2014

Three essays on envrionmental economics

Hocheol Jeon
Iowa State University

Three essays on environmental economics

by

Hocheol Jeon

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:

Joseph A. Herriges, Major Professor

Catherine L. Kling

Wallace Huffman

John Miranowski

Brent Kreider

Iowa State University

Ames, Iowa

2014

DEDICATION

I dedicate my dissertation to my father Hongsung Jeon who have devoted his entire life for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I would like to express the deepest appreciation to my major professor, Prof. Joseph A. Herriges, for his incredibly valuable guidance and encouragement during my years in graduate school. Without his guidance and persistent help, this dissertation would not have been possible.

I am grateful to my committee members, Prof. Kling, Prof. Huffman, Prof. Miranowski, and Prof. Kreider, for encouraging words, thoughtful criticisms, and valuable comments on further improvements.

Finally, I would like to extend my deepest gratitude to my wife, Jina Choi for her words of encouragement and sacrifices for me, and to my two sons, Jonghwan and Jongin, who are the joys of my life. To my parents, Hongsung Jeon and Yeonja Choi, and to parents-in-law, Ohsik Choi and Jaeseon Jeon, and to my brother, sisters and sister-in-law, I would like to thank you.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Overview

This dissertation is a collection of three studies that investigate welfare measurement, causation of addition emission, and effectiveness of policy in environmental economics. The first study focuses on the discrepancy from different data sources in nonmarket valuation. This study suggests a new valuation method to solve one common problem shown frequently in combining RP and SP data. The second study examines the significance of the relation between two critical concerns, obesity and vehicle emission. This study shows, even though the impact of obesity on gasoline consumption seems to be significantly large in either one side or aggregate data level, the significance is ambiguous or the magnitude of impact is not sufficiently large study after removing unobserved household characteristics using household-level data. The third study investigates the rebound effect in vehicle use, a critical parameter in cost-benefit analyses of increases in the corporate average fuel economy (CAFE) standards. The following illustrates the ideas and findings of three chapters contained within this dissertation.

The first study, "Combining Revealed and Stated Preference Data: A Latent Class Approach," proposes a new framework to combine revealed and stated preference data when the convergent validity assumption is not hold.

A substantial literature exists combining data from revealed preference (RP) and stated preference (SP) sources, aimed either at testing for the convergent validity of the two approaches used in nonmarket valuation or as a means of drawing on their relative strengths to improve the ultimate estimates of value. In doing so, it is assumed that convergence of the two elicitation approaches is an "all or nothing" proposition; i.e., the RP and SP data are either consistent with each other or they are not. The purpose of this paper is to propose an alternative framework that allows for possible divergence among individuals in terms of the consistency between their RP and SP responses. In particular, we suggest the use of the latent class approach to segment the population into two groups. The first group has RP and SP responses that are internally consistent, while the remaining group exhibits some form of inconsistent preferences. An EM algorithm is employed in an empirical application that draws on the moose hunting data set used in earlier combined RP and SP exercises. The empirical results suggest that somewhat less than half the sample exhibits inconsistent preferences. We also examine differences in welfare estimates drawn from the two classes.

The second study, "Does Obesity Matter for the Environment? Evidence from Vehicle Choices and Usage" studies the interesting link between obesity and vehicle emission, using unique household-level data.

The rising rate of obesity has become a prominent social concern in the U.S. and throughout the world. Several recent studies examine how obesity influences households' driving or vehicle choice behavior. While the results in prior studies are compelling, the studies suffer from two shortcomings. First, prior studies rely on aggregate data (national or county level), rather than individual or household level observations, potentially masking important factors determining individual choices for vehicles and driving. Second, while previous works able to establish a link between obesity and vehicle choice or driving, linking vehicle choice, in turn, to overall emissions requires information regarding vehicle miles driven. The objective of this study is to address these two limitations, using household observations from the Panel Study of Income Dynamics (PSID), jointly modeling the impact of obesity on the vehicle choice and vehicle miles traveled (VMT). In particular, we investigate the impact of obesity and overweight by employing both reduced-form (linear panel model) and structural model (joint discrete and continuous choice model). Our study shows that the prevalence of obesity in 2005 has remained at the 1981 level, and gasoline consumption would be 3% saved in the reduced-form approach. While the rate of overweight people in 2005 has remained at the 1981 level, only 1.6% less gasoline would be demanded using the structural-approach. Our empirical findings suggest that the comprehensive impact of obesity and overweight on gasoline consumptions is little or ambiguous in contrast to the results of prior studies considering either driving or vehicle choices.

The third study, "Vehicle Fuel Efficiency and the Rebound Effect: Evidence from U.S. Panel Data" examines the rebound effect of vehicle usage. This study revisits one of the classical issues in energy economics using U.S. panel data which have never used in this area.

The Corporate Average Fuel Economy (CAFE) standards are a centerpiece of the United States' efforts to control mobile source air pollution. In addition to being politically expedient relative to a more direct gasoline or carbon tax, they may also be more effective as a policy instrument, if, as the literature suggests, consumers are subject to the so-called "energy paradox," undervaluing the fuel cost savings from more fuel efficient vehicles. However, at the same time, there are also risks associated with CAFE standards. In particular, to the extent that individuals respond to increased fuel efficiency by driving more, a response known as the "rebound effect," the impact of the standards may be significantly diminished. A number of papers have sought to quantify the rebound effect, but the results have varied substantially. A key concern with this literature is that it is largely based on cross-sectional data sources, making it more difficult to control for the endogeneity of fuel economy. The purpose of this paper is to address these endogeneity concerns through the use of panel data techniques, drawing on data from the Panel Study of Income Dynamics (PSID). In contrast to prior studies using only cross-sectional data, we find the elasticity of vehicle miles traveled (VMT) with respect to both fuel economy and fuel price to be statistically significant. Our results show that a 1% increase in fuel prices or fuel economy (MPG) leads to a 0.41 to 0.67% increase in driving miles. We also examine heterogeneity of these elasticities across the income deciles. We find evidence that low income households are more responsive to changes in gasoline prices, but less sensitive to changes in fuel economy.

1.1.1 Dissertation Organization

The structure of the dissertation is organized as follows. The next three chapters consist of three independent papers regarding non-market valuation and energy economics, overall environmental economics. The dissertation closes with an overall summary of our findings and a general discussion of possible extensions.

CHAPTER 2. COMBINING REVEALED AND STATED PREFERENCE DATA: A LATENT CLASS APPROACH

2.1 Introduction

A substantial literature has emerged in the nonmarket valuation arena aimed at combining data from revealed preference (RP) and stated preference (SP) sources. The goal of such efforts vary. In some cases, the objective is to test the convergent validity of the RP and SP approaches [e.g., Azevedo et al. (2003), Huang et al. (1997), and Whitehead et al. (2010)]. In other instances, the two data sources are viewed as complementary, with RP data providing values grounded in individual behavior (rather than intentions), while SP data both expands on the range of variation in environmental amenities from what is observed in RP data and introduces experimental control over the impact of unobservable factors [e.g., von Haefen and Phaneuf (2008)]. To the extent that the RP and SP data are generated by the same underlying preferences, this approach argues that combining the two provides more accurate measures of value. Early examples along these lines include Cameron (1992) and Adamowicz et al. (1994), while more recent applications include Dosman and Adamowicz (2006) and Eom and Larson (2006). In either case, it is typically assumed that convergence between the RP and SP data sources is an "all or nothing" proposition; i.e., the RP and SP data are either consistent with each other or they are not. The purpose of this paper is to propose an alternative framework that allows for possible divergence among individuals in terms the consistency between their RP and SP responses. In particular, we suggest the use of latent class approach to segment the population into two groups. The first group has RP and SP responses that are internally consistent, while the remaining group exhibits some form of inconsistent preferences. Examining differences between the preferences of the two groups provides additional insights into the wedge between RP and SP responses. The framework also opens up the possibility of modeling class membership, along the lines employed by Boxall and Adamowicz (2002), with the goal of mitigating the behavior of those in the "inconsistent" class in subsequent RP/SP exercises.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the literature combining stated and revealed preference data. We then describe the proposed latent class model in Section 3, along with a description of the EM algorithm used in estimation. Section 4 presents a generated data experiment to illustrate the performance and characteristics of the model under different parameterizations, with particular attention paid to the size of the "inconsistent" class as a share of the population. These Monte Carlo exercises illustrate the impact, both in terms of parameter and welfare estimates, of ignoring discrepancies between the underlying RP and SP data generating processes, particularly when the "consistent" class is only a small share of the target population. We illustrate our framework in Section 5 using the Moose Hunting data first introduced by Adamowicz et al. (1997) in their RP/SP exercise, and subsequently employed by von Haefen and Phaneuf (2008). Our results indicate that nearly a half of the sample provided responses that suggest different RP and SP data generating processes and that welfare predictions are sensitive to the choice of which subgroup is used in valuing changes to the environment. The paper wraps up in Section 6 with a summary and conclusions.

2.2 The Literature on Combining RP/SP Data Sources

The idea of combining information from revealed preference and stated preference sources is by no means a new one, with papers appearing in the marketing, transportation, health and environmental economics literatures. In their recent review, Whitehead et al. (2008) note that the earliest efforts along these lines appeared in the transportation and marketing literatures nearly twenty-five years ago, with papers by Ben-Akiva and Morikawa (1990) and Ben-Akiva et al. (1994). Comparisons between RP- and SP- based welfare measures have, of course, been around for years in the environmental arena, including the pioneering goose hunting permit study by Bishop and Heberlein (1979). However, the objective of such comparisons was typically a convergent validity test, with the usual, though not universal, presumption being that the RP results were more reliable as they were based on actual behavior.[1]

The earliest efforts to explicitly combine the two sources in the environmental literature appeared somewhat later, with papers by Cameron (1992) and Adamowicz et al. (1994). These authors argued that RP and SP data should be viewed as complementary, rather than competing, sources of information. In particular, two key limitations of the revealed preference data are (a) insufficient variation in environmental amenities of interest and (b) the potential for the environmental amenities to be confounded with other observed or unobserved factors. Proposed environmental policy scenarios often involve changes that are outside of the range of historical environmental conditions, making extrapolation of preferences for such changes tenuous and dependent on strong assumptions regarding the form of individual preferences. More fundamentally, there may simply not be sufficient historical variation in the environmental attribute of interest to identify its impact on preferences. A related problem is that what variation is observed for an environmental amenity may be correlated with other observed or unobserved factors impacting consumer preferences, making it difficult to disentangle its causal effect on consumer behavior. Stated preference data, on the other hand, provides the researcher with greater control over the variation in environmental conditions presented to survey participants. In many cases, orthogonal treatments can be employed, though such treatments may be limited by the need to present realistic choice scenarios. von Haefen and Phaneuf (2008) highlight the fact that the experimental control associated with stated preference surveys can be used to isolate the causal impact of an environmental amenity on individual behavioral, avoiding problems of omitted variables bias encountered in stand-alone RP exercises. Eom and Larson (2006) illustrate the use of SP data, in combination with RP data, to identify non-use (or passive use) values that simply cannot be identified with RP data alone.

The major concern with stated preference data sources is that they might be susceptible to hypothetical bias. Revealed preference data can be used to “discipline” the stated preference responses with information on choices observed in the marketplace. One strategy is to rely primarily upon RP data to estimate the key preference parameters, such as the marginal utility

[1]See Randall (1994) and Azevedo et al. (2003) for alternative perspectives on the presumed reliability of RP results.

of income, leaving SP with the role of "filling-out" the marginal impacts of environmental amenities on individual preferences [e.g., von Haefen and Phaneuf (2003)]. Alternatively, if both sources are viewed as suspect, combining the two data sources may provide the best overall picture of consumer preferences.

The evidence regarding combining RP and SP data sources is mixed. Adamowicz et al. (1994) and Adamowicz et al. (1997), for example, find "...RP-SP parameter equality, once variance heterogeneity is accounted for, and ...that joint RP-SP models are superior to RP models alone." In contrast, von Haefen and Phaneuf (2008), using the same data as Adamowicz et al. (1997), reject consistency between the RP and SP responses, as do Azevedo et al. (2003) in a different setting. Both Jeon and Herriges (2010) and Whitehead et al. (2008) reject consistency between RP and SP responses in their respective studies, though the differences between the welfare measures derived from the RP and SP sources are not substantial. In all these studies, the tests for consistency are for the sample as a whole. In the next section, we outline a latent class model which estimates the proportion of the sample that exhibits inconsistent RP and SP preferences.

2.3 Model

This section begins by describing a single class joint model of RP and SP data in a repeated discrete choice setting. The structure of the model is similar to the one employed by von Haefen and Phaneuf (2008). The model is then extended using a latent class framework, allowing for some portion of the sample (s) to exhibit consistent RP and SP preferences, while the RP and SP parameters diverge for the remainder of the sample. As is typical of the recent literature on latent class models [e.g., Breffle et al. (2011); Evans and Herriges (2010); Kuriyama et al. (2010)], we propose estimating the parameters of the model using of an EM algorithm so as to avoid numerical difficulties often encounter with standard maximum likelihood estimation of latent class models [see, e.g., Train (2009)].

2.3.1 Combining RP and SP data

There are two common issues encountered when combining RP and SP recreation demand data. First, the relevant site attributes are generally different for the two data sources. Of particular concern in the context of the modeling RP choices is the fact that the analyst may observe only a subset of the choice attributes impacting an individual's decision. To the extent that there are unobserved choice attributes that are correlated with the attributes available to the researcher, steps must be taken to control for potential omitted variables bias. In contrast, stated preference choices can be thought of as providing the analyst with complete information on the relevant choice attributes, assuming of course that the SP study is well-designed and the respondents fully understand the instructions. To the extent that there are unobservable individual or site attributes impacting an individual's choices, the random assignment of observable treatment affects should avoid potential omitted variables bias. Second, given the differences in the decision making processes underlying the RP and SP data sources, there are likely to be differences in the unobservable factors impacting the corresponding decisions. These differences manifest themselves in differences between the scale parameters associated with the RP and SP portions of the model. Control for changes in the scale parameters of the two models is important in testing for consistency between the two data sources [see, e.g., Adamowicz et al. (1994) and Adamowicz et al. (1997)].

Starting with the revealed preference portion on the model, the data provide information on the number of times (n_{ij}^{RP}) individual i chose to visit each of j sites over the course of T_i trips.[2] The utility (U_{ijt}^{RP}) that individual i receives from choosing site j on trip t is assumed to be a linear function of observed (X_j^{RP}) and unobserved ($\tilde{X}_j^{RP}$) site specific attributes, travel costs to the site (p_{ij}), and an idiosyncratic error components ($\mu^{RP}\varepsilon_{ijt}$), where ε_{ijt} is an *iid*

[2]The model specified here is a site selection model, rather than a model that also characterizes the participation decision, as in the repeated logit framework of Morey et al. (1993). We focus on the site selection aspect of the individual's decision to be consistent with the earlier analysis of this same database by Adamowicz et al. (1997) and von Haefen and Phaneuf (2008).

Type I extreme value error term and μ^{RP} is the associated scale factor.[3] Formally,

$$\begin{aligned}
U_{ijt}^{RP} &= X_j^{RP}\beta^{RP} + \tilde{X}_j^{RP}\tilde{\beta}^{RP} + p_{ij}^{RP}\gamma^{RP} + \mu^{RP}\varepsilon_{ijt} \\
&= X_j^{RP}\beta^{RP} + \xi_j^{RP} + p_{ij}^{RP}\gamma^{RP} + \mu^{RP}\varepsilon_{ijt} \qquad (2.1) \\
&= \alpha_j^{RP} + p_{ij}^{RP}\gamma^{RP} + \mu^{RP}\varepsilon_{ijt} \qquad (2.2) \\
&= V_{ij}^{RP} + \mu^{RP}\varepsilon_{ijt}
\end{aligned}$$

where $V_{ij} = \alpha_j^{RP} + p_{ij}^{RP}\gamma^{RP}$, $\xi_j^{RP} \equiv \tilde{X}_j^{RP}\tilde{\beta}^{RP}$ and

$$\alpha_j^{RP} \equiv X_j^{RP}\beta^{RP} + \xi_j^{RP}. \qquad (2.3)$$

Absent any outside information, the impact of the observable factors X_{ij}^{RP} on individual choices cannot be directly disentangle from the impact of the unobservable factors summarized by ξ_j^{RP} in equation (2.1). Instead, only the parameters in (2.2) can be estimated, including the alternatives specific constants (ASC's) α_j^{RP}.[4] However, as suggested by Murdock (2006), a second stage regression can be use to identify β^{RP} by estimating equation (2.3) using fitted values for the alternative specific constants (i.e., the α_j^{RP}'s) and properly instrumenting for the X_{ij}^{RP}.

Turning to the stated preference data, the individuals are presented with a series of H choice scenarios, with each choice scenario involving K alternatives ($K = 3$ in the Moose Hunting data set). The utility U_{ikh}^{SP} that individual i associates with alternative k from choice scenario c is assumed to be a linear function of the designed characteristics for each of the choice alternatives (X_{ikh}^{SP}), the cost of the presented alternative (p_{ikh}), and an idiosyncratic error components ($\mu^{SP}\varepsilon_{ikh}$), where ε_{ikh} is an *iid* Type I extreme value error term and μ^{SP} is the associated scale factor. Formally

$$U_{ikh}^{SP} = X_{kh}^{SP}\beta^{SP} + p_{ikh}^{SP}\gamma^{SP} + \mu^{SP}\varepsilon_{ikh}. \qquad (2.4)$$

There are several features of (2.4) worth noting. First, there are no unobservable factors associated with the SP choice utilities, except of course those imbedded in the idiosyncratic

[3] Individual specific characteristics such as age, gender and education can also impact the site utilities, typically through interactions between individual and site characteristics. For now, we ignore these interaction effects for the sake of notational simplicity, but incorporate them later in both the Monte Carlo analysis and subsequent application.

[4] Of course, only $J - 1$ ASC's can be estimated, with one site's ASC normalized to zero.

error term. The random assignment of choice characteristics breaks the potential correlation between the observable treatments and any unobserved factors influencing the individual's decision. This is one of the key strengths of the stated preference approach. Second, while U_{ijt}^{RP} is constant over the choice alternatives (with, of course, the exception of the idiosyncratic error term), the utilities associated with the SP choices can vary substantially over the alternative choice occasions. This is a second key strength of the SP data.

Without further restrictions on the two sources of preference information, neither of the scale parameters μ^{RP} and μ^{SP} are identified and must be normalized to 1. The corresponding contribution of an individual to the likelihood function is then given by:

$$\mathcal{L}_i^{IC}\left(\theta^{IC}\right) = \left\{\prod_{j=1}^{J}\left[\frac{\exp(\alpha_j^{RP} + p_{ij}^{RP}\gamma^{RP})}{\sum_{m=1}^{J}\exp(\alpha_m^{RP} + p_{im}^{RP}\gamma^{RP})}\right]^{n_{ij}^{RP}}\right\} \cdot \prod_{h=1}^{H}\left[\prod_{k=1}^{K}\left\{\frac{\exp(X_{ikh}^{SP}\beta^{SP} + p_{ikh}^{SP}\gamma^{SP})}{\sum_{r=1}^{K}\exp(X_{irh}^{SP}\beta^{SP} + p_{irh}^{SP}\gamma^{SP})}\right\}^{1_{ikh}^{SP}}\right], \tag{2.5}$$

where $1_{ikh}^{SP} = 1$ if individual i chose alternative k in SP choice scenario h and equals zero otherwise and $\theta^{IC} \equiv (\alpha_\bullet^{RP}, \gamma^{RP}, \beta^{SP}, \gamma^{SP})$ denotes the parameter of the model, with $\alpha_\bullet^{RP} \equiv (\alpha_1^{RP}, \ldots, \alpha_{J-1}^{RP})$ denoting the complete vector of ASC's. The IC subscript (i.e., "inconsistent") on the likelihood function is used to indicate that this specification does not impose consistency between preferences underlying the RP and SP responses.

