.92</td>
<td>-0.15</td>
<td>0.16</td>
</tr>
<tr>
<td>Full-time worker</td>
<td>0.1983</td>
<td>0.0235</td>
<td>0.0911</td>
<td>2.18</td>
<td>0.03</td>
<td>0.02</td>
<td>0.38</td>
</tr>
<tr>
<td>Worker having children</td>
<td>0.0750</td>
<td>0.0081</td>
<td>0.0789</td>
<td>0.95</td>
<td>0.34</td>
<td>-0.08</td>
<td>0.23</td>
</tr>
<tr>
<td>Honolulu County</td>
<td>0.0229</td>
<td>0.0025</td>
<td>0.0894</td>
<td>0.26</td>
<td>0.80</td>
<td>-0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>Union member</td>
<td>0.5966</td>
<td>0.0424</td>
<td>0.2080</td>
<td>2.86</td>
<td>0.00</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Firm size 1 - 9</td>
<td>-0.3080</td>
<td>-0.0403</td>
<td>0.1043</td>
<td>-2.95</td>
<td>0.00</td>
<td>-0.51</td>
<td>-0.10</td>
</tr>
<tr>
<td>Firm size 10 - 499</td>
<td>-0.1843</td>
<td>-0.0210</td>
<td>0.0788</td>
<td>-2.34</td>
<td>0.02</td>
<td>-0.34</td>
<td>-0.03</td>
</tr>
<tr>
<td>Eligible for PHI</td>
<td>-0.5430</td>
<td>-0.0879</td>
<td>0.1880</td>
<td>-2.89</td>
<td>0.00</td>
<td>-0.91</td>
<td>-0.17</td>
</tr>
<tr>
<td>White-collar worker</td>
<td>0.1152</td>
<td>0.0124</td>
<td>0.0804</td>
<td>1.43</td>
<td>0.15</td>
<td>-0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>Constant</td>
<td>0.4596</td>
<td>0.0839</td>
<td>0.52</td>
<td>0.60</td>
<td>-1.27</td>
<td>2.19</td>
<td></td>
</tr>
</tbody>
</table>

Note: industry and year dummies are included as control variables.

Another familiar measure of the overall effectiveness of the model, coefficients of multiple determination, $R^2$, is not available in probit or logit regression. Instead several substitutes for the $R^2$ statistic are available called pseudo $R^2$s. These are used to evaluate the overall goodness of fit of the model, but no consensus has emerged on the single best measure. Also, there is no advice in the literature about which is preferred under what circumstances. Different pseudo $R^2$s give different results. Some regression analysts find these pseudo $R^2$ statistics to be of little utility (King, 1990). Others use these pseudo $R^2$ statistics extensively while evaluating model performance (Lewis-Beck and Skalaban, 1990). Therefore, these measures should be used as only rough guides without attributing great importance to any one precise figure.
McFadden (1974), McKelvey and Zavoina (1975), Aldrich and Nelson (1984) and Dhrymes (1986) derived the four most commonly used pseudo $R^2$s in literature. Many computer packages compute one or more pseudo $R^2$s. The STATA 8.2 computer package used in this study calculates McFadden’s pseudo $R^2$. It is defined as one minus the ratio of the full model log likelihood value denoted as $\ln L_1$ to the log likelihood value for the null model (i.e., with an intercept only) denoted as $\ln L_0$. Mathematically, it can be written as pseudo $R^2 = 1 - \frac{\ln L_1}{\ln L_0}$. The value of pseudo $R^2$ ranges from 0 to 1. The larger the value of Pseudo $R^2$, the better the model fits. Morrison (1972) argued that pseudo $R^2$s usually have lower values than the usual $R^2$ in simple regression models. Note that this value does not depend on the number of observations in the data set, $n$, while the other two measures of pseudo $R^2$ depend on the number of observations in the data set. Therefore, the value of this pseudo $R^2$ will never decline when irrelevant variables are added to the model. The value of McFadden’s pseudo $R^2$ is equal to 0.12 which indicates that the regression is much less than perfect but still highly significant.

The third column (3) presents the marginal probability effects of the corresponding coefficients contained in column (2). The results are broadly consistent with prior studies of the demand for health insurance. The results show that the demand for health insurance is negatively related to the after-tax price of health insurance in all five states. A ten cent increase in the after-tax price of health insurance will reduce the probability of purchasing health insurance by 2.1%. The elasticity of demand for health insurance with respect to the after-tax price is $-0.13^4$ in Hawai‘i. Therefore, a ten percent increase in the

---

4 Price elasticity is a ratio of the marginal change (i.e., per unit change) between the probability of being covered and the after-tax price of health insurance multiplied by the ratio of the averages of these variables.
The after-tax price of health insurance will reduce the probability of purchasing it by about 1.3% (i.e., 10% X-0.13=-1.3%). Different states have provided very different tax subsidies. Therefore, the after-tax price varies from $0.61 in a high-tax state like Hawai'i to $0.68 in a low-tax state like Florida. The price elasticity is calculated by the following formula: \( E_{tp} = \frac{\partial Y^*}{\partial X_{tp}} \left( \frac{tp}{Y^*} \right) = b_p * \phi(X\beta) * \frac{tp}{\Phi(X\beta)} \).

The price elasticity is evaluated at the average predicted probability of each state at the mean and the mean of the after-tax price of health insurance in that state. The after-tax price elasticity estimate is well within the range of estimates reported in the literature.

The findings also support a small positive income elasticity. The income variable is included in the regression after taking its natural logarithm. The worker's income has a small positive effect on the probability of purchasing health insurance. The elasticity of demand for health insurance with respect to income is 0.02. A ten percent increase in income will increase the probability of purchasing health insurance by 2.0%. This elasticity varies along the distribution curve. The closer \( Y^* \) or \( Z \) comes to the ceiling or floor, the smaller the value \( Y^*(1-Y^*) \) or \( f(Z) \), and the smaller the elasticity of income as compared to the middle of the distribution. If income increases the likelihood of buying health insurance, an increase of $5,000 in income from $40,000 to 45,000 would increase that likelihood more than an increase from $100,000 to 105,000. A high-income person would no doubt already have a high probability of buying health insurance, and a $5000 increase would do little to increase their already high probability. The same would hold for an increase in income from $0 to $5000. Since neither income is likely to be sufficient to purchase health insurance, the increase in income has little impact on the probability of
purchasing health insurance. Nevertheless, the income elasticity increases as the
employee's income increases. In the middle-range, however, the additional $5000 may
make the difference between being able to afford a health insurance and not being able to
afford a health insurance. The income elasticity estimate is also within the range of
estimates reported in the literature. The income elasticity is calculated by the following
formula: 

\[ E_j = \frac{\partial Y^*}{\partial X_j} \cdot \left(1 / Y^* \right) = b_j \cdot \phi(X\beta) / \Phi(X\beta). \]

The partial derivative of continuous variables reveals the change in probability for a
small change in X, or the change in the (slope of the) tangent line due to a one-unit
change in X at a particular value. It does not interpret the change in the logistic curve for
a one-unit change in the independent variable. The formula for the partial derivatives of
the logit function is given in the following equation.

\[ \frac{\partial Y^*}{\partial X_i} = b_i \cdot Y^* \cdot (1 - Y^*). \]

Simply multiply the logistic regression coefficient by the selected probability and 1
minus the probability. However, the partial derivative for the probit regression takes the
following form 

\[ \frac{\partial Y^*}{\partial X_i} = b_i \cdot \phi(X\beta) \quad \text{where} \phi(X\beta) \text{ is the density or height of the} \]

standard normal curve at the point of X\beta. Mathematically, the probability density at the
point X\beta can be presented by the following formula:

\[ \phi(X\beta) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{(X\beta)^2}{2} \right). \]

The density defined by the formula is highest at the X\beta score of 0, and successively
decreases as the X\beta deviates in either direction from 0. Both of the formulas for partial
derivatives nicely reveal the nonlinear effects of an independent variable on probabilities.
The effect of b (in terms of logged odds) translates into a different effect on the
probabilities depending on the level of Y* and \phi(X\beta). The effect will be at its maximum
when \( Y^* \) equals 0.5 or Z values are near 0. The closer \( Y^* \) or Z comes to the ceiling or floor, the smaller the value \( Y^*(1-Y^*) \) or \( \phi(X\beta) \), and the smaller the effect a unit change in X has on the probability. The partial derivatives for continuous variables and differences in predicted probabilities for dummy variables are calculated by using the STATA 8.2 computer package.

A partial derivative of the coefficient for an independent dummy variable approximates the group difference from the omitted group to the dummy variable group, in probabilities. However, the group difference in probabilities, like the partial derivative, varies with the point on the logistic curve or standard normal curve, the X values, and the Y values. However, the effect of dummy and continuous variables on predicted probabilities depends on the choice of starting point. Changes in probabilities will appear larger for points near the middle of the curve than near the floor or ceiling. A single coefficient cannot fully describe the relationship of a variable with probabilities, and the effect of a unit change on probabilities differs depending on the starting Z score and the values of the independent variables.

To calculate the change in probabilities for a dummy variable, take the mean of the dependent variable as the predicted probability for the omitted group. Translate the value into a Z score for the cumulative normal curve (i.e., the area below the Z score) using a table or computer function. Then add the probit coefficient for the dummy variable to this Z score and transform the new Z score sum back into a new probability using a table or computer function. The new probability minus the mean probability shows the difference
in predicted probabilities between the two groups. The same logic of calculating predicted probabilities holds for continuous variables. After translating the mean probability into a probit Z score, add the coefficient for the continuous variable and transform the probit back into a probability. The difference between the two probabilities gives the predicted change due to a one-unit change in the continuous variable.

The regression controls for age by including age and its squared terms as independent regressors. Both regressors are highly statistically significant at the 1% level of significance. The regression results suggest that the effect of age on the probability of being insured, controlling for other variables, is non-linear over the entire age range. Figure 1.4 presents a more detailed breakdown of the average probability of being insured by age. The average probability of being insured is plotted against age. The probability of being insured is measured along the vertical axis (i.e., y-axis) while the age of the worker is displayed on horizontal axis (i.e., x-axis). The shape of the curve follows the theoretically-expected U-shaped curve. The probability of being insured is high for children because of a high risk of sickness at early ages and because parents prefer to avoid health risks for their children. It begins to decrease as health stock improves as children grow up. It reaches its minimum at the age of 36 when health expenditures are the lowest. Then it starts to increase smoothly as health expenditures start to increase because health stock starts to deteriorate after age 36. The probability of being insured is also higher during the early 20s because those are child-bearing years. The thick curve represents the within-sample estimated curve while the dotted part of the curve at both ends is out-of-sample forecasting. The estimated model parameters are used to compute
out-of-sample forecasting. During the past decade the proportion of elderly people has
increased significantly in the US as a whole and in Hawai‘i as well. This trend is expected
to continue for a few more decades because of the aging and graying of baby-boomers.
This will increase the proportion of the population age 65 and over which uses the most
health care in the US. As a result, the pressure on the demand for health insurance and
health care will increase.

The level of education has a positive effect on the demand for health insurance. The
probability that a college graduate will purchase employer-provided health insurance is
1.3 percentage points higher than someone with less than a college education. The gender
of the employee also plays an important role in decision-making about the purchase of
health insurance. The probability that a male will buy employer-provided health
insurance is 0.8% less than a female buying the health insurance. The probability that a
married person will buy employer-provided health insurance is 4.3% more than an
unmarried person. These variables are highly significant.

To control for the ethnic preferences of Hawai‘i’s inhabitants, the variable race is
introduced in the regression. Hawai‘i is the most ethnically and racially diverse state in
the United States. The major races are Caucasians, Americans of Japanese descendent,
Polynesians, Native Hawai‘ians and others. Among major races, Americans of Japanese
descendent have the highest coverage rate. They are a significant portion of the
population in Hawai‘i. This is why the sign of the non-white population is positive. The
probability that a non-white person will buy employer-provided health insurance is
0.08% more than a white person. This variable is not statistically significant. The
probability that a full-time worker will buy employer-provided health insurance is 2.3%
more than a part-time worker. An employed person may incur a larger “time price” of
going to the physician than a part-time employed person. On the other hand, the full-time
employed might get cheaper health insurance due to the Prepaid Health Care Act.

Household composition effects are introduced by dummy variables which indicate
whether the earner lives in a household with children under age 18. The probability that
an earner living in a household with one or more children under age 18 will buy
employer-provided health insurance is 0.1% more than a worker living in a household
without children.
A region-specific supply effect is represented by the "Honolulu" variable. The population of Hawai'i is divided into two sub-populations: Honolulu and non-Honolulu. The people who live on Oahu are grouped into Honolulu County. The people of the neighboring Islands are grouped into non-Honolulu counties. More than 75% of the working population of Hawai'i lives on Oahu. The results show that people who live in Honolulu County are more likely to be insured as compared to those who live in the non-Honolulu counties. This fact reflects the supply-side effect of health services. Most of the hospitals are in Honolulu County and the majority of doctors serve in these hospitals or near the area. The residents of non-Honolulu may incur a larger cost of going to the physician than the residents of Honolulu. This makes the insurance more valuable for people living in Honolulu County because of lower travel costs. People living on the neighboring Islands have to fly to Honolulu to get some health services which costs them extra cash outlays for transportation, food, and lodging away from home. The longer travel also costs them extra time lost from work or home duties. These extra costs make insurance less valuable for people living on the neighboring Islands. This is why the sign of Honolulu is positive in the regression.

The probability that a unionized earner will buy employer-provided health insurance is 4.2% more than a non-unionized earner. Unions often try to negotiate benefits package which mostly include health insurance benefits. The results also indicate that there is a positive correlation between firm size and the demand for health insurance. In every state, employees of large firms are more likely to be insured than those who work for small firms. Three dummy variables are generated from the firm size variable: small-
scale firms (1 to 9 employees), medium-scale firms (10 to 499 employees) and large-scale firms (500 and over employees). The large-scale firms dummy is used as the reference group in the regression. There is a positive correlation between the size of the firm and the probability of being insured. The probability that an eligible earner for a public health care program will buy employer-provided health insurance is 8.8\% less than an ineligible earner. The reason is that the eligible earner has an incentive to try to substitute other public health care programs which are relatively cheaper than the private program. The probability that a white-collar earner will buy employer-provided health insurance is 1.2\% more than a blue-collar earner. There is concern that some of the independent variables may be jointly determined by health insurance. For example, there is a well-known positive correlation between income and benefits. Good jobs offer higher pay and better benefits. This is why white color employees are more likely to be insured. Similarly, individuals who desire health insurance may choose to work in occupations that are more likely to offer health insurance or choose to work full-time rather than part-time in order to be eligible for benefits.

Considerable fluctuations from year to year and industry to industry are noted by many economists even when income and prices are held constant. This is why a full set of year and major industry dummies are included in the regression in order to explicitly control for structural shifts over time and industry. These variables are included to eliminate bias in the regression coefficients but not reported in the table.
The size of a coefficient relative to its standard error provides the basis for a test of significance. The standard errors of the coefficients are given in column (4). The coefficient divided by its standard error is called the calculated Z-statistic. The Z-statistics for each corresponding coefficient are given in column (5). These Z-statistics are compared with the Z-distribution table given in any classic statistics textbook for evaluation of their significance. The observed p-value is given in column (6). The p-values indicate that all estimated coefficients are significantly different from zero at the 5% level of significance, except for a few of the variables discussed in the text above. The columns (7) and (8) report lower and upper bounds of 95% confidence interval of the corresponding odds ratio, respectively. The final column (9) represents the mean value of each variable.

Many health economists think that health status is a major factor in the decision of whether to buy health insurance. Self-declared health status is included in the regression and results are reported in table 1.4.
Table 1.4: Probit regression estimates for the demand for health insurance in Hawaii

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients (1)</th>
<th>dy/dx (2)</th>
<th>Std.Err (3)</th>
<th>Z (4)</th>
<th>P&gt;Z (5)</th>
<th>95% CI (6)</th>
<th>X-bar (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>After-tax-price</td>
<td>-1.8994</td>
<td>-0.2045</td>
<td>0.7510</td>
<td>-2.53</td>
<td>0.011</td>
<td>-3.371</td>
<td>-0.428</td>
</tr>
<tr>
<td>Log income</td>
<td>0.1886</td>
<td>0.0203</td>
<td>0.0588</td>
<td>3.21</td>
<td>0.001</td>
<td>0.073</td>
<td>0.304</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0335</td>
<td>-0.0036</td>
<td>0.0160</td>
<td>-2.09</td>
<td>0.037</td>
<td>-0.065</td>
<td>-0.002</td>
</tr>
<tr>
<td>Age square</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.0002</td>
<td>2.75</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>College or higher</td>
<td>0.1077</td>
<td>0.0118</td>
<td>0.0753</td>
<td>1.43</td>
<td>0.152</td>
<td>-0.040</td>
<td>0.255</td>
</tr>
<tr>
<td>Male</td>
<td>-0.0721</td>
<td>-0.0076</td>
<td>0.0783</td>
<td>-0.92</td>
<td>0.357</td>
<td>-0.226</td>
<td>0.081</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.3999</td>
<td>0.0418</td>
<td>0.0846</td>
<td>4.73</td>
<td>0.000</td>
<td>0.234</td>
<td>0.566</td>
</tr>
<tr>
<td>Non-white race</td>
<td>0.0228</td>
<td>0.0025</td>
<td>0.0803</td>
<td>0.28</td>
<td>0.776</td>
<td>-0.135</td>
<td>0.180</td>
</tr>
<tr>
<td>Full-time worker</td>
<td>0.1856</td>
<td>0.0217</td>
<td>0.0905</td>
<td>2.05</td>
<td>0.040</td>
<td>0.008</td>
<td>0.363</td>
</tr>
<tr>
<td>Worker having children</td>
<td>0.0649</td>
<td>0.0069</td>
<td>0.0786</td>
<td>0.83</td>
<td>0.409</td>
<td>-0.089</td>
<td>0.219</td>
</tr>
<tr>
<td>Honolulu County</td>
<td>0.0202</td>
<td>0.0022</td>
<td>0.0904</td>
<td>0.22</td>
<td>0.823</td>
<td>-0.157</td>
<td>0.197</td>
</tr>
<tr>
<td>Union member</td>
<td>0.6024</td>
<td>0.0423</td>
<td>0.2121</td>
<td>2.84</td>
<td>0.005</td>
<td>0.187</td>
<td>1.018</td>
</tr>
<tr>
<td>Firm size 1 - 9</td>
<td>-0.3070</td>
<td>-0.0398</td>
<td>0.1046</td>
<td>-2.94</td>
<td>0.003</td>
<td>-0.512</td>
<td>-0.102</td>
</tr>
<tr>
<td>Firm size 10 - 499</td>
<td>-0.1871</td>
<td>-0.0212</td>
<td>0.0789</td>
<td>-2.37</td>
<td>0.018</td>
<td>-0.342</td>
<td>-0.032</td>
</tr>
<tr>
<td>Eligible for PHI</td>
<td>-0.5674</td>
<td>-0.0927</td>
<td>0.1896</td>
<td>-2.99</td>
<td>0.003</td>
<td>-0.939</td>
<td>-0.196</td>
</tr>
<tr>
<td>White-collar worker</td>
<td>0.1138</td>
<td>0.0122</td>
<td>0.0809</td>
<td>1.41</td>
<td>0.159</td>
<td>-0.045</td>
<td>0.272</td>
</tr>
<tr>
<td>Very good health</td>
<td>-0.0716</td>
<td>-0.0078</td>
<td>0.0889</td>
<td>-0.8</td>
<td>0.421</td>
<td>-0.246</td>
<td>0.103</td>
</tr>
<tr>
<td>Good health</td>
<td>-0.0849</td>
<td>-0.0094</td>
<td>0.1008</td>
<td>-0.84</td>
<td>0.400</td>
<td>-0.282</td>
<td>0.113</td>
</tr>
<tr>
<td>Poor health</td>
<td>-0.3656</td>
<td>-0.0508</td>
<td>0.1476</td>
<td>-2.48</td>
<td>0.013</td>
<td>-0.655</td>
<td>-0.076</td>
</tr>
<tr>
<td>Constant</td>
<td>0.4464</td>
<td>0.6846</td>
<td>0.5</td>
<td>0.614</td>
<td>-1.287</td>
<td>2.180</td>
<td></td>
</tr>
</tbody>
</table>

Note: industry and year dummies are included as control variables.

The results presented in Table 1.4 follow the format of Table 1.3. The specification of the regression is exactly the same presented in Table 1.3 except for making health status an extra regressor. This variable is included in the regression because many health economists and policy makers think that health status is an important predictor for health care utilization. Therefore, there is a strong incentive for the worker with poor health to buy health insurance. This is why health status is included in empirical models of the demand for health insurance. The self-reported health status variable has four categories: excellent, very good, good and average health. In the original data set there were five categories excellent, very good, good, average and poor health. The last two categories average and poor health were merged due to very few numbers of observations in these
categories. One dummy variable is created for each category. The excellent health
category is used as the reference group. Economic theory presumes that healthier people
are less likely to be insured. So the expected signs for the health dummies were positive.
But the regression results surprisingly indicate the opposite sign to what was expected. In
contrast, Finkelstein (2002), using Canadian data for the year 1991 and 1994, find correct
signs for health dummies. There are various possible reasons for this incorrect sign, such
as influential data points, misspecification of functional form and endogeneity of the
health variable.

The next step is to explore the possible reasons for this incorrect sign. In the first step,
influential data points were detected and removed. The results still indicate that even
removing the influential data points from the data does not affect the sign. In the second
step, different functional forms of the regression were estimated. Again, the results seem
to indicate that the robustness of the sign pattern for these dummies remained stable and
opposite to what was expected by the theory. The study also tried estimating different
specifications of Equation (1.14). In one specification, I included only the most plausibly
exogenous variables, such as union membership, age categories, marital status, gender,
earner living in a family with children under 18, and educational attainment. In another
specification, I add those variables which are potentially endogenous, such as income,
occupation, full-time versus part-time and health status. Again, the results indicate that
the signs of these dummies remain opposite to what was predicted by theory. Although
the sign of self-reported health status is wrong it remains a strong predictor of the
demand for health insurance.
It is unclear from the literature how to address this endogeneity problem. The study has estimated Heckman probit and Biprobit models where health is used as an endogenous variable. A structural two-equation simultaneous model of non-linear demand for health insurance and linear health status functions were also estimated by using a two-stage method. In the first stage, an ordered probit of health status was estimated to predict health status. The health status was assumed to linearly depend on age, education, income and illness. Then the predicted values of health status were calculated. Following that the predicted values of health status were used as an independent variable in the demand for health insurance equation. For more details see computer program on page 89. However, both the elasticity estimates remain the same. Instrumental variable method is also used to control the endogeneity of the regressor. It is difficult to find a good instrumental variable to solve this endogeneity problem. The sign of the health status remains the same. Despite all efforts, the sign of the health status variable remained opposite to what is intuitively appealing. Health status also remains statistically significant. A number of interaction terms were also added in the regression to allow for the joint effects of different variables. These were insignificant and not reported in results. However, the price and income elasticity estimates remain robust.

1.8. Comparison of Results with Other Major Studies
Table 1.5 presents the price and income elasticity estimates of this study for the state of Hawai'i, four other selected states and the overall United States. The price and income elasticity estimates of Hawai'i can be directly compared with the estimates of other selected states and the United States as a whole because all these estimates are obtained
from almost the same model and exactly the same definitions of the variables. But, it is difficult to find a reasonable method of comparing the variability in these results because the definitions of the variables and the source of variation for price and income are not exactly the same in all studies. The nature of non-experimental data complicates the situation even more. These estimates are also sensitive to the methods used to estimate them. Nevertheless, these estimates give some idea of the responsiveness of the demand for health insurance due to price and income variables. Different empirical studies have used different variables as proxies for the price of health insurance due to data limitations. A wide range of price elasticity estimates are calculated by different empirical studies. Woodbury and Hamermesh (1992) reported very high price elasticities of -2 to -3. On the other hand, some previous studies found much lower elasticities, with an average coefficient of -0.16. Nevertheless, Farley and Wilenshy (1985) reported a -0.41 price elasticity that looks to be more than three times higher than what this study has calculated for the state of Hawai‘i. The income elasticities are generally positive but very close to zero. The elasticity estimates of price and income are summarized in Table 1.5.

Table 1.5:
Price and Income elasticities of demand for health insurance

<table>
<thead>
<tr>
<th>State</th>
<th>Price Elasticity</th>
<th>Income Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawai‘i</td>
<td>-0.13%</td>
<td>0.02%</td>
</tr>
<tr>
<td>Nevada</td>
<td>-0.52%</td>
<td>0.05%</td>
</tr>
<tr>
<td>Michigan</td>
<td>-0.36%</td>
<td>0.01%</td>
</tr>
<tr>
<td>California</td>
<td>-0.44%</td>
<td>0.06%</td>
</tr>
<tr>
<td>Florida</td>
<td>-0.58%</td>
<td>0.05%</td>
</tr>
<tr>
<td>United States</td>
<td>-0.34%</td>
<td>0.03%</td>
</tr>
</tbody>
</table>
The comparison of price elasticity for the four selected states and the overall United States with the state of Hawai'i indicates that the demand for health insurance in Hawai'i is very inelastic for private workers who are governed under PHCA. The price elasticity of demand for ESI is twice as elastic for exempt workers in Hawai'i. This large difference is most likely because of the Pre-paid Health Care Act (PHCA). In most cases, (except for Michigan) the income elasticity of health insurance for the other selected states and the overall United States also clearly exceeds that of Hawai'i. Previous studies of the U.S. demand for health insurance reports a wide range of price elasticities, from -0.16 to -3, which is much too wide to provide reliable guidance in assessing the effectiveness of a tax subsidy policy in Hawai'i. The results of this study indicate that the price elasticity of health insurance in Hawai'i is significantly lower than the other selected states and the overall United States. These elasticity estimates are also consistent with economic theory and empirical evidence from experimental data. When these elasticities are estimated with observations on private non-unionized full-time workers only, the results are not substantially different from those reported in Table 1.5.