The insight of von Haefen and Phaneuf (2008) is that, by combining the two data sources and imposing consistency in the underlying data generating processes, portions of the RP preferences parameters can now be identified. Specifically, assuming that $\beta^{RP} = \beta^{SP} = \beta^C$ and $\gamma^{RP} = \gamma^{SP} = \gamma^C$, the corresponding likelihood function becomes:

$$\mathcal{L}_i^{C}\left(\theta^{C}\right) = \left\{\prod_{j=1}^{J}\left[\frac{\exp(X_j^{RP}\beta^C + \xi_j^C + p_{ij}^{RP}\gamma^C)}{\sum_{m=1}^{J}\exp(X_j^{RP}\beta^C + \xi_j^C + p_{im}^{RP}\gamma^C)}\right]^{n_{ij}^{RP}}\right\} \cdot \prod_{h=1}^{H}\left[\prod_{k=1}^{K}\left\{\frac{\exp\left[\omega(X_{ikh}^{SP}\beta^C + p_{ikh}^{SP}\gamma^C)\right]}{\sum_{r=1}^{K}\exp\left[\omega(X_{irh}^{SP}\beta^C + p_{irh}^{SP}\gamma^C)\right]}\right\}^{1_{ikh}^{SP}}\right]. \tag{2.6}$$

where $\omega \equiv \mu^{RP}/\mu^{SP}$ is the ratio of RP and SP scale parameters and $\theta^C \equiv (\xi_\bullet^C, \gamma^C, \beta^C, \omega)$ and $\xi_\bullet^C \equiv (\xi_1^C, \ldots, \xi_{J-1}^C)$. Note that, unlike in the case when consistency was not imposed, we can now estimate the composite impact of the unobservable factors (i.e., the ξ_j^C's). Also note that in imposing consistency we are not requiring that the scale parameter be the same across the two data sources.

2.3.2 Latent Class Model

The standard approach in the literature is to estimate both the consistent and inconsistent models (i.e., using the likelihood functions in equations (2.6) and (2.5), respectively) and to choose between the two specifications based standard tests. The model being proposed in this paper is to consider an in-between approach, allowing for the possibility that individuals differ in terms of the consistency of their RP and SP responses. In particular, we adopt latent class model with two distinct groups: *Class C* in which individual exhibit consistent preference parameters across their RP and SP data sources as in depicted in (2.6) and *Class IC* in which individuals have disparate RP and SP parameters as depicted in (2.5). Class membership is not known to the analysts. Therefore, the overall likelihood function (i.e., unconditional on class membership) for individual i can be formulated as

$$\mathcal{L}_i(\theta) = s\mathcal{L}_i^C\left(\theta^C\right) + (1-s)\mathcal{L}_i^{IC}\left(\theta^{IC}\right) \tag{2.7}$$

where $s \in [0,1]$ is the probability of being in the consistent class and $\theta \equiv (\theta^C, \theta^{IC}, s)$ denotes the full set of parameters to be estimated. The class membership probability can be modeled as a function of individual characteristics, including the individuals socio-demographic or attitudinal characteristics [see, e.g., Boxall and Adamowicz (2002)]. The advantage of this approach is that, by understanding the factors that influence membership in the inconsistent class, researchers may be able target corrective measures to avoid the inconsistencies themselves. For now, however, we focus on the simpler case is which the probability of class membership is a constant.

Equation (2.7) can used directly to estimate all of the model's parameters, including the class membership probability s, by standard maximum likelihood techniques. However, latent class models are notoriously difficult to estimate directly. Instead, following the current practice in the latent class literature [e.g., Morey et al. (2006), Evans and Herriges (2010)], we employ an Expectation-Maximization (EM) algorithm. The next subsection briefly describes steps involved in the EM algorithm used in our applications.

2.3.3 EM algorithm

EM algorithms can be useful for maximizing a likelihood function when standard optimization procedures can be numerically challenging, which is often the case in the presence of latent variables and particularly the case in latent class models. In our framework, the latent variable is class membership c_i, which equals C if the individual belongs to the consistent class and equals IC if the individual belongs to the inconsistent class, with $Pr(c_i = C) = s$.

The EM algorithm is an iterative procedure, alternating between two steps: 1) Calculating an expectation as a function of the current iteration's parameter values and 2) maximizing that expectation with respect to the parameters of the model. Specifically, following the general notation in chapter 14 of Train (2009), let θ_t denote the value of the parameters at iteration t. To maximize (2.7) using the EM algorithm, we define a new function evaluated at θ_t that can be used to obtain the parameter vector's next iteration; i.e., θ_{t+1}. Specifically, let

$$\begin{aligned}\mathcal{E}(\theta|\theta_t) &\equiv \sum_{i=1}^{N} \left\{ h_{it}^{C} \log\left[s\mathcal{L}_i^C(\theta^C)\right] + h_{it}^{IC} \log\left[(1-s)\mathcal{L}_i^{IC}(\theta^{IC})\right] \right\} \\ &= \sum_{i=1}^{N} \left[h_{it}^{C} \log(s) + h_{it}^{IC} \log(1-s)\right] + \sum_{i=1}^{N} h_{it}^{C} \log\left[(\mathcal{L}_i^C(\theta^C)\right] + \sum_{i=1}^{N} h_{it}^{IC} \log\left[\mathcal{L}_i^{IC}(\theta^{IC})\right]\end{aligned} \tag{2.8}$$

where s is the share of the population in class C and h_{it}^c denotes the probability of membership in class c ($c = C, IC$) conditional on the individual's observed choices. Using Bayes rule:

$$h_{it}^c = h(c_i = c|y_\bullet, s_t) = \frac{s_t\mathcal{L}_i^c(\theta^c)}{s_t\mathcal{L}_i^C(\theta^C) + (1-s_t)\mathcal{L}_i^{IC}(\theta^{IC})} \tag{2.9}$$

where $y_\bullet$ denotes the full set of choices (i.e., the n_{ij}^{RP}'s and 1_{ikh}^{SP}'s). Forming this expectation represents the first step in the EM algorithm.

The second step involves maximizing $\mathcal{E}(\theta|\theta_t)$ with respect to θ. Conveniently, as can be seen in equation (2.8), $\mathcal{E}(\theta|\theta_t)$ is separable into three distinct components that can be independently maximized. In particular, maximizing $\mathcal{E}(\theta|\theta_t)$ with respect to s corresponds to maximizing

$$\mathcal{E}(s|\theta_t) = \sum_{i=1}^{N} \left[h_{it}^{C} \log(s) + h_{it}^{IC} \log(1-s)\right], \tag{2.10}$$

yielding

$$s_{t+1} = \frac{\sum_{i=1}^{N} h_{it}^{C}}{\sum_{i=1}^{N} (h_{it}^{C} + h_{it}^{IC})}. \tag{2.11}$$

Maximizing $\mathcal{E}(\theta|\theta_t)$ with respect to θ^c ($c = C, IC$) corresponds to maximizing

$$\mathcal{E}(\theta^c|\theta_t) = \sum_{i=1}^{N} h_{it}^{c} \log\left[(\mathcal{L}_i^c(\theta^c)\right], \tag{2.12}$$

which is just class-specific maximum likelihood estimation using h_{it}^{c} as weights. The updated parameters (i.e., θ_{t+1}^{c}) are the corresponding solutions to these maximizations; i.e.,

$$\theta_{t+1}^{c} = arg \max_{\theta^c} \sum_{i=1}^{N} h_{it}^{c} \log\left[(\mathcal{L}_i^c(\theta^c)\right]. \tag{2.13}$$

Thus, the steps for estimation of the latent class model using the EM algorithm are

1. Specify initial values for the share and coefficients in each class. We set $s_0 = 0.5$ and obtain θ_0^c for class using unweighted maximum likelihood for that class.

2. Calculate the probability of being in each class conditional on the observed choices using (2.9).

3. Update the share of class C using (2.11).

4. Update the parameters of each class by estimating weighted MLE using (2.13)

5. Repeat steps 2-4 until convergence.

2.4 Generated Data Experiments

In this section, we describe a series of generated data experiments designed to illustrate the latent class model introduced in Section 3. Particular attention is paid to the performance of the model given different sample sizes and the proportion of the population belonging to the consistent class, as well as the impact of erroneously assuming that this class proportion is either zero or 1. Throughout, the pseudo-data sets were structured so as to mimic the general structure of the data set used in the application in Section 5.

As described in previous section, each individual is assumed to belong to either the consistent class ($c_i = C$) or inconsistent class ($c_i = IC$), with $Pr(c_i = C) = s$. Using a slight generalization of the model from the previous section (i.e., incorporating interactions between site and individual characteristics), the RP and SP conditional utilities for individuals belonging to the consistent class are assumed to take the form:

$$\begin{aligned} U_{ijt}^{RP} &= X_j^{RP}\beta^C + Z_i X_j^{RP}\rho^C + p_{ij}^{RP}\gamma^C + \xi_j + \mu^{RP}\varepsilon_{ijt} \\ U_{ikh}^{SP} &= X_k^{SP}\beta^C + Z_i X_k^{SP}\rho^C + p_{ik}^{SP}\gamma^C + \mu^{SP}\varepsilon_{ikh} \end{aligned} \tag{2.14}$$

where Z_i denotes an individual characteristics such as age, gender or education. On the other hand, for individuals belongs to the inconsistent class, these conditional utilities are assumed to take the form:

$$\begin{aligned} U_{ijt}^{RP} &= X_j^{RP}\beta^{RP} + Z_i X_j^{RP}\rho^{RP} + p_{ij}^{RP}\gamma^{RP} + \xi_j + \mu^{RP}\varepsilon_{ijt} \\ U_{ikh}^{SP} &= X_k^{SP}\beta^{SP} + Z_i X_k^{SP}\rho^{SP} + p_{ik}^{SP}\gamma^{SP} + \mu^{SP}\varepsilon_{ikh} \end{aligned} \tag{2.15}$$

In the generated data experiments, we consider a total of 15 scenarios varying the scenarios along two dimensions:

1. The probability of membership in the consistent class, with $s \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$; and
2. The number of observations, with $N \in \{200, 500, 1000\}$.

In all of the scenarios, the number of alternatives available on each choice occasion is fixed in the RP and SP settings, with $J = 20$ and $K = 3$, respectively. The corresponding total number of choice occasions are likewise fixed for the RP and SP settings, with $T = 10$ and $H = 15$, respectively. Finally, for each scenario, 100 generated data sets were constructed.

The specific steps used to generate data sets are as follows:

1. The vector of individual characteristics (Z_i), site characteristics (X_j), and travel costs (p_{ij}) were drawn from the following distribution:

$$\begin{aligned} Z_i &\sim N(0,1) \\ X_j^{RP} &\sim N(0,1) \\ X_k^{SP} &\sim N(0,2) \\ p_{ij}^{RP} &\sim \log N(0,1) \\ p_{ik}^{SP} &\sim \log N(0,2) \\ \xi_j &\sim N(-2,0.05) \end{aligned}$$

2. Each individual in the sample was then randomly assigned to either the consistent class (i.e., $c_i = C$) or the inconsistent class (i.e., $c_i = IC$), with $Pr(c_i = C) = s$.

3. Depending upon the class to which they were assigned, either equations (2.14) or equations (2.15) were then used to generate the conditional utilities U_{ijt}^{RP} and U_{ikh}^{SP} for each choice occasion and alternative employing the following parameters:

   - $\beta^C = -2.0$;
   - $\rho^C = -3.0$;
   - $\gamma^C = -0.8$; and
   - $\omega = 0.4$

   for the consistent class and

   - $\beta^{RP} = -1.2$;
   - $\rho^{RP} = -0.7$;
   - $\gamma^{RP} = -1.8$;
   - $\beta^{SP} = -0.6$;
   - $\rho^{SP} = -0.5$; and
   - $\gamma^{SP} = -0.4$.

for the inconsistent class. For both classes, the error terms (i.e., ε_{ijt}'s and ε_{ikh}'s) were drawn from the Type I extreme value distribution.

4. Given the conditional utilities U_{ijt}^{RP} and U_{ikh}^{SP} for each choice occasion, the individual's choices (i.e., 1_{ijt}^{RP} and 1_{ikh}^{SP}) were then determined by the alternative yielding the highest utility.

For each generated sample, we estimate three different models:

- *Model 1*: The latent class model described in Section 4 and based on the likehood function in equation (2.7);
- *Model 2*: The fully inconsistent model based on the likelihood function in equation (2.5); and
- *Model 3*: The fully consistent model based on the likelihood function in equation (2.6).

We then compare and contrast the three models in terms of the implied welfare impact from closing the most popular site in the sample.

Table 2.1 summarizes the resulting parameter estimates for Model 1.[5] In particular, for each scenario (i.e., combination of s and N), the table reports the mean parameter estimates across the 100 replications, as well as the corresponding 5th and 95th percentile values. Since Model 1 is consistent with the underlying data generating process, it is not surprising that the mean parameter estimates are generally quite close to the true parameters. However, the estimates are less stable when the share of individuals in the consistent class (i.e., s) is quite small. This is to be expected since the estimation then relies on relatively few individuals to identify the parameters for the consistent class. Somewhat unexpected is the fact that the parameter estimates are not as varied at the other extreme (i.e., when $s = 0.9$).

Parameter estimates using the other two models (i.e., Models 2 and 3), are provided in Appendix Tables A.1 and A.2, respectively. Since these models are not consistent with the

[5]Estimates for the alternative specific constants α_j^{RP} and ξ_j^C are not reported in Table 2.1 for the sake of space, but are available from the authors upon request. Also, estimates for the parameters β_{RP} are obtained through a second stage regression based on the fitted alternative specific constants from the first stage and using the relationship in (2.3).

underlying data generating process, it is not surprising that they tend to yield greater departures from the underlying parameters of the model. In general, Model 2 performs relatively well when most of the population is drawn from the inconsistent class (e.g. $s = 0.1$), whereas Model 3 performs relatively well when most of the population is drawn from the consistent class (e.g., $s = 0.9$).

Perhaps more important than the performance of a model in terms of individual parameter estimates is its performance in estimating the welfare impacts of a proposed policy scenario. Table 2 summarizes the performance of the three models in terms of estimating the average welfare impact of two policy scenarios:

- *Scenario A*: Closure of site 1.
- *Scenario B*: Improvement in site quality for alternative 1. This corresponds to a fifty percent reduction in X_1^{RP}.

For the latent class model (i.e., Model 1), the appropriate welfare measure is a weighted average of the compensating variation from the consistent and inconsistent class models, with the weights being the corresponding class probabilities; i.e.,

$$CV \quad = \quad s \times CV^C + (1-s)CV^{RP} \tag{2.16}$$

where s is the probability of being in the consistent class, with CV^C and CV^{RP} denote the standard log-sum calculations based on the consistent class and inconsistent class RP parameter estimates, respectively.

In contrast, the standard approaches in the literature are to either not impose consistency across the RP and SP data source (as in Model 2), computing compensating variation based on the RP parameter estimates, or to impose consistency for all individuals (as in Model 3), computing compensating variation based on the constrained parameter estimates derived from the two data sources.

Table 2.2 summarizes the mean absolute percentage errors (MAPE) associated with these three approaches, i.e.

$$MAPE \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\text{True Welfare Loss}_i - \text{Welfare Loss Estimates}_i}{\text{True Welfare Loss}_i}\right| \tag{2.17}$$

For all six experiments, the MAPEs are generally lowest for the latent class model (i.e., Model 1), which should be the case since it is in accord with the underlying data generating process. For Scenario A (the closure of site 1), the MAPE's from the latent class model lie between 5% and 12%, with the errors diminishing as the available sample size increases. The errors are larger for both single class specifications.

The MAPE's are substantially larger for Scenario B, ranging from approximately 10 percent when $N = 1000$ and $s = 0.5$ to over 70 percent when $N = 200$ and $s = 0.10$. This pattern is not surprising. The larger errors for Scenario B are expected, since welfare calculation in this case depends crucially on estimates of β^{RP}, which are obtained from a second stage regression of only $J = 20$ site alternative specific constants on site attributes X_j^{RP}. The MAPE's are uniformly smallest for the latent class model when $s = 0.5$, with the population evenly divided between the inconsistent and consistent classes, effectively providing a more balanced bases for estimating the underlying class parameters. In contrast, when $s = 0.1$, only 10 percent of the sample is assumed to be from the consistent class, providing little information for gleaning the parameters of that class. As was the case for Scenario A, Scenario B generally yields higher MAPE's for the single class specifications. The consistent class model performs best as the proportion of individuals in the consistent class is largest (i.e., $s = 0.9$), whereas the inconsistent class model performs best as the portion of individuals in the inconsistent class is largest (i.e., $s = 0.1$).

2.5 Application

2.5.1 Data

To illustrate our proposed latent class model, we reconsider the Moose Hunting data used by both Adamowicz et al. (1994) and von Haefen and Phaneuf (2008) to examine the potential for combining RP and SP data sources. The data for this study was collected from a sample of 422 individuals drawn from moose hunting license holders living in the Canadian towns of Drayton Valley, Edson, Hinton, Edmonton, and Whitecourt. Individuals were initially contacted by mail, with a follow-up phone call inviting them to attend a meeting. Of the 422 hunters initially contacted, 312 individuals (74%) agreed to attend the meeting. Of the 312 hunters who confirmed attendance, 271 (87%) actually attended the meeting.[6]

The study provides both revealed preference (RP) and stated preference (SP) data. The RP data consists of reported moose hunting trips to 14 wildlife management units (WMUs) during 1992, as well as respondent socio-demographic characteristics. SP data takes the form of a choice experiment in which each respondent was presented with a series of 16 choice scenarios (i.e., $H = 16$) each including three alternatives (i.e., $K = 3$), with two of the alternatives involving hypothetical sites while the third alternative was an opt-out (i.e., not hunting) option.[7]

Table 4.2 reports summary statistics for both individual and site characteristics.The mean age of hunters in the sample was just under forty years, and they had an average of about 20 years of general hunting experience and about 16 years of experience hunting moose. More than half of hunters completed high school and most of them reported incomes in the ranges of \$20,000 to \$60,000. For both real (RP) and hypothetical (SP) sites, the alternatives are defined in terms of six attributes: travel cost, moose population, level of congestion, access within hunting area (no trail, cutlines or seismic lines), quality of road and the presence of forest activity (logging).

2.5.2 Results

A total of four models were estimated using the Moose Hunting data:

1. *SC-Consistent*: A single class (SC) model imposing consistency across the RP and SP data sources;

2. *SC-RP*: A single class model of preferences based only on the RP data;

3. *SC-SP*: A single class model of preferences based only on the SP data;

4. *LC*: A latent class model with a portion s belonging to the consistent class (denoted *LC-Consistent*) and a portion $(1 - s)$ belong to the inconsistent class (denoted by *LC-RP* and *LC-SP* for the revealed and stated preference components, respectively).

[6]See McLeod et al. (1993) for additional details regarding the sampling and data collecting procedures.

[7]In empirical setting, we include dummy variable for 'not hunting' (SP dummy) to capture impact of the opt-out option.

Tables 2.4 and 2.5 provide the resulting parameter estimates. Table 4 focuses on the core parameters; i.e., the class share s in the case of the latent class model, the relative RP/SP scale parameter ω identified only when consistency is imposed for a class, and the travel cost parameters (i.e., the γ's). Table 2.5 reports the main effect of site characteristics (i.e, the β's) and interactions between site characteristics and individual attributes (i.e., the ρ's).[8]

Starting with Table 2.4, the latent class model indicates that the population is roughly evenly divided between the consistent and inconsistent classes, with $s = 0.53$. Both the single and latent class models indicate a significant difference in scales between the RP and SP responses, with ω in the range of 0.18 to 0.22. This indicates that there is greater variability in the unobservable components of individual preferences in the case of SP data relative to RP data (i.e.; $\mu^{RP} < \mu^{SP}$). Finally, while all of the specifications yield negative and statistically significant travel cost coefficient, the γ's vary substantially. Cross-model comparisons of the estimated γ's is difficult, since the scale parameter differences between the RP and SP models cannot be estimated when consistency is not imposed. However, it does appear as though the latent class structure highlights the gap between consistent and inconsistent preferences. In particular, the marginal utility of income ($-\gamma$) is largest when it is imposed for only a portion of the population, rather than for the population as a whole. Or, to put it another way, the consistent class appears to consist of individuals whose choices are substantially influenced by price.

Turning to Table 2.5, note that there are two sets of parameters being presented. The first column of parameters are the main effects associated with the site characteristics; i.e., the β's in equation (1). For those models involving only the RP data, the β's can generally only be recovered in a second stage regression using the estimated ASC's (i.e., the α_j's) and equation (3).[9] However, with $J = 20$, the main effects for the eleven site characteristics used by von Haefen and Phaneuf (2008) cannot be reasonably estimated and are not reported here. The second set of parameters are the ρ's in equation (14), reflecting interactions between individual

[8]The parameter estimates reported here for the single class models have the same signs and are similar in magnitude to those reported in von Haefen and Phaneuf (2008), though the specifications differ in that von Haefen and Phaneuf incorporate a mixed logit structure.

[9]One exception is the main effect for the "unpaved" site access, since this characteristic varies across sites and individuals because individuals choose different roads to assess the sites.

and site characteristics. In general, these parameters vary substantially across the various RP and SP specifications, often changing signs and significance. The pattern of these parameters for the single class models are similar to those reported in von Haefen and Phaneuf (2008).

Interpreting the individual parameters in Table 5 is difficult. In order to illustrate the differences across the various models, we consider their implications in terms of welfare estimates for three specific scenarios employed by von Haefen and Phaneuf (2008) :

- Case 1: Closer site WMU #344.
- Case 2: Decrease moose population from more than 4 per day to 3-4 per day at WMU #348.
- Case 3: Increase moose population from less than 1 per day to 1-2 moose per day at WMU #344.

We assume fixed coefficients within a class we can use standard *log-sum term* for computing welfare change. Formally, we can write the deterministic component of utility as following;

$$\begin{aligned} V_{ijt} &= V(X_j, Q_j; \theta) \\ &= (\beta_0 + Z_i\beta_1)X_j + Q_j\beta_q + p_{ij}\gamma \end{aligned} \tag{2.18}$$

where Q_j denotes moose population at site j, and p_{ij} denotes travel costs from household's residence to site j, and X_j and Z_i represent the other site attributes and sociodemographic variables respectively. Compensating variation (CV) associated with a change of moose population from Q_j^0 to Q_j^1 is

$$CV_h(\beta^h) = \frac{1}{\beta_p}\left\{\log\left(\sum_{j=1}^{J}\exp\left[V(X_j, Q_j^1; \theta^h)\right]\right) - \log\left(\sum_{j=1}^{J}\exp\left[V(X_j, Q_j^0; \theta^h)\right]\right)\right\} \text{ for } h = \mathrm{C}, \mathrm{IC}^{\mathrm{RP}}, \mathrm{IC}^{\mathrm{SP}} \tag{2.19}$$

From (2.19), we can estimate three values of CV for latent class model, i.e. CV using estimates of consistent class, estimates of RP of inconsistent class, and SP of inconsistent class. Therefore, we consider two alternative strategies for construction welfare measure. The first strategy is to use CV of only consistent class. In this case, we consider individuals who responded same way between RP and SP as the respondents who said truthful preference. The second is to adopt weighted average of two values shown in (2.16).

Table 2.6 shows the results of welfare analysis. Although the estimates for case 1 are not qualitatively quite different, the estimate of combined model is slightly larger (in absolute sense) than only consistent class, however, smaller than weighted estimated with both RP and SP. Welfare results for case 2 and case 3 also have similar patterns to case 1.[10]

2.6 Conclusion

Revealed preference data (RP) are based on actual choices of respondents while stated preference data (SP) are collected in experimental or survey situations. Therefore, both have obvious advantages and limitations. The advantage of RP data is that they are the collection of real choices, which reflect their budget constraint and other variables. However, since they rely on historical data, variation of alternative attributes is limited and it makes difficult to analyze new policy beyond currently existing status. While SP data have much variation relying on experimental design, they obviously have hypothetical bias. To mitigate the limitations and get advantages from both data, combining revealed and stated preference data is common in recent environmental economics, marketing and transportation literature. Moreover, the data make it easy to estimate models with unobserved attributes without depending on additional econometric technique such as Murdock's two-stage estimation.

It, however, relies on underlying assumption that both data have common data-generating process. In other words, both data must have same coefficients. However, the assumption was often rejected in previous studies [Jeon and Herriges (2010), von Haefen and Phaneuf (2008)]. Combined RP/SP strategy is still used to compute welfare analysis in some prior studies due to strong points relative to either RP or SP data model even when the assumption is not satisfied. Although previous literature proposed to selectively use the parameter estimates from several different models using single data, it is *ad hoc* or implicitly rely on cross-equation restriction.

[10]For case 2 and 3, welfare estimates cannot be recovered since there is no variable estimates for the site attributes, i.e. no mean effect.

The purpose of this paper has been to introduce an alternative framework for combining revealed and stated preference data. The literature typically considers only two possible scenarios: either respondent's behaviors in RP and SP are consistent for everyone or they are consistent for no one. In this paper, we suggest a middle ground, using latent class approach to segment the population into two groups. The first group has RP and SP responses that are internally consistent, while the remaining group exhibits some form of inconsistent preferences. Moreover, as usual latent class model, we propose EM algorithm which is an iterative procedure to converge maximum likelihood estimation due to the numerical difficulty in empirical study.

We illustrated in our generated data experiments that ignoring discrepancy between real and hypothetical choices makes huge biases in estimating parameters and welfare analysis while our method shows much small bias. It implies that our method takes advantages from combining two data and controls convergent validity assumption as well.

Our empirical application, using moose hunting data in Alberta, Canada, provides evidences of heterogeneity from the individual's propensity to show differences between RP and SP data. As previous study [von Haefen and Phaneuf (2008)] using same data we used in current study pointed out explicitly that the convergent validity assumption are not satisfied, our proposed model also shows almost half of individuals responds different ways between RP and SP data. This difference makes different parameter estimates between two classes. Obviously, our model results in different welfare estimates to combined single class RP/SP model for several different welfare loss or gain scenarios. Our results imply that ignoring heterogeneous responses in two data source can mislead welfare analysis.

There is an unresolvable question which one of two latent class model estimates is better than the other. Even though weighted averages with RP show small error in our experiments, we suggest that researchers and policy makers choose either one based on empirical data. As mentioned in von Haefen and Phaneuf (2008), RP data usually have not enough variations which make a difficulty in estimation specially in the presence of unobservable. On the other hand, SP data has the limitations that they may behave differently in real trips. Therefore, in case that RP data have abundant variations and there is no identification problems, we recommend to use weighted RP estimates, otherwise weighted SP estimates.

Bibliography

Wiktor Adamowicz, J. Louviere, and M. Williams. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26, 1994.

Wiktor Adamowicz, Jeffrey Swait J., Peter Boxall, Jordon Louviere, and M. Williams. Perceptions vs. objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of Environmental Economics and Management*, 32, 1997.

Christopher Azevedo, Joseph A. Herriges, and C. Kling. Combining revealed and stated preference : consistency tests and their interpretations. *American Journal of Agricultural Economics*, 85:525–537, 2003.

M. Ben-Akiva and T. Morikawa. Estimation of switching models from revealed preferences and stated intentions. *Transportation Research A*, 24, 1990.

M. Ben-Akiva, M. Bradley, T. Morikawa, J. Benjamin, T. Novak, H. Oppewal, and V. Rao. Combining revealed and stated preferences data. *Marketing Letters*, 85(4):335–350, 1994.

Richard C. Bishop and Thomas A. Heberlein. Measuring values of extra market goods: Are indirect measures biased? *American Journal of Agricultural Economics*, 61(5):926–930, 1979.

P.C. Boxall and W.L. Adamowicz. Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and Resource Economics*, 23, 2002.

William Breffle, Edward Morey, and J. Thacher. A joint latent-class model : combining likert-scale preference statements with choice data to harvest preference heterogeneity. *Environmental and Resource Economics*, 50, 2011.

Trudy A. Cameron. Combining contingent valuation and travel cost data for the valuation of nonmarket goods. *Land Economics*, 68(3):302–317, 1992.

D. Dosman and Wiktor Adamowicz. Combining stated and revealed preference data to construct an empirical examination of intrahousehold bargaining. *Review of Economics of the Household*, 4, 2006.

Y.-S. Eom and D.M. Larson. Improving environmental valuation estimates through consistent use of revealed and stated preference information. *Journal of Environmental Economics and Management*, 52, 2006.

Keith Evans and Joseph A. Herriges. Rounding in recreation demand model: A latent class count model. *ISU working paper*, 2010.

Ju-Chin Huang, Timothy C. Haab, and John C. Whitehead. Willingness to pay for quality improvements: Should revealed and stated preference data be combined? *Journal of Environmental Economics and Management*, 34(3):240–255, 1997.

Yongsik Jeon and Joseph A. Herriges. Convergent validity of contingent behavior responses in models of recreation demand. *Environmental and Resource Economics*, 45, 2010.

Koichi Kuriyama, W. Michael Hanemann, and James R. Hilger. A latent segmentation approach to a kuhn-tucker model : An application to recreation demand. *Journal of Environmental Economics and Management*, 60, 2010.

K. McLeod, Peter C. Boxall, Wiktor L. Adamowicz, M. Williams, and Jordon J. Louviere. The incorporation of non-timber goods and services in integrated resource management. *Department of Rural Economy, University of Alberta*, 1993.

Edward Morey, Robert D. Rowe, and Michael Watson. A repeated nested-logit model of atlantic salmon fishing. *American Journal of Agricultural Economics*, 75(3):578–592, 1993.

Edward Morey, Jennifer Thacher, and William Breffle. Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model. *Environmental and Resource Economics*, 34(1):91–115, 2006.

Jennifer Murdock. Handling unobserved site characteristics in random utility models of recreation demand. *Journal of Environmental Economics and Management*, 51, 2006.

Allen Randall. A difficulty with the travel cost method. *Land Economics*, 70, 1994.

Kenneth Train. *Discrete Choice Methods with Simulation, 2nd edition.* Cambridge University Press, 2009.

Roger H. von Haefen and Daniel J. Phaneuf. Incorporating observed choice into construction of welfare measures from random utility models. *Journal of Environmental Economics and Management*, 45, 2003.

Roger H. von Haefen and Daniel J. Phaneuf. Identifying demand parameters in the presence of unobservables : A combined revealed and stated preference approach. *Journal of Environmental Economics and Management*, 56, 2008.

J.C. Whitehead, Pattanayak S., Van Houtven, and GelsoB. Combining revealed and stated preference data to estimate the nonmarket value of ecological services : an assessment of the state of the science. *Journal of Economic Survey*, 22, 2008.

J.C. Whitehead, D.J. Phaneuf, C.F. Dumas, J. Herstine, J. Hill, and B. Buerger. Convergent validity of revealed and stated behavior with quality change: A comparison of multiple and single site demands. *Environmental and Resource Economics*, 45, 2010.

Table 2.1 Generated Data Experiments - Model 1 Parameter Estimates

Class	Parameter	TRUE s	TRUE values	N=200 Mean	N=200 5%	N=200 95%	N=500 Mean	N=500 5%	N=500 95%	N=1000 Mean	N=1000 5%	N=1000 95%
		0.10	0.10	0.24	0.06	0.89	0.19	0.07	0.89	0.21	0.07	0.90
		0.25	0.25	0.31	0.19	0.75	0.30	0.20	0.72	0.28	0.23	0.61
	s	0.50	0.50	0.50	0.43	0.58	0.50	0.45	0.54	0.50	0.47	0.54
		0.75	0.75	0.75	0.68	0.82	0.75	0.71	0.78	0.75	0.73	0.78
		0.90	0.90	0.90	0.85	0.94	0.90	0.88	0.93	0.90	0.88	0.92
		0.10	0.40	0.38	0.22	0.56	0.38	0.24	0.49	0.38	0.24	0.45
		0.25	0.40	0.39	0.26	0.49	0.39	0.24	0.45	0.39	0.35	0.43
	ω	0.50	0.40	0.40	0.36	0.45	0.40	0.37	0.43	0.40	0.38	0.42
		0.75	0.40	0.40	0.36	0.45	0.40	0.37	0.42	0.40	0.39	0.41
		0.90	0.40	0.40	0.37	0.44	0.40	0.37	0.42	0.40	0.39	0.41
		0.10	-2.00	-2.24	-3.35	-1.43	-2.16	-2.74	-1.58	-2.10	-2.74	-1.69
		0.25	-2.00	-2.05	-2.61	-1.59	-2.05	-2.61	-1.73	-2.02	-2.24	-1.83
Consistent	β^C	0.50	-2.00	-2.01	-2.32	-1.75	-2.01	-2.18	-1.82	-1.99	-2.12	-1.87
		0.75	-2.00	-2.01	-2.25	-1.77	-2.01	-2.15	-1.86	-2.00	-2.09	-1.91
		0.90	-2.00	-2.01	-2.24	-1.81	-2.00	-2.14	-1.88	-2.01	-2.10	-1.92
		0.10	-1.20	-1.29	-2.06	-1.03	-1.24	-1.90	-1.05	-1.30	-1.99	-1.10
		0.25	-1.20	-1.23	-1.76	-0.90	-1.21	-1.98	-0.79	-1.22	-1.39	-1.06
	β^{RP}	0.50	-1.20	-1.21	-1.47	-1.04	-1.19	-1.34	-1.08	-1.21	-1.33	-1.09
		0.75	-1.20	-1.33	-2.27	-1.10	-1.21	-1.36	-1.08	-1.22	-1.36	-1.11
		0.90	-1.20	-1.82	-3.48	-1.00	-1.32	-1.98	-1.00	-1.24	-1.45	-1.05
Inconsistent		0.10	-0.60	-0.62	-0.89	-0.53	-0.62	-0.81	-0.57	-0.62	-0.80	-0.57
		0.25	-0.60	-0.61	-0.78	-0.53	-0.61	-0.74	-0.53	-0.61	-0.66	-0.56
	β^{SP}	0.50	-0.60	-0.60	-0.70	-0.53	-0.60	-0.65	-0.57	-0.60	-0.63	-0.57
		0.75	-0.60	-0.61	-0.75	-0.50	-0.61	-0.69	-0.55	-0.60	-0.66	-0.56
		0.90	-0.60	-0.64	-0.91	-0.44	-0.61	-0.73	-0.48	-0.61	-0.69	-0.54
		0.10	-3.00	-2.48	-4.19	-0.70	-2.64	-3.58	-0.77	-2.64	-3.30	-0.75
		0.25	-3.00	-2.71	-3.47	-0.80	-2.73	-3.24	-0.89	-2.85	-3.15	-1.32
Consistent	γ^C	0.50	-3.00	-2.94	-3.32	-2.61	-2.95	-3.19	-2.81	-2.98	-3.11	-2.87
		0.75	-3.00	-3.01	-3.21	-2.76	-3.01	-3.17	-2.88	-2.99	-3.09	-2.90
		0.90	-3.00	-3.02	-3.20	-2.82	-3.01	-3.16	-2.89	-3.00	-3.10	-2.92
		0.10	-0.70	-1.00	-3.29	-0.64	-0.93	-2.94	-0.66	-1.02	-3.02	-0.67
		0.25	-0.70	-0.88	-2.80	-0.62	-0.88	-3.03	-0.65	-0.80	-1.26	-0.67
	γ^{RP}	0.50	-0.70	-0.73	-1.25	-0.61	-0.73	-0.80	-0.64	-0.71	-0.76	-0.66
		0.75	-0.70	-0.71	-0.88	-0.54	-0.69	-0.78	-0.61	-0.71	-0.78	-0.65
		0.90	-0.70	-0.73	-1.12	-0.35	-0.72	-0.89	-0.59	-0.72	-0.84	-0.61
Inconsistent		0.10	-0.50	-0.64	-1.27	-0.46	-0.59	-1.22	-0.47	-0.60	-1.27	-0.48
		0.25	-0.50	-0.59	-1.25	-0.46	-0.57	-1.18	-0.47	-0.54	-0.92	-0.48
	γ^{SP}	0.50	-0.50	-0.53	-0.80	-0.43	-0.52	-0.58	-0.46	-0.51	-0.54	-0.47
		0.75	-0.50	-0.50	-0.63	-0.36	-0.50	-0.57	-0.44	-0.50	-0.54	-0.45
		0.90	-0.50	-0.52	-0.72	-0.33	-0.51	-0.61	-0.40	-0.50	-0.57	-0.45

Table 2.1: Generated Data Experiments - Model 1 Parameter Estimates (cont'd)

Class	Parameter	TRUE s	TRUE values	N=200 Mean	N=200 5%	N=200 95%	N=500 Mean	N=500 5%	N=500 95%	N=1000 Mean	N=1000 5%	N=1000 95%
Consistent	ρ^C	0.10	-0.80	-1.05	-1.86	-0.63	-0.94	-1.77	-0.66	-0.94	-1.77	-0.73
		0.25	-0.80	-0.90	-1.73	-0.70	-0.89	-1.76	-0.72	-0.85	-1.23	-0.76
		0.50	-0.80	-0.81	-0.95	-0.73	-0.81	-0.86	-0.75	-0.80	-0.84	-0.77
		0.75	-0.80	-0.80	-0.88	-0.74	-0.80	-0.84	-0.77	-0.80	-0.83	-0.77
		0.90	-0.80	-0.80	-0.86	-0.75	-0.80	-0.83	-0.77	-0.80	-0.83	-0.77
Inconsistent	ρ^{RP}	0.10	-1.80	-1.67	-1.92	-0.92	-1.69	-1.86	-0.83	-1.66	-1.84	-0.78
		0.25	-1.80	-1.73	-1.94	-0.89	-1.72	-1.86	-0.89	-1.75	-1.85	-1.57
		0.50	-1.80	-1.79	-1.98	-1.62	-1.79	-1.90	-1.70	-1.80	-1.86	-1.74
		0.75	-1.80	-1.83	-2.08	-1.63	-1.80	-1.93	-1.69	-1.81	-1.92	-1.71
		0.90	-1.80	-1.79	-2.20	-1.29	-1.83	-2.05	-1.67	-1.83	-2.05	-1.67
	ρ^{SP}	0.10	-0.40	-0.39	-0.44	-0.31	-0.39	-0.42	-0.34	-0.39	-0.42	-0.32
		0.25	-0.40	-0.39	-0.43	-0.33	-0.39	-0.43	-0.34	-0.39	-0.42	-0.36
		0.50	-0.40	-0.40	-0.45	-0.35	-0.40	-0.43	-0.36	-0.40	-0.42	-0.38
		0.75	-0.40	-0.40	-0.48	-0.33	-0.40	-0.45	-0.36	-0.40	-0.43	-0.37
		0.90	-0.40	-0.42	-0.53	-0.28	-0.40	-0.48	-0.33	-0.40	-0.46	-0.35