---

5 Hawai'i -0.16 vs U.S. -0.46 for all earners including self-employed, government workers, and private workers. Hawai'i -0.12 vs U.S. -0.42 for private non-unionized full-time workers only.
1.9. Policy Implications
The government tax-subsidy will have a lesser effect on health insurance coverage in Hawai‘i as compared to other selected states and the overall United States because the Prepaid Health Care Act has significantly reduced the price elasticity of demand for ESI in Hawai‘i. The tax-subsidy policy will be almost three times less effective in Hawai‘i than nationwide because the price elasticity of the demand for ESI in Hawai‘i is almost three times lower than the overall United States. It also has important implications for other public policies, which depend on the responsiveness of the demand for ESI in Hawai‘i.

These elasticity estimates also provide interesting insights about insurance premiums in Hawai‘i. Health insurance premiums will increase more rapidly in Hawai‘i than in the other selected states and the overall United States due to a significant difference in the price elasticity of demand for ESI in Hawai‘i since the Prepaid Health Care Act regulation reduces the substitutability of other options. For example, the environmental tax on cars in Japan has pushed up car prices in Japan and their insurance premiums are much higher than the United States and other OECD countries. Estimates of the factors affecting demand for health insurance can be used by policy-makers to formulate more precise policies for coverage in Hawai‘i.
1.10. Conclusion
This study has explored the effect of the Prepaid Health Care Act on the demand for employer-sponsor health insurance (ESI) in Hawai‘i using CPS data for the years 1995 - 2003. This regulation requires that the employer must provide health insurance benefits to full-time employees and pay for at least half of the premium. Spouses and other family members are exempt from this Act as well as government employees, part-time workers and those who are self-employed. This regulation reduces flexibility in the choice of health insurance by full-time workers which causes the demand for health insurance to be restrictive and less price elastic. The principal finding of this study is that the Prepaid Health Care Act has significantly reduced the price and income elasticities of health insurance for full-time private workers in Hawai‘i. The results indicate that price, income and other substitutes for private health insurance have an impact on the worker’s decision to purchase ESI. The price and income elasticities of the demand for ESI for private workers in Hawai‘i are -0.13 and 0.02, respectively. Both elasticities are substantially lower than other comparable states and the overall United States. In particular, the price elasticity is almost three times lower than the overall United States. The price elasticity of the exempt group in Hawaii is -0.18 which is significantly higher than full-time workers in Hawai‘i. This price elasticity gap between full-time workers and the exempt group is not very prominent in other comparable states and the overall United States. In this analysis, the demand for ESI in Hawai‘i for full-time private workers and all other workers is estimated separately for comparison. Moreover, the demand for ESI is estimated for four other comparable states which do not have this kind of regulation. The
demand for ESI in the overall United States is also estimated separately. However, in Hawai‘i, the price elasticity of exempted workers is relatively higher than for full-time workers in Hawai‘i but significantly lower than the overall United States. The elasticity estimates for full-time workers and exempt workers are not significantly different in other selected states and the overall United States when they are compared within selected states. More importantly, the elasticity difference for exempted workers in Hawai‘i and other states is not as dramatic as in the case of full-time private workers. The price elasticity for other comparable selected states is even higher than the average national estimates except for Michigan. The price elasticity for the overall United States is -0.34 which is well within the range of other comparable studies. Most of the literature reports that it is negative and significantly smaller than one in absolute value.

This result has very important policy implications for tax policies and insurance premiums. The lower price elasticity of ESI will put upward pressure on insurance premiums in the future. To affect coverage, the tax subsidy policy will be three times less effective in Hawai‘i than on the mainland because of the lower price elasticity. These elasticities are of considerable practical importance to public health policy-makers because these are the basis for calculations of the effect of tax-subsidies on employer-provided health insurance. These elasticity estimates also provide interesting insights about insurance premiums in Hawai‘i. Premiums will increase more rapidly in Hawai‘i than in other states and the overall United States due to a significant difference in the price elasticity of demand for ESI. The Prepaid Health Care Act regulation has reduced the substitutability among other options of health insurance. Just as the environmental tax
on cars in Japan has pushed up car prices in Japan and pushed their insurance premiums much higher than the United States and other OECD countries.

The results also provide interesting insights into the role of demographic variables on the demand for health insurance. The results show that people who live on Oahu are more likely to be insured as compared to those who live on the neighboring islands because of the availability of medical services. Most of the hospitals and doctors are available on Oahu which makes health insurance more valuable for residents of Oahu. The sign of self-reported health status is opposite than what was expected, but it remains a strong predictor of the demand for health insurance.

Young children should have relatively moderate premiums, the middle-aged group should have the lowest premiums, and the elderly/disabled group should have the highest premiums since health expenditures have a U-shaped curve in age. The demand for health insurance is expected to increase because the proportion of elderly workers with comparatively high probabilities of sickness is increasing in Hawai‘i over time.

The tax subsidy serves as a powerful instrument to extend coverage for workers but not enough to ensure 100% coverage. The effectiveness of a tax subsidy is substantially reduced in Hawai‘i as compared to other selected states and the overall United States because of the PHCA. The effect of a tax subsidy heavily depends upon empirical model specifications. If the model specification does not correspond to the underlying
behavioral structure that derives the demand for health insurance, policy recommendations based on the estimation may not have the desired effects.

The results of the study also support a small positive income elasticity. The worker's income has a small positive effect on the probability of purchasing health insurance. Income elasticity of demand for ESI of full-time private workers is 0.02 for Hawai'i and 0.03 for the overall United States. The income elasticity of demand for ESI of exempted workers is 0.03 for Hawai'i and 0.05 for the overall United States. Obviously, it is higher for exempted workers in Hawai'i as well as in the overall United States, but the difference between these two estimates is relatively smaller.

The findings of the study also support the hypothesis that health insurance is a normal good and a rise in the tax subsidy increases the probability of purchasing ESI. The tax-subsidy is more attractive to high wage workers as compared to low wage workers because high wage workers pay higher marginal tax rates. By comparison, low wage workers benefit the most from the public health care programs because they are more likely to be eligible. The study shows that after-tax-price, income and other socio-economic determinants significantly influence the purchase of ESI. The demand for ESI is expected to increase in the future because the average age of workers is increasing over time. The tax-subsidy serves as a powerful tool to extend coverage for employees but is not sufficient to achieve 100% coverage in Hawai'i.
1.11. Bibliography


## 1.12. Appendix

### Table 1.1A: Probit regression estimates for the demand for health insurance in US

<table>
<thead>
<tr>
<th>Variables</th>
<th>coefficients (1)</th>
<th>dy/dx (2)</th>
<th>Std.Err (3)</th>
<th>Z (4)</th>
<th>P&gt;Z (5)</th>
<th>95% CI (6)</th>
<th>X bar (7)</th>
<th>X bar (8)</th>
<th>X bar (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>After-tax-price</td>
<td>-2.9375</td>
<td>-0.5842</td>
<td>0.0734 -40.02</td>
<td>0</td>
<td>-3.081</td>
<td>-2.794</td>
<td>0.643</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log income</td>
<td>0.1775</td>
<td>0.0353</td>
<td>0.0062 28.52</td>
<td>0</td>
<td>0.165</td>
<td>0.190</td>
<td>10.195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.0638</td>
<td>-0.0127</td>
<td>0.0020 -32.26</td>
<td>0</td>
<td>-0.068</td>
<td>-0.060</td>
<td>37.885</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age square</td>
<td>0.0009</td>
<td>0.0002</td>
<td>0.0000 36.34</td>
<td>0</td>
<td>0.001</td>
<td>0.001</td>
<td>1602.290</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College or higher</td>
<td>0.2499</td>
<td>0.0505</td>
<td>0.0077 32.43</td>
<td>0</td>
<td>0.235</td>
<td>0.265</td>
<td>0.549</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.1536</td>
<td>-0.0298</td>
<td>0.0081 -18.96</td>
<td>0</td>
<td>-0.169</td>
<td>-0.138</td>
<td>0.636</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently married</td>
<td>0.4027</td>
<td>0.0776</td>
<td>0.0087 46.29</td>
<td>0</td>
<td>0.386</td>
<td>0.420</td>
<td>0.427</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-white race</td>
<td>-0.2459</td>
<td>-0.0538</td>
<td>0.0099 -24.82</td>
<td>0</td>
<td>-0.265</td>
<td>-0.226</td>
<td>0.160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time worker</td>
<td>0.0843</td>
<td>0.0173</td>
<td>0.0093 9.1</td>
<td>0</td>
<td>0.066</td>
<td>0.102</td>
<td>0.802</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worker having children</td>
<td>0.1610</td>
<td>0.0316</td>
<td>0.0083 19.46</td>
<td>0</td>
<td>0.145</td>
<td>0.177</td>
<td>0.419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honolulu County</td>
<td>0.5093</td>
<td>0.0736</td>
<td>0.0469 10.87</td>
<td>0</td>
<td>0.417</td>
<td>0.601</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Union member</td>
<td>0.4585</td>
<td>0.0694</td>
<td>0.0295 15.54</td>
<td>0</td>
<td>0.401</td>
<td>0.516</td>
<td>0.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm size 1 - 9</td>
<td>-0.6394</td>
<td>-0.1632</td>
<td>0.0107 -59.97</td>
<td>0</td>
<td>-0.660</td>
<td>-0.619</td>
<td>0.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firm size 10 - 499</td>
<td>-0.3030</td>
<td>-0.0620</td>
<td>0.0080 -37.95</td>
<td>0</td>
<td>-0.319</td>
<td>-0.287</td>
<td>0.423</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible for PHI</td>
<td>-0.4976</td>
<td>-0.1274</td>
<td>0.0211 -23.63</td>
<td>0</td>
<td>-0.539</td>
<td>-0.456</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White-colar worker</td>
<td>0.2264</td>
<td>0.0447</td>
<td>0.0084 27.1</td>
<td>0</td>
<td>0.210</td>
<td>0.243</td>
<td>0.473</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.7960</td>
<td>0.0888</td>
<td>20.22 1.622</td>
<td>0</td>
<td>1.622</td>
<td>1.970</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: industry and year dummies are included as control variables.
### Table 1.2A: Probit regression estimates for the demand for health insurance in US

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients (1)</th>
<th>dy/dx (2)</th>
<th>Std.Err (3)</th>
<th>Z (4)</th>
<th>P&gt;Z (5)</th>
<th>95% CI</th>
<th>X bar (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>After-tax-price</td>
<td>-2.8587</td>
<td>-0.5649</td>
<td>0.0737</td>
<td>-38.77</td>
<td>0</td>
<td>-3.003</td>
<td>-2.714</td>
</tr>
<tr>
<td>Log income</td>
<td>0.1715</td>
<td>0.0339</td>
<td>0.0062</td>
<td>27.45</td>
<td>0</td>
<td>0.184</td>
<td>0.159</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0600</td>
<td>-0.0119</td>
<td>0.0020</td>
<td>-30.27</td>
<td>0</td>
<td>-0.064</td>
<td>-0.056</td>
</tr>
<tr>
<td>Age square</td>
<td>0.0009</td>
<td>0.0002</td>
<td>0.0000</td>
<td>35.38</td>
<td>0</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>College or higher</td>
<td>0.2309</td>
<td>0.0463</td>
<td>0.0077</td>
<td>29.84</td>
<td>0</td>
<td>0.216</td>
<td>0.246</td>
</tr>
<tr>
<td>Male</td>
<td>-0.1599</td>
<td>-0.0308</td>
<td>0.0081</td>
<td>-19.69</td>
<td>0</td>
<td>-0.176</td>
<td>-0.144</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.3958</td>
<td>0.0758</td>
<td>0.0087</td>
<td>45.39</td>
<td>0</td>
<td>0.379</td>
<td>0.413</td>
</tr>
<tr>
<td>Non-white race</td>
<td>-0.2314</td>
<td>-0.0501</td>
<td>0.0099</td>
<td>-23.28</td>
<td>0</td>
<td>-0.251</td>
<td>-0.212</td>
</tr>
<tr>
<td>Full-time worker</td>
<td>0.0779</td>
<td>0.0158</td>
<td>0.0093</td>
<td>8.4</td>
<td>0</td>
<td>0.060</td>
<td>0.096</td>
</tr>
<tr>
<td>Worker having children</td>
<td>0.1592</td>
<td>0.0310</td>
<td>0.0083</td>
<td>19.2</td>
<td>0</td>
<td>0.143</td>
<td>0.175</td>
</tr>
<tr>
<td>Honolulu County</td>
<td>0.5204</td>
<td>0.0741</td>
<td>0.0470</td>
<td>11.08</td>
<td>0</td>
<td>0.428</td>
<td>0.612</td>
</tr>
<tr>
<td>Union member</td>
<td>0.4538</td>
<td>0.0684</td>
<td>0.0295</td>
<td>15.37</td>
<td>0</td>
<td>0.396</td>
<td>0.512</td>
</tr>
<tr>
<td>Firm size 1 - 9</td>
<td>-0.6431</td>
<td>-0.1635</td>
<td>0.0107</td>
<td>-60.21</td>
<td>0</td>
<td>-0.664</td>
<td>-0.622</td>
</tr>
<tr>
<td>Firm size 10 - 499</td>
<td>-0.3049</td>
<td>-0.0620</td>
<td>0.0080</td>
<td>-38.09</td>
<td>0</td>
<td>-0.321</td>
<td>-0.289</td>
</tr>
<tr>
<td>Eligible for PHI</td>
<td>-0.4875</td>
<td>-0.1236</td>
<td>0.0211</td>
<td>-23.1</td>
<td>0</td>
<td>-0.529</td>
<td>-0.446</td>
</tr>
<tr>
<td>White-colar worker</td>
<td>0.2182</td>
<td>0.0429</td>
<td>0.0084</td>
<td>26.06</td>
<td>0</td>
<td>0.202</td>
<td>0.235</td>
</tr>
<tr>
<td>Very good health</td>
<td>-0.1022</td>
<td>-0.0205</td>
<td>0.0090</td>
<td>-11.41</td>
<td>0</td>
<td>-0.120</td>
<td>-0.085</td>
</tr>
<tr>
<td>Good health</td>
<td>-0.2315</td>
<td>-0.0492</td>
<td>0.0099</td>
<td>-23.35</td>
<td>0</td>
<td>-0.251</td>
<td>-0.212</td>
</tr>
<tr>
<td>Poor health</td>
<td>-0.2806</td>
<td>-0.0638</td>
<td>0.0156</td>
<td>-18</td>
<td>0</td>
<td>-0.311</td>
<td>-0.250</td>
</tr>
<tr>
<td>Constant</td>
<td>1.8289</td>
<td>0.0891</td>
<td>20.53</td>
<td>0</td>
<td>1.654</td>
<td>2.004</td>
<td></td>
</tr>
</tbody>
</table>

Note: Industry and year dummies are included as control variables.

### Stata Program

/* Chapter 1: Employer-sponsored Health Insurance*/
clear
log using c:\cps\chapter1\privateemp.log, replace
set more 1
set mem 1000m
#delimit;

use c:\cps\nbermar9403drop.dta, clear;
keep id recnum mcare mchm champ covhi chhi covgh health ernush incern taxinc faminc emcont p state msafp cslyr fnumper
  earner emlzsz mrgtax year age sex race _ educ grdatn marstat income hhseq cntyfp
  unmem wgt indmly ocmly nberf taxp
  nbertaxrate nberfica taxp;
keep if year>=1995;
/*

mvdecode _all, mv(-1);
mvdecode _all, mv(-2);
mvdecode _all, mv(-3);
mvdecode _all, mv(-4);
mvdecode _all, mv(-9);
*/

/*Insured=1 Uninsured=0*/
gen public= (mcare==1 | mcaid==1 | chmc==1 | champ==1);
gen private= (covhi==1 | chhi==1 | chhi==2);
gen insured= (public==1 | private==1);
gen uninsured=insured==0;

gen empbased=(covgh==1 | chhi==2);
gen hawaii=state==95;
gen navada=state==88;
gen michigan=state==34;
gen california=state==93;
gen florida=state==59;

gen year95=year==1995;
gen year96=year==1996;
gen year97=year==1997;
gen year98=year==1998;
gen year99=year==1999;
gen year00=year==2000;
gen year01=year==2001;
gen year02=year==2002;
gen year03=year==2003;

gen female=(sex==2);
gen male=sex==1;
gen white=race==1;
gen nonwhite=race==1;

gen element=_educ<=6 if year==1990 | year==1991;
replace element=grdatn<=33 if year==1992;
gen middle=( _educ==7 | _educ==8 | _educ==9 ) if year==1990 | year==1991;
replace middle=(grdatn==34 | grdatn==35) if year==1992;
gen leshigh=(element==1 | middle==1);
gen high=( _educ==10 | _educ==11 | _educ==12 | _educ==13 ) if year==1990 | year==1991;
replace high=(grdatn==36 | grdatn==37 | grdatn==38 | grdatn==39) if year==1992;
gen somecol=( _educ==14 | _educ==15 | _educ==16 ) if year==1990 | year==1991;
replace somecol=(grdatn==40 | grdatn==41 | grdatn==42) if year==1992;
gen collegep=( _educ==17 | _educ==18 & _educ==.) if year==1990 | year==1991;
replace collegep=(grdatn==43 | grdatn==44 | grdatn==45 | grdatn==46) if year>=1992;
gen college=(somecol==1 | collegep==1);

gen manage=occmly==1;
gen pro=occmly==2;
gen tech=occmly==8 if year==2003;
replace tech=occmly==3 if year<=2002;
gen sales=occmly==4 if year==2003;
replace sales=occmly==4 if year<=2002;
gen clerk=occmly==5 if year==2003;
replace clerk=occmly==5 if year <=2002;
gen service=occmly==3 if year==2003;
replace service=(occmly==6 | occmly==7 | occmly==8) if year <=2002;
gen labor=occmly==10 if year==2003;
replace labor=occmly==11 if year<=2002;
gen whitecol=(pro==1 | manage==1 | clerk==1 | sales==1);

gen privatemp=clslyr==1;
recode clslyr 1/4=1;
recode clslyr 5 6 =2;
*self-employed if clslyr==2;
gen selfemp=clslyr==2;
gen employee=clslyr==1;

gen married=(marstat==1 | marstat==2 | marstat==3);

gen child=age<19;
gen adult=(age>=19 & age<=64);

sort year hhseq;
by year hhseq: gen sumchild=sum(child);
by year hhseq: gen totchild=sumchild[_N];
gen children=totchild>0;
gen partime=ernush<=19 & ernush>=1;
gen fultime=ernush<=99 & ernush>=20;
gen unemp=ernush<=0;

recode health 4 5 =4;
gen health1=health==1;
gen health2=health==2;
gen health3=health==3;
gen health4=health==4;

gen honolulu=(state==95 & msafp==3320);
gen union=unmem==1;
gen emplsz9=emplsz==1;
gen emplsz499=(emplsz==2 | emplsz==3 | emplsz==4);
gen emplsz500p=(emplsz==5 | emplsz==6);
gen malemarried=male*married;
gen y=income;
gen fpl=faminc/p;
gen eligible=(fpl<=1);
gen nwgt=wgt/100;
gen agesq=age^2;
gen lny=ln(y);
gen tst=nberstaxrate/100;
gen tf=mrgtax/100;
gen tsm=;
replace tsm=0.0765 if incem>O & incem<82000;
replace tsm=0.0145 if incem>82000;
gen tp=(1-tsm-tf-tst)/(1+tsm);
su empbased tst tf tsm tp if hawaii==1;
su empbased tst tf tsm tp if navada==1;
su empbased tst tf tsm tp if michigan==1;
su empbased tst tf tsm tp if california==1;
su empbased tst tf tsm tp if florida==1;
keep if privatemp==1;
su insured tp lny age agesq college male married nonwhite fultime children 1 honolulu
union emplsz9 emplsz499
  emplsz500p selfemp eligible whitecol health1 health2 health3 health4 tsm tf tst
[aw=nwgt]
  if hawaii==1 & incem>0 & tp>0 & tp==.

su insured tp lny age agesq college male married nonwhite fultime children 1 honolulu
union emplsz9 emplsz499
  emplsz500p selfemp eligible whitecol health1 health2 health3 health4 tsm tf tst
[aw=nwgt] if incem>0;

xi:probit insured tp lny age agesq college male married nonwhite fultime children 1 honolulu
union emplsz9
emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00
year01 year02 year03 [pw=nwgt]
  if hawaii==1 & incem>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fultime children 1 honolulu
union emplsz9
emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00
year01 year02 year03 [pw=nwgt]
  if hawaii==1 & incem>0, robust cluster(hhseq);
xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
  emplsz499 selfemp eligible whitecol i.indmly i.health year96 year97 year98 year99
year00 year01 year02 year03
  [pw=nwgt] if hawaii==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
  emplsz499 selfemp eligible whitecol i.indmly i.health year96 year97 year98 year99
year00 year01 year02 year03
  [pw=nwgt] if hawaii==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
  emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00
year01 year02 year03 [pw=nwgt]
  if navada==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
  emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00
year01 year02 year03 [pw=nwgt]
  if navada==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
  emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00
year01 year02 year03 [pw=nwgt]
  if michigan==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
xi:probit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if michigan==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if michigan==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if california==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if california==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if california==1 & incern>0, robust cluster(hhseq);

xi:dprobit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if california==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children honolulu union emplsz9
   emplsz499 selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if florida==1 & incern>0, robust cluster(hhseq);
xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if florida==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if florida==1 & incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xi:probit insured tp lny age agesq college male married nonwhite fulltime children
honolulu union emplsz9
emplsz499 selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
if incern>0, robust cluster(hhseq);

xilogit insured tp lny age agesq college male married nonwhite fulltime children
honolulu emplsz
selfemp eligible whitecol i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if hawaii==1, cluster(hhseq);
mfx compute if e(sample);

xi:logit insured tp lny age agesq college male married nonwhite fultime totchild honolulu union emplsz
   selfemp eligible whitecol i.health i.indmly year96 year97 year98 year99 year00 year01 year02 year03 [pw=nwgt]
   if hawaii==1, cluster(hhseq);
mfx compute if e(sample);

fitstat;
prvalue if e(sample), x(age=0 agesq=0) rest(mean);
prvalue if e(sample), x(age=1 agesq=1) rest(mean);
prvalue if e(sample), x(age=5 agesq=25) rest(mean);
prvalue if e(sample), x(age=10 agesq=100) rest(mean);
prvalue if e(sample), x(age=15 agesq=225) rest(mean);
prvalue if e(sample), x(age=20 agesq=400) rest(mean);
prvalue if e(sample), x(age=25 agesq=625) rest(mean);
prvalue if e(sample), x(age=30 agesq=900) rest(mean);
prvalue if e(sample), x(age=35 agesq=1225) rest(mean);
prvalue if e(sample), x(age=40 agesq=1600) rest(mean);
prvalue if e(sample), x(age=45 agesq=2025) rest(mean);
prvalue if e(sample), x(age=50 agesq=2500) rest(mean);
prvalue if e(sample), x(age=55 agesq=3025) rest(mean);
prvalue if e(sample), x(age=60 agesq=3600) rest(mean);
prvalue if e(sample), x(age=65 agesq=4225) rest(mean);
prvalue if e(sample), x(age=70 agesq=4900) rest(mean);
prvalue if e(sample), x(age=75 agesq=5625) rest(mean);
prvalue if e(sample), x(age=80 agesq=6400) rest(mean);
prvalue if e(sample), x(age=85 agesq=7225) rest(mean);
prvalue if e(sample), x(age=90 agesq=8100) rest(mean);
prvalue if e(sample), x(age=95 agesq=9025) rest(mean);

col;
recode health 1 2 3 =1;
recode health 4 5=2;
gen heal=health==1;

xi: heckprob insured tp lny age agesq high college male married nonwhite fultime totchild honolulu union emplsz
whitecol selfemp i.indmly year95 year96 year97 year98 year99 year00 year01 year02 year03 [iw=nwgt] if hawaii==1,
    select (heal = lny age agesq high college male) cluster(hhseq);

    gen ind1=indmly==1;
    gen ind2=indmly==2;
    gen ind3=indmly==3;
    gen ind4=indmly==4;
    gen ind5=indmly==5;
    gen ind6=indmly==6;
    gen ind7=indmly==7;
    gen ind8=indmly==8;
    gen ind9=indmly==9;
    gen ind10=indmly==10;
    gen ind11=indmly==11;
    gen ind12=indmly==12;
    gen ind13=indmly==13;
    gen ind14=indmly==14;
    gen ind15=indmly==15;

    biprobit insured heal ( tp lny age agesq high college male married nonwhite fulltime
totchild honolulu union emplsz
    selfemp whitecol ind1 ind2 ind3 ind4 ind5 ind6 ind7 ind8 ind9 ind10 ind11 ind12 ind13 ind14) [iw=nwgt] if hawaii==1;

*end of file;
Essay Two

Demand for Health Care in Hawai'i

Abstract
The demand for physician visits in Hawai'i is estimated using the Medical Expenditure Panel Survey (MEPS) and the National Health Interview Survey (NHIS) from 1996 to 2002. The effects of health insurance, age, and cost sharing on the demand for physician visits in Hawai'i are examined. The results of the study show that the demand for physician visits is slightly higher in Hawai'i (i.e., 3.51) than in the United States (i.e., 3.42) because of higher coverage and because it has larger proportion of people 65 and older than nationwide. Higher demand for health care will put upward pressure on the cost of health care in Hawai'i.

Health insurance is a very significant determinant of the demand for physician visits. Covered people make more frequent visits to physicians than the uninsured because they have lower out-of-pocket costs. The number of physician visits responds negatively to changes in the amount paid out-of-pocket. Reductions in the level of cost sharing increase the demand for physician visits. Three alternative models are used to estimate the pure price response: the Poisson model, the Negative Binomial model and the Multinomial Logit model.

These three models suggest that price elasticities for physician visits are in the -0.1 to -0.2 range. These values are consistent with those in the lower range of the non-experimental literature which vary from -0.1 to -2.1. The RAND experimental study reported that elasticity estimate is -0.2. The magnitude of the price elasticity of physician
visits should impact the generosity of health insurance. If the price elasticity is higher, then the insurance coverage should be less generous. These results also show that increased family income positively affects physician visits. Depending on the insurance plan, people with higher incomes have 20 percent to 40 percent more physician visits than low-income people. The main reason for the rise in health care costs is not increased life expectancy but the invention of expensive medical technology, surgical procedures and blockbuster pills.
2.1. Introduction
The purpose of this chapter is to uncover important determinants of the demand for health care in Hawai'i. The cost of health care has increased to the point where it has created financial hardship for many individuals and families who would like to purchase health care services. Expenditures on health care have been growing at an average rate of 6.4% per year for the last four decades, which is outpacing inflation and the growth of family income. This trend is expected to continue for at least a few more decades because as people age, they use increasingly more health care resources. Hawai'i has the nation’s largest percentage of elderly residents. Therefore, this rise in the cost of health care is expected to be more severe in Hawai'i than in other states. Currently, 15.2% of gross national product is spent on health care as compared to 13.4% in 1994, 9.2% in the 1980s, 6% in the 1960s, 4% in the 1920s, and 2% in the 1900s (Department of Labor, various reports). Health expenditures reached an average of $5440 per person in 2002, up $419 from 2001. Health care spending is expected to comprise 18.4% of GDP in 2014. The most commonly cited reasons for this rapid increase in health expenditures are an increase in the demand for health care due to the aging of the population and the expending technological possibilities in the health care service sector. This study analyzes the effect of the aging population on the demand for health care.

The demand for health care is increasing because the proportion of the population that is elderly is steadily growing over time. It is common wisdom that as society gets older its health stock deteriorates faster and society needs more health care. The proportion of the population aged 65 and over has grown steadily from 8.1% in 1950 to 11.3% in 1980 and
13% in 2000. Presently, Hawai‘i has the nation’s largest proportion of elderly people. It is expected to rise to 18% in 2020 (DBEDT, 2001). On the other hand, the proportion of the population under the age of 5 has steadily fallen from 10.8% in 1950 to 6.3% in 2000, with the exception of the baby boom growth period during the 1950s and 1960s. Therefore, during the past two decades the gap between the proportion of the population aged 65 and over and the younger population has widened substantially. This intergenerational shift in the age structure has significantly changed in the past few decades. Per capita health care spending for people age 65 or older tends to average three times that for younger people. It is widely believed among health economists that the aging of the population is a major driver of the annual growth in the demand for health care in the United States. Until a century or so ago, retirement was a rare phenomenon virtually all workers died while still in the labor force. Now almost all workers retire in their 50’s or 60’s and die in their 80’s. A remarkable rate of increase in life expectancy since 1950, due to rapid innovations and improvements in medical technology, has also significantly increased the demand for health care. All these factors add to medical costs.

Over the past few decades, a rapid increase in the demand for health care has made it financially difficult for the current health care system to meet all of its obligations. The impending retirement of the baby boom generation will make it even harder for the present health care system to sustain itself in the future. In response to these growing health care costs, policy makers are being confronted with the choice of whether to reduce health care benefits or to increase the health care taxes borne by younger workers. In 1972, the Federal government amended the Social Security Act that allowed states to
require nominal co-payments on optional services. In 1982 the cost sharing ability of the states was extended under the Tax Equity and Fiscal Responsibility Act. The effect of these policy proposals depends critically on the price and income elasticities of demand for health care. If the price elasticity of the demand for health care is large, then an increase in the price of health care will significantly reduce the demand for health care. It also tells us that a price increase on the part of the provider would result in a net decrease in their revenues. On the other hand, if the price elasticity of the demand for health care is small, then an increase in the price of health care will insignificantly reduce the demand for health care and increase the net revenues of providers. Similarly, income elasticity of demand for health care provides information about demand responses by different income groups. If the income elasticity of demand for health care is high, an increase in cost will lead to an undesirably large decrease in the demand for health care by low-income households. A high income elasticity of demand for health care also means that the share of expenditures on health care will increase as family income grows.

So far, there is no study that has estimated the elasticities of the demand for health care in Hawai'i with respect to its price and the consumer's income. However, there are a few studies that have estimated the price elasticity of the demand for health care in the United States and other countries. A wide variety of empirical methods and data sources are used for the estimation of the price elasticity of demand for health care in different countries. These studies exhibit a wide variety of price and income elasticities estimates. The price and income elasticities of demand for health care in the United States range from -0.1 to -1.8, and from 0 to 1.6 respectively, which are much too wide to provide reliable guidance
to policy-makers in assessing the effectiveness of alternative public policy proposals for health care financing in Hawai‘i. The estimates are also useful for predicting how various policy changes relating to price and income are likely to affect the amount of health care demanded by the population of Hawai‘i. This study will provide more reliable and accurate estimates of price and income elasticities of the demand for health care in Hawai‘i.

The consumer’s demand for health care provided by general practitioner doctors is estimated in this chapter by using the Medical expenditures survey data for the years 1996-2002. The price and income elasticities of the demand for health care are calculated. Finally, it will examine other factors that affect the demand for health care. The study will pay special attention to selectivity and endogeneity problems. The demand for other health care such as hospital stays, dental care, preventive care and ambulatory services are not studied. This choice is made in order to retain a relatively homogeneous type of service which is visits to the doctor.

The remainder of essay two has six sections which are organized as follows. The next section summarizes the previous literature on the demand for health care. It is organized in a step-by-step progression from the simplest models to the most complex. Following that, a section presents a simple theoretical framework and derives the demand for health care that underlies the entire essay two. The third section specifies the empirical model, discusses the expected results and the estimation procedure. It also provides definitions for variables used in the analysis. The next section contains the main empirical results of
this essay. The second-to-last section of this essay compares the results of this study with other related studies. In the last section, broader conclusions are drawn from the empirical results.

2.2. Literature Review
This section will provide an overview of the empirical literature on the price and income elasticities of the demand for health care. There are several studies that have estimated the demand for health care in the United States and other countries. There are a few studies that have estimated the demand for different states of the United States, such as Feldstein (1973), Fuchs and Kramer (1972). However, there is no study that has estimated the demand for health care in Hawai‘i.

The majority of health care studies have used three types of data: experimental, quasi-experimental and observational survey data. This study uses observational data for Hawai‘i from MEPS and NHIS. Previous health care studies have used numerous models and methods to estimate the demand for health care. Most of these studies have used a model-based approach to estimate the demand for health care, while some studies used only a descriptive approach. The most commonly used models are: linear probability models, binomial probit models, binomial logit models, multinomial probit models, multinomial logit models, nested multinomial logit models, Poisson models, Negative Binomial models, one-part models, two-part models, and latent models. This study uses the Poisson, Negative Binomial and multinomial logit models.
In the literature, the demand for health care is measured in either physical or monetary units. The physical units are the number of physician visits made by the consumer during a particular time period. Most of these studies used demand for health care in physical units, such as Cameron, et al (1988), Pohlmeier and Ulrich (1995), Deb and Trivedi (1997), Munkin and Trivedi (2003) and Gerdtham (1997). The monetary unit is the share of medical expenditures per unit of time. Other studies have measured the demand for health care in monetary units, such as Duan et al (1983), Manning et al (1987), Keeler et al (1988), and McCall et al (1991). This study will measure the demand for health care in physical units.

These studies exhibit a wide variety of price elasticity estimates which range from -0.04 to -2.0. The price elasticity of the demand for health care is low for basic health care services but higher for certain types of preventive care services. This is not surprising because the number of available substitutes for preventive care is greater than for basic care. As the price of preventive care increases, consumers substitute nutritional supplements and healthy foods, for example. In addition, preventive care may be considered as more of a luxury than a necessity. Therefore, individuals may reduce their demand for preventive care as the price of such care increases.

Estimates of the income elasticity of demand for health care are in the range from 0 to 1.8. The positive sign indicates that as income increases the demand for health care services will also increase. There is a great amount of literature on the income elasticity of the demand for health care. Theoretically, the effect of income on the demand for
health care should be positive but small in magnitude. If consumers have access to free health care, changes in income should not affect their ability to obtain health care. The empirical estimates in the literature are consistent with this theory. Studies based on time series data tend to report higher income elasticities while studies based on cross-sectional data tend to report lower income elasticities. The larger elasticities from time series studies reflect the incorporation of the effects of technical change in the field of health care. Some studies have used experimental data to estimate income elasticity, but their findings are consistent with observational studies. For details see Taylor and Wilensky (1983) and Holmer (1984).

Phelps and Newhouse (1974) were the first to estimate the demand for health care using microeconomic data. Their estimates were based on the 1963 United States Household Survey data. They extended Grossman's (1972) model to estimate demand equations for hospital days and physician's visits separately. They used coinsurance rates as a source of price variation in both equations. They used the OLS and 2SLS methods to estimate demand, hospital days and physicians' visits. They found price and income elasticities of demand for both are below -0.2 and 0.6 respectively, which is very small relative to other studies. However, their study is outdated, whereas this study will use the most recent data, the 2001 National Health Interview Survey.

Heller (1982) also estimated the demand for health care. He used the number of consultations as a measure for the amount of health care demanded. Musgrove (1983) used medical expenditures as a measure for the amount of health care demanded. Gertler
et al (1987) estimated the demand for health care for different income groups and found that it is highly elastic with respect to price, but this elasticity decreases as income increases. Phelps and Newhouse (1974) used coinsurance rates as a source of variation and estimated a price elasticity of -0.12. Scitovsky and Snyder (1972) used experimental data to estimate the physician's visit price elasticity of demand at -0.14 which is much lower than -1 which was reported by Davis and Russell (1972). Sahn, Younger and Genicot (2003) used data for rural Tanzania and found that own-price elasticity for private clinics is -1.69, which is much higher than reported by previous studies. Rosett and Huang (1973) used individual-level observational data for the year 1960 to estimate a tobit regression for physician's visits. They used out-of-pocket costs as a source of price variation. They reported an overall elasticity of demand for health care equal to -1.5, which is significantly larger than other studies. Wedig (1988) used observational data from the 1980 National Medical Care Utilization and Expenditure Survey. He used out-of-pocket costs as the source of price variation and reported the physician's visit price elasticity of demand for health care ranges from -0.16 to -0.35, depending on health status. People with fair or poor health were found less price responsive than those who reported their health to be good or excellent. The latest studies show that the price elasticity for physician's visits is much higher than reported by studies conducted during the 1970s. The wide range of these estimates make it difficult for policy-makers to rely on any single estimate.

Manning et al (1987) also estimated the demand for health care using experimental data. He used coinsurance as the price of health care in his model. He estimated the share of
medical expenditures as a dependent variable. He found that the price elasticity of the demand for health care is in the range from -0.17 to -0.22. This means a 100% increase in user price will cause the demand to fall between 17% and 22% depending upon the elasticity. These estimates are consistent with other studies which used actual data. His study found that introducing cost-sharing by coinsurance has reduced the utilization of health care relative to free care at the point of delivery.

Bitran and McInnes (1993) estimated the demand for health care in the Dominican Republic and El Salvador. They used household-level micro data collected by the Health Care Financing department of Latin America for the year 1989. They estimated the price elasticity of demand for health care is -0.5. They used descriptive analysis and multivariate analysis approaches to calculate the price and income elasticities for both countries. Sahn, et al. (2003) used observational micro data to estimate the demand for health care in rural Tanzania. They used the multinomial logit model.

Acton (1975, 1976) used sample data from England for the year 1996 to estimate the number of visits for various types of care. These services are funded by the public tax system and have no monetary price at the point of delivery. Therefore, he used travel and waiting time as a proxy for the price of health care. Average waiting time for cataract surgery is 245 days in England. He estimated time-price elasticity for private physician visits as -0.25 and outpatient visits (in the hospital) as -0.958. Coffey (1983) and Mueller and Monheit (1988) also reported a time-price elasticity for health care of -0.1, which is much lower than that reported by Acton. Janssen (1992) used data from the National
Health Interview Survey from the Netherlands' Central Bureau of Statistics to estimate the demand for health care. He found time-price elasticity in the range of -0.09 to -0.14. He found that the time-price elasticity of the demand for health care is negative and small. This suggests that people are not very sensitive to changes in time-prices. These studies used data for publicly provided health care which has no monetary price at the point of delivery. In the United States this care is provided through the market. Therefore, the data on the out-of-pocket expenditures for the doctor’s visit is available. This study will use it as a monetary price for this service.

2.3 How Health Insurance Affects Demand for Health Care

A special factor that complicates analysis of the demand for medical care is health insurance. Once a person is covered by health insurance, the effective price to consumers to consume medical care is substantially lower than what the providers receive at the time of the service. The difference is paid by the insurer. Even though the premium charged by the insurance company will eventually recover all of the costs of that insurance, including medical care purchased through the plan, the net result of insurance still leads to a lower relevant price for decision-making by the consumer. Consumers pay a lower price for health care at the time of purchase than what the sellers receive. This lower price provides an incentive to consumers to consume more health care than they would if they had to pay the full price. The providers of this health care are paid full price, however, receiving out-of-pocket payments from the patient and reimbursement from the insurer. The consequence of this system is that the quantity of health care demanded will increase.
The effect of insurance on the demand for health care can be explained by Figure 2.1. The equilibrium price and quantity of health care demanded are $P_e$ and $Q_e$ respectively. Once a person purchases health insurance, the consumer price at the point of service will decrease. The decrease in consumer price at the delivery point depends upon the type of coverage he has. If the consumer has comprehensive coverage, then the consumer price at the service point will be zero. The quantity demanded for health care will increase from $Q_e$ to $Q_1$ after the purchase of insurance. The increase in demand for health care from $Q_e$ to $Q_1$ is due to the consumer moral hazard problem. Consumer moral hazard is a change in the attitude of the consumer to over-consume the quantity of health care due to insurance against the full costs of such care. The increase in quantity demanded of health care due to moral hazard depends in part on the difference between the cost of the care and the value individuals place on that care. The value consumers place on the care is measured by the vertical distance below the demand curve which is less than the full cost of that care, $P_e$. If this true, insurance causes a loss in consumer welfare because of the
insurance-induced increase in consumption, whose costs are passed back to the consumer in the form of higher insurance premiums. The welfare loss is thus incurred by the enrollee at the time the insurance premium is paid, not when care is received. The welfare loss is shown by the shaded area ABQ₁ in the Figure 2.1. More welfare to society could be obtained by shifting the resources used by this excess demand to other alternative uses (Pauly, 1968).

This welfare loss can be reduced by introducing co-payments or coinsurance. If the consumer has to pay C proportion of the price of health care as a co-payment, then the consumer’s price will become CP, not zero as in the case of full coverage. He will consume health care Q₂ which is less than Q₁ but still larger than Qₑ. As C approaches zero (full coverage), the welfare loss caused by the over-purchase of medical care increases more and more. The welfare loss also depends upon the elasticity of demand for health care for any given level of insurance coverage. If the demand for health care is more elastic then the welfare loss is larger for any given level of price change and vice versa. Therefore, it is less desirable for society to provide health insurance.

Provider moral hazard might also exist in the case of fee-for-service remuneration methods because doctors have a financial incentive to provide care in excess of that which would be provided by trading with fully informed consumers (Evans, 1974). This is known as supplier-induced demand. It will shift the demand for health care outward. This is discussed later.
The moral hazard problem can be controlled to some extent by introducing user charges or coinsurance. It can also be controlled by government intervention in the health care market. This is why the health care market is subject to many regulations. Some of these policies have attempted to substitute regulation for competition by regulating entry and investment or price. The optimal coinsurance policy will make patients pay for care up to the point where the marginal gains from less risk-sharing are just offset by the marginal benefits from reduced provision of low-valued care.

Evidence on consumer moral hazard is observed in many countries. This is why a number of countries have introduced co-payments that put some financial burden on the consumer in order to discourage the unnecessary use of health care or doctor’s time. Some major empirical studies that estimated the effect of moral hazard on health care were conducted by Fuchs (1996), Feldstein (1973), Feldman and Dowd (1991) and Manning and Marquis (1996). They found that as the level of cost-sharing (in this case the co-payment) increases, the demand for health care become more responsive to the price. These estimates have convinced policy makers to increase coinsurance rate and deductibles to control the moral hazard problem.

2.4. Conceptual Framework
In this section, the economic behavior of a representative worker is explained regarding his health care demand. Demand for a doctor’s visit is observed by the number of visits an individual is willing to obtain in cases of need as a function of health care price, the characteristics of the individual (for example, perception of need, income, location, insurance coverage), and the characteristics of the providers (for example, prices,
location, quality). The worker's and provider's characteristics determine the health care utilization pattern. Assuming other factors that influence the demand for a doctor's visit are constant, the demand for health care is sensitive to the out-of-pocket price. Other factors that also might affect the demand for health are used as controls variables such as income, age, health status, education, region etc. In this section, an individual's demand for health care is derived by maximizing his utility function subject to his budget constraint. The individual making the decision is a globally risk-averse person whose welfare depends upon his health status (H) and the consumption of all other goods and services (X). The socio-economic factors and personal characteristics of the worker (Z) also affect his utility.

Mathematically, the utility function can be written as: $U = U(X, H; Z)$. The key goal of health care is to improve health level. The health level improves as more and more units of health care (M) are consumed. An individual spends his income (Y) on medical care (M) and all other goods and services (X). His budget constraint can be written as: $Y \geq P_x X + P_m M$. Where $P_m$ is the out-of-pocket payment to the doctor, which is considered to be the price of health care, and $P_x$ is the price of non-health goods and services.

The demand for health care is shown by how much health care the worker will consume at different prices. Holding everything else constant, the amount of medical care, M, is inversely related to the price of health care, $P_m$. This negative relationship exists mainly due to two reasons. First, the marginal productivity of health care is diminishing in producing health. Second, the marginal utility of health is diminishing in producing
utility. The probability of an individual i's demand for medical care can be obtained as follows:

$$\text{Prob}_i = f (\text{Price of health care, Income, Illness level, other socioeconomic factors}) \ldots (2.1).$$

This probability can be used to obtain aggregate projections of resource needs at the county or state level. Once the probabilities are known for both groups of the population, then the demand for care can be obtained as follows:

Total Demand for Doctor's Visit = (population of group i)* (Probability of illness)*Prob$_i$.

For example, if the size the Hawaiian population is 1.2 million, the proportion of frequent and infrequent users are .33 and .36, respectively, their annual illness and injury rate is 2.5 cases per capita, and the probability of seeking care is 40 percent for infrequent users and 90 percent for frequent users, then the annual quantity demanded of doctor's visits will be $(1.2 \times 0.33 \times 2.5 \times 0.4) + (1.2 \times 0.36 \times 2.5 \times 0.9).$ These estimates can help policymakers find the number of doctors required to satisfy the demands of the community for health care.

2.5. Determinants of the Demand for Health Care

The demand for health care is the quantity of health care that an individual is willing to obtain at different price levels. It is determined by a number of factors, such as the price of health care, the income of the individual, insurance coverage, health status, location, illness, waiting time, quality of health care and population size, etc. The price of health care is the main determinant of demand that deserves special attention. If all other factors influencing demand are held constant, then price will be negatively related to the quantity demanded because a rational individual prefers to pay less rather than more for a particularly quantity of health care. The quality of health care has a positive effect on the
demand for health care. Similarly, people prefer to wait less time rather than more to see a particular health care provider, so the length of time the user waits before being seen by the health professional also influences the demand for health care. The travel time and waiting time at the health facility might also affect the demand for health care. The greater the distance the patient has to travel, the less will be the demand for health care. Therefore, an increase in either the cash price or the time cost of medical services should reduce the amount of health care demand.

The second most important factor in the determination of the demand for health care is the worker’s income. A rise in income is expected to cause an increase in the consumption of most goods, including health care. An increase in hourly wage rate increases an individual’s income but also increases the opportunity cost of time. The net effect of wage increases on the demand for medical care depends on the relative importance of the income or substitution effects. Also, workers with higher incomes often have better health insurance coverage than those with lower incomes. This implies that more income means more demand for health care. On the other hand, workers with higher incomes get sick less often than low-income people because they can afford healthier food and a better living environment. However, income and health status are positively correlated. It leads to less health care use. These two effects work in opposite directions and weaken the overall effect of income. Therefore, the effect of changes in income on the demand for health care cannot be accurately predicted in advance. In general, health is a normal good and the worker’s income has a positive effect on the demand for health care. Heller (1976), like many others, found that an increase in income
does not significantly alter the demand for health care. The income variable is included in the regression after taking its natural logarithm.

There are too many other variables that affect health care utilization behavior. The age of the worker is another important factor that affects the demand for health care.

Theoretically, there is a U-shaped relationship between age and health care use. Age is used as a proxy for the depreciation rate of health capital. Young children are more vulnerable to disease and accidents. Working age people visit a general practitioner less frequently because of their high health stock and experience. Elderly people are expected to have relatively high visiting rates due to lower health stock. The proportion of elderly people is increasing because baby boomers are approaching retirement age. Therefore, the demand for health care is expected to increase in the future. Illness or need positively affects the demand for health care. There is interdependence between illness and age. Therefore an interaction term between illness and age is used in the regression.

There are two hypotheses about the effect of education on the demand for health care which predict opposite effects on the demand for health care. The first hypothesis is that more educated people are more aware of the existence of treatment options as well as of the benefits that the treatments can have on their health status. As a consequence, they are more likely to seek care when ill. An alternative hypothesis is that more educated people have milder health problems because they consume more preventive care, have higher income and better living conditions and food. As a consequence, they are less likely to get ill and therefore use less health care. These two effects work in opposite directions.
and weaken the overall effect of education on the demand for health care. Education is highly correlated with many of the variables, particularly income and urbanization but it might have an independent effect on the demand for health care as well. Therefore, an education variable will be included in the regression model but the effect of education is unclear.

Many needs are sex specific, such as pregnancy in specific age groups that impacts only women. Therefore, women generally experience greater morbidity and demand more health care. But it is difficult to generalize about the effect of sex on the demand for health care. Family size might also affect the demand for health care. Health insurance increases access to health care. Therefore, it might have positive effect on the demand for health care. A dummy variable set at 1 for those with insurance is included in the regression. A dummy variable set at 1 for those reporting a long-standing illness or physical disability is also included in the regression.

Health status plays a critical role in determining the demand for health care because people seek care when they have health problems. Health status is negatively related to the demand for health care. Individuals who visit the physician at higher rates are more likely to have health problems discovered and may perceive their health to be poorer than those who do not visit the physician very frequently. Manning et al (1982) concluded that health status explains most of the variation in regression models of health care utilization. The five-point health status scale; excellent, very good, good, fair and poor, is used as an
independent variable in this study. The problem with this variable is that it is difficult to measure.

Individuals who are eligible for Medicaid, Medicare and other public insurance programs may use more health care than those who are not eligible for these public insurance programs. A dummy variable is set to 1 for eligible people to allow for this effect. The distribution of medical resources in urban and rural areas is also an important issue. To capture this regional effect, a dummy variable is set to 1 if person lives in an MSA and zero otherwise.

A number of obvious physical conditions affect health care use, such as pregnant women and young children. The number of children in a household is included in the regression to control for this effect. However, information about pregnant women is not available in MEPS. Therefore that variable will not be included in the model.

There are various other factors which also affect the demand for health care, such as geographic location, environment, ways of living, the genetic stock of the population, assets, travel and waiting time, etc. Data on these variables is not available so these will not be included in the regression.