Table 2.2 Generated Data Experiments: Welfare Performance of Model

		Class Ratio	Latent Class	Single Class	
Scenario	N	(s)	Weighted	Consistant	Inconsistent
A	200	0.10	10.11	21.64	44.49
		0.25	11.82	29.99	38.46
		0.50	11.69	35.88	32.42
		0.75	11.14	28.12	30.61
		0.90	11.79	17.23	32.56
	500	0.10	7.07	19.90	45.04
		0.25	7.50	28.42	39.31
		0.50	6.68	34.56	33.84
		0.75	7.75	29.47	30.93
		0.90	7.63	16.26	32.34
	1000	0.10	4.90	19.20	46.44
		0.25	4.94	26.57	40.44
		0.50	5.35	32.81	34.70
		0.75	5.45	25.90	32.69
		0.90	5.28	13.99	34.00
B	200	0.10	70.37	128.93	56.16
		0.25	65.26	401.35	68.26
		0.50	21.37	60.02	75.25
		0.75	24.30	54.80	96.97
		0.90	29.00	32.91	109.65
	500	0.10	36.24	118.71	56.37
		0.25	35.74	92.42	65.99
		0.50	14.03	141.91	78.15
		0.75	27.42	274.40	190.96
		0.90	14.60	27.67	109.58
	1000	0.10	30.72	110.55	56.81
		0.25	13.08	81.49	66.47
		0.50	9.50	88.28	74.95
		0.75	10.11	67.33	123.67
		0.90	12.78	24.34	248.24

Table 2.3 Summary Statistics

Variables		Mean	Std. Dev	Minimum	Maximum
Socioeconomic	Age (year)	39.63	10.71	18	70
Attributes	Income ($)	51,722	22,809	10,000	110,000
	Total number of trip	3.62	5.68	0	41
	General hunting experience (year)	20.19	10.24	2	51
	Moose hunting experience (year)	16.88	9.87	1	49
	Edmonton resident dummy[a]	0.45	0.48	0	1
	High school diploma dummy	0.91	0.27	0	1
Site Attributes	Travel cost ($)	219.71	101.69	88.64	558.92
	Moose population(effects coded)[b]				
	less than 1 moose per day	0.14	0.52	-1	1
	1-2 moose per day	0.5	0.63	-1	1
	3-4 moose per day	0.07	0.46	-1	1

[a] Edmonton is unique urban region in this data set, which is relatively far from hunting area.
[b] Seeing or hearing moose or seeing fresh sign such as tracks browse or droppings. McLeod et al. (1993)

Table 2.4 Core Parameter Estimates

Parameter	Model	Est.	t-stat
Class Share (s)	Latent Class (LC)	**0.53**	11.6
RP/SP Scale (μ)	Single Class (SC) - Consistent	**0.22**	6.68
	LC-Consistent	**0.18**	10.57
Travel Cost	SC-Consistent	**-1.65**	-6.97
	SC-RP	**-1.51**	-22.24
	SC-SP	**-0.42**	-28.59
	LC-Consistent	**-3.57**	-14.1
	LC-RP	**-1.02**	-6.73
	LC-SP	**-0.34**	-9.00

Table 2.5 Parameter Estimates for Site Characteristics

		Main		Interaction Effect					
				Gen Hunt Exp		Edmonton		HS diploma	
Parameter	Model	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
Unpaved[a]	SC-Consistent	0.41	0.70	-1.19	-0.71	**0.51**	1.98	-0.45	-0.97
	SC-RP	**0.79**	2.14	**-1.05**	-2.26	0.00	0.00	**-0.83**	-2.37
	SC-SP	-0.02	-0.08	-0.33	-0.84	0.11	1.70	0.04	0.20
	LC-Consistent	-0.35	-0.11	-2.34	-1.52	0.83	1.28	0.50	0.15
	LC-RP	2.13	0.20	0.76	0.25	-0.01	0.00	-3.61	-0.34
	LC-SP	-0.19	-0.14	0.37	0.31	0.08	0.54	0.10	0.08
No Trail	SC-Consistent	-1.73	-0.89	0.44	0.16	**-1.93**	-3.17	1.44	0.75
	SC-RP	-	-	0.02	0.00	0.31	0.00	0.21	0.00
	SC-SP	-0.38	-1.28	0.02	0.05	**-0.45**	-5.52	0.33	1.22
	LC-Consistent	0.49	0.05	-0.93	-0.20	**-4.28**	-4.54	0.19	0.02
	LC-RP	-	-	-0.77	0.00	0.35	0.00	1.87	0.00
	LC-SP	-1.03	-0.52	1.04	0.70	-0.25	-1.00	0.53	0.28
Old Trail	SC-Consistent	1.12	0.96	-1.95	-1.12	**1.89**	4.26	0.01	0.01
	SC-RP			-2.00	0.00	1.28	0.00	0.40	0.00
	SC-SP	0.17	0.57	0.12	0.20	0.15	1.65	0.12	0.55
	LC-Consistent	0.70	0.17	**-5.40**	-2.45	**4.41**	9.00	0.05	0.01
	LC-RP	-	-	**3.22**	2.00	0.03	0.09	**-5.49**	-5.20
	LC-SP	0.50	0.27	-0.49	-0.33	-0.24	-1.12	0.27	0.15
4WD Trail	SC-Consistent	0.65	0.88	1.76	0.94	0.20	0.54	-0.50	-0.85
	SC-RP	-	-	2.06	0.00	-0.35	0.00	0.23	0.00
	SC-SP	0.32	1.21	0.07	0.13	**0.20**	2.38	-0.30	-1.30
	LC-Consistent	-0.84	-0.16	**6.89**	2.99	-0.32	-0.53	0.36	-0.53
	LC-RP	-	-	-5.53	0.00	1.02	0.00	11.09	0.00
	LC-SP	0.83	0.53	-0.84	-0.66	0.24	1.28	-0.60	-0.41
No Hunters	SC-Consistent	**2.61**	2.29	-3.57	-1.55	-0.12	-0.27	1.10	1.31
	SC-RP	-	-	-0.75	0.00	0.52	0.00	-0.10	0.00
	SC-SP	**0.56**	2.37	-0.79	-1.36	0.04	0.40	0.24	1.18
	LC-Consistent	1.19	0.12	-1.07	-0.20	-0.22	-0.22	2.95	0.31
	LC-RP	-	-	-2.49	0.00	0.39	0.00	3.18	0.00
	LC-SP	1.09	0.59	-1.69	-1.04	0.02	0.11	-0.04	-0.02
On ATV	SC-Consistent	-0.93	-1.31	0.39	0.22	**0.91**	2.76	-0.73	-1.25
	SC-RP	-	-	-1.49	0.00	1.05	0.00	-0.19	0.00
	SC-SP	-0.38	-1.34	0.22	0.36	0.09	0.99	0.05	0.18
	LC-Consistent	-0.19	-0.03	2.07	0.62	**1.61**	2.58	-2.04	-0.36
	LC-RP	-	-	-4.97	0.00	0.78	0.00	6.34	0.00
	LC-SP	-1.15	-0.54	1.51	0.85	0.11	0.52	0.47	0.22
No logging	SC-Consistent	-0.21	-0.61	1.57	1.13	0.10	0.39	0.03	0.13
	SC-RP	-	-	**1.81**	3.22	**-0.41**	-2.52	0.00	0.01
	SC-SP	0.05	0.29	0.31	1.09	0.01	0.17	-0.09	-0.48
	LC-Consistent	-0.08	-0.02	3.55	3.84	-0.05	-0.16	-0.53	-0.12
	LC-RP	-	-	-7.78	-1.76	**-1.50**	-2.37	**10.90**	2.14
	LC-SP	0.27	0.16	-0.45	-0.49	0.00	0.01	-0.12	-0.07

Table 2.5: Parameter Estimates for Site Characteristics (cont'd)

		Main		Interaction Effect					
				Gen Hunt Exp		Edmonton		HS diploma	
Parameter	Model	Est.	t-stat	Est.	t-stat	Est.	t-stat	Est.	t-stat
< 1 Moose	SC-Consistent	**-5.94**	-5.41	1.64	1.08	-0.04	-0.10	0.13	0.32
	SC-RP	-	-	**1.68**	2.45	0.26	1.29	0.46	1.05
	SC-SP	**-1.00**	-2.97	-0.24	-0.45	-0.03	-0.34	-0.19	-0.62
	LC-Consistent	**-7.37**	-2.29	2.97	1.20	**-2.10**	-2.05	-0.77	-0.24
	LC-RP	-	-	1.49	0.56	6.10	0.00	3.95	0.47
	LC-SP	-1.03	-0.55	-0.65	-0.44	-0.09	-0.39	0.01	0.00
1-2 Moose	SC-Consistent	-0.49	-0.87	-2.72	-1.71	**1.64**	4.87	0.19	0.50
	SC-RP	-	-	**-3.37**	-4.83	**2.50**	13.91	0.31	0.74
	SC-SP	-0.04	-0.09	-0.04	-0.07	-0.09	-0.91	0.05	0.13
	LC-Consistent	-0.57	-0.27	0.99	0.86	**2.15**	4.53	-0.06	-0.03
	LC-RP	-	-	**-5.90**	-2.55	8.25	0.00	2.62	0.37
	LC-SP	0.01	0.01	-0.44	-0.26	-0.26	-1.08	0.07	0.03
3-4 Moose	SC-Consistent	**1.67**	2.37	1.03	0.70	-0.29	-0.82	0.31	0.53
	SC-RP	-	-	0.70	0.98	0.32	1.80	0.30	0.99
	SC-SP	0.31	0.84	0.31	0.56	0.01	0.08	0.08	0.23
	LC-Consistent	3.32	1.38	1.24	0.94	0.01	0.02	-0.39	-0.16
	LC-RP	-	-	-2.25	-0.93	4.31	0.00	5.71	2.55
	LC-SP	0.13	0.06	0.69	0.39	0.08	0.27	0.09	0.04
SP	SC-Consistent	**-5.99**	-2.93	-3.65	-0.82	-1.35	-1.40	-0.49	-0.31
outside dummy	SC-RP	-	-	-	-	-	-	-	-
	SC-SP	**-1.45**	-11.14	**-0.81**	-2.57	**-0.31**	-6.31	-0.12	-1.20
	LC-Consistent	**-9.76**	-5.81	**10.46**	3.09	0.91	1.62	-1.10	-0.72
	LC-RP	-	-	-	-	-	-	-	-
	LC-SP	-0.82	-0.18	**-9.60**	-4.75	**-0.73**	-3.01	-0.12	-0.03

Boldface indicated statistical significance at the 5% level. We exclude one site attribute, 'On foot' (Encounters with other hunters on foot), which is used in von Haefen and Phaneuf (2008) since 'On foot' has perfectly same value as 'No Hunter', which make perfect multicolliearity problem.

[a] Unpaved site characteristics varies across sites and individual because individuals choose different roads to assess the sites.

Table 2.6 The results of Welfare analysis

Model	Case1	Case2	Case3
Single Class: Consistent	-3.46	-9.48	99.76
Single Class: RP	-3.76	-	-
Latent Class: Only consistent Class	-3.18	-3.90	72.27
Latent Class : weighted with RP	-4.31	-	-

CHAPTER 3. DOES OBESITY MATTER FOR THE ENVIRONMENT? EVIDENCE FROM VEHICLE CHOICES AND DRIVING

3.1 Introduction

The rising rate of obesity has become a prominent social concern in the U.S. and throughout the world. A recent report of the National Coalition on Health Care (NCHC) shows that 35.7 % of U.S. adults are obese in 2009-2010 (NCHS Data Brief 2012). It is a well-known fact that obesity causes public health problems, such as high blood pressure, heart disease, and a number of other adverse health conditions. These health problems increase medical costs to society as well. Cawley and Meyerhoefer (2012), for example, argue that obesity is associated with $2,741 (in 2005 dollars) per person higher annual medical health care costs in the U.S.

A number of recent studies suggest the societal impacts of obesity also extend to the environmental arena. The studies show that obesity can be an additional indirect factor leading to increased gasoline consumption through several different channels. Broadly, the literature focuses on the positive relation between obesity, vehicle miles traveled (VMT), and the share of light trucks among noncommercial vehicles.[1] Specifically, the literature analyzing how obesity influences households' driving or vehicle choice behavior can be classified into three categories: (1) An engineering approach, (2) A focus on the relationship between VMT and obesity, and (3) A focus on the relationship between vehicle choices and obesity.

The first group of studies examine the mechanical relationship between weight and vehicle fuel efficiency. Jacobson and McLay (2006) show that obesity and overweight people increase fuel consumption of noncommercial vehicles by 0.8 % due to the resulting additional weight in vehicles. This corresponds to approximately one billion gallons of additional gasoline consumed

[1] Figure 3.1 shows the trend and relation between variables.

in the U.S. on an annual basis. Similarly, Dannenberg (2004) indicate that the weight increase of U.S. passengers account for up to a 2.4 % increase in fuel consumption by the U.S. airline industry.

The second group of studies examine how obesity and overweight affect vehicle use and vice versa. Courtemanche (2011) examines the sequential relation that lower fuel prices cause people to use their vehicles more frequently, in turn, causing individuals to become more obese. Using cross-sectional individual level data from the Behavioral Risk Factor Surveillance System (BRFSS), he found that a 1% increase in gasoline price would lead to a 10% reduction in the rate of obesity and overweight. Jacobson et al. (2011) argue a positive relationship exists between the number of miles driven by each licensed driver (VMT/LD) and adult obesity with a six-year time lag. They estimate a one-mile increase in driving by all licensed drivers would result in a 2.2% increase in the adult obesity rate six years later. However, it is likely to be a spurious correlation because they do not control for other factors such as income, which may affect this relationship and the corresponding time trend. For example, the video game market has been growing tremendously over the last few decades. Nowadays, many adults and children enjoy playing video games rather than participating in physical exercise. This would also cause an increase in overweight and obesity. Without controlling for such factors, the causal relationship between driving patterns and obesity can be difficult to discern.

Finally, Li et al. (2011) take this argument a step further by suggesting that obese and overweight individuals also contribute to emissions through their vehicle choices. They estimate the impact of obesity on vehicle choice by adopting the BLP-type aggregate data logit model with county-level annual sales data. They find that new vehicles purchased by individuals, who are obese or overweight, are, on average, less fuel efficient than those selected by others in the general population. Their simulation results show that, had the U.S. rates for overweight and obesity in 2005 had remained instead at their 1981 levels, there would have been a fuel savings of 8.6%. However, in calculating these fuel savings, they implicitly assume that the VMT of overweight and obesity people is equal to the VMT of the other individuals in the population.

While the results in earlier works are compelling, the analysis suffers from two shortcomings.[2] First, researchers rely upon aggregate data (national level time series or county level cross section data), rather than household level observations, which potentially mask important factors determining vehicle choices and VMT. Second, while prior studies investigate a relationship between obesity and either vehicle choices or vehicle miles traveled (VMT), linking vehicle choices to overall gasoline consumption (or emissions) requires information regarding vehicle use. In other words, the analysis implicitly assumes that obese individuals drive as much as other individuals in the population. However, obesity may lead to a more sedentary lifestyle, which could offset the individual's choice for a less fuel efficient vehicle, resulting in no net increase in transportation-related emissions. On the other hand, if obese individuals use their vehicle more frequently, rather than say biking or walking, the overall impact of obesity on overall emissions could be further exacerbated. In either case, ignoring the impact of weight on vehicle use would provide an incomplete picture. As West (2004) points out, "... an unobserved household characteristic that affects the utility of miles driven in a particular vehicle bundle is likely to affect both its probability of selection and its intensity of use" (p. 740). The corresponding demand equations for vehicle purchases or VMT are likely to yields biased estimates.

The goal for this study is to address the two limitations of the existing literature by drawing on the unique household level panel data set provided by the Panel Study of Income Dynamics (PSID). To the best of our knowledge, the PSID is unique in providing vehicle ownership, vehicle usage, and BMI information in a panel data setting. Using these data, we employ two different empirical strategies: (1) a reduced-form fixed effects model and (2) a structural discrete and continuous model of vehicle choice and VMT.

First, we employ the linear panel fixed effects model. In contrast to the data used in prior studies, the PSID is a panel data set that can be used to ameliorate omitted variable bias by removing time invariant unobserved variables through the use fixed effects. In the reduced-form approach, we estimate the effect of obesity and overweight on VMT, fuel economy, and gasoline consumption, providing a more comprehensive analysis of the relationship between obesity

[2] We leave the first category of prior studies, i.e., mechanical relationship between fuel efficiency and weight of passengers, as an area for engineers, focusing instead on the remaining relationships.

and either VMT or fuel economy than previous studies. OLS estimates suggest that obesity and overweight have significant impacts on VMT, fuel economy, and gasoline consumption. Specifically, the OLS results suggest that overweight (BMI $>$ 25) households drive 8.2% more, own 2.3% less fuel efficient vehicles, and consume 9.4% more gasoline than non-overweight households. However, fixed effects model leads to substantially different conclusions, namely that the overweight households do not have a significant difference with normal households in terms of either VMT or gasoline consumption, and they own just 0.7% less fuel efficient vehicles. At the same time, obesity (BMI $>$ 30) remains a significant factor in the fixed effects model. Obese households drive 8.4% more and consume 6.3% more gasoline than non-obese (including overweight and normal) households. Based on our fixed effect model results, we implement very simple simulations with similar conditions with Li et al. (2011). Our simulation suggests that if the rate of obesity (BMI $>$30) in 2005 (35.1 %) had remained at the 1981 level (15.0 %), gasoline consumption in 2005 is 26,484 trillion btu instead of 27,309 trillion btu, i.e., approximately 3.0 % savings in gasoline consumption. Our findings suggest that obesity has an effect to reduce gasoline consumption, but the magnitude of savings is not as large as prior studies predict. Finally, in addition to the basic fixed effects model, we also estimate difference-in-difference (DID) and instrumental variable models as other methods to control endogeneity of obesity and overweight people.

As a structural approach, we adopt a joint discrete and continuous econometric model. There are two types of discrete and continuous models used in prior studies. One set of studies adopts a two-stage estimator developed by Dubin and McFadden (1984). Train (1986) and West (2004) estimate the consumer's behavior on vehicle purchases and their utilization by adopting DM's two-stage method. One main advantage of the DM method is relatively easy to implement. However, the DM approach is inefficient because the two portions of the discrete and continuous model are estimated separately, despite being driven by a common, underlying behavioral model. The other approach used in the literature is a single-step (i.e., simultaneously) estimated model of vehicle choice and usage. For example, Bento et al. (2009) estimate the distributional impact of gasoline tax on both vehicle purchases and VMT, while Roth (2012) examines the equivalence between a simple fuel economy standard and a feebate.

Spiller (2012) has recently extended this general modeling approach to better accommodate a households choice of not only individual vehicles, but the portfolio of vehicles the individuals choose to hold. The advantage of the single-step approach in estimation is it better reflects the obvious joint nature of the vehicle purchased by a household and the vehicle miles traveled by the household. However, estimation can be challenging when there are a large number of alternatives and a nonlinear indirect utility function. Bento et al. (2009) employ the repeated discrete-continuous choice model and the Bayesian estimation technique to accommodate the large number of vehicle choices available to consumers. Based on the results of the structural model, we investigate the impact of overweight (BMI $>$ 25) on gasoline demand. If the rate of overweight (BMI $>$ 25) has remained at the 1981 level (47.1%) rather than 67.3%, the gasoline consumption in 2005 is 26,870 trillon btu, a 1.6% savings which is even smaller than our reduced-form results.

The remainder of the paper is organized as follows. In Section 2 we describe the PSID data used in our empirical analysis, including key summary statistics. Section 3 details the empirical model and estimation strategy, and Section 4 presents the empirical results. In Section 5, we present our conclusion.

3.2 Data

To assess the impact of obesity on gasoline demand through both vehicle choices and utilization using household level, we collect the data from several sources. As a main source of data, we use the Panel Study of Income Dynamics (PSID). In addition, vehicle characteristics (e.g., fuel economy, wheelbase, etc.), come from Ward's Automotive Yearbooks, EPA fuel economy, American Chamber of Commerce Researcher's Association(ACCRA) Cost Of Living Index (COLI), National Automobile Dealers Association (NADA) used car price data, and other web sources.

Our data for household characteristics, including vehicle ownership, gasoline expenditure, and the Body Mass Index (BMI) for both the head and spouse come from the PSID.[3] The PSID

[3]Since the PSID does not provide BMI directly, we computed it using information on height and weight and the formula, $\frac{\text{weight(lb)}}{\text{height(in)}^2} \times 703$.

began in 1968 with a nationally representative sample of U.S. households and is the longest running national panel study in the world. The PSID collects information on employment, income, wealth, health, education, and numerous other socio-demographic characteristics. Until 1997, these data were collected on an annual basis, but has subsequently been conducted on a biannual basis. Of particular interest to the current study is the fact, starting in 1999, the PSID has collected details on vehicle-ownership, including manufacturer, model, and model year, as well as gasoline expenditure.[4] Therefore, our sample periods are the recent 6-surveys, 1999, 2001, 2003, 2005, 2007, and 2009. The reason we use the PSID is, to the best of our knowledge, it is unique in providing household-level panel data for both vehicle ownership and BMI.[5]

To construct vehicle attributes, we use Ward's Automotive Yearbook (1982-2009), which includes a wide range of vehicle characteristics, such as fuel economy, wheelbase, length, width, and horsepower by make, model, and model year. For most vehicles, we use EPA's fuel economy data set. However, the EPA data only covers vehicles built in 1984 onwards. Ward's Automotive Yearbook was used to construct fuel economy data for vehicle models older than 1984.

There are two main costs that affect vehicle choices and driving miles-vehicle purchase costs and driving costs. Even though several different concepts have been used in prior studies, we adopt the rental rate and per-mile operating cost suggested by Bento et al. (2009). The rental rate is calculated as $r_{ij} = D_j + \rho P_j + 0.85 I^A_{ij}$, where $(D_j, P_j, I^A_{ij}, \rho)$ denote the depreciation rate, vehicle prices, annual insurance cost, and real interest rate, respectively. The depreciation and vehicle price are calculated using NADA Used Car Guide data.[6] Insurance costs vary with vehicle manufacturer, model, prices, age, region, and other factors. However, there does not exist data to reflect all factors. Therefore, we use state-level average insurance premiums from

[4]The PSID includes the number of vehicles each household owns or leases, and detailed model information on up to three vehicles.

[5]The PSID data is open for public use and has been widely employed for a variety of studies in labor and environmental economics. The PSID provides manufacturer, model year, and types (car, utility, pickup, and van) related to information on vehicle and current state related to address. However, for more precise information, we obtained the data for restricted use, which the PSID distributes confidentially through special contracts. Among confidential data, for this study, we have acquired information on specific vehicle model (e.g., Toyota Camry) and geospatial data regarding the individual household's place of residence.

[6]Depreciation is calculated as difference of current and next year real used car prices. We use MSRP a vehicle price.

the National Association of Insurance Commissioners (NAIC) and proportion by vehicle make and model computed from *Insurance.com* to obtain insurance premiums varying by region and model.[7]

We use yearly average of Daily Treasury Real Long-Term Rates for the interest rate. To calculate per-mile operating cost, $p_{ij}^{M} = (p_{i}^{gas}/MPG_{j}) + 0.15I_{ij}^{M}$, we use CBSA level gasoline prices from the American Chamber of Commerce Researchers Association (ACCRA) data. Since the ACCRA data cover around 300 regions, these provide precise variations by regions. Figure 3.2 shows the distribution of gasoline price by regions and years. Additionally, in contrast to the National Household Travel Survey (NHTS) widely used in prior studies [Bento et al. (2009); Spiller (2012); Roth (2012)], the PSID does not provide vehicle miles traveled. Instead, the PSID provides each household's gasoline expenditures, as completed in the Consumer Expenditure Survey (CES) used in other studies [Feng et al. (2005); West (2004)]. Using the MPG of a household's vehicles, we compute driving miles as VMT = (Gasoline Expenditure/Gasoline Price) $\times$ MPG. In this case, using an accurate measure of gasoline prices is important.[8]

Table 3.1 presents summary statistics by number of vehicles each household owns after cleaning the data. The households which own one or two vehicles account for around 71.3% of the PSID sample. As one might expect, many zero- and one-vehicle households belong to low income households, while multi-vehicle households typically have higher incomes. The rate of obesity and overweight is around 70%, very close to the national level for obesity and overweight in 2009 (73%, according to the Centers for Disease Control and Prevention). The rate of obesity does not differ substantially across the number of vehicles owned, multi-vehicle households do exhibit a relatively large percentage of overweight people.

[7] As Bento et al. (2009) reported, assigning 85% for rental price and 15% to operating cost is followed by insurance company's suggestion.

[8] The NHTS is the authoritative source for national data on travel behaviors. Recently, NHTS released data for 2001 and 2009. The NHTS VMT estimates (BESTMILE) are based on odometer reading, self-report annual miles, and model year. Household's average monthly VMT in 2001 and 2009 is 1765.6 and 1654.2 miles, respectively. Even though our estimates in 2001 and 2009 are slightly lower than the NHTS estimates, they are close and share the same trends to our estimates.

3.3 Empirical Model

Our goal is to identify the causal effects that obesity affects on gasoline consumption through vehicle choices and driving. To assess the impact of obesity, we investigate both reduced and structural-form approaches.

3.3.1 Reduced-Form Approach

There are several issues that arise to establish a causal linkage between obesity (and overweight), and both vehicle purchases and usage. One of the key concerns is the confounding effect of unobserved time-invariant factors on vehicle choices, such as individual education levels or possibly commuting distance to work. Instrumental variable method is widely used to control these unobserved individual characteristics. However, as Angrist and Pischke (2009) note, it is often difficult to find good instruments. In the current paper, we start with a fixed effects panel data model to control for time invariant unobserved variables which may lead to omitted variable bias. In our application, vehicle choices are correlated with miles traveled because both discrete and continuous choices share common characteristics [West (2004)]. For example, consumers who purchase vehicles with high fuel efficiency generally drive more because the per-mile operating cost is low.[9] This model is then expanded to control for the possible endogeneity of obesity and overweight through the use of instrumental variables. Specifically, we use as instruments sibling's average BMI, participation in the food stamp program, and the proportion of eating-out costs. In addition, we also implement standard difference-in-difference by dealing with obesity and overweight as a treatment.

First, we describe a standard fixed effects model to assess the impact of obesity that controls for unobserved individual time-invariant heterogeneity through the inclusion of individual fixed effects and time fixed effects.[10] Starting with the specification for fuel economy, we have

$$\ln(MPG_{it}) = \alpha_0 + \alpha D_{it} + \beta_1 \ln(\text{GasPrice}_{it-h_i}) + \beta_2 \ln(y_{it} - r_{it}) + \mathbf{X}_{it}\boldsymbol{\gamma} + \varphi_i + \delta_t + \varepsilon_{it} \quad (3.1)$$

[9] This behavioral response is called 'rebound effect' or 'take-back effect' and is studied further in the next chapter.

[10] In estimation of all reduced-form models, we drop households, which do not own any vehicle because the reduced-form model cannot control zero VMT and zero MPG appropriately. In other words, it can distort estimates. However, the structural-form model, which will be discussed, includes zero-vehicle households because the discrete choice model basically incorporates the option to "not own".

where D_{it} is an indicator variable that equals one, if the head of household i is obese or overweight at time t, and GasPrice_{it-h_i} denotes the price of gasoline when the household's vehicle was purchased,[11] y_{it} denotes household income, r_{it} denotes the vehicle rental rate, $\mathbf{X}_{it}$ is a vector of household i's characteristics at time t,[12] φ_i is a household specific fixed effects, δ_t denotes year fixed effects, and ε_{it} is an error term. The parameter α captures the causal effect of interest, i.e., obesity and overweight, fuel efficiency.

The basic specification for VMT and Gasoline Demand is similar, with

$$\ln(Y_{it}) = \alpha_0 + \alpha D_{it} + \beta_1 \ln(\text{OperatingCost}_{it}) + \beta_2 \ln(y_{it} - r_{it}) + \mathbf{X}_{it}\boldsymbol{\gamma} + \varphi_i + \delta_t + \varepsilon_{it} \qquad (3.2)$$

where Y_{it} is the outcome of interest (i.e., VMT and gasoline consumption) for household i at time t, and $\text{OperatingCost}_{it} \equiv \text{GasPrice}_{it}/\text{MPG}_{it}$).

A basic problem with the models in (3.1) and (3.2) is the obesity and overweight dummy variables may be endogenous. We consider four approaches to handling this issue. First, we estimate a difference-in-difference (DID) model utilizing the longitudinal nature of our PSID data. The treatment group in our DID framework is the set of households not overweight in the pre-treatment period but obese or overweight in the post-treatment periods, and the control group is not obese or overweight households in both pre and post treatment period. The DID estimation is basically the same as the fixed effects model [Angrist and Pischke (2009)]. The only difference between fixed effects model and DID in our application is how to define control and treatment groups. In the fixed effects model, we simply define obese or overweight people as treatment, others are the control group. However, in DID framework, we restrict people, who are non-obese and non-overweight, in both pre- and post-periods into the control group, but assign people who are non-obese or non-overweight in the pre-treatment period and obese or overweight in the post-treatment into the treatment group.

Second, we estimate instrumental variable versions of our models. The natural instrument for obesity and overweight for each household is sibling's BMI, which can be constructed using using the PSID data. Additionally, we also employ the ratio for eating-out costs over food

[11] In the case of households with more than one vehicle, we use an average value of two or three gasoline prices at the time of purchase.

[12] Specifically, $\mathbf{X}_{it}$ includes number of family, number of adults, number of vehicles, and age of vehicles.

consumption, and the participation in food stamp program similar to Li et al. (2011).[13] Using these instruments, we estimate the impact of obesity, α, by classical 2SLS using instruments.

However, since D_{it} has a binary nature, we can write the linear functional form by the taking expected value conditional on exogenous variables

$$\begin{aligned} E(\ln(Y_{it})|\mathbf{W}_{it}, \mathbf{Z}_{it}) &= \alpha E(D_{it}|\mathbf{W}_{it}, \mathbf{Z}_{it}) + \mathbf{W}_{it}\boldsymbol{\tau} + E(\varepsilon_{it}|\mathbf{W}_{it}, \mathbf{Z}_{it}) \\ &= \alpha \mathbf{P}(D_{it} = 1|\mathbf{W}_{it}, \mathbf{Z}_{it}) + \mathbf{W}_{it}\boldsymbol{\tau} \end{aligned} \tag{3.3}$$

where $\mathbf{W}_{it} = \{\ln(\text{OperatingCost}_{it}), \ln(y_{it} - r_{it}), \ln(\text{GasPrice}_{it-h_i}), \mathbf{Year}_t, \mathbf{X}_{it}\}$, i.e., all exogenous variables, and $\mathbf{P}(D_{it} = 1|\mathbf{W}_{it}, \mathbf{Z}_{it})$ is a probability that household i become obese or overweight conditional on exogenous variables because D_{it} is binary. There are two ways to estimate equation (3.3), leading to our third and fourth variations on the model in our basic models. First, we can use a simple two-step procedure, (i) estimate $\mathbf{P}(D_{it} = 1|\mathbf{W}_{it}, \mathbf{Z}_{it})$ by probit and obtain the first-stage fitted probabilities, $\hat{\Phi}_{it} = \Phi(\hat{\pi}_0 + \mathbf{X}_{it}\hat{\pi}_1 + \mathbf{Z}_{it}\hat{\pi}_2)$, and (ii) run OLS using $\hat{\Phi}_{it}$ in place of D_{it}. However, applying OLS directly with fitted probability cannot guarantee consistency [Wooldridge (2010)]. Another way is to use the fitted probability as an instrumental variable of D_{it}. In other words, as previous, (i) at the first stage, get the fitted probabilities, $\hat{\Phi}_{it}$ by probit estimation, and (ii) run 2sls, i.e. instrumental variable estimation using $\hat{\Phi}_{it}$ from the first stage as instrumental variable.

3.3.2 Structural-Form Approach

We next discuss our structural model approach based on each household's utility maximization. As discussed, since the vehicle choices and miles traveled are correlated by unobserved household characteristics, the structural model should capture the relation to avoid bias in estimation. There are several recent studies to analyze the consumer's behavior on vehicles and driving choices using joint estimation to control household unobserved characteristics [e.g., Bento et al. (2009); Feng et al. (2005); Roth (2012); Spiller (2012); and West (2004)]. We adopt the model developed by Bento et al. (2009), which employs a full-information one-step structural approach. The model assumes households face T_i choice occasions, i.e., in line with

[13] They use the ratio of average price of fast food over the average food price at home, and the participation rate in the food stamp program as instruments for obesity rate in counties.

mixed logit with repeated choices. The conditional indirect utility function is defined from the log-linear VMT demand function as,

$$U_{itj} = \begin{cases} -\frac{1}{\lambda_i}\exp\left(-\lambda_i\left(\frac{y_i}{T_i}\right)\right) + \mathbf{H}_i^{\varphi}\varphi_i + \mu_i\varepsilon_{it0} & \text{if } j=0, \\ \\ -\frac{1}{\lambda_i}\exp\left(-\lambda_i\left(\frac{y_i}{T_i} - r_{ij}\right)\right) - \frac{1}{\beta_i}\exp\left(\alpha_i\mathbf{H}_i^{\alpha}\mathbf{X}_j^{\alpha} + \beta_{ij}p_{ij}^{M}\right) & \\ \qquad + \tau_i\mathbf{H}_i^{\tau}\mathbf{X}_j^{\tau} + \mu_i\varepsilon_{itj} & \text{if } j=1,...,J \end{cases} \tag{3.4}$$

$$\text{where} \quad \beta_i = -\exp(\tilde{\beta}_i)$$

$$\lambda_i = \exp(\tilde{\lambda}_i)$$

and where (p_{ij}^M, r_{ij}, y_i) are vehicle per-mile operating cost, rental rate of vehicle, and income of household i's for jth vehicle, respectively, $\mathbf{H}_i^{\alpha}$ is household characteristics, $\mathbf{X}_j$ is vehicle characteristics, and ε_{itj} is an error component capturing unobservable determinants of the households decision, assumed an *i.i.d.* extreme value random variable.[14]

The other exogenous variables, except rental price and per-mile operating cost, that may affect the household's choices are included in $\alpha_i\mathbf{H}_i^{\alpha}\mathbf{X}_j^{\alpha}$. Above all, the key variable for this study, obesity, and interaction with vehicle characteristics are included as

$$\alpha_i\mathbf{H}_i^{\alpha}\mathbf{X}_j^{\alpha} \equiv \alpha_1 D_i + \sum_{k=1}^{5}\alpha_{2(k)i}\cdot\text{vintage}_{jk} + \alpha_3\text{WB}_j + \alpha_4\cdot(\text{WB}_j \times \text{Head Age}_i) + \alpha_5\cdot(\#\text{ of Adults}) \tag{3.5}$$

Following Hausman (1981), we derive the VMT demand equation by applying Roys identity.

$$\text{VMT}_{itj} = \exp\left(\alpha_i\mathbf{H}_i^{\alpha}\mathbf{X}_j^{\alpha} + \beta_i p_{ij}^{M} + \lambda_i\left(\frac{y_i}{T_i} - r_{ij}\right)\right) + \eta_{itj}, \text{for } j=1,...,J \tag{3.6}$$

where η_{itj} represents an idiosyncratic error assumed independent across the J alternatives with zero mean and standard deviation, $\sigma_i = \exp(\sigma_i^*)$, such that $\eta_{itj} \sim N(0,\sigma_i)$. Therefore, the combined likelihood function is

$$L_i = \prod_{t=1}^{T_i}\left[\prod_{j=0}^{J}\left(\frac{\exp(V_{ij})}{\sum_k \exp(V_{ik})}\right)^{1_{itj}}\prod_{j=1}^{J}\log\left(\Phi(VMT_{itj},\sigma_i|j)^{1_{itj}}\right)\right] \tag{3.7}$$

[14] For the discrete choices model, we have to define choice sets, i.e., discrete choice alternatives following Bento et al. (2009). We divide 10 vehicle classes (compact, luxury compact, midsize, fullsize, luxury mid/full size, small SUV, large SUV, small truck, large truck, and mini van), 5 vintages (new, 1-2 years, 3-6 years, 7-11 years, and 12-18 years), and 7 manufacturers (Ford, Chrysler, GM, Honda, Toyota, Other Asians, and European).

where 1_{itj} is an indicator function that equals 1 if household i chooses j at tth choice occasion, and 0 otherwise. We adopt random parameters for all variables to reflect heterogeneous taste and flexible specification. These random parameters and nonlinear specifications make estimation via maximum simulated likelihood implausible. As Bento et al. (2009) suggested, we employ Bayesian procedure using the Markov Chain Monte Carlo (MCMC) technique.[15]

3.4 Empirical Results

3.4.1 Reduced-Form Approach

The results of basic reduced-form model are presented in Tables 3.3, 3.2, and 3.4. Observations are weighted using the PSID survey weights. For comparisons, we report the results of OLS estimation as well. It is well-known the OLS estimation can yield biased estimates due to omitted variable, when the unobserved household time-invariant variables are correlated with other variables. In our application, if the unobservable variables, such as diet, lifestyle, and intensity of physical activity are correlated with the obesity and overweight variables, OLS regression may result in inconsistent estimates. We report the results of a variety of additional results that assess the overall robustness of the estimates using other alternative variables to represent obesity and overweight, such as BMI, and height and weight as well.

Panels A and B of Table 3.2, show the estimates of core parameters of pooled OLS and fixed effects model of the fuel economy equation. The parameters we are interested in are obesity and overweight. Column (1) shows that overweight (BMI $>$ 25) is a significant effect in both OLS and fixed effects model. On the other hands, obesity does not have a statistically significant effect in fixed effects model while OLS estimate is a significant negative value. Column (3) include both obesity and overweight, therefore, obesity parameter measure an additional marginal effect over overweight. In both OLS and fixed effects model, there is no statistically significant effect of obesity. Although obesity and overweight dummies give intuitively attractive measures, these variables lose some information by changing 0 or 1. Columns (4), (5), and (6) present the result of the estimation with BMI, height, and weight. In particular, column

[15] The detailed estimation algorithm is described in appendix of Bento et al. (2009) and ch12 of Train (2009).