2.6. Model Specification and Methodology
Andersen (1968) was the first researcher who made a comprehensive model for health care demand. According to his model, the person’s decision to buy health care depends on three types of variables, such as: enabling variables, need variables and predisposing
variables. The enabling variables are the coinsurance rate and an individual’s income. The need variables are disability in days, self-perceived health status and level of illness. The predisposing variables are demographic characteristics, social structure, health beliefs and tastes. It is important to note that health insurance reduces the price of health care for the consumer in the budget constraint since he pays the coinsurance rate specified in his insurance policy instead of paying the market price per unit of care. Therefore, the consumer will purchase more health care units and less other goods with insurance than without it. It may be due to a moral hazard problem that the individuals seek more care when they are insured, or it may be due to supplier-induced demand in which the physicians provide more care when they are well reimbursed.

The health is valuable to the individual as other goods and services and it enters into his utility function with other good and services. It is assumed that additional health care services improve the person’s health. The person’s decision to buy health care also depends on the relative the price of health care, the individual’s income, tastes etc. The theory of demand for health care suggests that workers’ consumption of health care will be negatively associated with their coinsurance rate, given that premiums are unrelated to utilization. The opportunity cost of seeking medical care is approximated by the sum of the price paid to use the service. If demand for health care is inelastic in its price, it will increase the expenditures on health care if an increase in medical care price relative to other prices goes up.
Normally the supply of health care service is captured in its market price. However, the true price of physician services is often not known. Either it is not reported in the data set or distorted because of health insurance or public provision of health services, or other institutional constraints. Health care services are highly subsidized in the U.S. and in most other countries by welfare programs, by government-provided health services, by social security systems, and by private insurance systems. All these programs reduce consumers' direct cash costs. These subsidies reduce the importance of prices as an allocative device. Despite all of these price distortions, demand for health care responds to insurance-induced variation in price (coinsurance or out-of-pocket price).

Many poor individuals are insured through public programs. It might affect their response to the demand for health care from rich individuals. This is why an eligibility variable is included in the regression. Health status is one of the most predominant explanatory variables for health care use.

Approximately, 35% of the population report no physician visits during the survey year, 40% say one or two or three, 25% say four or more visits. There are two approaches to deal with excess zeros; zero-inflated models and hurdle or two-part models. The zero-inflated or with zeros model gives more weight to the probability of observing zeros. The two-part model is an attempt to correct the problem with non-users in the one-part model by separating behavior into two.
The count data models calculate the conditional probability distribution of the dependent count variable as a function of explanatory variables. These models estimate how changes in an explanatory variable affect various aspects of the outcome distribution, and predictions of the outcome distribution for given values of the explanatory variables. The count dependent variable takes non-negative values \( Y = 0, 1, 2, \ldots \) without an explicit upper limit. The distribution of the dependent variable is highly skewed to the right with a long tail and has large proportions of zeros on the other end of the distribution. These are basic properties required for the application of count data regressions. There are several estimation techniques suitable for estimation using data with the above-mentioned characteristics: the Poisson, the Negative binomial, the zero altered negative binomial and the two-part models. These are described in the following section.

A Poisson Regression Model

The Poisson regression model is considered to be a benchmark model for count data as the ordinary linear model is considered as a benchmark for real-valued continuous data. It is more suitable for the demand for health care model than the ordinary linear model because it explicitly recognizes the non-negative integer character of the dependent variable. Moreover, the ordinary linear model gives an inconsistent estimator of \( \beta \) if true data is generated by the Poisson process (Winkelmann, 2003, pg 63). Also, the Poisson regression model automatically takes care of the Heteroskedasticity problem which is most common in count data.

The standard univariate Poisson regression model makes the following three assumptions. First, each value of a dependent variable \( Y_i \) given \( X_i \) follows a Poisson
distribution. Second, the mean parameter, \( \lambda_i \), is a function of explanatory variables \( X_i \) and an unknown vector of parameters \( \beta \). Third, each value of \( Y_i \) is independently distributed from other values of \( Y_j \) where \( j=1, 2, 3...n \).

The dependent variable \( Y_i \) measures the number of times an individual visits a physician during the past 12 months. The length of time interval (i.e. 12 month) is the same for each individual. A physician’s visit depends upon the number of explanatory variables, \( X_i \). By combining the first two assumptions we get a conditional probability density function of a Poisson-distributed random variable \( Y_i \):

\[
\Pr(Y_i = y_i / X_i = x_i) = \frac{\lambda_i^y \exp(-\lambda_i)}{y_i!}, \quad y_i=0, 1, 2..... \quad (2.2)
\]

Any length of time can be normalized to unity without a loss of generality. The expected value of \( Y_i \) per time interval conditional on the explanatory variable \( X_i \) can be denoted by \( \lambda_i \). Mathematically, it can be written as: \( \mathbb{E}(Y_i/X_i)=\lambda_i=\exp(X_i\beta) \). Or in log-linear form \( \ln \lambda_i = X_i\beta \) (Winkelmann, 2003, pg 8). This transformation ensures that the estimated \( \lambda_i \) is positive for all possible combinations of parameters and explanatory variables. The mean of the Poisson distribution is always equal to its variance. Therefore, the above mentioned transformation can be written as, \( \mathbb{E}(Y_i/X_i)=\text{Var}(Y_i/X_i)=\lambda_i=\exp(X_i\beta) \). This is why the Poisson regression model is characterized by a single parameter model. The parameter \( \lambda \) depends on the exogenous variables, the \( X \) vector.

The column vector of unknown parameters \( \beta \) can be estimated by the maximum likelihood method since each \( y_i \) is independently distributed. The estimators are robust,
asymptotically efficient and follow the normality property. As long as the mean function \( \lambda \) is correctly specified, the estimator for \( \beta \) remains consistent even if the variance does not equal the mean. The mean \( \lambda \) is not equal to the variance only if the true distribution is not a Poisson distribution. However, the model must be correctly specified in order to apply the efficiency results for maximum likelihood estimators.

The marginal effects for a representative individual can be written as:

\[
\frac{E(Y_i/X_i)}{\partial X_{ij}} = \exp(X_i\beta)\beta_j = E(Y_i/X_i)\beta_j \quad j=1, 2,..., k
\]

It is more common and simpler to consider the relative change in \( E(Y_i/X_i) \) associated with a small change in \( X_{ij} \) because it is constant and equal to the coefficient \( \beta_j \):

\[
\frac{E(Y_i/X_i)(Y_i/X_i)}{\partial X_{ij}} = \beta_j
\]

If \( X_j \) is in logarithmic form, \( \beta_j \) has the interpretation of an elasticity: a percentage change in \( E(Y_i/X_i) \) per percentage change in \( X_{ij} \). If \( X_j \) is a dummy variable then, \( \beta_j \) gives the approximate relative impact of the dummy variable on \( E(Y_i/X_i) \) (Winkelmann, 2003, page 69).

Alternative models provide better estimators if the restriction of equal mean and variance is not satisfied. The most commonly used alternative models are the log-linear, Binomial and Negative Binomial models. The log-linear model does not allow zero counts, so it is not appropriate for this data since there are around 37% of the observations are zero counts. The negative Binomial regression is discussed in the following section.
A Negative Binomial Regression Model

A negative binomial regression model is more appealing than the Poisson regression model for physician visits because sample data shows that the conditional variance of the physician visits exceeds the conditional mean. Mathematically it is called overdispersion, i.e. $E(Y_i/X_i) < \text{Var}(Y_i/X_i)$. This overdispersion may be due to unobservable heterogeneity in physician visits. The negative binomial regression model is more general than the Poisson regression model because it allows overdispersion. The negative binomial regression model is also suitable for a large proportion of zero visits and a positively skewed distribution of health care use.

The negative binomial regression model is derived from a compound Poisson and Gamma distribution for the random variable $Y_i$. The density function of the negative binomial distribution is derived by adding an error term to the conditional mean of the Poisson distribution. $\lambda_i = \exp(X_i \beta) \exp(\varepsilon)$ or $\ln \lambda_i = X_i \beta + \varepsilon$. Where $\exp(\varepsilon)$ follows a gamma distribution with mean one and variance $\alpha$. Substituting the value of $\lambda_i$ into the Poisson density function given in equation (2.2) and integrating $\varepsilon$ out of the expression yields the following the probability density function of the negative binomial model:

$$\Pr(Y_i = y_i / \alpha, \lambda) = \frac{\Gamma(\alpha + Y_i)}{\Gamma(\alpha) \Gamma(Y_i + 1)} \left( \frac{\alpha}{\lambda + \alpha} \right)^\alpha \left( \frac{\lambda}{\lambda + \alpha} \right)^{Y_i}, \ y_i = 0, 1, 2, \ldots$$  \hspace{1cm} (2.3)

where $\Gamma(.)$ = gamma function. The mean of the negative binomial distributed random variable $Y_i$ is, $E(Y_i / \alpha, \lambda) = \lambda$ and variance $\text{Var}(Y_i / \alpha, \lambda) = \lambda + \alpha \lambda^2$. It is obvious that the conditional variance is always greater than the conditional mean. The introduction of an extra parameter $\alpha$ permits the conditional mean to differ from the conditional variance.
The Poisson regression model is a special case of the negative binomial regression model in which the variance parameter $\alpha$ is equal to zero.

**A Two-Part or Hurdle Regression Model**

The essence of the two-part, or Hurdle, regression model is that it decomposes one observed random variable $Y_i$ (physician visits) into two observed random variables—worker's decision to visit a physician or not, $Y_i > 0$ and their decision about how often $Y_i > 0$? The decision-making process is completed in two stages. In the first stage, workers make a decision whether to use health care services or not during a given time interval. This can be interpreted as a splitting mechanism which divides workers into non-users and potential users. In the second stage, workers make a decision about how many times they want to visit a physician conditional on its being positive at the first stage. These two stages, the participation decision (0, 1) and the positive count (1, 2, 3,...) are different stochastic processes. In the first stage, it is the patient who decides whether to visit the physician while in the second stage it is mainly the physician who decides the intensity of the treatment. Therefore, the model has two equations. The first is a probit equation for the dichotomous event of having zero versus positive visits:

$$Y_i = \Pr(\text{non-users}) = F(Z_i \delta)$$

where $F(Z_i \delta)$ is a cumulative distribution function. This first equation separates users from nonusers. The second equation is a data count model (i.e. a Poisson or a negative binomial model) explained above. The Poisson or the negative binomial density could be used to model the utilization of health care services. The mean of the distribution $\lambda$ is conditional on another set of variables, $X_i$. 
The study has used two models to estimate the demand for physician visits: a multinomial logit model and a data count model. For the multinomial logit model, the data is divided into three categories: nonusers, infrequent users and frequent users. The appeal of this division is driven by the important empirical fact that there is a sharp dichotomy between the use of the demand for health care by healthy and ill workers. At one extreme, perfectly healthy workers do not use health care at all, meaning zero physician visits. At the other extreme, ill workers very frequently use health care. Approximately 35% of the population report no physician visits during the survey year, 40% say one or two or three, 25% say four or more visits. The multinomial logit models are used when there are more than two outcomes of the dependent variables. The individual must belong to only one of the three categories and the labeling of the categories is arbitrary. Let $Y$ denote a random variable taking on the values 0, 1 and 2 representing three types of people: nonusers, infrequent users and frequent users respectively. Let $X$ denote a set of all the conditional variables. The coefficients tell us that how the probabilities will respond due to a change in $X$. The response probabilities can be written as follows:

$$P(Y = j / X) = \frac{\exp(X\beta_j)}{1 + \sum_{h=1}^{2} \exp(Xh\beta_h)}, \quad j = 1, 2 \ldots (2.4)$$

where $\beta$ is a $Kx1$ vector of unknown parameters. Because the response probabilities must sum to unity,

$$P(Y = 0 / X) = 1 \left[ \frac{1}{1 + \sum_{h=1}^{2} \exp(Xh\beta_h)} \right] \ldots \ldots (2.5)$$

The partial effects for this model are complicated. For a continuous $X_k$, we can write
Where \( \beta_{hk} \) is the \( k \)th element of \( \beta_h \) and \( g(X, \beta) = 1 + \sum_{h=1}^{2} \exp(X\beta_h) \). Equation (2.6) shows that even the direction of the effect is not determined entirely by \( \beta_{hk} \). Since the density of \( Y \) is fully specified for a given \( X \), estimation of the multinomial logit model is best carried out by the maximum likelihood method. The problem with this model is that there is no data on effective prices for nonusers of health care because they make no expenditures on health care. The year dummies were included in the regression to control for technological advances in health care. Most of the studies show that health status is a very important determinant of utilization of health care services. Therefore, health status is also included in the regression.

2.7. Data

The data used in this study were not specifically collected for the purpose of estimating this model. Instead, the data on relevant variables were extracted from the Medical Expenditure Panel Survey (MEPS) from 1996 to 2002 and the National Health Interview Survey (NHIS) from 1995 to 2001. NHIS and MEPS are both nationally representative, very well organized comprehensive health surveys. MEPS started in 1996 and have continued to collect information about health care use, medical costs and insurance coverage. The target population of NHIS is all residents of the United States, while the target population of MEPS is a 25% random sample of NHIS. The target population is stratified into 358 geographic regions known as primary sampling units (PSUs). Once a household is selected for inclusion in the survey, the questionnaire is administered in two stages. The first stage is an in-person interview with one respondent. The second stage is
in person or by phone. The sample size and weighted numbers for Hawai'i and the United States are given in Table 2.1.

<table>
<thead>
<tr>
<th>Year</th>
<th>Obs</th>
<th>Weighted Obs</th>
<th>Weighted Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>382</td>
<td>1178335</td>
<td>20482</td>
</tr>
<tr>
<td>1997</td>
<td>361</td>
<td>1181607</td>
<td>34551</td>
</tr>
<tr>
<td>1998</td>
<td>230</td>
<td>1181578</td>
<td>24072</td>
</tr>
<tr>
<td>1999</td>
<td>326</td>
<td>1186700</td>
<td>24618</td>
</tr>
<tr>
<td>2000</td>
<td>326</td>
<td>1198120</td>
<td>25096</td>
</tr>
<tr>
<td>2001</td>
<td>366</td>
<td>1277322</td>
<td>33556</td>
</tr>
<tr>
<td>2002</td>
<td>319</td>
<td>1065071</td>
<td>34928</td>
</tr>
<tr>
<td>Ave 7 Year</td>
<td>330</td>
<td>1181248</td>
<td>28186</td>
</tr>
</tbody>
</table>

On average, Hawai'i has 330 observations per year for seven years while the United States has 28186 observations.

Table 2.2 presents the list of the variables used in the estimation along with their weighted means, standard deviations, and minimum and maximum values. The first column of the table reports the variable names used in the regression. The weighted means of the corresponding variables are reported in the second column. The means of the dummy variables equal the proportion of cases with a value of 1, and can be interpreted as a probability. The standard deviation and minimum and maximum values are reported in the subsequent columns respectively.
Table 2.2: Summary Statistic of the Raw Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>physician Visits</td>
<td>3.51</td>
<td>5.78</td>
<td>0</td>
<td>170</td>
</tr>
<tr>
<td>out-of-pocket price</td>
<td>14.87</td>
<td>52.3</td>
<td>0</td>
<td>518</td>
</tr>
<tr>
<td>family income</td>
<td>51879</td>
<td>24919.99</td>
<td>98</td>
<td>198212</td>
</tr>
<tr>
<td>age</td>
<td>34.98</td>
<td>19.99</td>
<td>0</td>
<td>88</td>
</tr>
<tr>
<td>agesq</td>
<td>1223.6</td>
<td>1302.62</td>
<td>0</td>
<td>7744</td>
</tr>
<tr>
<td>female</td>
<td>0.512</td>
<td>0.499</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>college</td>
<td>0.4</td>
<td>0.26</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>married</td>
<td>0.45</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>white</td>
<td>0.31</td>
<td>0.35</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>health1 (Excellent)</td>
<td>0.3</td>
<td>0.44</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>health2 (Very good)</td>
<td>0.32</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>health3 (good)</td>
<td>0.26</td>
<td>0.42</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>health4 (fair)</td>
<td>0.09</td>
<td>0.25</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>health5 (poor)</td>
<td>0.03</td>
<td>0.13</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>eligible</td>
<td>0.11</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>family size</td>
<td>3.1</td>
<td>2.76</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>military</td>
<td>0.08</td>
<td>0.12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>insured</td>
<td>0.9</td>
<td>0.14</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>union</td>
<td>0.12</td>
<td>0.25</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>msa</td>
<td>0.83</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: The summary statistics are calculated by using 2310 observations.

The average number of physician visits in Hawai'i is 3.51 visits per year, which is slightly higher than in the United States as a whole, which is 3.44. The data also show that the average age in Hawai'i is approximately 35, which is slightly higher than the average age nationwide of 34.3. The percentage of the covered population in Hawai'i is 90% which is much higher than the nationwide coverage of 85%. Union membership in Hawai'i is 12% which is higher than the 11% in the United States as a whole.

2.8. Empirical Results

The number of physician visits means the total number of times that individuals went to the physician during the past 12 months. These visits do not include inpatient visits. On average, 35% of the population of Hawai'i did not make any visits to a physician during the last year which is higher than the national average of 32%. However, a larger proportion of the population of Hawai'i 40%, makes 1 - 3 visits per year. This is higher
than the national average of 37%. The proportion of the population of Hawai’i that makes four or more visits is 25% which is lower than the national average 30%. For more details see Table 2.3 for direct and model-based estimates.

Table 2.3: Distribution of Population by Doctor Visits

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DV0</td>
<td>34.61%</td>
<td>24.63%</td>
<td>31.72%</td>
<td>28.71%</td>
</tr>
<tr>
<td>DV13</td>
<td>40.35%</td>
<td>48.15%</td>
<td>37.20%</td>
<td>42.11%</td>
</tr>
<tr>
<td>DV4+</td>
<td>25.04%</td>
<td>27.22%</td>
<td>29.65%</td>
<td>29.18%</td>
</tr>
</tbody>
</table>

The model-based estimates are calculated using a multinomial logit model. The three categories of physician visits are: nonusers, infrequent users (1-3), and frequent users (4 or more). The explanatory variables are out-of-pocket price, family income, age, age squared, a gender dummy, education, marital status, race, health status, eligibility for public health programs, family size, a dummy for active duty in the armed forces, insurance status, a dummy for union membership and a dummy for MSA.

On average, each individual in Hawai’i makes 3.51 visits per year which is slightly higher than in the United States at 3.44. For more details see Table 2.4.

Table 2.4: Average Number of Doctor Visits

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DV13</td>
<td>1.76</td>
<td>1.56</td>
<td>1.74</td>
<td>1.63</td>
</tr>
<tr>
<td>DV4+</td>
<td>10.36</td>
<td>9.76</td>
<td>9.32</td>
<td>9.10</td>
</tr>
<tr>
<td>Ave</td>
<td>3.51</td>
<td>3.21</td>
<td>3.44</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Average visits include all people. The model based estimates are calculated using a Negative Binomial regression model. The explanatory variables and model specification are the same as that used for multinomial logit model.
There is a strong positive correlation between age and physician visits. On average, the elderly population (i.e., 65 and older) visited a physician more frequently than children and working age people. Children 0-18 visited a physician’s office an average of 2.5 visits per year while working age people, 19–64, visited a physician’s office 1.52 times per year and those 65 years and older visited a physician’s office an average of 8.5 times per year. Therefore, a physician’s visit function is quadratic in age.

### Table 2.5: Average number of doctor Visits by Age

<table>
<thead>
<tr>
<th>Age</th>
<th>Hawaii</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DV13</td>
<td>DV4+</td>
</tr>
<tr>
<td>Age 0-18</td>
<td>1.94</td>
<td>8.21</td>
</tr>
<tr>
<td>age 19-64</td>
<td>1.70</td>
<td>9.89</td>
</tr>
<tr>
<td>age 65+</td>
<td>1.98</td>
<td>11.28</td>
</tr>
<tr>
<td>Model Based Est.</td>
<td>Age 0-18</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>age 19-64</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>age 65+</td>
<td>2.25</td>
</tr>
</tbody>
</table>

The model-based estimates are calculated by the Negative Binomial model. The explanatory variables and model specification are the same as that used in the multinomial logit model.

Health insurance has a positive effect on physicians’ visits. Insured individuals visit a physician more frequently than the uninsured. Direct and model-based estimates show that covered people make more visits than the uninsured. For details see Table 2.6.

### Table 2.6: Average Number of Doctor Visits by Insurance

<table>
<thead>
<tr>
<th></th>
<th>Hawaii</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Est.</td>
<td>DV13</td>
</tr>
<tr>
<td>Uninsured</td>
<td>1.60</td>
<td>8.45</td>
</tr>
<tr>
<td>Insured</td>
<td>1.80</td>
<td>10.19</td>
</tr>
<tr>
<td>Uninsured</td>
<td>1.71</td>
<td>7.89</td>
</tr>
<tr>
<td>Insured</td>
<td>2.05</td>
<td>11.20</td>
</tr>
</tbody>
</table>

124
The model-based estimates are calculated by the Negative Binomial model. The explanatory variables and model specification are the same as that used in the multinomial logit model.

2.9. Further Research

In this study, health is an exogenous variable. In fact, health is determined by health care use. Therefore health should be used as an endogenous variable and there should be a separate equation that determines health status. The model will consist of two equations determined simultaneously:

\[ H^* = \alpha Z + \nu \]
\[ Y = \beta X + \delta H^* + \varepsilon \]

The first equation can be interpreted as either a production function or a demand function of health. In both cases \( H^* \) is the desired health status determined by individual and family characteristics. The second equation resembles the utilization equation. A more comprehensive general equilibrium can also be constructed which would have separate equations for physicians, patients, medical drugs and insurance. These cannot be tested given limited data.
2.10. Conclusion
The demand for physician visits in Hawai'i is estimated using MEPS and NHIS data from 1996-2002. The effect of cost sharing, age and health insurance on the demand for health care is estimated. The results show that the demand for health care is slightly higher in Hawai'i than in the United States as a whole at 3.51 and 3.42, respectively. Physician visits negatively respond to changes in the amount paid out-of-pocket. A reduction in the level of cost sharing increases the demand for physician visits. Three alternative models are used to estimate the pure price response. These three models suggest that price elasticities for health care are in the -0.1 to -0.2 range. These values are consistent with those in the lower range of the non-experimental literature which vary from -0.1 to -2.1.

The magnitude of the price elasticity of health care should make a difference in the generosity of health insurance. If the price elasticity is higher, then insurance coverage should be less generous. The results also show that increased income positively affects physician visits. Depending on the insurance plan, higher income people have 20 percent to 40 percent more physician visits than low income people. The main reason for the rise in health care costs is not increased life expectancy but the invention of expensive medical technology, surgical procedures and blockbuster pills.

The effect of health insurance on the demand for health care and health status is also examined. The estimates of this study suggest that health insurance has a positive effect on the demand for physician visits. Insured people visit physicians more frequently than the uninsured. Persons 65 and over had the highest number of physician visits, 8.7 visits per year. Females had a higher visit rate when compared to males. However, the highly
skewed distribution of physician visits, with a large number of non-users, makes it difficult to choose a single model which will give reliable estimates.
2.11. Bibliography


Ro, Kong-Hyun, (1972). “Patient Characteristics, Hospital Characteristics and Hospital Use” in Fuchs, V. R. (Ed), “Essays in the Economics of Health and Medical Care” NBER.