(5) shows that ln(BMI) has a significant negative effect in both OLS and fixed effects model. In addition, column (5) includes (BMI/100) and $(\text{BMI}/100)^2$ to reflect non-linearity, the estimates for these two variables are significant as well. Column (6) reports the result of estimation with height and weight which are statistically significant in OLS but only weight is significant in fixed effects model. Even though our results have the same direction with Li et al. (2011), the size of the impact on fuel economy is relatively small. We will discuss and compare our results with Li et al. (2011) below. Table 3.3 presents the estimation results of VMT regression. In contrast to the results in fuel economy regression, overweight has a significant impact on VMT in OLS but not in fixed effects model. On the other hand, a effect for obesity in column (2) shows statistically significant estimates for both models. Moreover, the fixed effects model in column (3) shows obesity has a significant marginal effect, whereas, only the overweight estimate is statistically significant in OLS. Columns (4), (5), and (6) show any measures for obesity and overweight, such as ln(BMI), $(\text{BMI}/100)^2$, height, and weight, are not statistically significant in fixed effects model. Above results for VMT and fuel economy imply obese household drive 8.4% more than non-obese households, and households with overweight head own 0.7% less fuel efficient vehicles than non-overweight households.[16] As a result of combining driving and fuel efficiency, the estimation for gasoline consumption equation provides a final consequence. In Table 3.4, panel A: OLS results show all measures, obesity (BMI $>$ 30), overweight (BMI $>$ 25), ln(BMI), (BMI/100), $(\text{BMI}/100)^2$, and ln(height) have statistically significantly positive effects on gasoline consumption. However, fixed effects model suggests households with obese head consume gasoline more around 6 % than non-obese households.

To check robustness and compare results with the fixed effects model, we estimate the impact of obesity and overweight through a variety of methods. First, by utilizing the nature of the panel data, we implement difference-in-difference (DID). As discussed above, DID is widely used in recent empirical works to measure treatment effects or policy impact. In our case, treatment is obesity and overweight. The treatment group can be defined as people who were not obese or overweight during the pre-treatment period, but become obese or overweight in the

[16] In log-linear model, the parameter of dummy variable can be interpreted as percent change because of $\frac{Y(D=1)-Y(D=0)}{Y(D=0)} = \exp(\alpha) - 1 \approx \alpha$ at 0.

post-treatment period. On the other hand, non-obese or non-overweight households in both pre and post treatment period are assigned as a control group. In our application, households need a time period to adjust their vehicles, i.e., fuel efficiency, (or sometimes driving) after treatment since changing vehicles comes with a large number of costs. Therefore, we restrict treatment group as households, which are non-obese or non-overweight in 1999, but obese or overweight in 2003; while we compare the dependent variables between 1999 and 2009. That is to say, we set 1999 as the pre-treatment period and 2009 as the post-treatment period. In Table 3.5, DID results show that obesity and overweight are not statistically significant for all dependent variables at the 5 % significance level. One ongoing concern in the literature on obesity is endogeneity of obesity and overweight. A conventional way used in previous studies to overcome is the instrumental variable approach [Cawley and Meyerhoefer (2012); Li et al. (2011)]. We adopt sibling's average BMI, proportion of eating-out cost to total food cost, and whether a household receives food-stamps used in previous literature as instruments for obesity and overweight. However, as indicated in the previous section, due to the binary nature of obesity and overweight dummies, applying 2sls using instruments is possibly inefficient. Therefore, we investigate the impact of obesity and overweight through several different approaches using IV. Table 3.6 reports the estimates of IV regression. In all cases, obesity and overweight have significant effects just on ln(MPG). However, IV estimates are much larger in magnitude than fixed effects and OLS. In addition, several prior studies report the endogeneity of fuel economy in estimation of the VMT [Linn (2013); Frondel and Vance (2013)]. This study adopts two kind of instruments for fuel economy; (1) gasoline prices at the time the vehicle was purchased and its interactions with household characteristics, (2) gasoline guzzler tax. The results of IV regression with both instruments, i.e., for obesity and overweight, and fuel economy, are presented in Table 3.7. There is no statistically significant estimates, which are very large values. As Frondel and Vance (2013) noted, gasoline prices at the time of purchasing vehicles for fuel economy suffer from weak instrumental problem. Further, although gas guzzler tax is highly correlated with fuel economy, the variable is also possibly endogeneous because it can be correlated with observed and unobserved variables, such as vehicle prices and the comport of vehicle. On the other hand, our estimates cannot be matched with those for prior studies because they use different units of data (e.g., the rate of obesity and overweight) and models. However, our results roughly compare with Li et al. (2011). They report that the MPG elasticity to obesity and overweight rate is -0.289, i.e., 1 % increase in obesity and overweight rate brings with 0.289 % decrease in fuel economy. Moreover, they show that if the rate of obesity and overweight in 2005 (67 %), had remained at the 1981 level (47 %), the average fuel economy for new vehicles demanded in 2005 would be increased from 22.99 to 24.98 miles per gallon. However, based on our results that obese and overweight people own 0.7 % less fuel efficient vehicle, we derive only a slightly increase in MPG, i.e., 23.02 instead of 22.99 from simple calculations. Li et al. (2011) also suggest that approximately 8.6% savings in gasoline consumption can be obtained through the improvement of fuel efficiency by a decrease in the rate of obesity and overweight.[17] From the results in Table 3.4, overweight (BMI >25) does not have a significant effect on gasoline consumption. With the same simple calculation and if the rate of obesity (BMI >30) in 2005 (35.1 %) had remained at the 1981 level (15.0 %), the gasoline consumption in 2005 is 26,484 trillion btu instead of 27,309 trillion btu, i.e., approximately 3.0 % savings in gasoline consumption.[18]

3.4.2 Structural-Form Approach

We use each year's data for estimation rather than pooled data from 1999 to 2009 because the full sample includes too many observations to estimate the model with random parameters. Decomposing the sample allows us to understand the change of preferences over time. The parameters in the structural model cannot be interpreted directly because of non-linearity in contrast to a reduced linear form model. Therefore, we compute marginal effects of obesity and overweight by using the following simple equations.

$$\frac{\left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 1) \cdot \mathrm{MPG}_j\right] - \left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 0) \cdot \mathrm{MPG}_j\right]}{\left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 0) \cdot \mathrm{MPG}_j\right]} \times 100 \tag{3.8}$$

[17] Li et al. (2011) assume that annual VMT is 12,000 miles and constant for all households regardless of the rate of obesity.

[18] Source: the rate of obesity http://www.cdc.gov/nchs/data/hestat/obesity_adult_07_08/obesity_adult_07_08.pdf, and fuel consumption http://www.eia.gov/totalenergy/data/monthly/pdf/sec2_11.pdf

$$\frac{\text{VMT}_{ij^*}(D_i = 1) - \text{VMT}_{ij^*}(D_i = 0)}{\text{VMT}_{ij^*}(D_i = 0)} \times 100 \tag{3.9}$$

$$\frac{\left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 1) \cdot (\text{VMT}_{ij}(D_i = 1)/\text{MPG}_j)\right] - \left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 0) \cdot (\text{VMT}_{ij}(D_i = 0)/\text{MPG}_j)\right]}{\left[\sum_{j=1}^{J} \mathbf{P}_{ij}(D_i = 0) \cdot (\text{VMT}_{ij}(D_i = 0)/\text{MPG}_j)\right]} \times 100 \tag{3.10}$$

where D_i represents dummy variable for obesity or overweight. Equations (3.8), (3.9), and (3.10) are marginal effects of obesity or overweight on fuel economy, VMT, and gasoline consumption, respectively. In particular, equation (3.10) for gasoline consumption reflects the comprehensive impact of obesity or overweight through the change of both VMT and fuel economy. Table 3.8 reports the average effects of overweight (BMI $>$ 25) for the joint discrete and continuous model. In comparison with the results of the reduced-form approaches, the impact of overweight on fuel economy is relatively small, but it has larger effects on VMT. To compare with the reduced-form approaches and Li et al. (2011), we implement the same simulations as the reduced-form approaches. Even though our results provide different results according to survey year, based on the results of 2003 year, we can compute the impact of overweight (BMI $>$ 25) on gasoline demand. If the rate of overweight (BMI $>$ 25) remained at the 1981 level (47.1%) rather than 67.3%, the gasoline consumption in 2005 is 26,870 trillon btu, a 1.6% savings, which is even smaller than our reduced-form results.

3.5 Conclusions

The rising rate of obesity has contributed large social and economic problems over the last several decades in the U.S. and throughout the world. It is well-known that the high prevalence of obesity is accompanied by several negative health problems such as high blood pressure, diabetes, or heart disease. Besides these common negative effects, recent studies examine the impact of obesity and overweight on the environment through vehicle emissions. There are two channels to increase gasoline consumption - low fuel economy and more vehicle use. There are two main concerns in the previous research. First, employing aggregate types of data, such as the national level time series or county level cross section data, makes difficult to control for unobserved individual or household characteristics. Second, prior studies focus on one side of impacts, i.e VMT or fuel economy. Two channels on gasoline consumption may cause conflicting or synergistic actions.

The objective of this study is to provide a complete picture for the impact of obesity on gasoline consumption by adopting household level observations from the Panel Study of Income Dynamics (PSID). We investigate the impact of obesity by employing both reduced-form (linear panel model) and structural model (joint discrete and continuous choice model). Based on our estimates, we implement simple simulations in line with Li et al. (2011). First, for the reduced-form model, if the prevalence of obesity in 2005 has remained at the 1981 level, gasoline consumption would be 3% saved. On the other hand, the structural model shows that if the rate of overweight in 2005 has remained at the 1981 level, only 1.6% less gasoline would be demanded. Our empirical findings suggest that the comprehensive impact of obesity and overweight on gasoline consumptions are minimal or ambiguous in contrast to the results of prior studies considering either driving or vehicle choices.

Bibliography

Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics - An Empiricists Companion*. Princeton University Press, 2009.

Antonio Bento, L. Goulder, M. Jacobsen, and R. von Haefen. Distributional and efficiency impacts of increased u.s. gasoline taxes. *American Economic Review*, 99, 2009.

John Cawley and Chad Meyerhoefer. The medical care costs of obesity: An instrumental variables approach. *Journal of Health Economics*, 31, 2012.

C. Courtemanche. A silver living? the connection between gasoline prices and obesity. *Journal of Health Economics*, 49, 2011.

Burton D.C. Jackson R.J. Dannenberg, A.L. Economic and environmental costs of obesity, the impact on airlines. *American Journal of Preventive Medicine*, 27(3):264, 2004.

J.A. Dubin and D.L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2):345–362, 1984.

Ye Feng, Don Fullerton, and Li Gan. Vehicle choices, miles driven, and pollution policies. *NBER Working Paper*, 2005.

Manuel Frondel and Colin Vance. Fuel taxes versus effiency standard - an instrumental variable approach. *Working Paper*, 2013.

J. A. Hausman. Exact consumers surplus and deadweight loss. *American Economic Review*, 71(4):662–676, 1981.

S. Jacobson and L. McLay. The economic impact of obesity on automobile fuel consumption. *JEngineering Economist*, 51, 2006.

Sheldon H. Jacobson, Douglas M. King, and Rong Yuan. A note on the relationship between obesity and driving. *Transport Policy*, 18(5):772–776, 2011.

Shanjun Li, Yanyan Liu, and Junjie Zhang. Lose some, save some: Obesity, automobile demand, and gasoline consumption. *Journal of Environmental Economics and Management*, 61, 2011.

Joshua Linn. The rebound effect for passenger vehicles. *RFF Working Paper*, 2013.

Kevin Roth. The unintended consequences of uncoordinated regulation: Evidence from the tranportation sector? *Working Paper*, 2012.

Elisheba Spiller. Household vehicle bundle choice and gasoline demand: A discrete-continous approach. *working paper*, 2012.

Kenneth Train. *Qualitative Choice Analysis*. MIT Press, Cambridge, MA, 1986.

Kenneth Train. *Discrete Choice Methods with Simulation, 2nd edition*. Cambridge University Press, 2009.

S.E. West. Distributional effects of alternative vehicle pollution control policies. *Journal of Public Economics*, 88, 2004.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2010.

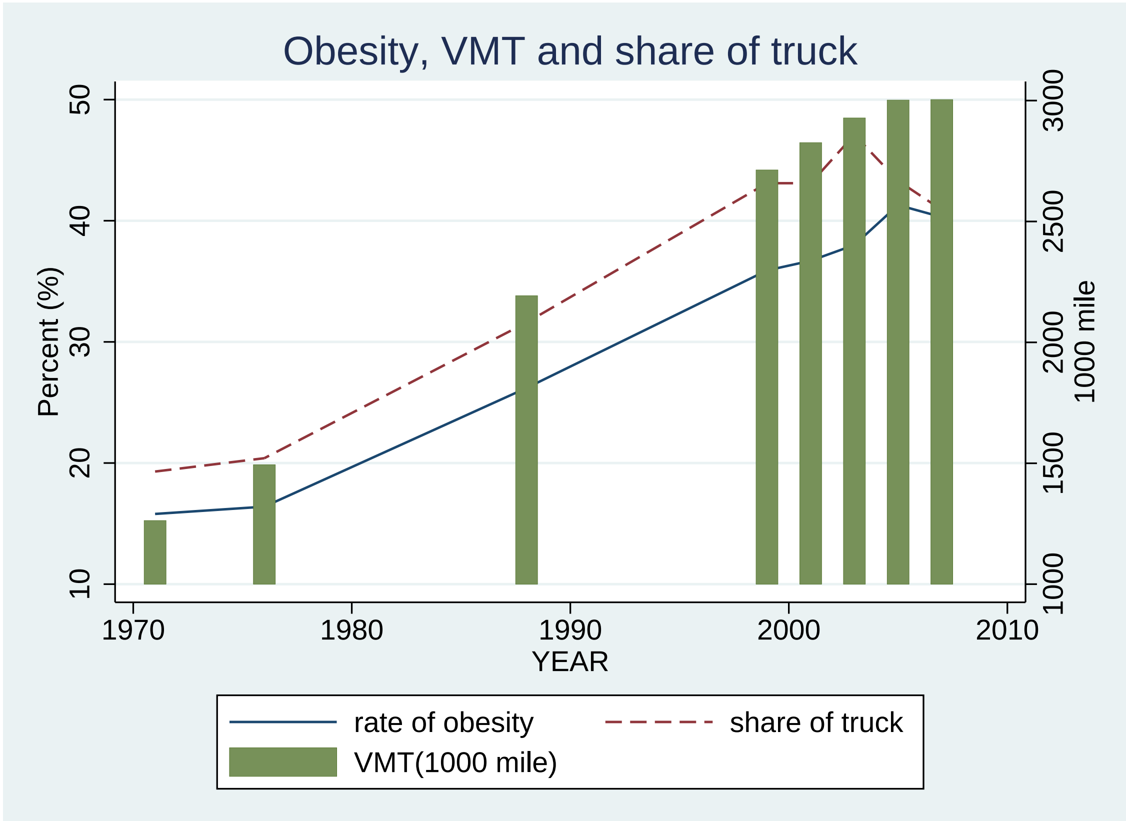

Figure 3.1 The rate of obesity, Share of Truck, and VMT, Data Source: Behavioral Risk Factor Surveillance System (BRFSS), National Transportation Statistics of Bureau of Transportation Statistics, and EPA Light-duty automotive technology and fuel economy trends: 1975 Through 2012

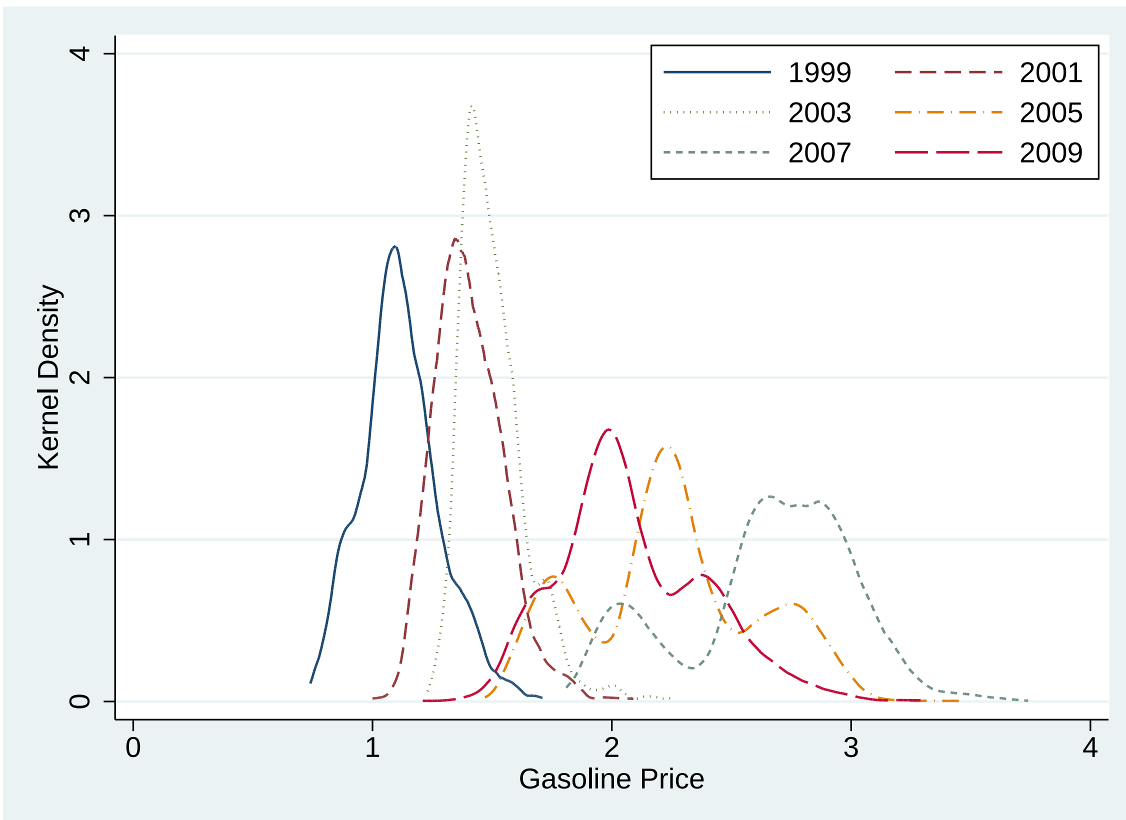

Figure 3.2 The Distribution of Gasoline Prices: 1999-2009

Table 3.1 Summary Statistics

	HH with 0 Vehicle	HH with 1 Vehicle	HH with 2 Vehicle	HH with 3 Vehicle
Household Size	2.28	2.17	3.03	3.33
	(1.59)	(1.39)	(1.31)	(1.30)
Number of Adults	1.43	1.44	2.03	2.35
	(0.69)	(0.64)	(0.52)	(0.69)
VMT (1000 miles)	-	12.94	20.65	26.42
	-	(10.89)	(15.08)	(22.15)
Average MPG	-	20.64	19.60	19.47
	-	(4.10)	(3.05)	(2.57)
Rate of Obesity (BMI > 30)	0.31	0.29	0.28	0.28
	(0.46)	(0.45)	(0.45)	(0.45)
Rate of Overweight (BMI > 25)	0.63	0.65	0.74	0.75
	(0.48)	(0.48)	(0.44)	(0.44)
Age of Head	45.08	43.84	43.98	46.38
	(19.17)	(16.94)	(14.03)	(11.68)
Years of Education (years)	11.42	13.04	13.47	13.34
	(2.79)	(2.58)	(2.56)	(2.54)
Household Income < $30,000	0.73	0.37	0.09	0.05
	(0.44)	(0.48)	(0.29)	(0.22)
Household Income $30,000 - $60,000	0.19	0.39	0.27	0.20
	(0.40)	(0.49)	(0.44)	(0.40)
Household Income $60,000 - $75,000	0.08	0.24	0.64	0.75
	(0.27)	(0.43)	(0.48)	(0.44)
Household Income > $75,000	0.05	0.16	0.49	0.62
	(0.21)	(0.36)	(0.50)	(0.49)
Households (share)	6,257 (17.4%)	12,419 (34.5%)	13,279 (36.8 %)	4,085 (11.3 %)

Note: The standard deviations are reported in the parenthesis.