2.12. Appendix

Table 2.1A

<table>
<thead>
<tr>
<th>Poisson regression</th>
<th>Number of obs =</th>
<th>1082</th>
</tr>
</thead>
</table>

Log pseudo-likelihood = -3.41E+08

Wald chi2(17) = 71.42

Prob > chi2 = 0

| Doctor visits | Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---------------|-------|-----------|-------|-------|-------------------|
| Out-of-pocket price | -0.518581 | 0.053198 | -9.75 | 0.000 | -0.6228472 -0.414315 |
| Income        | 0.025702 | 0.0185277 | 1.39  | 0.165 | -0.0106121 0.062015 |
| Age           | 2.86E-05 | 0.0046882 | 0.01  | 0.995 | -0.00916 0.009217 |
| Age square    | 9.23E-05 | 0.0000478 | 1.93  | 0.053 | -1.25E-06 0.000186 |
| Female        | 0.26576  | 0.0294663 | 9.02  | 0.000 | 0.2080076 0.323513 |
| Currently married | 0.03777  | 0.0353396 | 1.07  | 0.285 | -0.0314939 0.107035 |
| White race    | 0.188202 | 0.0473917 | 3.97  | 0.000 | 0.0953157 0.281088 |
| Very good health | 0.123932 | 0.0368532 | 3.36  | 0.001 | 0.0517009 0.196163 |
| Good health   | 0.300178 | 0.0396848 | 7.56  | 0.000 | 0.2223967 0.377958 |
| Fair Health   | 0.547641 | 0.0551494 | 9.93  | 0.000 | 0.4395501 0.655732 |
| Poor health   | 0.943609 | 0.1496976 | 6.3   | 0.000 | 0.8502066 1.23701  |
| Eligible for PHI | 0.068266 | 0.0762196 | 0.9   | 0.37  | -0.0811212 0.217654 |
| Family size   | -0.028457 | 0.0120541 | -2.36 | 0.018 | -0.0520829 -0.004832 |
| Active duty in military insured | 0.115612 | 0.0500768 | 2.31  | 0.021 | 0.0174628 0.21376 |
| Number of days ill | 0.007534 | 0.0007705 | 9.78  | 0.000 | 0.0060239 0.009044 |
| Living in MSA | 0.112617 | 0.0304657 | 3.7   | 0.000 | 0.0529049 0.172328 |
| cons          | 0.337952 | 0.2005995 | 1.68  | 0.092 | -0.0552158 0.73112  |

Note: Standered Errors are adjusted on duid
### Table 2.2A

Negative binomial regression

| Doctor visits          | Coef.   | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|------------------------|---------|-----------|-------|------|----------------------|
| Out-of-pocket price    | -0.490653 | 0.053926  | -9.1  | 0    | -0.596346 -0.38496   |
| Income                 | 0.018108  | 0.01666   | 1.09  | 0.277| -0.014545 0.050761   |
| Age                    | -0.002749 | 0.004561  | -0.6  | 0.547| -0.011688 0.006191   |
| Age square             | 0.000128  | 4.72E-05  | 2.7   | 0.007| 3.49E-05 0.00022     |
| Female                 | 0.25702   | 0.029366  | 8.75  | 0    | 0.199463 0.314577    |
| Currently married      | 0.066196  | 0.031342  | 2.11  | 0.035| 0.004767 0.127626    |
| White race             | 0.205201  | 0.043544  | 4.71  | 0    | 0.119856 0.290546    |
| Very good health       | 0.118626  | 0.037514  | 3.16  | 0.002| 0.045099 0.192152    |
| Good health            | 0.264931  | 0.039552  | 6.7   | 0    | 0.187411 0.342451    |
| Fair Health            | 0.547189  | 0.05257   | 10.41 | 0    | 0.444153 0.650224    |
| Poor health            | 0.942766  | 0.16428   | 5.74  | 0    | 0.620783 1.264749    |
| Eligible for PHI       | 0.090757  | 0.0857    | 1.06  | 0.29 | -0.077211 0.258726   |
| Family size            | -0.026692 | 0.011768  | -2.27 | 0.023| -0.049756 -0.003628  |
| Active duty in military| 0.080813  | 0.0426    | 1.9   | 0.058| -0.002682 0.164307   |
| insured                | 0.201194  | 0.056364  | 3.57  | 0    | 0.090722 0.311666    |
| Number of days ill     | 0.011367  | 0.000769  | 14.78 | 0    | 0.009859 0.012875    |
| Living in MSA          | 0.101117  | 0.030112  | 3.36  | 0.001| 0.042099 0.160136    |
| cons                   | 0.43276   | 0.180271  | 2.4   | 0.016| 0.079437 0.780804    |

Inalpha -0.71872 0.037339 -0.791902 -0.645537
alpha 0.487376 0.018198 0.452983 0.524381
Stata Program

/* Chapter 2: Demand for Health Care */
clear
log using c:\cps\chapter2mepsnhisnew.log, replace
set more 1
set mem 1000m
#delimit ;
use "d:\MEPSNHIS\MEPSNHIS9602new.dta", clear;
replace year=1995 if year==95;
replace year=1996 if year==96;
replace year=1997 if srvy_yr==1997;
replace year=1998 if srvy_yr==1998;
replace year=1999 if srvy_yr==1999;
replace year=2000 if srvy_yr==2000;
replace year=2001 if srvy_yr==2001;
gen wt=.;
replace wt=wtdper96 if myear==1996;
replace wt=wtdper97 if myear==1997;
replace wt=wtdper98 if myear==1998;
replace wt=perwt99f if myear==1999;
replace wt=perwt00f if myear==2000;
replace wt=perwt01f if myear==2001;
replace wt=perwt02p if myear==2002;
gen varstrata=.;
replace varstrata=varstr96 if myear==1996;
replace varstrata=varstr97 if myear==1997;
replace varstrata=varstr98 if myear==1998;
replace varstrata=varstr99 if myear==1999;
replace varstrata=varstr00 if myear==2000;
replace varstrata=varstr01 if myear==2001;
replace varstrata=varstr if myear==2002;
gen varpsua=.;
replace varpsua=varpsu96 if myear==1996;
replace varpsua=varpsu97 if myear==1997;
replace varpsua=varpsu98 if myear==1998;
replace varpsua=varpsu99 if myear==1999;
replace varpsua=varpsu00 if myear==2000;
replace varpsua=varpsu01 if myear==2001;
replace varpsua=varpsu if myear==2002;
svysset [pweight=wt], strata(varstrata) psu(varpsua);
gen dv=.;
replace dv=(obdrv96 + opdrv96) if myear==1996;
replace dv=(obdrv97 + opdrv97) if myear==1997;
replace dv=obdrv98 + opdrv98 if myear==1998;
replace dv=obdrv99 + opdrv99 if myear==1999;
replace dv=obdrv00 + opdrv00 if myear==2000;
replace dv=obdrv01 + opdrv01 if myear==2001;
replace dv=obdrv02 + opdrv02 if myear==2002;
replace dv=0 if dv=.;
gen insured=.;
replace insured=1 if ( uninsurd==2 & myear==1996 & insured==.);
replace insured=1 if ( unins97==2 & myear==1997 & insured==.);
replace insured=1 if ( unins98==2 & myear==1998 & insured==.);
replace insured=1 if ( unins99==2 & myear==1999 & insured==.);
replace insured=1 if ( unins00==2 & myear==2000 & insured==.);
replace insured=1 if ( unins01==2 & myear==2001 & insured==.);
replace insured=0 if insured==. & myear==2002;
gen dv1=(dv>0 & dv<=3);
gen dv2=dv>3 & dv-=.;
gen status=.;
replace status=0 if dv==0;
replace status=1 if dv1==1;
replace status=2 if dv2==1;
replace status=0 if status==.;
sort myear;
by myear: tab status;
by myear: tab status if hawaii==1;
by myear: tab status [iweight=wt];
by myear: su dv [iweight=wt] if hawaii==1;
by myear: su dv [iweight=wt] if dv>1 & dv<=3;
by myear: su dv [iweight=wt] if hawaii==1 & dv>1 & dv<=3;
by myear: su dv [iweight=wt] if dv>4 & dv-=.;
by myear: su dv [iweight=wt] if hawaii==1 & dv>4 & dv-=.;
by myear: su dv [iweight=wt];
by myear: su dv [iweight=wt] if hawaii==1;
gen agel=.;
replace agel=age if myear==1996;
replace agel=age if myear==1997 & agel==.;
replace agel=age98x if myear==1998 & agel==.;
replace agel=age99x if myear==1999 & agel==.;
replace agel=age00x if myear==2000 & agel==.;
replace agel=age01x if myear==2001 & agel==.;
replace agel=age02x if myear==2002 & agel==.;
drop if agel<0;
gen agecat=recode(agel, 18, 64, 100);
sort myear;
by myear: tab agecat status [iweight=wt], row;
by myear: tab agecat status [iweight=wt] if hawaii==1, row;
tab agecat status [iweight=wt], row;
tab agecat status [iweight=wt] if hawaii==1, row;
su dv if dv>0 & dv<=3 & agecat==18;
su dv if hawaii==1 & dv>0 & dv<=3 & agecat==18;
su dv if dv>4 & agecat==18;
su dv if hawaii==1 & dv>4 & agecat==18;
su dv if dv>0 & dv<=3 & agecat==64;
su dv if hawaii==1 & dv>0 & dv<=3 & agecat==64;
su dv if dv>4 & agecat==64;
su dv if hawaii==1 & dv>4 & agecat==64;
su dv if dv>0 & dv<=3 & agecat==100;
su dv if hawaii==1 & dv>0 & dv<=3 & agecat==100;
su dv if dv>4 & agecat==100;
su dv if hawaii==1 & dv>4 & agecat==100;
sort myear;

by myear:tab insured status [iweight=wt], col;
by myear:tab insured status [iweight=wt] if hawaii==1, col;
tab insured status [iweight=wt], col;
tab insured status [iweight=wt] if hawaii==1, col;
tab health status [iweight=wt], col;
tab health status [iweight=wt] if hawaii==1, col;
su dv if dv>0 & dv<3 & insured==0;
su dv if hawaii==1 & dv>0 & dv<3 & insured==0;
su dv if dv>=4 & insured==0;
su dv if hawaii==1 & dv>=4 & insured==0;
su dv if insured==0;
su dv if hawaii==1 & insured==0;
su dv if dv>0 & dv<3 & insured==1;
su dv if hawaii==1 & dv>0 & dv<3 & insured==1;
su dv if dv>=4 & insured==1;
su dv if hawaii==1 & dv>=4 & insured==1;
su dv if insured==1;
su dv if hawaii==1 & insured==1;
gen p=.;
replace p=(obdtch96 + opvtch96)/dv if myear==1996 & p==.;
replace p=(obdtch97 + opvtch97)/dv if myear==1997 & p==.;
replace p=(obdtch98 + opvtch98)/dv if myear==1998 & p==.;
replace p=(obdtch99 + opvtch99)/dv if myear==1999 & p==.;
replace p=(obdtch00 + opvtch00)/dv if myear==2000 & p==.;
replace p=(obdtch01 + opvtch01)/dv if myear==2001 & p==.;
replace p=0 if p==.;
gen outofpocketperv=.;
replace outofpocketperv=(obdslf96 + opvslf96)/dv if myear==1996 & outofpocketperv=.;
replace outofpocketperv=(obdslf97 + opvslf97)/dv if myear==1997 & outofpocketperv=.;
replace outofpocketperv=(obdslf98 + opvslf98)/dv if myear==1998 & outofpocketperv=.;
replace outofpocketperv=(obdslf99 + opvslf99)/dv if myear==1999 & outofpocketperv=.;
replace outofpocketperv=(obdslf00 + opvslf00)/dv if myear==2000 & outofpocketperv=.;
replace outofpocketperv=(obdslf01 + opvslf01)/dv if myear==2001 & outofpocketperv=.;
replace outofpocketperv=0 if outofpocketperv=.;
gen outofpockettot=totslf01;
gen eprice=outofpocketperv/p;
gen agesq=age1^2;
gen male=(sex==1);
gen female=(sex==2);
gen lesshigh=hidegyr<=2;
gen high=hidegyr=3;
gen college=educ==16 & (educ==97 | educ==98 | educ==99 | educ==.);
gen married=.;
replace married=marry96x==1 & myear==1996;
replace married=marry97x==1 & myear==1997;
replace married=marry98x==1 & myear==1998;
replace married=marry99x==1 & myear==1999;
replace married=marry00x==1 & myear==2000;
replace married=marry01x==1 & myear==2001;
replace married=marry02x==1 & myear==2002;
replace married=0 if married==.;
drop if (marryOlx==8 | marryOlx==9);

gen white=.;
replace white=1 if racex==5 & myear<=2001;
replace white=1 if racex==1 & myear==2002;
replace white=0 if white==.;
gen nonwhite=(racex==5 & myear<=2001);
replace nonwhite=(racex==1 & myear==2002);
gen familysize=famsze96 if myear==1996;
replace familysize=famsze97 if familysize==. & myear==1997;
replace familysize=famsze98 if familysize==. & myear==1998;
replace familysize=famsze99 if familysize==. & myear==1999;
replace familysize=famsze00 if familysize==. & myear==2000;
replace familysize=famsze01 if familysize==. & myear==2001;
replace familysize=famsze02 if familysize==. & myear==2002;
drop if familysize==1;
*inapplicable is -1 or otherwise 1-18 family members;
*gen military=didserve==1;
*gen daysill=(ddnwrk31 + ddnwrk42 + ddnwrk53 + ddnscl31 +ddnscl42 +
  ddnscl53 );
*recode daysill -34/0 =0;
gen health1=rtehlth1==1 & myear==1996;
replace health1=rthlth31==1 & myear>1997;
gen health2=rtehlth1==2 & myear==1996;
replace health2=rthlth31==2 & myear>1997;
gen health3=rtehlth1==3 & myear==1996;
replace health3=rthlth31==3 & myear>1997;
gen health4=rtehlth1==4 & myear==1996;
replace health4=rthlth31==4 & myear>1997;
gen health5=rtehlth1==5 & myear==1996;
replace health5=rthlth31==5 & myear>1997;
drop if (rtehlth1<0 | rtehlth1==.) & myear==1996;
drop if (rthlth31<0 | rthlth31==.) & myear>1997;
gen eligible=.;
replace eligible=1 if povcat<=2 & myear==1996 & eligible==.;
replace eligible=1 if povcat97<=2 & myear==1997 & eligible==.;
replace eligible=1 if povcat98<=2 & myear==1998 & eligible==.;
replace eligible=1 if povcat99<=2 & myear==1999 & eligible==.;
replace eligible=1 if povcat00<=2 & myear==2000 & eligible==.;
replace eligible=1 if povcat01<=2 & myear==2001 & eligible==.;
replace eligible=0 if eligible==.;
gen msa=.;
replace msa=1 if msa96==1 & myear==1996 & msa==.;
replace msa=1 if msa97==1 & myear==1997 & msa==.;
replace msa=1 if msa98==1 & myear==1998 & msa==.;
replace msa=1 if msa99==1 & myear==1999 & msa==.;
replace msa=1 if msa00==1 & myear==2000 & msa==.;
replace msa=1 if msa01==1 & myear==2001 & msa==.;
replace msa=1 if msa02==1 & myear==2002 & msa==.;
replace msa=0 if msa==.;
gen y=wagepnx if myear==1996;
replace y=wagep97x if y==. & myear==1997;
replace y=wagep98x if y==. & myear==1998;
replace y=wagep99x if y==. & myear==1999;
replace y=wagep00x if y==. & myear==2000;
replace y=wagep01x if y==. & myear==2001;
drop if y<=0;  
gen lny=ln(y);  
gen worker=y>0;  
su dv p eprice outofpocketperv outofpockettot y lny agel agesq female 
mommed married college white health1 health2 health3 health4 health5 
eligible insured msa;  
su dv p eprice outofpocketperv outofpockettot y lny agel agesq female 
mommed married college white health1 health2 health3 health4 health5 
eligible insured msa if hawaii==1;  
poisson dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt], robust cluster(dupersid);  
poisgof;  
poisgof, pearson;  
nbreg dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt], robust cluster(dupersid);  

*heckman dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt], select(dv = lny age agesq female 
mommed married white health2 health3 health4 health5 eligible familysize 
insured msa) cluster(dupersid);  
*mfx compute;  
mlogit status outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt], base(0) robust cluster(dupersid);  
mfx compute, predict(outcome(0));  
mfx compute, predict(outcome(1));  
mfx compute, predict(outcome(2));  

poisson dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt] if hawaii==1 , robust cluster(dupersid);  
poisgof;  
poisgof, pearson;  
nbreg dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt] if hawaii==1, robust cluster(dupersid);  

*heckman dv outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt] if hawaii==1, select(dv = lny age agesq female 
mommed married white health2 health3 health4 health5 eligible familysize 
insured msa) cluster(dupersid);  
mfx compute;  
mlogit status outofpocketperv agel agesq female married college white 
health2 health3 health4 health5 
eligible insured msa [pw=wt]if hawaii==1, base(0) robust 
cluster(dupersid);  
mfx compute, predict(outcome(0));  
mfx compute, predict(outcome(1));  
mfx compute, predict(outcome(2));  

*end of file;
Essay Three

The Effect of the Prepaid Health Care Act on Labor Force Utilization and Labor Sorting in Hawai’i

Abstract
This chapter examines the impact of the Prepaid Health Car Act (PHCA) of 1974 on labor force utilization and labor market sorting in Hawai’i using empirical methods. The Hawai’i PHCA requires private-sector employers to provide health insurance to their full-time workers. However, certain classes of workers are exempt from this regulation including part-time employees (i.e., less than 20 hours per week), self-employed workers, family workers, public-sector workers and employees working under a collective bargaining contract. We hypothesize that the PHCA will cause employment to shift from the exempt class to the regulated class of workers and that among regulated employees, utilization of labor will rise. Using four decades of data from the Current Population Survey (CPS) 1963-2004, we produce direct estimates (weighted tabulations) and model-based estimates (multinomial logit regressions) of the distribution of the labor force by hours employed and across sectors. These estimates are produced for Hawai’i before and after the implementation of the 1974 regulation. Using the workers in the overall United States as a control group, we implement a difference-in-difference approach to sweep out economy-wide social and structural change. The results indicate a modest shift in the labor force with more workers in the regulated class and greater utilization of labor among full-time workers than would have otherwise occurred.
3.1. Introduction

Many economists and public policy makers think that the Prepaid Health Care Act causes bimodal labor force utilization and labor sorting from the exempt to the non-exempt sector in Hawai‘i. The underlying rationale of this hypothesis is as follows: the Prepaid Health Care Act (PHCA) requires that employers must provide health insurance benefits to full-time employees and contribute at least half of the premium. At the same time, workers employed less than 20 hours per week do not qualify for health insurance benefits from their employer. Gruber and McKnight (2003) reported that the employers’ contribution to the premium is over 80% of the total premium. Also, employers’ health insurance premium contributions were estimated to be 5.6% of private industry’s total compensation in 2001. The Prepaid Health Care Act creates a great incentive for employers to employ more part-time rather than full-time workers in order to avoid health insurance costs. On the other hand, employers would like to utilize the utmost hours of full-time employees without damaging their health. Employers would reduce the number of workers who work 20 to 35 hours per week in order to reduce per hour employee benefit costs. If other things remain the same, a worker who works 20 to 35 hours per week is the most expensive worker from an employer’s point of view. This rationale will create a bimodal distribution of hours worked in Hawai‘i due to the Prepaid Health Care Act. For more details see figure 3.1.
If health insurance is valuable to the workers then they would opt to switch from the exempt sector to the non-exempt sector in order to obtain health insurance benefits.

Very little systematic effort has been made to assess the validity of this claim. The main purpose of this essay is two-fold. First, to test whether the Prepaid Health Care Act causes bimodal labor force utilization in Hawai'i and second, to test whether the Prepaid Health Care Act causes labor sorting from the exempt to the non-exempt sector.

Hawai'i established a unique health policy under the Prepaid Health Care Act (PHCA) which was passed on September 2, 1974 and became effective on January 1, 1975. It was suspended by Supreme Court order on October 5, 1981. PHCA was ineffective for fifteen months until on January 14, 1983 when congress granted a special exemption from
ERISA that allowed PHCA to resume. PHCA was passed the same year as the federal Employee Retirement Income Security Act (ERISA) which superceded all state laws related to employee benefits. Data for the survey years 1982 and 1983 was discarded because the law was not in effect during these two years.

Hours worked is affected by various factors such as wages, time, sex, age, race, education, marital status and industry. Average weekly hours worked has decreased over time from 43 hours in 1963 to 40 hours in 2004. This is mainly because the industrial structure of the economy has changed over time from manufacturing to services. On average males work seven hours per week more than the females. This gender gap is closing over time but it is still significant. Average weekly hours worked increases with age up to 50, and then declines. Studies show that more educated people work more hours than less educated people. On average, married people work more hours per week than unmarried people. Average hours worked also differs significantly from industry to industry. Individuals employed in the manufacturing sector work more hours than those in the service sector. There are other factors which may also affect hours worked, but they are not considered here because the data was too thin, particularly before the Prepaid Health Care Act.

A matching data technique is used to control for heterogeneity and bias in the estimates. The data is one-to-one matched by the above mentioned eight variables to control the affect of these variables on the hours worked per week. One-to-one matching means selecting a single observation from the Hawai'i sample to match each observation taken

---

1 29 U.S.C. 1144 §514 (b) 5(A).
from the sample of the rest of the United States. This matched sampling approach was introduced by Neyman (1923) and was popularized by Rubin (1974) and Holland (1986). The matching technique will also reduce unobserved heterogeneity.

The remainder of essay three has eight sections which are organized as follows. Section 3.2 summarizes the previous literature on hours worked. It is organized in a step-by-step progression from the simplest models to the most complex. In section 3.3 the method of analysis is explained. In section 3.4 the data and its sources are explained. The main empirical results of this chapter are explained in Section 3.5. Section 3.6 compares the results of this study with other related studies. In the last section, broader conclusions are drawn from the empirical results.

3.2. Literature Review

This section provides an overview of the literature on labor force utilization. There are only two major studies that have examined the effect of the Prepaid Health Care Act on hours worked in Hawai‘i. Both of these studies are reviewed in detail in the following paragraphs. There are several studies that have estimated hours worked in the United States and other countries; these are also reviewed briefly in this section. Most of these studies have used observational survey data collected by the United States Bureau of Labor Statistics or the United States Census Bureau. This study also uses the Current Population Survey March Supplement data collected by the United States Bureau of Labor Statistics from 1963 to 2004 with the exception of twelve years of data from 1968 to 1976 and 1982 and 1983 when Hawai‘i’s data was grouped with other western states and law was not ineffect.
The first study that examined the effects of the Prepaid Health Care Act on hours worked in Hawai'i was an empirical study conducted by Thurston (1997). He examined the effects of the Prepaid Health Care Act on employment, wages and health insurance coverage. Two different data sets were used: the United States Censuses of Population and Housing Survey for 1970 and Current Population Surveys March Supplement for 1990-1993. Thurston used the 1970 Census data for the pre-program period calculations and he pooled the 1990-1993 Current Population Surveys March supplement for the post-program period calculations. It was concluded in his paper that the number of part-time workers in Hawai'i increased significantly after the Prepaid Health Care Act was implemented. These two data sets were collected for different purposes and use different definitions of hours worked. Also, Census labor market data is considered inferior to the Current Population Survey data. Simple tabulations were used to draw conclusions and did not control for other factors which could affect hours worked. Thus, his conclusion might have been affected. The present study uses consistent data and controls for other factors which affect the hours worked.

The second major study is an experimental study conducted by Sherstynuk (2004) in which she explains the theoretical rationale for the Hawai'i-type partial mandate. Her study predicted that the Hawai'i-type partial mandate would cause an increase in the number of part-time workers. Later her predictions were confirmed by experimental data. She made a number of simplifying assumptions which do not hold true in the real world. Sherstynuk’s study assumed that the labor market is competitive and there is no
productivity loss from a worker working two part-time jobs or one full-time job. The Hawai‘i-type partial mandate would reduce employers’ costs by hiring two part-time workers instead of one full-time worker. Since part-time workers are not eligible for health insurance benefits, their employers can save insurance costs by hiring two part-time workers instead of one full-time worker. In the real world, the labor market is not perfect and part-time workers are not as productive as full-time workers. This empirical study does not confirm her theoretical predictions and experimental results.

In the United States, there are a number of studies on hours worked. A few of the major studies are reviewed in the following section. Bell and Freeman (2000) examined the relationship between hours worked and wage-inequality in the United States and Germany. Micro data from both countries were used to examine this relationship. It was concluded that in the United States workers work more hours than in Germany because wage-inequality is higher in the United States than Germany. They found that the difference in working hours is more significant in those industries which have more wage-inequality. It was concluded that higher wage-inequality causes the laborer to work more hours.

Kirkland (2000) examined the question of how many hours per week Americans work at their paying jobs. She used two survey data sets—the Current Population Survey (CPS) and Current Employment Statistics (CES) from 1964 to 1999. Both data sets indicate that there is a long-term downward trend in the average hours worked. The downward trend in hours worked is more dramatic in trade and service industries than in the
manufacturing industry. In fact, the average weekly working hours in the manufacturing industry has increased over the same period of time from 40 to 42 hours.

McGrattan and Rogerson (2004) examined how the number of average weekly hours worked per person in the United States has changed since World War II. They concluded that the overall average number of hours worked per person has been roughly constant; but they found a significant shift in hours worked from males to females, from older workers to younger workers and from single to married workers. The reason for this reallocation is an increase in relative wages of female to males, an increase in Social Security benefits to retired workers, and changes in family structure.

Johnson (2002) examined the hours worked in the United States and compared it with the rest of the world. Using International Labor Organization (ILO) data for the year 2001-2002, he found that workers in the United States work more hours per week than in any other country in the industrialized world. Annual hours worked per person in the United States has increased from 1900 to 1980 hours between 1990 and 2000 while this number has decreased in other industrialized nations during the same time period.

3.3. Method and Methodology
This section provides an overview of the methods used in this chapter to examine the effect of the Prepaid Health Care Act on hours worked and labor force utilization in Hawai'i. There is no well-established theoretical and conceptual framework found in the literature for this kind analysis. The following three methods are used with matched and
unmatched data – 1) simple tabulation method, 2) simple difference-in-difference tabulation analysis and 3) multinomial logit difference-in-difference regression analysis.

The method of matching data is popular in evaluating social programs because it uses observed explanatory variables to adjust for differences in outcomes unrelated to treatment that gives rise to selection bias. Roy (1951) and Quandt (1972) applied this technique in economic studies. The matched sampling method is mainly used to remove selection bias due to the matching variables. This technique also reduces unobserved heterogeneity and increases precision and robustness.

There are various approaches to control for bias due to observable variables in the matching literature. Common approaches are one-to-one matching on covariates and propensity scores. One-to-one matching means selecting a single observation from Hawai’i to match each observation from the rest of the United States by year, age, sex, race, martial status, education, wage and industry. This approach was introduced by Neyman (1923) and was popularized by Rubin (1974) and Holland (1986).

The difference-in-difference approach compares the percentage of workers moving from the exempt sector to non-exempt sector. The time period from 1963 to 1967 is used as Hawai’i’s own control in the pre-legislation period. The rest of the United States that did not adopt the legislation is used as a control group to account for the structural and industrial changes in the economy. Other things remaining constant, the difference in the before and after period for Hawai’i gives the effect of the Prepaid Health Care Act. If the
state’s characteristics change over time, then the difference between the before and after period for Hawai‘i yields a biased estimate. This bias can be removed by subtracting the difference of the before and after period for the other 49 states. In other words, the difference in the before and after period for Hawai‘i minus the difference in the before and after period for the other 49 states of the United States yields the difference-in-difference estimate.

A simple difference-in-difference analysis does not yield the best estimator in terms of efficiency of the most precise inference because a simple difference-in-difference analysis does not utilize all the other factors that affect the hours worked. This is why multinomial logit regression analysis is also used for further investigation. The multinomial logit regression estimates might also be biased if the time trend followed in the hours worked is different in Hawai‘i and the other 49 states.

3.4. Data
This section describes the data sources and its descriptive statistics. Data used in this study is extracted from the Current Population Survey (CPS) March Supplements for the years 1963-2004, excluding 1968-1976 and 1982 and 1983. These ten years are excluded because during these years Hawai‘i’s data was grouped with other western states and there is no way to separate Hawai‘i’s data from those states. The remaining two years are excluded because during these years the law was not in effect. The CPS is a household survey but individual level data is used in this study. The sample for this study is limited to labor force participants who worked positive hours for pay. This restriction is required to calculate the average number of hours worked. This study has also excluded employed
individuals who worked variable hours during the reference period and not in the age group from 19 to 64.

The CPS is one of the oldest, largest and most well-recognized nationally representative labor surveys in the United States. The Census Bureau has been conducting the CPS on a monthly basis since the 1940s to provide up-to-date information on the labor force and demographic characteristics of the US population. Initially, the CPS was designed to collect up-to-date facts about the number of Americans who are employed, unemployed or not in the labor force. Over the years, the survey has been expanded into many areas such as earnings and hours worked. In addition to the employment facts, since 1963 the CPS March Supplement started collecting information on hours worked.

Currently, the CPS interviews around 99,000 housing units, or 300,000 individuals, monthly. The subjects are scientifically selected from 792 primary sampling units comprising 2007 counties and independent cities with coverage in every state. The sample is selected in different stages. In the first stage of the sampling, the U.S is divided into 792 primary sampling units (PSUs). Every PSU falls within the boundary of a state and is comprised of a metropolitan area or a large county, or a group of smaller counties. They are then grouped into strata. The strata are constructed so that they are as homogeneous as possible with respect to the labor force and other social and economic characteristics. One PSU is randomly selected per stratum. In the second stage of sampling, a sample of four housing units within the sample PSUs is drawn. A selected housing unit is interviewed for 4 consecutive months dropped from of the sample for the
next 8 months, and is brought back in the following 4 months and then dropped from the sample. From year to year the sample size of the CPS varies but the size has been significantly increased since 2002 to improve state-level estimates. The CPS sample is based on the civilian non-institutional population of the United States. A person 15 years old or over residing in a scientifically selected household is interviewed to collect information about the whole household.

Table 3.1 presents the number of observations used in the estimation for all three variables. The first column of the table explains the variable names and survey years for each variable. In subsequent columns, the number of observations for Hawai‘i and the United States are reported. The last two columns of the table present matched and unmatched number of observations for the United States. The sample is not equally distributed between the before and after periods; the after period sample is much larger than the before period because of many more years and much larger sample size in recent years.

<table>
<thead>
<tr>
<th>Table 3.1: Sample size used in this study from CPS 1963-2004</th>
<th>Hawaii</th>
<th>US(matched)</th>
<th>US(unmatched)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before PHCA (Last Week All Jobs: 1963-67)</td>
<td>356</td>
<td>2190</td>
<td>144287</td>
</tr>
<tr>
<td>After PHCA (Last Week All Jobs: 1977-2004)</td>
<td>14654</td>
<td>69082</td>
<td>1507600</td>
</tr>
<tr>
<td>Last Week Main Job: 1977-2004</td>
<td>7483</td>
<td>33843</td>
<td>710525</td>
</tr>
<tr>
<td>Last Year All Jobs (single employer): 1977-2004</td>
<td>16833</td>
<td>91814</td>
<td>1759087</td>
</tr>
</tbody>
</table>

One-to-one matching means selecting a single observation from Hawai‘i to match each observation from the rest of the United States by year, age, sex, race, marital status, education, wage and industry. There were a total of 33 years from 1963 to 2004 excluding 9 years from 1968 to 1976 when Hawai‘i’s data were grouped with other western states. Nine age categories were used: 19-24, 25-29, 30-34, 35-39, 40-44, 45-49,
50-54, 55-59 and 60-64. There were two categories of sex: male and female. Two categories of race, white and non-white, were used. Two categories of marital status, married and unmarried, were used. There were six categories of education: elementary, middle, high school, some college education and college graduates. Thirteen wage categories - 1-4999, 5000-9999, 10000-14999, 15000-19999, 20000-24999, 25000-29999, 30000-34999, 35000-39999, 40000-44999, 45000-49999, 50000-54999, 55000-59999, 60000 and over, were used. There were fourteen industry categories: 1) agriculture, 2) mining, 3) construction, 4) manufacturing, 5) trade, 6) transportation and utility, 7) information, 8) financial activity, 9) professional and business services, 10) educational and health services, 11) leisure and hospitality services, 12) other services, 13) public administration, 14) armed Forces.

3.5. Empirical Results
The number of hours per week workers in the United States work at their paying job is examined by three variables used in the March CPS: “hours”, “emush” and “hrslyr”. The variable “hours” reports the usual weekly hours worked on all jobs held during the survey’s reference period; “emush” is the variable which represents the usual weekly hours worked on the main job held during the survey’s reference period, and the variable “hrslyr” reports the usual weekly hours worked on all jobs held during last year. All three variables are based on workers’ reports on hours worked, not the employers’ reports on the employees’ hours of work.
Non-exempt workers are private workers who usually work 20 or more hours a week. In contrast, exempt workers are private workers who usually work less than 20 hours per weeks, self-employed workers or government employees. There were 59.3% non-exempt workers in Hawai‘i before the Prepaid Health Care Act. That percentage was substantially lower than the nationwide figure of 65.7% in matched data and 69.6% in unmatched data. Over time, this has increased to 64.8% in Hawai‘i which is a 5.5% increase. The number of non-exempt workers has decreased in the rest of the United States to 65.1% in matched data and slightly increased to 70.4% in unmatched data after the Prepaid Health Care Act. This is 5.5% increase in the non-exempt workers in Hawai‘i and marginal decrease by a 0.6% in the rest of the United States in matched data and a 0.8% increase in the rest of the United States in unmatched data. In other words, there is a significant change in Hawai‘i across the exempt and non-exempt sectors while there is no significant change in the rest of the United States in matched and unmatched data. These results indicate that the increase in the proportion of non-exempt workers might be due to the Prepaid Health Care Act. For more details see table 3.2.
Table 3.2: Distribution of Workers in Hawaii and the rest of the United States by employment status

<table>
<thead>
<tr>
<th></th>
<th>Hawaii weighted No.</th>
<th>US (matched) weighted No.</th>
<th>US (unmatched) weighted No.</th>
<th>Percentage</th>
<th>Percentage</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before PHCA (Last Week all Jobs CPS: 1963-67)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt</td>
<td>372,656</td>
<td>1,368,094</td>
<td>105,946,560</td>
<td>40.71%</td>
<td>34.34%</td>
<td>30.41%</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>68,693</td>
<td>372,398</td>
<td>32,108,397</td>
<td>7.50%</td>
<td>9.35%</td>
<td>9.22%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>474,064</td>
<td>2,243,717</td>
<td>210,319,559</td>
<td>51.79%</td>
<td>56.32%</td>
<td>60.37%</td>
</tr>
<tr>
<td>Total</td>
<td>915,403</td>
<td>3,984,209</td>
<td>348,374,516</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>After PHCA (Last Week all Jobs CPS: 1977-2004)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt</td>
<td>4,470,439</td>
<td>37,612,662</td>
<td>850,550,698</td>
<td>35.21%</td>
<td>34.90%</td>
<td>29.64%</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>1,562,339</td>
<td>17,307,838</td>
<td>374,864,641</td>
<td>12.30%</td>
<td>16.06%</td>
<td>13.06%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>6,664,859</td>
<td>52,844,123</td>
<td>1,644,000,000</td>
<td>52.49%</td>
<td>49.04%</td>
<td>57.29%</td>
</tr>
<tr>
<td>Total</td>
<td>12,697,637</td>
<td>107,764,623</td>
<td>2,869,400,000</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Last Week main Job CPS: 1977-2004</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt</td>
<td>2,219,504</td>
<td>18,192,067</td>
<td>397,566,719</td>
<td>34.49%</td>
<td>33.50%</td>
<td>27.21%</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>702,922</td>
<td>8,177,123</td>
<td>157,103,872</td>
<td>10.92%</td>
<td>15.06%</td>
<td>10.75%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>3,512,173</td>
<td>27,936,193</td>
<td>906,220,197</td>
<td>54.58%</td>
<td>51.44%</td>
<td>62.03%</td>
</tr>
<tr>
<td>Total</td>
<td>6,434,599</td>
<td>107,764,623</td>
<td>1,460,900,000</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Last Week all Jobs (single employer, CPS: 1977-2004)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt</td>
<td>4,159,138</td>
<td>34,175,705</td>
<td>776,475,273</td>
<td>33.42%</td>
<td>32.65%</td>
<td>27.49%</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>1,381,132</td>
<td>16,698,346</td>
<td>311,429,284</td>
<td>11.10%</td>
<td>15.95%</td>
<td>11.03%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>6,905,643</td>
<td>53,789,528</td>
<td>1,736,200,000</td>
<td>55.49%</td>
<td>51.39%</td>
<td>61.48%</td>
</tr>
<tr>
<td>Total</td>
<td>12,445,912</td>
<td>104,663,579</td>
<td>2,824,100,000</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

The difference-in-difference estimates cannot be obtained for the other two variables—the main job last week and all jobs last year—since both the variables were introduced after the Prepaid Health Care Act. However, the pattern of labor force utilization is very similar in all three variables. In Hawai‘i an increase in the proportion of workers who work 20 to 35 hours per week is observed after the Prepaid Health Care Act from 7.5% to 12.3%. The proportion of this group in the rest of the United States has increased even more than in Hawai‘i. The slower growth in this labor force group in Hawai‘i might be due to the Prepaid Health Care Act.

The following group of six pie charts given in figure 3.2 graphically represents these estimates for the years 1963-1967 and 1978-2004 for Hawai‘i and rest of the United
States with matched and unmatched data. The pie charts compare estimates of exempt and non-exempt sectors in Hawai'i and the rest of the United States before and after the Prepaid Health Care Act was implemented. These pie charts indicate that there is moderate labor sorting from exempt to non-exempt sectors because of the Prepaid Health Care Act. For more details see appendix A of this chapter.
Similarly, the proportion of exempt workers in Hawai'i has decreased from 40.7% to 35.2%. The proportion of exempt-type workers in the rest of the United States has marginally increased from 34.3% to 34.9% in matched data and marginally decreased from 30.4% to 29.6% in unmatched data during the same time period. Therefore, it can be concluded from the previous results that moderate labor sorting from the exempt to the non-exempt sector has occurred in Hawai'i because of the Prepaid Health Care Act.

The percentage of all workers employed less than 20 hours per week has increased from 5.8% to 6.8% after the Prepaid Health Care Act a 1.0% increase. At the same time, the percentage of all workers employed less than 20 hours per week in the rest of the United States has increased from 7.6% to 9.4% in matched data and 6.3% to 7% in unmatched data during the same time. This increase is 1.8% and 0.7% in the rest of the United States in matched and unmatched data respectively. The growth in all workers employed less than 20 hours per week is 40% higher in the rest of the United States than in Hawai'i in matched data.

Figure 3.3 presents a distribution of private workers in Hawai'i before and after the Prepaid Health Care Act. The percentage of private workers in Hawai'i who work less than 20 hours per week has decreased slightly from 6.1% to 6.0% after the Prepaid Health Care Act which is a 0.1% decrease.
Figure 3.3: Distribution of hours worked in Hawai‘i by the private sector workers age 19 to 64 before and after the Prepaid Health Care Act.

The percentage of government workers in Hawai‘i who are employed less than 20 hours per week has marginally increased from 6.46% to 7.01% after the Prepaid Health Care Act, a 0.55% increase. After the Prepaid Health Care Act, the percentage of full-time government workers decreased slightly. The average number of hours worked per week by full-time government workers in Hawai‘i remained the same at 40.1 hours per week. There was no significant change in the distribution of government employees in Hawai‘i after the Prepaid Health Care Act. For more details see appendix 1A of this chapter.

The percentage of self-employed workers who worked less than 20 hours per week has increased from 11.74% to 13.93% after the Prepaid Health Care Act which is a 2.2% increase. In addition, the percentage of full-time self-employed workers has decreased by
the same amount. The average number of hours worked per week by full-time workers has slightly decreased from 43.8 hours to 43 hours per week.

The percentage of all workers employed 20 to 35 hours per week has increased from 7.5% to 12.3%, a 4.8 percent point increase for this type of worker. At the same time, the percentage of all workers employed for 20 to 35 hours per week in the rest of the United States has increased from 9.4% to 16.1% during the same time a 6.7% increase. Therefore, the rate of change in this type of worker in the rest of the United States is higher than in Hawai'i. Economic theory predicts the opposite. It predicts that the employer will hire more part-time workers to avoid the health insurance cost or hire more workers who work more than 35 hours per week to reduce the per hour fixed benefit cost.

After the Prepaid Health Care Act, the percentage of full-time workers increased. The average number of hours worked per week by full-time workers increased slightly from 40.8 hours to 41 hours per week.

Table 3.3 presents the direct differences between Hawai'i and the rest of the United States as well as the multinomial logit difference regression estimates. Direct estimates of differences show that there are more exempt workers in Hawai'i by 0.3 percentage points in matched data and 5.6 percentage points in unmatched data as compared to the rest of the United States. Model-based estimates of differences show that the percentage of exempt workers in Hawai'i is higher by 1.9 percentage points in matched data and 3.7 percentage points in unmatched data as compared to the rest of the United States. For all
workers employed 20 to 35 hours per week the percentage in Hawai‘i is lower: 3.8 percentage points in matched data and 0.8 percentage points in unmatched data. The number of workers who work 36 hours per week is higher in Hawai‘i by 3.5 percentage points in matched data while unmatched data shows a 4.8 percentage points lower than the rest of the United States. Model-based results are very close to the direct estimates for Hawai‘i and the rest of the United States. Both of the other variables show similar patterns. For more details see table 3.3.

Table 3.3: Difference between Hawaii and the rest of the United States (all Workers)

<table>
<thead>
<tr>
<th>HI-US Difference (Matched Data)</th>
<th>HI-US Difference (Unmatched Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Week all Jobs CPS:1977-2004</td>
<td>Direct Estimates</td>
</tr>
<tr>
<td>Exempt</td>
<td>0.31%</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>-3.76%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>3.45%</td>
</tr>
<tr>
<td>Last Week main Job CPS:1977-2004</td>
<td>Exempt</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>-4.14%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>3.14%</td>
</tr>
<tr>
<td>Last Year all Jobs (single employer, CPS:1977-2004)</td>
<td>Exempt</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>-4.85%</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>4.10%</td>
</tr>
</tbody>
</table>

Table 3.4 presents the direct differences and multinomial logit difference regression estimates between Hawai‘i and the rest of the United States for private workers only. Direct estimates of the difference show that the number of private workers who work less than 20 hours per week in Hawai‘i are lower than the rest of the United States by 2.87 percentage points in matched data and 0.05 percentage points in unmatched data. Model-based estimates of difference show that private workers who work less than 20 hours per week in Hawai‘i are fewer than the rest of the United States by 2.17 percentage points in matched data and 0.55 percentage points higher in unmatched data. The percentage of
private workers employed 20 to 35 hours per week has also decreased by 3.97 percentage points in matched data while unmatched data shows a slight increase in this group. The number of workers who work 36 hours per week or more in Hawai‘i is higher by 6.82 percentage points in direct difference estimates and 5.51 percentage points in model-based estimates of matched data while unmatched data shows a slight decrease in this group. The direct and model-based results for other both variables are very close to the first variable. For more details see table 3.4.

<p>| Table 3.4: Difference between Hawaii and the remainder of the United States (Private Workers) |
|---------------------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Hi-US Difference(Matched Data) | Hi-US Difference(Unmatched Data) |</p>
<table>
<thead>
<tr>
<th>Direct Estimates</th>
<th>Model Based</th>
<th>Direct Estimates</th>
<th>Model Based</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Last Week all Jobs CPS:1977-04</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hours 1-19</td>
<td>-2.87%</td>
<td>-2.17%</td>
<td>-0.05%</td>
</tr>
<tr>
<td>hours 20-35</td>
<td>-3.97%</td>
<td>-3.33%</td>
<td>0.49%</td>
</tr>
<tr>
<td>Full-time 36+</td>
<td>6.82%</td>
<td>5.51%</td>
<td>-0.46%</td>
</tr>
<tr>
<td><strong>Last Week main Job CPS:1977-04</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hours 1-19</td>
<td>-2.85%</td>
<td>-1.50%</td>
<td>0.32%</td>
</tr>
<tr>
<td>hours 20-35</td>
<td>-4.79%</td>
<td>-3.37%</td>
<td>1.37%</td>
</tr>
<tr>
<td>Full-time 36+</td>
<td>7.66%</td>
<td>4.87%</td>
<td>-1.66%</td>
</tr>
<tr>
<td><strong>Last Year all Jobs (single employer, CPS:1977-04)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hours 1-19</td>
<td>-2.42%</td>
<td>-1.53%</td>
<td>0.26%</td>
</tr>
<tr>
<td>hours 20-35</td>
<td>-5.58%</td>
<td>-4.50%</td>
<td>1.15%</td>
</tr>
<tr>
<td>Full-time 36+</td>
<td>7.99%</td>
<td>6.03%</td>
<td>-1.41%</td>
</tr>
</tbody>
</table>

A simple difference-in-difference analysis is used to calculate the effect of the Prepaid Health Care Act on the exempt and non-exempt sectors and hours worked. State characteristics change over time, so the difference of the before and after period for Hawai‘i minus the difference of the before and after period for the other 49 states yields the difference-in-difference estimate. These estimates are given in table 5. The results might be biased if the time trend for hours worked in Hawai‘i and the rest of the United States is not the same.
Table 3.5, which follows, presents the direct difference-in-difference estimates and pooled multinomial logit difference-in-difference regression estimates for matched and unmatched data. The regression analysis allows the model to control for demographic and other factors, such as age, gender, race, wage rate, union, establishment size, industry and year which are not controlled for in the direct difference-in-difference estimates. The impact of the Prepaid Health Care Act is evaluated by estimating the following multinomial logit regression:

\[ \text{Empstatus} = \beta_0 + \beta_1(\text{Hlbefore}) + \beta_2(\text{Hlafter}) + \beta_2(\text{USafter}) + X\delta + \varepsilon \]  

Where “Empstatus” has three categories; workers who are exempt from the Prepaid Health Care Act, those who work 20-35 hours per week, and 36 hours or more worked. Hlbefore is an indicator variable for the state of Hawai'i before the Prepaid Health Care Act was implemented. “Hlafter” is an indicator variable for the state of Hawai'i after the Prepaid Health Care Act was implemented. The “USafter” variable is an indicator variable for all 49 states of the United States except Hawai'i. The variable X includes all the other regressors, such as age, age square, education level, gender, race, wage income, 32 year dummies and 13 industry dummies.

### Table 3.5: Difference-in-Difference Estimates of Hours Worked in Hawaii (All workers)

<table>
<thead>
<tr>
<th>The remainder of the United States as a control group</th>
<th>Difference-in-Difference Estimates</th>
<th>Difference-in-Difference Model Based Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Matched</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt hours 20-35</td>
<td>-6.06%</td>
<td>-8.05%</td>
</tr>
<tr>
<td>Exempt Full-time 36+</td>
<td>-1.91%</td>
<td>-0.45%</td>
</tr>
<tr>
<td></td>
<td>7.97%</td>
<td>8.50%</td>
</tr>
<tr>
<td><strong>Unmatched</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exempt hours 20-35</td>
<td>-4.73%</td>
<td>-4.41%</td>
</tr>
<tr>
<td>Exempt Full-time 36+</td>
<td>0.96%</td>
<td>4.30%</td>
</tr>
<tr>
<td></td>
<td>3.78%</td>
<td>0.11%</td>
</tr>
</tbody>
</table>

Table 3.5 presents direct and model-based estimates of hours worked in Hawai'i by all workers. Matched and unmatched data is used for this estimation. The second column of
Table 3.5 presents direct difference-in-difference estimates while the final column shows a model-based difference-in-difference estimate of the Prepaid Health Care Act on hours worked in Hawai‘i. The direct difference-in-difference estimates of matched data show that the percentage of exempt workers has decreased by 6.06 percentage points in Hawai‘i. The model-based difference-in-difference estimates also show an 8.05 percentage point decrease in the exempt workers in Hawai‘i. The results also show that workers who work from 20 to 35 hours per week has decreased by 1.91 percentage points in direct estimates and by 0.45 percentage points in model-based estimates. Meanwhile, the proportion of full-time workers has increased by 7.98 percentage points in the direct estimates and 8.5 percentage points in the model-based estimates. The results for the unmatched data indicate unstable signs and are smaller in magnitude. For more details see the last three rows of Table 3.5. These findings confirm the hypothesis that the Prepaid Health Care Act will decrease the proportion of exempt workers in Hawai‘i. The results also show that non-exempt workers have increased more in Hawai‘i as compared to the other 49 states.

Table 3.6 presents direct and model-based estimates of hours worked in Hawai‘i by private workers. Matched and unmatched data is used for this estimation. The second column of Table 3.6 presents direct difference-in-difference estimates while the final column shows the model-based difference-in-difference estimate of the Prepaid Health Care Act’s impact on hours worked in Hawai‘i. The direct difference-in-difference estimates of matched data show that the percentage of private workers who work less than 20 hours per week decreased by 1.35 percentage points in Hawai‘i. The model-based
difference-in-difference estimates also show a 1.58 percentage point decrease in private workers who work less than 20 hours per week in Hawai‘i. The results also show that the percentage of workers who work from 20 to 35 hours per week has decreased by 2.92 percentage points in direct estimates and a 1.01 percentage points in model-based estimates while the proportion of full-time workers has increased by 4.26 percentage points in the direct estimates and 2.59 percentage points in the model based-estimates. The results of unmatched data are not stable. For more details see Table 3.6. These findings confirm the conclusion from Tables 3.3 and 3.4 that the Prepaid Health Care Act has modestly decreased the most expensive labor group (20-35 hours work per week).

<table>
<thead>
<tr>
<th>Table 3.6: Difference-in-Difference Estimates of Hours Worked in Hawaii (Private workers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The rest of the United States as a control group</td>
</tr>
<tr>
<td><strong>Matched</strong></td>
</tr>
<tr>
<td>Hours 1-19</td>
</tr>
<tr>
<td>Hours 20-35</td>
</tr>
<tr>
<td>Full-time 36+</td>
</tr>
<tr>
<td><strong>Unmatched</strong></td>
</tr>
<tr>
<td>Hours 1-19</td>
</tr>
<tr>
<td>Hours 20-35</td>
</tr>
<tr>
<td>Full-time 36+</td>
</tr>
</tbody>
</table>

The multinomial logit difference-in-difference model specification used in Table 3.6 is the same as used in Table 3.5 and explained above.

### 3.6. Comparison with other Studies
There are only two major studies which have examined the effect of the Prepaid Health Care Act on hours worked in Hawai‘i. The first study was an empirical study conducted by Thurston (1997). He examined the effects of the Prepaid Health Care Act on
employment, wages and health insurance coverage. He used two different data sets, the
U.S. Censuses of Population and Housing for the year 1970 and Current Population
Surveys March supplement for the years 1990-93. He used the year 1970 Census data for
the pre-program period calculations and the Current Population Surveys March
supplement data for the post-program period calculations. He concluded in his empirical
paper that the number of part-time workers in Hawai‘i has increased significantly after
the Prepaid Health Care Act was implemented. These two data sets are collected for
different purposes and use different definitions of hours worked. Also Census labor
market data is considered inferior to the Current Population Surveys. This is why his
conclusion might be affected.

The second major theoretical study later supported by experimental data was conducted
by Sherstyuk (2004). She explained the theoretical rationale for why the Hawai‘i-type
partial mandate would cause an increase in the number of part-time workers. She later
confirmed her predictions with experimental data. She made a number of simplifying
assumptions which are not true in the real world. This may be why her predictions and
experimental results are contradicted in this empirical study.

The results of this study indicate that the overall number of private workers who work
less than 20 hours per week has decreased by 1.35 percentage points using direct
estimates and 1.58 percentage points by using the model-based estimates. At the same
time, the number of self-employed individuals who work less than 20 hours per week has
increased by more than 2 percentage points. This partially cancels out the overall effect
on the number of part-time workers who work less than 20 hours per week. Since there are three times as many private workers as self-employed workers, the impact is not fully canceled out. However, the number of part-time government workers has marginally increased by 0.5%. The reason for inconsistency with Thurston (1997) might be due to the length of data. Thurston (1997) used only four years data from 1990 to 1993 while this study has used 33 years of data from 1963 to 2004 excluding 1968 to 1976.