Table 3.2 Average Effects of Obesity and Overweight on Fuel Economy

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS			
Obesity (BMI $>$ 30)		-0.013***	-0.004			
		(0.003)	(0.005)			
Overweight (BMI $>$ 25)	-0.023***		-0.021***			
	(0.003)		(0.004)			
ln(BMI)				-0.029***		
				(0.005)		
(BMI/100)					-0.224***	
					(0.033)	
$(\text{BMI}/100)^2$					0.116***	
					(0.019)	
ln(Height)						-0.058***
						(0.007)
ln(Weight)						-0.394***
						(0.027)
$\ln(\text{GasPrice}_{t-h})$	0.016**	0.016**	0.016*	0.016**	0.002	0.018**
	(0.008)	(0.008)	(0.010)	(0.008)	(0.007)	(0.008)
ln(Income - Rental Price)	-0.023***	-0.024***	-0.023***	-0.024***	-0.024***	-0.019***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
			Panel B: Fixed Effects Model			
Obesity (BMI $>$ 30)		-0.006	-0.006			
		(0.004)	(0.004)			
Overweight (BMI $>$ 25)	-0.007**		-0.007*			
	(0.004)		(0.004)			
ln(BMI)				-0.011**		
				(0.005)		
(BMI/100)					-0.103**	
					(0.049)	
$(\text{BMI}/100)^2$					0.051**	
					(0.026)	
ln(Weight)						-0.034**
						(0.016)
ln(Height)						-0.031
						(0.070)
$\ln(\text{GasPrice}_{t-h})$	0.006	0.006	0.006	0.006	0.002	0.006
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
ln(Income - Rental Price)	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 3.3 Average Effects of Obesity and Overweight on VMT

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS			
Obesity (BMI > 30)		0.048**	0.015			
		(0.023)	(0.028)			
Overweight (BMI > 25)	0.082***		0.076***			
	(0.021)		(0.026)			
ln(BMI)				0.094**		
				(0.042)		
(BMI/100)					0.803***	
					(0.287)	
$(BMI/100)^2$					-0.420**	
					(0.170)	
ln(Height)						0.174***
						(0.061)
ln(Weight)						-0.010
						(0.230)
ln(GasPrice/MPG)	-0.382***	-0.376***	-0.382***	-0.376***	-0.378***	-0.391***
	(0.055)	(0.055)	(0.059)	(0.055)	(0.055)	(0.056)
ln(Income - Rental Price)	0.114***	0.116***	0.114***	0.116***	0.116***	0.113***
	(0.017)	(0.017)	(0.020)	(0.017)	(0.017)	(0.017)
			Panel B: Fixed Effects Model			
Obesity (BMI > 30)		0.084**	0.082**			
		(0.041)	(0.041)			
Overweight (BMI > 25)	0.020		0.014			
	(0.038)		(0.038)			
ln(BMI)				0.022		
				(0.050)		
(BMI/100)					0.506	
					(0.468)	
$(BMI/100)^2$					-0.283	
					(0.242)	
ln(Weight)						0.183
						(0.182)
ln(Height)						-0.200
						(0.749)
ln(GasPrice/MPG)	-0.420***	-0.421***	-0.421***	-0.420***	-0.421***	-0.421***
	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)	(0.080)
ln(Income - Rental Price)	0.091***	0.091***	0.091***	0.091***	0.091***	0.090***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively. Fixed effects model include # of family, # of adults, # of vehicle, vehicle age fixed effects, and year fixed effects. Pooled OLS include all variables in fixed effects model and head age, head education and city size by population fixed effects as well.

Table 3.4 Average Effects of Obesity and Overweight on Gasoline Demand

	(1)	(2)	(3)	(4)	(5)	(6)
			PanelA: OLS			
Obesity (BMI > 30)		0.061***	0.025			
		(0.018)	(0.023)			
Overweight (BMI > 25)	0.094***		0.084***			
	(0.016)		(0.020)			
ln(BMI)			0.320***	0.116***		
			(0.045)	(0.031)		
(BMI/100)					0.940***	
					(0.211)	
$(\text{BMI}/100)^2$					-0.485***	
					(0.124)	
ln(Height)						0.212***
						(0.045)
ln(Weight)						0.122
						(0.173)
ln(GasPrice/MPG)	0.320***	0.328***	0.320***	0.327***	0.325***	0.304***
	(0.041)	(0.041)	(0.045)	(0.041)	(0.041)	(0.042)
ln(Income - Rental Price)	0.114***	0.116***	0.114***	0.116***	0.116***	0.111***
	(0.013)	(0.013)	(0.015)	(0.013)	(0.013)	(0.013)
			Panel B: Fixed Effects Model			
Obesity (BMI > 30)		0.063**	0.061*			
		(0.031)	(0.031)			
Overweight (BMI > 25)	0.022		0.017			
	(0.028)		(0.028)			
ln(BMI)				0.017		
				(0.039)		
(BMI/100)					0.285	
					(0.355)	
$(\text{BMI}/100)^2$					-0.160	
					(0.183)	
ln(Weight)						0.120
						(0.135)
ln(Height)						-0.079
						(0.535)
ln(GasPrice/MPG)	0.164***	0.163***	0.163***	0.164***	0.164***	0.163***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
ln(Income - Rental Price)	0.082***	0.083***	0.082***	0.083***	0.083***	0.082***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 3.5 Average Effects of Obesity and Overweight: Difference-in-Difference

	ln(MPG)	ln(VMT)	ln(Gas Demand)
Overweight (BMI > 25)	-0.027*	-0.006	0.010
	(0.015)	(0.110)	(0.098)
Obesity (BMI > 30)	0.011	0.014	-0.012
	(0.094)	(0.019)	(0.089)

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 3.6 Average Effects of Obesity and Overweight: Instrumental Variable Approach 1

	IV	Probit-2SLS	Probit-OLS
		Overweight (BMI > 25)	
ln(VMT)	0.070	-0.250	-0.317
	(0.045)	(0.307)	(0.394)
ln(MPG)	-0.094**	-0.088***	-0.049**
	(0.039)	(0.033)	(0.022)
ln(Gas Demand)	-0.142	-0.079	-0.103
	(0.310)	(0.228)	(0.293)
		Obesity (BMI > 30)	
ln(VMT)	0.048	-0.386	-0.380
	(0.051)	(0.381)	(0.382)
ln(MPG)	-0.074**	-0.067**	-0.030*
	(0.034)	(0.034)	(0.017)
ln(Gas Demand)	-0.165	-0.197	-0.197
	(0.270)	(0.284)	(0.284)
Instruments	Sibling's BMI	Sibling's BMI	Sibling's BMI
for Obesity or Overweight	Food-Stamp	Food-Stamp	Food-Stamp
	Eating-Out Cost	Eating-Out Cost	Eating-Out Cost

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 3.7 Average Effects of Obesity and Overweight: Instrumental Variable Approach 2

	IV1	IV2	Probit-IV1	Probit-IV2
		Overweight (BMI > 25)		
ln(VMT)	-0.313	-0.376	-0.119	-0.160
	(0.319)	(0.419)	(0.253)	(0.294)
ln(Gas Demand)	-0.127	-0.182	0.067	0.144
	(0.244)	(0.316)	(0.190)	(0.220)
		Obesity (BMI > 30)		
ln(VMT)	-0.258	-0.370	-0.215	-0.271
	(0.313)	(0.364)	(0.325)	(0.364)
ln(Gas Demand)	-0.104	-0.196	-0.051	0.075
	(0.237)	(0.275)	(0.243)	(0.277)
Instruments	Sibling's BMI	Sibling's BMI	Sibling's BMI	Sibling's BMI
for Obesity or Overweight	Food-stamp	Food-stamp	Food-stamp	Food-stamp
	Eating-out Cost	Eating-out Cost	Eating-out Cost	Eating-out Cost
for MPG	$P^g_{it-h_i}$	Gas Guzzler Tax	$P^g_{it-h_i}$	Gas Guzzler Tax
	Interactions		Interactions	

Notes: Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 3.8 Average Effects of Overweight: Structural-Form Model

	1999	2001	2003	2005	2007	2009
MPG (%)	-0.001	0.000	0.000	-0.002	-0.002	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
VMT (%)	13.740	14.479	15.591	23.109	10.097	15.924
	(4.852)	(3.467)	(3.661)	(3.014)	(3.625)	(3.808)
Gasoline (%)	8.118	8.404	8.999	13.554	5.773	9.244
	(2.857)	(2.012)	(2.113)	(1.769)	(2.081)	(2.207)

CHAPTER 4. VEHICLE FUEL EFFICIENCY AND THE REBOUND EFFECT: EVIDENCE FROM U.S. PANEL DATA

4.1 Introduction

Global warming represents a profound environmental threat to our planet and mobile sources of air pollution represent a key contributor to this problem. Indeed, the transportation sector as a whole ranks second as a source of global carbon emissions, accounting for roughly 28% of annual emission of carbon dioxide (EPA Inventory of U.S. Greenhouse Gases and Sinks 2014).[1] U.S. efforts to reduce the carbon footprint of its transportation sector have centered around the Corporate Average Fuel Economy (CAFE) standards. These standards require vehicle manufacturers to meet a sales-weighted, fleet-wide, fuel economy benchmark. However, an unintended consequence of CAFE standards is people may drive more as vehicle fuel efficiency increases. This so-called 'rebound effect' potentially offsets much of the gains in the reduction of energy consumption sought through the CAFE standards.

The rebound effect can be divided into two components: direct and indirect rebound effects. Specifically, the direct effect consists of total substitution and a 'partial' income effect stemming from a change in fuel price or fuel efficiency. The 'partial' income effect captures the increased or decreased usage of the vehicles owned by an individual in response to a price or fuel efficiency change. The remaining indirect rebound effect corresponds to the changes of the energy consumption on other energy using product, i.e. other transportation or appliances.

Ideally, both direct and indirect rebound effects should be considered to capture the entire rebound effect. However, as it is difficult to estimate the indirect effect, most studies in the literature focus only on quantifying the direct effect, an approached followed in this paper as

[1] Source: http://www.epa.gov/climatechange/ghgemissions/usinventoryreport.html

well.

While there is already substantial literature on the rebound effect, there remains no consensus regarding its magnitude. Estimates of the rebound vary substantially, changing with the empirical methods employed, the types of or time periods covered by the available data, and the specific definition used to characterize the rebound effect. Several different types of data sets have been used in prior studies, including cross-sectional household level data, aggregate panel data, and time-series data, with only a single German data set drawing on household level panel data [Frondel et al. (2008, 2012); Frondel and Vance (2013)]. The reported VMT elasticity estimates of the rebound effect from studies using household level data range from 0.1 to 0.9, while those from studies using aggregate data are in the range from 0.03 to 0.34.

One of the main concerns in estimating the rebound effect is the potential endogeneity of the fuel economy variable, with fuel economy potentially correlated with unobserved household and/or vehicle characteristics. Two approaches have dominated the literature in dealing with this issue: (1) the use of instrumental variables for fuel economy and (2) a joint discrete and continuous choice model along the lines of Dubin and McFadden (1984) that explicitly accounts for the correlation between vehicle miles traveled and vehicle choices. The difficulty with the instrumental variable approach is the availability of suitable instruments, while the discrete-continuous choice model approach depends upon structural assumptions. This paper addresses the endogeneity issue, instead, through the use of panel data techniques, controlling for unobserved, but time invariant household factors with fixed effects. Data for the analysis are provided by the Panel Study of Income Dynamics (PSID), yielding a household level panel data base for 1999 through 2011. For comparison, we implement several different estimation strategies, including pooled OLS, random and fixed effects models, and an instrumental variable model.

Finally, although the rebound effect generally refers to the VMT response to the introduction of more efficient vehicles, i.e., the elasticity of VMT with respect to fuel economy, there are other definitions of the rebound effect used in the literature. For example, studies measure the rebound effect in terms of the elasticity of VMT with respect to gasoline price or per-mile operating cost. An important assumption underlying these alternative definitions is consumers react the same way to a decrease in gasoline prices versus an improvement in fuel economy. In other words, they do not care about where the reduction of costs come. Studies that have tested this hypothesis do not provide a consistent conclusion [Gillingham (2012); Greene (2012); Frondel et al. (2012); Small and Dender (2007)]. To examine this hypothesis, we estimate the rebound effect based on various definitions. While our estimates of the rebound effect vary somewhat with definition used, from 0.58 to 0.80, we cannot reject the hypothesis that the response to fuel price and fuel economy are the same. Finally, we examine the heterogeneous responses to the change of price or fuel efficiency across income deciles. We find a statistically significant difference in the rebound effect for low and high income households.

The remainder of this paper is organized as follows. Section 2 presents a review of the recent literature estimating the rebound effect. Section 3 details the data used in this study, while Section 4 describes the empirical model specifications. Section 5 presents estimation results. The paper summarizes in section 6 with conclusions.

4.2 Related Literature

A number of papers have appeared in the literature estimating the elasticity of vehicle miles traveled (VMT) or fuel used with respect to fuel price or fuel economy. These elasticities are broadly accepted as alternative measures of the 'rebound effect.' However, the magnitude of the estimated rebound effect varies with the definition, empirical modeling techniques employed, and the type of data used [EPA and DOT (2012); Gillingham et al. (2014); Sorrell et al. (2009)]. This section begins with an overview of the competing definitions for the rebound effect appearing in the literature, followed by a summary of results, methodologies, and types of data used in the literature. Sorrell and Dimitropoulos (2008) provide a relatively recent and more extensive summary of the rebound effect literature.

4.2.1 Defining the Rebound Effect

The definition for the rebound effect can be largely categorized into four types of elasticities, (1) an elasticity of vehicle traveled miles (VMT) with respect to fuel economy, (2) an elasticity of VMT with respect to per-mile operating cost, (3) an elasticity of fuel demand with respect to fuel economy, and (4) an elasticity of VMT with respect to fuel price. Formally, the first definition of the rebound effect is based on the canonical relation, i.e., higher fuel economy leads to more driving.

Definition 1 *Elasticity of VMT with respect to fuel economy:*

$$\eta_{\mu}(\text{VMT}) = \frac{\partial \ln(\text{VMT})}{\partial \ln(\mu)} \tag{4.1}$$

where μ is fuel economy (MPG).

Although definition 4.1 is the most natural, it assumes consumer's choice of fuel economy is exogenous.

A second definition is based on per-mile operating cost, capturing the impact of both fuel efficiency and fuel price.[2]

Definition 2 *Elasticity of VMT with respect to per-mile operating cost (p^M):*

$$\eta_{p^M}(\text{VMT}) = -\frac{\partial \ln(\text{VMT})}{\partial \ln(p^M)} \tag{4.2}$$

where per-mile operating cost is defined as fuel price divided by fuel economy (MPG), p^{gas}/μ.

There are two key assumptions underlying definition 4.2. First, as was the case for definition 4.1, fuel economy is treated as exogenous. Second, the impact of the change in fuel economy on VMT is the same as the impact of changing fuel prices, i.e., the reduction of per-mile operating cost due to an increase in fuel economy is the same as those due to a decrease in fuel prices. There is evidence in the literature suggesting this assumption does not always hold [Linn (2013)].

The third and fourth definitions are as follows:

Definition 3 *Elasticity of Gasoline Demand with respect to fuel prices:*

$$\eta_{p^{\text{gas}}}(\text{Gasoline Demand}) = -\frac{\partial \ln(\text{Gasoline Demand})}{\partial \ln(p^{\text{gas}})} \tag{4.3}$$

[2]Note the rebound effect in this case is the negative of the elasticity of VMT with respect to fuel cost. The sign change is standard in the literature, making the signs comparable across the various definitions of the rebound effect.

Definition 4 *Elasticity of VMT with respect to fuel prices:*

$$\eta_{p^{\text{gas}}}(\text{VMT}) = -\frac{\partial \ln(\text{VMT})}{\partial \ln(p^{\text{gas}})} \tag{4.4}$$

where p^{gas} is fuel price.

Frondel et al. (2008) show that the four definitions are equivalent, if the following assumptions hold: (1) fuel price are exogenous, (2) fuel efficiency does not depend on fuel price, i.e., μ is not a function of p^{gas}, and (3) individuals respond to a change in fuel prices and fuel economy equivalently (i.e., measuring their impact solely in terms of the associated change in operating costs).[3] In our empirical analysis, we explicitly test this third assumption.

4.2.2 Evidence in the Literature

The techniques employed in recent literature on the rebound effect can be broadly divided into two groups: (1) reduced-form linear models (employing pooled OLS, panel estimation, and IV regression approaches) versus (2) structural (joint discrete and continuous choice) models. The structural models capture the discrete choices made by individuals in terms of which (and how many) vehicles to own and the continuous choice of how many vehicle miles to travel [Bento et al. (2009); Roth (2012); Spiller (2012); West (2004)]. Joint modeling of these two decisions accounts for the bias stemming from unobserved household characteristics that simultaneously influence both vehicle choice and VMT. The joint models typically rely on household level cross-sectional data (e.g., the National Household Travel Survey, NHTS, or the Consumer Expenditure Survey, CEX). Examples of structural models include Bento et al. (2009), who estimate the gasoline price or per-mile operating cost elasticity of VMT to be in the range from 0.34 to 0.74, and Spiller (2012) and Roth (2012), with rebound effect estimates of 0.45 and 0.62, respectively. These three studies use 2001 and 2009 NHTS data. West (2004), using 1997 CEX data, shows the largest value, 0.87.

Table 4.1 provides a summary of recent rebound effect estimates categorized by definition, data sources and estimation methods. Reduced-form studies use a variety of techniques and data sources. A series of studies [Frondel et al. (2008, 2012); Frondel and Vance (2013)] estimate

[3] Li et al.(forthcoming) show that gasoline prices are exogenous.

the rebound effect using German household panel data (German Mobility Panel) through a variety of estimation strategies, including random effects models, fixed effects models, and quantile regression. The estimates of the rebound effect in their studies generally have larger values than for other studies (ranging from 0.42 to 0.90), although this is attributed, in part, to better public transportation and higher fuel price in Germany [Gillingham et al. (2014)]. Linn (2013) estimates fuel efficiency elasticity with household cross-sectional data (NHTS) using instrumental variables (IV) regression to control for the endogeneity of fuel economy. The best estimate by Linn (2013) is 0.44, larger than the most OLS-based estimates.

Finally, estimates of the rebound effect in studies using aggregate data, such as Small and Dender (2007) using aggregate panel data and Hughes et al. (2008) using aggregate time series data, vary between 0.03 and 0.22, falling at the lower end of the estimates in Table 4.1. This may stem, in part, from attenuation bias, due to ignored heterogeneity in the household characteristics when relying on aggregate data.