Sherstyuk (2004) assumed that the labor market is competitive and there is no productivity loss from a worker working two part-time jobs or one full time job. The Hawai’i-type partial mandate would reduce employers’ costs by hiring two part-time workers instead of one full-time worker, since part-time workers are not eligible for health insurance benefits. In the real world, the labor market is not perfect and part-time workers are not as productive as full time workers. Therefore, contradictory results are possible.

3.7. Further Research
This study assumed that hours worked are exogenously determined. In the real world, hours worked cannot be analyzed in isolation from wage rate and health insurance benefits. It should also be examined in the context of employment decisions, which determine the worker’s opportunity set for obtaining health insurance benefits. A simultaneous equation model should be generated which will consider the employment decision and equations for the wage rate and hours worked.
3.8. Conclusion
This essay examines whether the Prepaid Health Care Act of 1974 has caused bimodal labor force utilization and labor sorting from the exempt to the non-exempt sector in Hawai‘i. Simple tabulation and multinomial logit regression model are used with matched and unmatched data to examine these concerns. The matched data method is preferred over the unmatched data method because it removes heterogeneity and bias from the estimates and has relatively more stable results. A difference-in-difference analysis also confirms that the theoretical predictions regarding the Prepaid Health Care Act cause bimodality in the hours worked and cause labor sorting from the exempt sector to the non-exempt sector in Hawai‘i. If the time trend for hours worked in Hawaii and the rest of the United States is not similar, difference-in-difference analyses might give biased results.

After the Prepaid Health Care Act was implemented, the overall distribution of the average number of hours worked changed moderately. Also, the direct difference-in-difference estimates from matched data show that the number of part-time private workers decreased by 1.35 percentage points in Hawai‘i. Similarly, the model-based difference-in-difference estimates from matched data show that the decrease is 1.58 percentage points which is slightly bigger than the direct difference-in-difference estimates. The unmatched data also show the same pattern but the model-based estimates are a little different. For more details see Table 3.6. This result contradicts the theoretical prediction and empirical findings of Thurston (1997). However, the number of self-employed individuals who work less than 20 hours per week has increased by 2.19%. Government workers who worked less than 20 hours per week have slightly increased
during the same period (by 0.55%). Therefore, the estimates from unmatched data for overall part-time workers show a slight decrease in the number of workers who work less than 20 hours per week.

The difference-in-difference estimates show that Hawai'i has a significantly lower proportion of exempt workers than the rest of the United States. The proportion of exempt workers in Hawai'i decreased from 40.71% to 35.21% after the Prepaid Health Care Act was implemented, while this proportion did not change more than one percent in matched and unmatched data for the rest of the United States. Therefore, it can be concluded that labor sorting from exempt to non-exempt sector has occurred in Hawai'i because of the Prepaid Health Care Act.

The percentage of all workers employed 20 to 35 hours per week has increased from 7.5% to 12.3%, a 4.8% increase in these types of workers. At the same time, the percentage of all workers employed for 20 to 35 hours per week in the rest of the United States has also increased from 9.4% to 16.1% during the same time, a 6.7% increase. The percentage of all private workers employed for 20 to 35 hours per week in Hawai'i as compared to the rest of the United States has decreased by 2.9% in direct estimates and 1.01% in the model based estimates in the matched data. A similar but unstable pattern is observed in unmatched data.
3.9. Bibliography


3.10. Appendix A

Figure 3.1A: Distribution of hours worked in Hawai’i by the government employees age 19 to 64 before and after the Prepaid Health Care Act.

Figure 3.2A: Distribution of hours worked in Hawai’i by the self-employed age 19 to 64 before and after the Prepaid Health Care Act.
Figure 3.3A: Distribution of hours worked in Hawai‘i and the United States by the private sector workers age 19 to 64 before the Prepaid Health Care Act.

Figure 3.4A: Distribution of hours worked in Hawai‘i by the private sector workers age 19 to 64 after the Prepaid Health Care Act.
Figure 3.5A: Distribution of hours worked in Hawai'i and United States by the self-employed age 19 to 64 before the Prepaid Health Care Act.

Figure 3.6A: Distribution of hours worked in Hawai'i and the United States by the self-employed age 19 to 64 after the Prepaid Health Care Act.
### Table 3.1A: Multinomial Logit Regression for Hawaii (Matched Data)

| empstatus | Variable           | Coefficient | Std. Err. | z     | P>|z|  | Lower lim | Upper lim |
|-----------|--------------------|-------------|-----------|-------|-----|-----------|------------|
| Hours 20-35 | Age                | 0.1452      | 0.0269    | 5.40  | 0.00 | 0.0925    | 0.1979     |
|           | Age squared        | -0.0018     | 0.0003    | -5.38 | 0.00 | -0.0024   | -0.0011    |
|           | college            | -0.4685     | 0.0964    | -4.86 | 0.00 | -0.6575   | -0.2796    |
|           | male               | -0.0481     | 0.0941    | -0.51 | 0.61 | -0.2326   | 0.1363     |
|           | Currently married  | -0.0066     | 0.1047    | -0.06 | 0.95 | -0.2119   | 0.1986     |
|           | Non-white          | -0.0143     | 0.1033    | -0.14 | 0.89 | -0.2168   | 0.1882     |
|           | Union              | 0.7041      | 0.3341    | 2.11  | 0.04 | 0.0492    | 1.3590     |
|           | Union-unknown      | 0.3163      | 0.1258    | 2.51  | 0.01 | 0.0697    | 0.5629     |
|           | Constant           | -2.1353     | 0.5797    | -3.68 | 0.00 | -3.2715   | -0.9992    |
| Hours 36+  | Age                | 0.3292      | 0.0243    | 13.53 | 0.00 | 0.2815    | 0.3769     |
|           | Age squared        | -0.0038     | 0.0003    | -13.00| 0.00 | -0.0044   | -0.0033    |
|           | college            | -0.2654     | 0.0876    | -3.03 | 0.00 | -0.4372   | -0.0937    |
|           | male               | 0.6587      | 0.0849    | 7.76  | 0.00 | 0.4922    | 0.8251     |
|           | Currently married  | 0.1949      | 0.0961    | 2.03  | 0.04 | 0.0065    | 0.3832     |
|           | Non-white          | 0.1716      | 0.0952    | 1.80  | 0.07 | -0.0150   | 0.3582     |
|           | Union              | 0.7542      | 0.3169    | 2.38  | 0.02 | 0.1330    | 1.3754     |
|           | Union-unknown      | 0.2988      | 0.1164    | 2.57  | 0.01 | 0.0707    | 0.5268     |
|           | Constant           | -4.5284     | 0.5226    | -8.66 | 0.00 | -5.5527   | -3.5041    |

**Predicted values**
- Exempt: 0.04993
- Hours 20-35: 0.16468
- Hours 36+: 0.77539

### Table 3.2A: Multinomial Logit Regression for the United States (Matched Data)

| empstatus | Variable           | Coefficient | Std. Err. | z     | P>|z|  | Lower lim | Upper lim |
|-----------|--------------------|-------------|-----------|-------|-----|-----------|------------|
| Hours 20-35 | Age                | 0.1137      | 0.0144    | 7.92  | 0.00 | 0.0856    | 0.1419     |
|           | Age squared        | -0.0013     | 0.0002    | -7.28 | 0.00 | -0.0017   | -0.0010    |
|           | college            | -0.3901     | 0.0452    | -8.63 | 0.00 | -0.4786   | -0.3015    |
|           | male               | 0.1468      | 0.0463    | 3.17  | 0.00 | 0.0561    | 0.2375     |
|           | Currently married  | -0.1489     | 0.0563    | -2.65 | 0.01 | -0.2592   | -0.0386    |
|           | Non-white          | 0.0248      | 0.0730    | 0.34  | 0.73 | -0.1183   | 0.1679     |
|           | Union              | 0.0094      | 0.1869    | 0.05  | 0.96 | -0.3570   | 0.3758     |
|           | Union-unknown      | -0.1275     | 0.0548    | -2.33 | 0.02 | -0.2349   | -0.0201    |
|           | Constant           | -1.4296     | 0.2682    | -5.33 | 0.00 | -1.9552   | -0.9040    |
| Hours 36+  | Age                | 0.3140      | 0.0133    | 23.53 | 0.00 | 0.2879    | 0.3402     |
|           | Age squared        | -0.0037     | 0.0002    | -21.58| 0.00 | -0.0040   | -0.0033    |
|           | college            | -0.4515     | 0.0417    | -10.82| 0.00 | -0.5332   | -0.3897    |
|           | male               | 1.3317      | 0.0414    | 32.14 | 0.00 | 1.2505    | 1.4129     |
|           | Currently married  | -0.1334     | 0.0539    | -2.48 | 0.01 | -0.2389   | -0.0278    |
|           | Non-white          | 0.0705      | 0.0671    | 1.05  | 0.29 | -0.0610   | 0.2021     |
|           | Union              | 0.0148      | 0.1717    | 0.09  | 0.93 | -0.3218   | 0.3514     |
|           | Union-unknown      | 0.1092      | 0.0506    | 2.16  | 0.03 | 0.0100    | 0.2084     |
|           | Constant           | -4.4804     | 0.2487    | -18.01| 0.00 | -4.9679   | -3.9930    |

**Predicted values**
- Exempt: 0.071648
- Hours 20-35: 0.197991
- Hours 36+: 0.72026

---

173
Table 3.3A: Multinomial Logit Regression for Hawaii (Matched Data)

| empstatus | Variable   | Coefficient | Std. Err. | z      | P>|z| | Lower lim | Upper lim |
|-----------|------------|-------------|-----------|--------|-------|-----------|-----------|
| Hours 20-35 | Age        | 0.2296      | 0.0427    | 5.37   | 0.00  | 0.1458    | 0.3134    |
|           | Age square | -0.00027    | 0.0005    | -5.20  | 0.00  | -0.0037   | -0.0017   |
|           | college    | -0.6574     | 0.1597    | -4.12  | 0.00  | -0.9704   | -0.3443   |
|           | male       | -0.2092     | 0.1504    | -1.39  | 0.16  | -0.5040   | 0.0857    |
|           | Currently married | -0.3185 | 0.1574 | -2.02 | 0.04 | -0.6270 | -0.0099 |
|           | Non-white  | -0.0178     | 0.1699    | -0.11  | 0.92  | -0.3449   | 0.3093    |
|           | Union      | 2.3810      | 1.0192    | 2.34   | 0.02  | 0.3833    | 4.3786    |
|           | Union-unknown | -0.1086 | 0.3138 | -0.35 | 0.73 | -0.7236 | 0.5064 |
|           | Constant   | -2.3445     | 0.7160    | -3.27  | 0.00  | -3.7478   | -0.9412   |
| Hours 36+  | Age        | 0.4956      | 0.0396    | 12.51  | 0.00  | 0.4179    | 0.5733    |
|           | Age square | -0.00057    | 0.0005    | -11.85 | 0.00  | -0.0067   | -0.0048   |
|           | college    | -0.5522     | 0.1503    | -3.67  | 0.00  | -0.8467   | -0.2576   |
|           | male       | 0.6073      | 0.1379    | 4.40   | 0.00  | 0.3370    | 0.8777    |
|           | Currently married | -0.0672 | 0.1467 | -0.46 | 0.65 | -0.3546 | 0.2203 |
|           | Non-white  | 0.3102      | 0.1585    | 1.96   | 0.05  | -0.0006   | 0.6209    |
|           | Union      | 2.7923      | 1.0065    | 2.77   | 0.01  | 0.8196    | 4.7650    |
|           | Union-unknown | 0.2648 | 0.2837 | 0.93  | 0.35 | -0.2913 | 0.8209 |
|           | Constant   | -6.8542     | 0.6675    | -10.27 | 0.00  | -8.1625   | -5.5458   |

Predicted values

<table>
<thead>
<tr>
<th>empstatus</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exempt</td>
<td>0.03902</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>0.13924</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>0.81984</td>
</tr>
</tbody>
</table>

Table 3.4A: Multinomial Logit Regression for the United States (Matched Data)

| empstatus | Variable   | Coefficient | Std. Err. | z      | P>|z| | Lower lim | Upper lim |
|-----------|------------|-------------|-----------|--------|-------|-----------|-----------|
| Hours 20-35 | Age        | 0.1475      | 0.0216    | 6.82   | 0.00  | 0.1051    | 0.1899    |
|           | Age square | -0.0017     | 0.0003    | -6.10  | 0.00  | -0.0022   | -0.0011   |
|           | college    | -0.4489     | 0.0694    | -6.47  | 0.00  | -0.5849   | -0.3129   |
|           | male       | -0.0156     | 0.0734    | -0.21  | 0.83  | -0.1595   | 0.1263    |
|           | Currently married | -0.3153 | 0.0869 | -3.63 | 0.00 | -0.4857 | -0.1449 |
|           | Non-white  | -0.0139     | 0.0976    | -0.14  | 0.89  | -0.2052   | 0.1775    |
|           | Union      | -0.0326     | 0.2349    | -0.14  | 0.89  | -0.4929   | 0.4277    |
|           | Union-unknown | -0.2471 | 0.1360 | -1.82 | 0.07 | -0.5136 | 0.0195 |
|           | Constant   | -1.2327     | 0.3425    | -3.60  | 0.00  | -1.9040   | -0.5614   |
| Hours 36+  | Age        | 0.4414      | 0.0206    | 21.41  | 0.00  | 0.4010    | 0.4818    |
|           | Age square | -0.0051     | 0.0003    | -19.74 | 0.00  | -0.0056   | -0.0046   |
|           | college    | -0.6732     | 0.0660    | -10.21 | 0.00  | -0.8025   | -0.5440   |
|           | male       | 1.3557      | 0.0655    | 20.70  | 0.00  | 1.2273    | 1.4841    |
|           | Currently married | -0.3559 | 0.0863 | -4.12 | 0.00 | -0.5251 | -0.1868 |
|           | Non-white  | 0.1169      | 0.0901    | 1.30   | 0.19  | -0.0597   | 0.2934    |
|           | Union      | 0.2368      | 0.2125    | 1.11   | 0.27  | -0.1796   | 0.6532    |
|           | Union-unknown | 0.2210 | 0.1241 | 1.78  | 0.08 | -0.0222 | 0.4641 |
|           | Constant   | -6.0349     | 0.3285    | -18.37 | 0.00  | -6.6787   | -5.3911   |

Predicted values

<table>
<thead>
<tr>
<th>empstatus</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exempt</td>
<td>0.05403384</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>0.17299134</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>0.77118482</td>
</tr>
</tbody>
</table>
### Table 3.5A: Multinomial Logit Regression for Hawaii (Matched Data)

| empstatus  | Variable         | Coefficient | Std. Err. | z     | P>|z| | Lower lim | Upper lim |
|------------|------------------|-------------|-----------|-------|-----|----------|-----------|
| Hours 20-35| Age              | 0.2107      | 0.0335    | 6.29  | 0.00 | 0.1450   | 0.2763    |
|            | Age square       | -0.0026     | 0.0004    | -6.36 | 0.00 | -0.0034  | -0.0018   |
|            | college          | -0.4152     | 0.1179    | -3.52 | 0.00 | -0.6462  | -0.1841   |
|            | male             | 0.1539      | 0.1195    | 1.29  | 0.20 | -0.0803  | 0.3882    |
|            | Currently married| -0.1375     | 0.1231    | -1.12 | 0.26 | -0.3789  | 0.0103    |
|            | Non-white        | -0.0561     | 0.1254    | -0.45 | 0.66 | -0.3019  | 0.1897    |
|            | Union             | 1.8834      | 0.5349    | 3.52  | 0.00 | 0.8350   | 2.9317    |
|            | Union-unknown    | 0.2681      | 0.1518    | 1.77  | 0.08 | -0.0294  | 0.5657    |
|            | Constant         | -2.0779     | 0.5764    | -3.60 | 0.00 | -3.2077  | -0.9482   |
| Hours 36+  | Age              | 0.4505      | 0.0315    | 14.31 | 0.00 | 0.3888   | 0.5122    |
|            | Age square       | -0.0053     | 0.0004    | -13.88| 0.00 | -0.0060  | -0.0045   |
|            | college          | -0.3233     | 0.1106    | -2.92 | 0.00 | -0.5401  | -0.1066   |
|            | male             | 0.9968      | 0.1114    | 8.94  | 0.00 | 0.7784   | 1.2152    |
|            | Currently married| 0.1394      | 0.1159    | 1.20  | 0.23 | -0.0877  | 0.3665    |
|            | Non-white        | 0.1147      | 0.1189    | 0.96  | 0.34 | -0.1184  | 0.3478    |
|            | Union             | 2.0557      | 0.5170    | 3.98  | 0.00 | 1.0424   | 3.0689    |
|            | Union-unknown    | 0.2931      | 0.1441    | 2.03  | 0.04 | 0.0107   | 0.5755    |
|            | Constant         | -5.9493     | 0.5454    | -10.91| 0.00 | -7.0182  | -4.8804   |

#### Predicted values

- Exempt: 0.02899967
- Hours 20-35: 0.146096896
- Hours 36+: 0.82073136

### Table 3.6A: Multinomial Logit Regression for the United States (Matched Data)

| empstatus  | Variable         | Coefficient | Std. Err. | z     | P>|z| | Lower lim | Upper lim |
|------------|------------------|-------------|-----------|-------|-----|----------|-----------|
| Hours 20-35| Age              | 0.0796      | 0.0170    | 4.69  | 0.00 | 0.0463   | 0.1128    |
|            | Age square       | -0.0009     | 0.0002    | -4.15 | 0.00 | -0.0013  | -0.0005   |
|            | college          | -0.4431     | 0.0538    | -8.23 | 0.00 | -0.5486  | -0.3377   |
|            | male             | 0.2894      | 0.0598    | 4.84  | 0.00 | 0.1723   | 0.4066    |
|            | Currently married| -0.1844     | 0.0596    | -3.09 | 0.00 | -0.3011  | -0.0876   |
|            | Non-white        | -0.1019     | 0.0840    | -1.21 | 0.23 | -0.2665  | 0.0626    |
|            | Union             | 0.2825      | 0.2597    | 1.09  | 0.28 | -0.2265  | 0.7916    |
|            | Union-unknown    | -0.2960     | 0.0660    | -4.49 | 0.00 | -0.4253  | -0.1667   |
|            | Constant         | 0.1184      | 0.2718    | 0.44  | 0.66 | -0.4143  | 0.6512    |
| Hours 36+  | Age              | 0.3646      | 0.0164    | 22.25 | 0.00 | 0.3325   | 0.3967    |
|            | Age square       | -0.0042     | 0.0002    | -20.00| 0.00 | -0.0046  | -0.0038   |
|            | college          | -0.5399     | 0.0514    | -10.50| 0.00 | -0.6407  | -0.4391   |
|            | male             | 1.7083      | 0.0551    | 31.03 | 0.00 | 1.6004   | 1.8162    |
|            | Currently married| -0.1858     | 0.0591    | -3.14 | 0.00 | -0.3018  | -0.0699   |
|            | Non-white        | -0.0738     | 0.0794    | -0.93 | 0.35 | -0.2294  | 0.0818    |
|            | Union             | 0.5331      | 0.2456    | 2.17  | 0.03 | 0.0518   | 1.0144    |
|            | Union-unknown    | 0.1439      | 0.0632    | 2.28  | 0.02 | 0.0202   | 0.2677    |
|            | Constant         | -4.7209     | 0.2637    | -17.90| 0.00 | -5.2378  | -4.2040   |

#### Predicted values

- Exempt: 0.04430046
- Hours 20-35: 0.19110776
- Hours 36+: 0.76039178
Table 3.7A: Multinomial Logit Regression for Hawaii all workers (Matched Data)

| empstatus | Variable             | Coefficient | Std. Err. | z      | P>|z|  | Lower limit | Upper limit |
|-----------|----------------------|-------------|-----------|--------|------|-------------|-------------|
| Hours 20-35 | Age                  | 0.1452      | 0.0269    | 5.40   | 0.00 | 0.0925      | 0.1979      |
|           | Age suaqre           | -0.0018     | 0.0003    | -5.38  | 0.00 | -0.0024     | -0.0011     |
|           | college               | -0.4685     | 0.0964    | -4.86  | 0.00 | -0.6575     | -0.2796     |
|           | male                  | -0.0481     | 0.0941    | -0.51  | 0.61 | -0.2326     | 0.1363      |
|           | Currently married     | -0.0066     | 0.1047    | -0.06  | 0.95 | -0.2119     | 0.1986      |
|           | Non-white             | -0.0143     | 0.1033    | -0.14  | 0.89 | -0.2168     | 0.1882      |
|           | Union                 | 0.7041      | 0.3341    | 2.11   | 0.04 | 0.0492      | 1.3590      |
|           | Union-unknown         | 0.3163      | 0.1258    | 2.51   | 0.01 | 0.0697      | 0.5629      |
|           | Hawaii after phca     | 0.8305      | 0.3062    | 2.72   | 0.01 | 0.2323      | 1.4287      |
|           | Constant              | -2.1353     | 0.5797    | -3.68  | 0.00 | -3.2715     | -0.9992     |
| Hours 36+ | Age                  | 0.3292      | 0.0243    | 13.53  | 0.00 | 0.2815      | 0.3769      |
|           | Age suaqre           | -0.0038     | 0.0003    | -13.00 | 0.00 | -0.0044     | -0.0033     |
|           | college               | -0.2654     | 0.0876    | -3.03  | 0.00 | -0.4372     | -0.0937     |
|           | male                  | 0.6587      | 0.0849    | 7.76   | 0.00 | 0.4922      | 0.8251      |
|           | Currently married     | 0.1949      | 0.0961    | 2.03   | 0.04 | 0.0065      | 0.3832      |
|           | Non-white             | 0.1716      | 0.0952    | 1.80   | 0.07 | -0.0160     | 0.3682      |
|           | Union                 | 0.7542      | 0.3169    | 2.38   | 0.02 | 0.1330      | 1.3754      |
|           | Union-unknown         | 0.2988      | 0.1164    | 2.57   | 0.01 | 0.0707      | 0.5268      |
|           | Hawaii after phca     | 0.3535      | 0.2634    | 1.34   | 0.18 | -0.1627     | 0.6989      |
|           | Constant              | -4.5284     | 0.5226    | -8.66  | 0.00 | -5.5527     | -3.5041     |

<table>
<thead>
<tr>
<th>variable</th>
<th>dy/dx</th>
<th>Std. Err.</th>
<th>z</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>X bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exempt</td>
<td>Age</td>
<td>-0.0141</td>
<td>0.0011</td>
<td>-13.34</td>
<td>-0.0162</td>
<td>0.0120</td>
</tr>
<tr>
<td></td>
<td>Age suaqre</td>
<td>0.0002</td>
<td>0.0000</td>
<td>12.81</td>
<td>0.0001</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>college</td>
<td>0.0145</td>
<td>0.0042</td>
<td>3.44</td>
<td>0.0062</td>
<td>0.0227</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>-0.0256</td>
<td>0.0041</td>
<td>-6.21</td>
<td>-0.0337</td>
<td>-0.0175</td>
</tr>
<tr>
<td></td>
<td>Currently married</td>
<td>-0.0076</td>
<td>0.0046</td>
<td>-1.65</td>
<td>-0.0167</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>Non-white</td>
<td>-0.0068</td>
<td>0.0047</td>
<td>-1.43</td>
<td>-0.0160</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>Union</td>
<td>-0.0063</td>
<td>0.0079</td>
<td>-3.34</td>
<td>-0.0417</td>
<td>-0.0108</td>
</tr>
<tr>
<td></td>
<td>Union-unknown</td>
<td>-0.0013</td>
<td>0.0047</td>
<td>-2.83</td>
<td>-0.0224</td>
<td>-0.0041</td>
</tr>
<tr>
<td></td>
<td>Hawaii after phca</td>
<td>-0.0238</td>
<td>0.0173</td>
<td>-1.37</td>
<td>-0.0677</td>
<td>0.0102</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td>Age</td>
<td>-0.0226</td>
<td>0.0021</td>
<td>-10.84</td>
<td>-0.0266</td>
<td>-0.0185</td>
</tr>
<tr>
<td></td>
<td>Age suaqre</td>
<td>0.0003</td>
<td>0.0000</td>
<td>9.88</td>
<td>0.0002</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>college</td>
<td>-0.0298</td>
<td>0.0073</td>
<td>-4.10</td>
<td>-0.0441</td>
<td>-0.0156</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>-0.0926</td>
<td>0.0074</td>
<td>-12.51</td>
<td>-0.1072</td>
<td>-0.0781</td>
</tr>
<tr>
<td></td>
<td>Currently married</td>
<td>-0.0063</td>
<td>0.0080</td>
<td>-3.30</td>
<td>-0.0419</td>
<td>-0.0107</td>
</tr>
<tr>
<td></td>
<td>Non-white</td>
<td>-0.0247</td>
<td>0.0084</td>
<td>-2.95</td>
<td>-0.0411</td>
<td>-0.0083</td>
</tr>
<tr>
<td></td>
<td>Union</td>
<td>-0.0023</td>
<td>0.0193</td>
<td>-0.12</td>
<td>-0.0402</td>
<td>0.0356</td>
</tr>
<tr>
<td></td>
<td>Union-unknown</td>
<td>0.0047</td>
<td>0.0097</td>
<td>0.48</td>
<td>-0.0143</td>
<td>0.0237</td>
</tr>
<tr>
<td></td>
<td>Hawaii after phca</td>
<td>-0.0592</td>
<td>0.0191</td>
<td>3.10</td>
<td>-0.0217</td>
<td>-0.0967</td>
</tr>
<tr>
<td>Hours 36+</td>
<td>Age</td>
<td>0.0367</td>
<td>0.0023</td>
<td>15.99</td>
<td>0.0322</td>
<td>0.0412</td>
</tr>
<tr>
<td></td>
<td>Age suaqre</td>
<td>-0.0004</td>
<td>0.0000</td>
<td>-14.85</td>
<td>-0.0005</td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td>college</td>
<td>0.0154</td>
<td>0.0082</td>
<td>1.88</td>
<td>-0.0007</td>
<td>0.0314</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>0.1183</td>
<td>0.0081</td>
<td>14.53</td>
<td>0.1023</td>
<td>0.1342</td>
</tr>
<tr>
<td></td>
<td>Currently married</td>
<td>0.0339</td>
<td>0.0089</td>
<td>3.80</td>
<td>0.0164</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td>Non-white</td>
<td>0.0315</td>
<td>0.0094</td>
<td>3.38</td>
<td>0.0131</td>
<td>0.0498</td>
</tr>
<tr>
<td></td>
<td>Union</td>
<td>0.0286</td>
<td>0.0206</td>
<td>1.39</td>
<td>-0.0117</td>
<td>0.0689</td>
</tr>
<tr>
<td></td>
<td>Union-unknown</td>
<td>0.0085</td>
<td>0.0107</td>
<td>0.80</td>
<td>-0.0124</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td>Hawaii after phca</td>
<td>0.0354</td>
<td>0.0251</td>
<td>1.41</td>
<td>0.0137</td>
<td>0.0846</td>
</tr>
</tbody>
</table>

176
Table 3.8A: Multinomial Logit Regression for US, all workers (Matched Data)

<table>
<thead>
<tr>
<th>Hours worked on all jobs last week</th>
<th>Number of obs = 48350</th>
<th>Log pseudo-likelihood = -36280.776</th>
<th>Wald chi2(18) = 4049</th>
</tr>
</thead>
<tbody>
<tr>
<td>empstatus variable</td>
<td>Coefficient</td>
<td>Std. Err.</td>
<td>z</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0143</td>
<td>0.0028</td>
<td>5.07</td>
</tr>
<tr>
<td>Age suaqre</td>
<td>-0.0013</td>
<td>0.0002</td>
<td>-7.28</td>
</tr>
<tr>
<td>college</td>
<td>-0.3502</td>
<td>0.0441</td>
<td>-7.94</td>
</tr>
<tr>
<td>male</td>
<td>0.1474</td>
<td>0.0463</td>
<td>3.18</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.0505</td>
<td>0.0496</td>
<td>1.02</td>
</tr>
<tr>
<td>Non- white</td>
<td>0.0377</td>
<td>0.0729</td>
<td>0.52</td>
</tr>
<tr>
<td>Union</td>
<td>0.0348</td>
<td>0.1686</td>
<td>0.19</td>
</tr>
<tr>
<td>Union-unknown</td>
<td>-0.1489</td>
<td>0.0541</td>
<td>-2.75</td>
</tr>
<tr>
<td>US after phca</td>
<td>0.4920</td>
<td>0.1316</td>
<td>3.74</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0684</td>
<td>0.1583</td>
<td>0.43</td>
</tr>
<tr>
<td>Hours 36+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0353</td>
<td>0.0026</td>
<td>13.53</td>
</tr>
<tr>
<td>Age suaqre</td>
<td>-0.0037</td>
<td>0.0002</td>
<td>-21.58</td>
</tr>
<tr>
<td>college</td>
<td>-0.3580</td>
<td>0.0406</td>
<td>-8.82</td>
</tr>
<tr>
<td>male</td>
<td>1.3294</td>
<td>0.0414</td>
<td>32.13</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.4442</td>
<td>0.0465</td>
<td>9.55</td>
</tr>
<tr>
<td>Non- white</td>
<td>0.1019</td>
<td>0.0672</td>
<td>1.52</td>
</tr>
<tr>
<td>Union</td>
<td>0.0804</td>
<td>0.1712</td>
<td>0.47</td>
</tr>
<tr>
<td>Union-unknown</td>
<td>0.0141</td>
<td>0.0494</td>
<td>0.29</td>
</tr>
<tr>
<td>US after phca</td>
<td>0.2484</td>
<td>0.1131</td>
<td>2.20</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0551</td>
<td>0.1402</td>
<td>-0.39</td>
</tr>
<tr>
<td>variable</td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
</tr>
<tr>
<td>Exempt</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.0191</td>
<td>0.0009</td>
<td>-21.28</td>
</tr>
<tr>
<td>Age suaqre</td>
<td>0.0002</td>
<td>0.0000</td>
<td>19.43</td>
</tr>
<tr>
<td>college</td>
<td>0.0324</td>
<td>0.0031</td>
<td>10.40</td>
</tr>
<tr>
<td>male</td>
<td>-0.0751</td>
<td>0.0030</td>
<td>-25.49</td>
</tr>
<tr>
<td>Currently married</td>
<td>0.0096</td>
<td>0.0035</td>
<td>2.70</td>
</tr>
<tr>
<td>Non- white</td>
<td>-0.0042</td>
<td>0.0045</td>
<td>-0.94</td>
</tr>
<tr>
<td>Union</td>
<td>-0.0010</td>
<td>0.0119</td>
<td>-0.08</td>
</tr>
<tr>
<td>Union-unknown</td>
<td>-0.0041</td>
<td>0.0034</td>
<td>-1.21</td>
</tr>
<tr>
<td>US after phca</td>
<td>0.0567</td>
<td>0.0109</td>
<td>2.81</td>
</tr>
<tr>
<td>Hours 20-35</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.0280</td>
<td>0.0015</td>
<td>-18.42</td>
</tr>
<tr>
<td>Age suaqre</td>
<td>0.0003</td>
<td>0.0000</td>
<td>16.91</td>
</tr>
<tr>
<td>college</td>
<td>0.0025</td>
<td>0.0047</td>
<td>0.54</td>
</tr>
<tr>
<td>male</td>
<td>-0.1710</td>
<td>0.0043</td>
<td>-39.46</td>
</tr>
<tr>
<td>Currently married</td>
<td>-0.0047</td>
<td>0.0066</td>
<td>-0.71</td>
</tr>
<tr>
<td>Non- white</td>
<td>-0.0064</td>
<td>0.0074</td>
<td>-0.86</td>
</tr>
<tr>
<td>Union</td>
<td>-0.0007</td>
<td>0.0176</td>
<td>-0.04</td>
</tr>
<tr>
<td>Union-unknown</td>
<td>-0.0358</td>
<td>0.0055</td>
<td>-6.57</td>
</tr>
<tr>
<td>US after phca</td>
<td>-0.0547</td>
<td>0.0129</td>
<td>-2.36</td>
</tr>
<tr>
<td>Hours 36+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0470</td>
<td>0.0017</td>
<td>27.14</td>
</tr>
<tr>
<td>Age suaqre</td>
<td>-0.0005</td>
<td>0.0000</td>
<td>-25.06</td>
</tr>
<tr>
<td>college</td>
<td>-0.0349</td>
<td>0.0054</td>
<td>-6.48</td>
</tr>
<tr>
<td>male</td>
<td>0.2460</td>
<td>0.0048</td>
<td>51.72</td>
</tr>
<tr>
<td>Currently married</td>
<td>-0.0049</td>
<td>0.0076</td>
<td>-0.64</td>
</tr>
<tr>
<td>Non-white</td>
<td>0.0106</td>
<td>0.0084</td>
<td>1.26</td>
</tr>
<tr>
<td>Union</td>
<td>0.0016</td>
<td>0.0199</td>
<td>0.08</td>
</tr>
<tr>
<td>Union-unknown</td>
<td>0.0399</td>
<td>0.0063</td>
<td>6.29</td>
</tr>
<tr>
<td>US after phca</td>
<td>-0.0496</td>
<td>0.0159</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
Stata Program
/* Chapter 3: Hours worked and labor sorting*/
clear
log using c:\cps\matchedcps6304new.log, replace
set more 1
#delimit;

use c:\cps\hius6304.dta, clear;
destring _year, replace;
replace hours=. if _year>=1963 & _year<=1967 & hours==99;
*gen edu=_educ if _year<1991;
*replace edu=grdatn if _year>=1992;
gen hawaii=state==95 if _year>=1977 & _year<=2004;
*replace hawaii=state==99 if _year>=1968 & _year<=1976;
replace hawaii=state==86 if _year>=1963 & _year<=1967;
gen us=hawaii==1;
gen phcabefore=_year<=1967;
gen phcaafter=_year>=1977 & _year<=2004;
drop if (_class==4 | _class==5 | _class==9);
gen privat=_class==1;
gen govt=_class==2;
gen selfemp=_class==3;
*drop if (_clslyr==4 | _clslyr==5 | _clslyr==9);
*gen privat=_clslyr==1;
*gen govt=_clslyr==2;
*gen selfemp=_clslyr==3;
gen union=(unmem==1 & _year>=1983);
gen unionunknown=_year<1983;
replace unionunknown=0 if _year>=1983;
drop if (hours<=0 | hours==.);

gen exempt=(partime19==1 & govt==1 & selfemp==1);
gen empstatus=1 if partime19==1;
replace empstatus=2 if partime2035==1;
replace empstatus=3 if fultime==1;
drop if empstatus==.;
gen nwgt=wgt/100;
drop if nwgt<0;
svyset [iweight=nwgt], psu(hhid);
/*
*keep if nuemp==1;
tab empstatus [iweight=nwgt] if hawaii==1 & age>=19 & age<=64 & hours>0 & hours<98 & phcabefore==1;
tab empstatus [iweight=nwgt] if hawaii==1 & age>=19 & age<=64 & hours>0 & hours<98 & phcaafter==1;
tab empstatus [iweight=nwgt] if us==1 & age>=19 & age<=64 & hours>0 & hours<98 & phcabefore==1;
*/
tab empstatu [iweight=nwgt] if us==1 & age>=19 & age<=64 & hours>0 &
hours<=98 & phcaafter==1;

*keep if hawaii==1 & age>=19 & age <=64;
*keep _year sex age _race edu marstat indmly;
*save c:\cps\cps_shawain6304, replace;
*keep if age>=19 & age <=64;
*save c:\cps\cps_us6304, replace;

gen year63=_year==1963;
gen year64=_year==1964;
gen year65=_year==1965;
gen year66=_year==1966;
gen year67=_year==1967;
gen year77=_year==1977;
gen year78=_year==1978;
gen year79=_year==1979;
gen year80=_year==1980;
gen year81=_year==1981;
gen year82=_year==1982;
gen year83=_year==1983;
gen year84=_year==1984;
gen year85=_year==1985;
gen year86=_year==1986;
gen year87=_year==1987;
gen year88=_year==1988;
gen year89=_year==1989;
gen year90=_year==1990;
gen year91=_year==1991;
gen year92=_year==1992;
gen year93=_year==1993;
gen year94=_year==1994;
gen year95=_year==1995;
gen year96=_year==1996;
gen year97=_year==1997;
gen year98=_year==1998;
gen year99=_year==1999;
gen year00=_year==2000;
gen year01=_year==2001;
gen year02=_year==2002;
gen year03=_year==2003;
gen year04=_year==2004;

gen agesq=age^2;
gen female=(sex==2);
gen male=sex==1;
gen white=_race==1;
gen nonwhite=_race==1;

gen element=_educ<=6 if _year<=1990;
replace element=grdatn<=33 if _year>1992;
gen middle=( _educ==7 | _educ==8 | _educ==9 ) if _year<=1990 ;
replace middle=(grdatn==34 | grdatn==35) if _year>1992;
gen lesshigh=(element==1 | middle==1);  
gen high=( _educ==10 | _educ==11 | _educ==12 | _educ==13 ) if _year==1990;
replace high=(grdatn==36 | grdatn==37 | grdatn==38 | grdatn==39) if _year>1992;
gen somecol=( _educ==14 | _educ==15 | _educ==16 ) if _year<=1990;
replace somecol=(grdatn==40 | grdatn==41 | grdatn==42) if _year>1992;
gen collegep=( _educ==17 | _educ==18 & _educ==.) if _year<=1990;
replace collegep=(grdatn==43 | grdatn==44 | grdatn==45 | grdatn==46) if _year>=1992;
gen college=(somecol==1 | collegep==1);
gen married=(marstat==1 | marstat==2 | marstat==3);
gen child=age<19;
gen adult=(age>=19 & age<=64);
*gen honolulu=(hawaii==1 & msafp==3320);
*gen emplsz9=emplsz==1;
*gen emplsz499=(emplsz==2 | emplsz==3 | emplsz==4);
*gen emplsz500p=(emplsz==5 | emplsz==6);
*gen emplszunknown=_year<1988;
*gen malemarried=male*married;
*gen y=income;
*gen fpl=faminc/p;
*gen eligible=(fpl<=1);
drop if (_year==1982 | _year==1983);
x:probit exempt age agesq college male married nonwhite fultime union ununionunknown phcaafter i.indmly
year63 year64 year65 year66 year67 year77 year78 year79 year80 year81
year84 year85 year86 year87
year88 year89 year90 year91 year92 year93 year94 year95 year96 year97
year98 year99 year00 year01 year02 year03[pw=nwgt]
if hawaii==1, robust;
x:dprobit exempt age agesq college male married nonwhite fultime union ununionunknown phcaafter i.indmly
year63 year64 year65 year66 year67 year77 year78 year79 year80 year81
year84 year85 year86 year87
year88 year89 year90 year91 year92 year93 year94 year95 year96 year97
year98 year99 year00 year01 year02 year03[pw=nwgt]
if hawaii==1, robust;
x:probit exempt age agesq college male married nonwhite fultime union ununionunknown phcaafter i.indmly
year63 year64 year65 year66 year67 year77 year78 year79 year80 year81
year84 year85 year86 year87
year88 year89 year90 year91 year92 year93 year94 year95 year96 year97
year98 year99 year00 year01 year02 year03[pw=nwgt]
, robust;
x:dprobit exempt age agesq college male married nonwhite fultime union ununionunknown phcaafter i.indmly
year63 year64 year65 year66 year67 year77 year78 year79 year80 year81
year84 year85 year86 year87
year88 year89 year90 year91 year92 year93 year94 year95 year96 year97
year98 year99 year00 year01 year02 year03[pw=nwgt]
, robust;
col;
*gen phcabefore=_year<=1967;
*gen phcaafter=_year>=1977 & _year<=2004;
gen hiphcabefore=hawaii*phcabefore;
gen hiphcaafter=hawaii*phcaafter;
gen usphcabefore=us*phcabefore;
gen usphcaafter=us*phcaafter;
gen b4=(usphcabefore + usphcaafter);
gen b3=hiphcaafter;
gen b2=(hiphcabefore + hiphcaafter);
gen b1=(hiphcabefore + hiphcaafter + usphcaafter);
gen b0=(hiphcabefore + hiphcaafter + usphcbefore + usphcaafter);
su b0 b1 b2 b3 b4 hiphcabefore hiphcaafter usphcbefore usphcaafter if (b2==1 | b4==1);
mlogit empstatus age college male married nonwhite hiphcaafter  
[pweight=nwgt]if b2==1, base(1) robust;
mfx compute, predict(outcome(1));
mfx compute, predict(outcome(2));
mfx compute, predict(outcome(3));

mlogit empstatus age college male married nonwhite usphcaafter  
[pweight=nwgt]if b4==1, base(1) robust;
mfx compute, predict(outcome(1));
mfx compute, predict(outcome(2));
mfx compute, predict(outcome(3));

mlogit empstatus age college male married nonwhite b1 b2 b3  
[pweight=nwgt]if (b2==1 | b4==1), base(1) robust;
mfx compute, predict(outcome(1));
mfx compute, predict(outcome(2));
mfx compute, predict(outcome(3));

mlogit empstatus age college male married nonwhite hiphcaafter  
[pweight=nwgt]if phcaafter==1, base(1) robust;
mfx compute, predict(outcome(1));
mfx compute, predict(outcome(2));
mfx compute, predict(outcome(3));

col

tab _year if age >=19 & age<=64 & privat==1;
tab _year if age >=19 & age<=64 & govt==1;
tab _year if age >=19 & age<=64 & selfemp==1;
tab _year if hawaii==1 & age >=19 & age<=64 & privat==1;
tab _year if hawaii==1 & age >=19 & age<=64 & govt==1;
tab _year if hawaii==1 & age >=19 & age<=64 & selfemp==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & privat==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & govt==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & selfemp==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & privat==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & govt==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & selfemp==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>1 & hours<=98;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>1 & hours<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>1 & hours<=98 & (privat==1 & fultime==1);
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>1 & hours<=98;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>1 & hours<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>1 & hours<=98 & (privat==1 & fultime==1);
drop if govt==1;
drop if unmem==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & (selfemp==1 | partime==1);

drop if govt==1;
drop if unmem==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98;
tab _year [iweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98 & (selfemp==1 | partime==1);
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98 & (selfemp==1 | partime==1);

drop if govt==1;
drop if unmem==1;
keep if nuempl==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & (selfemp==1 | partime==1);
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<=98;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & privat==1 & fultime==1;
tab _year [iweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & (selfemp==1 | partime==1);

drop if govt==1;
drop if unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcbefore==1 & privat==1 & fultime==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & privat==1 & fultime==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcbefore==1 & (selfemp==1 | partime==1);
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & (selfemp==1 | partime==1);
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcbbefore==1 & privat==1 & fultime==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & privat==1 & fultime==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcbbefore==1 & (selfemp==1 | partime==1);
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & (selfemp==1 | partime==1);

drop if govt==1;
drop if unmem==1;
drop if _year<=1993;
tab ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98 & privat==1 & fultime==1;
tab ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98 & (selfemp==1 | partime==1);
tab ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98 & privat==1 & fultime==1;
tab ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98 & (selfemp==1 | partime==1);
drop if govt==1;
drop if unmem==1;
tab hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & privat==1 & fultime==1;
tab hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & (selfemp==1 | partime==1);
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & privat==1 & fultime==1;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<=98 & (selfemp==1 | partime==1);
svymean hours if (privat==1 & fultime==1), by (_year);
svymean hours if govt==1 & fultime==1, by (_year);
svymean hours if selfemp==1 & fultime==1, by (_year);
svymean hours if partime==1, by (_year);
svymean hours if hawaii==1 & privat==1 & fultime==1, by (_year);
svymean hours if hawaii==1 & govt==1 & fultime==1, by (_year);
svymean hours if hawaii==1 & selfemp==1 & fultime==1, by (_year);
svymean hours if hawaii==1 & partime==1, by (_year);
*drop if _year<=1976;
tab _year [iw=nwgt] if age >=19 & age<=64;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64;
tab _year [iw=nwgt] if age >=19 & age<=64 & nuemp==0;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & nuemp==0;
tab _year [iw=nwgt] if age >=19 & age<=64 & nuemp==1;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & nuemp==1;
tab _year [iw=nwgt] if age >=19 & age<=64 & nuemp==1 & privat==1 & fultime==1;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & nuemp==1 & privat==1 & fultime==1;
tab _year [iw=nwgt] if age >=19 & age<=64 & ernush==1 & ernush-=.;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush==1 & ernush-=.;
tab _year [iw=nwgt] if age >=19 & age<=64 & privat==1 & fultime==1;
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & privat==1 & fultime==1;
tab _year [iw=nwgt] if age >=19 & age<=64 & ernush==1 & ernush-=. & (privat==1 | fultime==1);
tab _year [iw=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush==1 & ernush-=. & (privat==1 | fultime==1);
tab hrslyr [aweight=nwgt] if age >=19 & age<=64;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64;
tab hrslyr [aweight=nwgt] if age >19 & age<=64 & nuemp==1;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & nuemp==1;
tab hrslyr [aweight=nwgt] if age >19 & age<=64 & nuemp==1 & nuemp==.;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & nuemp>=1 & nuemp==.;

tab ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & privat==1;
tab ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & govt==1;
tab ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & selfemp==1;

tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr==1 & hrslyr<=97 & phcaafter==1 & nuemp==1 & privat==1;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr==1 & hrslyr<=97 & phcaafter==1 & nuemp==1 & govt==1;
tab hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr==1 & hrslyr<=97 & phcaafter==1 & nuemp==1 & selfemp==1;

tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & govt==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & selfemp==1;

184
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcaafter==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & selfemp==1 & parttime==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcaafter==1 & selfemp==1 & parttime==1 & unmem==1;
tab hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if phcabefore==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=97 & phcaafter==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=97 & phcabefore==1 & privat==1 & fultime==1 & unmem==1;
tab hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=97 & phcaafter==1 & privat==1 & fultime==1 & unmem==1;
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98, nogr gen(x fx);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1, nogr gen(fx1) at(x);
label var fx0 "PHCAbefore"
label var fx1 "PHCAafter"
line fx0 fx1 x, sort ytitle(density);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98, nogr gen(x fx);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1, nogr gen(fx1) at(x);
label var fx0 "United States"
label var fx1 "Hawaii"
line fx0 fx1 x, sort ytitle(density);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98, nogr gen(x fx);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1, nogr gen(fx1) at(x);
label var fx0 "United States"
label var fx1 "Hawaii"
line fx0 fx1 x, sort ytitle(density);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & privat==1, nogr gen(x fx);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1 & privat==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & privat==1, nogr gen(fx1) at(x);
label var fx0 "PHCAbefore"
label var fx1 "PHCAafter"
line fx0 fx1 x, sort ytitle(density);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & govt==1, nogr gen(x fx);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1 & selfemp==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & selfemp==1, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & selfemp==1, nogr gen(x fx);
kdensity hours [aweight=nwgt] if age >=19 & age<=64 & hours>0 & hours<=98 & phcabefore==1 & selfemp==1, nogr gen(fx0) at(x);
kdensity hours [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hours>0 & hours<=98 & phcaafter==1 & selfemp==1, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98, nogr gen(x fx);
kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=98, nogr gen(fx0) at(x);
kdensity ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=98, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & privat==1, nogr gen(x fx);
kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & privat==1, nogr gen(fx0) at(x);
kdensity ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=97 & privat==1, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & govt==1, nogr gen(x fx);
kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & govt==1, nogr gen(fx0) at(x);
kdensity ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=97 & govt==1, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & selfemp==1, nogr gen(x fx);
kdensity ernush [aweight=nwgt] if age >=19 & age<=64 & ernush>0 & ernush<=97 & selfemp==1, nogr gen(fx0) at(x);
kdensity ernush [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & ernush>0 & ernush<=97 & selfemp==1, nogr gen(fx1) at(x);
label var fx0 "United States";
label var fx1 "Hawaii";
line fx0 fx1 x, sort ytitle(density);

keep if nuemp==1;
kdensity hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<=98, nogr gen(x fx);
kdensity hrslyr [aweight=nwgt] if us==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<98, nogr gen(fxO) at(x);
kdensity hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<98, nogr gen(fx1) at(x);
lable var fxO "United States";
lable var fx1 "Hawaii";
line fxO fx1 x, sort ytitle(density);
keep if nuemp==1;
kdensity hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & privat==1, nogr gen(x fx);
kdensity hrslyr [aweight=nwgt] if us==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & privat==1, nogr gen(fxO) at(x);
kdensity hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<98 & privat==1, nogr gen(fx1) at(x);
lable var fxO "United States";
lable var fx1 "Hawaii";
line fxO fx1 x, sort ytitle(density);
keep if nuemp==1;
kdensity hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & govt==1, nogr gen(x fx);
kdensity hrslyr [aweight=nwgt] if us==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & govt==1, nogr gen(fxO) at(x);
kdensity hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<98 & govt==1, nogr gen(fx1) at(x);
lable var fxO "United States";
lable var fx1 "Hawaii";
line fxO fx1 x, sort ytitle(density);
keep if nuemp==1;
kdensity hrslyr [aweight=nwgt] if age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & selfemp==1, nogr gen(x fx);
kdensity hrslyr [aweight=nwgt] if us==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<97 & selfemp==1, nogr gen(fxO) at(x);
kdensity hrslyr [aweight=nwgt] if hawaii==1 & age >=19 & age<=64 & hrslyr>0 & hrslyr<98 & selfemp==1, nogr gen(fx1) at(x);
lable var fxO "United States";
lable var fx1 "Hawaii";
line fxO fx1 x, sort ytitle(density);
svymean _wklywg if _esr==1, by (_year);
svymean _wklywg if hawaii==1 & _esr==1 , by (_year);
gen unemployed=_esr==3;
svymean unemployed, by (_year);
svymean unemployed if hawaii==1, by (_year);
gen partime=hours<=19 & hours>=1;
gen fultime=hours>=20 & hours<=99;
svymean partime, by (_year);
svymean partime if hawaii==1, by (_year);
svymean fultime, by (_year);
svymean fultime if hawaii==1, by (_year);
*end of file;