4.3 Data

This study employs the Panel Study of Income Dynamics (PSID), a longitudinal survey of a representative sample of U.S. households that began in 1968. Starting in 1999, the PSID collected data biannually.[4] The survey collects a wide range of economic, sociological, and psychological variables. Perhaps most importantly for our purposes, beginning in 1999, the PSID includes information on vehicles each household holds, with the specific make, model, and year for each vehicle (e.g., Honda Civic 2010).[5] For this reason, we use the biannual samples beginning in 1999 continuing through 2011. After dropping observations with missing information, the final data included 40,447 observations.[6]

Based on specific vehicle model information from the PSID, we constructed vehicle characteristics, including fuel economy, for each vehicle in the database using combining information

[4] All observations are weighted using the PSID sampling weight.

[5] Although the PSID provides the information on vehicle year, make, and manufacturer publicly, the specific models are provided only under a restricted use contract. Such a contract was acquired for the purpose of this study.

[6] Appendix describe data sources and process to construct the final data set in detail.

from the U.S. Environmental Protection Agency (EPA) and WARD's Automotive Yearbook.[7] Finally, the American Chamber of Commerce Researchers' Association (ACCRA) data are used to obtain gasoline prices varying with region (CMSA level) and month. The ACCRA provides only quarterly average gasoline prices by CMSA. To construct monthly gasoline prices, we compute monthly variations within each quarter from EIA monthly state average gasoline prices, and then apply these values to the ACCRA data. Additionally, since the ACCRA does not cover all CMSA shown in the PSID, we impute the gasoline prices using EIA gasoline prices at the state level.

Table 4.2 lists summary statistics by years. Household characteristics do not vary much across the sample years. Average family income in the PSID sample is somewhat higher than the national median. This is due, in part, to the fact we have restricted our sample to households that own at least one vehicle, thus excluding many households in the lower tail of the income distribution. The reported income has large standard variations because the total family income in the PSID includes actual loss, i.e., negative values.[8]

Related to vehicle variables, note the fuel economy was generally declining from 1999 through 2005, but subsequently has increased. National representative statistics by the Bureau of Transportation Statistics also show the overall fuel economy of new light-duty vehicles in same period did not improve much (0.7 MPG), while the fuel economy for new vehicles did increase substantially (2 MPG). At the same time, gasoline prices vary significantly across years, from a low in 1999 of \$1.01 per gallon to \$3.77. Gasoline prices did drop sharply in 2009, following the start of the Great Recession. VMT have generally declined over the sample period, although there was an increase in 2009, perhaps reflecting an increase in driving relative to flying in response to the Great Recession.

---

[7] The website, fueleconomy.org supported by EPA and Department of Energy(DOE) provides miles per gallon (MPG) by make, model, and year for vehicles between 1985 to present. However, our sample includes models older than 1985. For the vehicles 1982 to 1984, we obtained the MPG from WARD's Yearbook.

[8] In addition, the PSID includes a few extreme values which make a slightly large average household income. However, the median values show our data are close to national values. The nominal median income for our 2011 sample is \$58,000, slightly higher than the national median income, \$50,046.

4.4 Empirical Model

The empirical models adopted in this paper are based on a simple linear specification used in prior studies. In particular, the basic linear regression takes the form,

$$\ln(\mathrm{VMT}_{it}) = \alpha + \beta_1 \ln(\mathrm{MPG}_{it}) + \beta_2 \ln(p_{it}^{\mathrm{gas}}) + \mathbf{D}_i'\gamma + \mathbf{X}_{it}'\delta + \lambda Z_{it} + \tilde{\varepsilon}_{it} \tag{4.5}$$

where i and t are used to index household and time, respectively. The vector $\mathbf{D}_i$ includes time-invariant household characteristics, such as education or regional location, while the vector $\mathbf{X}_{it}$ includes time-variant household characteristics, such as income. Finally, Z_{it} is vehicle class fixed effect to control for the correlation between fuel economy and other broad vehicle characteristics.

One concern with estimating the model in (4.5) is any unobserved household characteristics (say $\mathbf{D}_i^u$) are implicitly included in the error term $\tilde{\varepsilon}_{it}$. Formally, the error term in equation (4.5) can be rewritten as $\tilde{\varepsilon}_{it} = \mathbf{D^u}_i'\gamma^u + \varepsilon_{it}$ where $\mathbf{D^u}_i$, denotes a vector of unobserved household characteristics, and ε_{it} is the idiosyncratic error term. To the extent that these unobservables are correlated with both vehicle choices, fuel efficiency and vehicle miles traveled, estimates of the rebound effect (i.e., β_1) will suffer from omitted variables bias. Dubin and McFadden (1984) proposed a two-stage discrete-continuous model using the conditional expectation correction method to control for such unobserved household characteristics in the context of household appliance and energy usage.

In the current paper, we instead take advantage of the panel structure of the PSID data to control for unobserved individual effects through the use of a fixed effects model. Specifically, we can rewrite our simple linear model as follows:

$$\begin{aligned} \ln(\mathrm{VMT}_{it}) &= \alpha + \beta_1 \ln(\mathrm{MPG}_{it}) + \beta_2 \ln(p_{it}^{\mathrm{gas}}) + \mathbf{D}_i'\gamma + \mathbf{D^u}_i'\gamma^u \quad && + \mathbf{X}_{it}'\delta + \lambda Z_{it} + \varepsilon_{it} \\ &= \alpha_i + \beta_1 \ln(\mathrm{MPG}_{it}) + \beta_2 \ln(p_{it}^{\mathrm{gas}}) && + \mathbf{X}_{it}'\delta + \lambda Z_{it} + \varepsilon_{it} \end{aligned} \tag{4.6}$$

where the individual fixed effect, $\alpha_i \equiv \alpha + \mathbf{D}_i'\gamma + \mathbf{D^u}_i'\gamma^u$, captures both observed and unobserved time-invariant variables. Ignoring individual (or household) unobserved variables lead to bias [Angrist and Pischke (2009)]. With this panel data structure, the omitted variable bias is eliminated implicitly through demeaning of the regression equation; i.e.,[9]

$$\begin{aligned} \ln(VMT_it) - \overline{\ln(VMT_i)} &= \beta_1[\ln(\text{MPG}_{it}) - \overline{\ln(\text{MPG}_i)}] + \beta_2(p_{it}^{\text{gas}} - \overline{p_i^{\text{gas}}}) + (\mathbf{X}_{it} - \overline{\mathbf{X}}_i)'\delta \\ &\quad +\lambda(Z_{it} - \overline{Z_i}) + (\varepsilon_{it} - \overline{\varepsilon_i}) \\ \triangle \ln(\text{VMT}_{it}) &= \beta_1 \triangle \ln(\text{MPG}_{it}) + \beta_2 \triangle p_{it}^{\text{gas}} + \triangle \mathbf{X}_{it}' \delta + \lambda \triangle Z_{it} + \triangle \varepsilon_{it} \qquad (4.7) \end{aligned}$$

where $\triangle \ln(\text{VMT}_{it}) = \ln(VMT_it) - \overline{\ln(VMT_i)}$ and so on. In the time-demeaned equation (4.7), the unobserved time-invariant variables captured by α_i are eliminated.

In summary, when estimating the rebound effect, there are two primary sources for omitted variable bias identified in the literature, (1) unobserved household characteristics and (2) unobserved vehicle characteristics, potentially correlated with fuel economy variable [Linn (2013)]. In our preferred model, we control these two sources by including vehicle class fixed effects and adopting the individual fixed effect model in a panel data setting. We also allow for time-period effects, i.e., year dummy variables to account for any aggregate individual-invariant time variables.

4.5 Empirical Results

Table 4.3 provides estimates of the rebound effects, based on empirical models of definition 1. As noted above, our empirical models include several fixed effects, such as vehicle class, city size, year, and vehicle age, to control for household and vehicle characteristics. Along with base models from equations (4.5) and (4.6), i.e., pooled OLS, random effects (RE), and fixed effects (FE) models, we also implement an IV regression, as well as random and fixed effects models with IV. In particular, we adopt the instruments proposed by Linn (2013); i.e., fuel price at the time the vehicle was purchased and interaction of this fuel price with household characteristics, such as household size, number of adults, vehicle age, number of vehicle, income, and education level fixed effects.

[9] Another way to correct omitted variable bias from unobserved variable is an instrumental variable (IV) approach. Linn (2013) proposes to use the gasoline price when the vehicle was purchased and its interactions with household characteristics as instrument variables. In the following section, we also report results based on the IV method following Linn (2013).

Columns 2 through 4 in Table 4.3 provide the baseline OLS, RE, and FE model estimates, respectively, employing the full sample. The estimated rebound effect using definition 1 (corresponding to the coefficient β_1 on $\ln(MPG)$ in equation (4.5) falls within a very narrow range (0.53 to 0.56) and is statistically significant at a 1% level for all three models. Our estimates are generally larger than the values in the recent literature, but lie well within the range found in the literature.[10]

Instead, if definition 2 is used to characterize the rebound effect (corresponding to the coefficient $-\beta_2$ on $\ln(p_{it}^{gas})$ in equation (4.5), the estimated rebound effect for our three baseline models lies between 0.61 and 0.64, and the estimates are statistically significant at a 1% level. This falls well within the range (from -0.23 to -0.80) given by Graham and Glaister (2002) from their extensive review of previous studies. As noted in our literature review section, the use of definition 2 for the rebound effect is equivalent to the measure based on definition 1, if the gasoline prices are exogenous and the consumer's response to a 1% change in gasoline prices are equivalent with 1% change in fuel economy. The evidence supporting the hypothesis ($H_0 : \beta_1 = -\beta_2$) is controversial in the literature. Some studies report no difference between the two rebound measures [e.g., Frondel et al. (2012); Small and Dender (2007)]. However, other studies find fuel economy has a large significant effect on VMT [Gillingham (2012)], while gasoline prices do not have a statistically significant impact on VMT [Li et al.(forthcoming) and Linn (2013)].[11] We test the linear hypothesis that the elasticity of VMT with respect to fuel economy is same as the the negative elasticity of VMT with respect to gasoline prices. The test results are presented in the last row of Table 4.3. Our results show that the equivalence of coefficients of logged fuel economy and logged gasoline prices. Additionally, the estimates for IV, RE, and FE with IV are presented in columns 5 to 7 of Table 4.3. According to the results of Linn (2013), the Angrist-Pischke first-stage F-test statistics (F = 7.25) reject the null hypothesis as a weak instrument, overidentifying restriction test, Hansen's J test (J statistics

[10] In our applications, a Hausman test rejects the hypothesis that the random effects model is valid. Therefore, our preferred model is fixed effects model.

[11] Li et al.(forthcoming) and Linn (2013) use cross-sectional data, NHTS. One possible explanation for insignificant estimates is that the cross-sectional data do not have sufficient variation, even though gasoline prices vary by region and quarter. Figure C.1 shows the distribution for gasoline prices in 2010 and 1991 to 2010 at state-level. The standard deviation of gasoline prices in 2010 is 0.16, while 1991 to 2010 is 0.66.

= 114.96) shows the overidentifying restriction is not valid. Above all, the estimates for logged MPG are not statistically significant. The results for regression show that the IV approach is not valid in our study using PSID.

Results from other definitions of the rebound effect are shown in Table 4.4. The estimates of the rebound effect lie in the range of 0.61 to 0.64 for definition 2, 0.61 to 0.73 for definition 3, and 0.67 to 0.76 for definition 4. All estimates are statistically significant at the 1% level. For the results of the estimation based on definition 1, our estimates for other definitions are generally larger than the estimates in prior studies even though these fall in the wide range of the previous literature, except definition 4.[12]

Another issue in the estimation of the rebound effect is the different behaviors between the households who own single vehicle and more than one vehicle. Tables 4.5 and 4.7 present the estimates from samples including households with only one vehicle, and Tables 4.6 and 4.8 are the results from the multi-vehicle holding household sample. Since households with more than one vehicle can increase their use of their more efficient vehicle when the gasoline prices rise, they can more easily maintain the VMT, even with higher gasoline prices. In other words, the elasticity of VMT to gasoline price for households with multi-vehicles should be lower than households with a single vehicle. Our results are consistent with this intuition and the results from prior studies [Zia Wadud et al. (2010); Frondel et al. (2012)].

Finally, we include interaction terms by income decile in fuel economy or price variables of equation 4.5 and 4.6 to investigate the heterogeneous responses to the changing prices or fuel efficiency across the income distribution, as West (2004). Table 4.9 presents the results of the heterogeneous rebound effect across income groups according to the definitions and models. The last three rows show test statistics for the hypothesis that the response to the change of fuel economy (definition 1), per-mile operating cost (definition 2), and gasoline prices (definitions 3 and 4) are the same between low (1st, 2nd, and 3rd decile) and high (8th, 9th, and 10th decile) income household. Our test results reject the equivalence of the coefficient for low income households with high income households in most cases. These results are consistent with prior

[12] As shown in Table 4.1, the estimates from other studies are 0.22 to 0.87 for definition 2, 0.03 to 0.90 for definition 3, and 0.23 to 0.70 for definition 4.

studies, which show the low income households respond more intensely to gasoline prices than high income households [Zia Wadud et al. (2010); West (2004)]. The estimates for definition 3 and 4 show the same patterns, which low income households reveal a higher rebound effect than high income households. On the other hand, the rebound effect in definitions 1 and 2 is lower for low income households. The intuitive reason why elasticities of fuel economy (definition 1) and per-mile operating cost (definition 2) show opposite patterns to gasoline price elasticities (definitions 3 and 4) is, in the long-run, high income households can change their vehicles into more efficient ones when they need to drive more. However, it is not easy for low income households, due to budget limit.

4.6 Conclusion

The policy-makers in the United States have adopted Corporate Average Fuel Economy (CAFE) standards to reduce greenhouse gas (GHG) rather than other policies, particulary gasoline tax. The reason they adhere to CAFE standards may be more politically expedient. Moreover, they argue the CAFE standards can be more effective than increasing gasoline prices to reduce gasoline consumption because consumers undervalue the gasoline costs they will pay later, relative to the vehicle prices they pay now, the so-called 'energy paradox'. However, people who oppose CAFE standards, many economists, assert increasing fuel efficiency raises driving miles, known as the rebound effect, while higher gasoline prices work to reduce driving for all vehicles. Although several studies investigate the rebound effect, the results are not consistent. In other words, the rebound effect in the literature has a wide range.

To estimate the rebound effect, i.e., the elasticity of fuel economy to VMT, one concern is to control the endogeneity of the fuel economy variable possibly correlated with household and vehicle unobserved variables. This paper adopts panel estimation using the panel study of income dynamics (PSID) to control unobserved household characteristics. Our preferred estimate, fixed effects model, is 0.56 - larger than those of recent other studies even though it lies within the range commonly accepted in the literature. Additionally, panel data have more variations, time and spatial, in gasoline prices compared with cross-section data, which have only spatial variations. Sufficient variations allow accurate estimation in gasoline price elasticity. While recent studies using cross-section data show that the gasoline price elasticity is relatively small and not statistically significant, we find that the gasoline price elasticity is statistically significant and larger than fuel economy elasticity. An additional reason the estimates in the literature have a huge range is that several different empirical definitions are used with an assumption the impact of gasoline prices on VMT are the same as fuel economy. We estimate the rebound effect based on various definitions. The range of the rebound effect across definitions in fixed effects model is from 0.41 to 0.67, which is not as wide as the estimates in prior studies. Furthermore, our test cannot reject the hypothesis which both gasoline price and fuel economy have same effects on VMT. Finally, we test the difference of response to the change of gasoline price or fuel economy by income decile as well. We find evidence that low income households are more reactive to a change in gasoline prices; whereas, less sensitive to changes in fuel economy.

Bibliography

Joshua D. Angrist and Jorn-Steffen Pischke. *Mostly Harmless Econometrics - An Empiricists Companion*. Princeton University Press, 2009.

Antonio Bento, L. Goulder, M. Jacobsen, and R. von Haefen. Distributional and efficiency impacts of increased u.s. gasoline taxes. *American Economic Review*, 99, 2009.

J.A. Dubin and D.L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2):345–362, 1984.

EPA and DOT. Joint technical support document: Final rulemaking for 2017-2025 light-duty vehicle greenhouse gas emission standards and corporate average fuel economy standards. *EPA-420-R-12-901*, 2012.

Manuel Frondel and Coline Vance. Re-identifying the rebound: What about asymmetry? *Energy Journal*, 34, 2013.

Manuel Frondel, Jorg Peters, and Colin Vance. Identifying the rebound: Evidence from a german household panel. *Energy Journal*, 29(4):154–163, 2008.

Manuel Frondel, Nolan Ritter, and Colin Vance. Heterogeneity in the rebound effect: Further evidence for germany. *Energy Economics*, 34, 2012.

Kenneth Gillingham. Selection on anticipated driving and the consumer response to changing gasoline prices. *Yale University Working Paper*, 2012.

Kenneth Gillingham, David Rapson, and Gernot Wagner. The rebound effect and energy efficiency policy. *Yale University Working Paper*, 2014.

Daniel J. Graham and Stepehn Glaister. The rebound effect and energy efficiency policy. *Journal of Transport Economics and Policy*, 36(1):1–26, 2002.

David L. Greene. The rebound effect and energy efficiency policy. *Energy Policy*, 41, 2012.

Jonathan E. Hughes, Christopher R. Knittel, and Daniel Sperling. Evidence of a shift in the short-run price elasticity of gasoline demand. *The Energy Journal*, 29, 2008.

Shanjun Li, Joshua Linn, and Erich Muehlegge. Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy*.

Joshua Linn. The rebound effect for passenger vehicles. *RFF Working Paper*, 2013.

Kevin Roth. The unintended consequences of uncoordinated regulation: Evidence from the tranportation sector? *Working Paper*, 2012.

Kenneth A. Small and Kurt Van Dender. Fuel efficiency and motor vehicle travel: The declining rebound effect. *Energy Journal*, 28(1):25–51, 2007.

Steve Sorrell and John Dimitropoulos. The rebound effect: Microeconomic definitions, limitations and extentions. *Ecological Economics*, 65(3):636, 2008.

Steve Sorrell, John Dimitropoulos, and Matt Sommerville. Empirical estimates of the direct rebound effect: A review. *Energy Policy*, 37, 2009.

Elisheba Spiller. Household vehicle bundle choice and gasoline demand: A discrete-continous approach. *working paper*, 2012.

S.E. West. Distributional effects of alternative vehicle pollution control policies. *Journal of Public Economics*, 88, 2004.

Zia Zia Wadud, Daniel J. Graham, and Robert B. Noland. Gasoline demand with heterogeneity in household responses. *Energy Journal*, 31(1):47, 2010.

Table 4.1 The Estimates of Rebound Effect in the Recent Literature

Type of elasticity	Study	Data	Method	Estimates
Definition 1 η_{μ}(VMT)	Frondel et al. (2012)	Germany Household level Panel (1997-2009)	RE	0.42
	Linn (2013)	U.S. Household level Cross-sectional (2001)	IV	0.44
Definition 2 η_{p^M}(VMT)	Small and van Dender (2007)	U.S. State-level Panel (1966-2001)	3SLS	SR: 0.05 LR: 0.22
	Frondel et al. (2012)	Germany Household level Panel (1997-2009)	RE	0.46
	Bento et al. (2009)	U.S. Household level Cross-sectional (2001)	Structural	0.74
	West (2004)	U.S. Household level Cross-sectional (1997)	Structural	0.87
Definition 3 η_{μ}(GasolineDemand)	Frondel et al. (2012)	Germany Household level Panel (1997-2009)	RE	0.90
	Bento et al. (2009)	U.S. Household level Cross-sectional (2001)	Structural	0.35
	Hughes et al. (2008)	U.S. Aggregate Time series (1975-1980)	OLS	0.21-0.34
	Hughes et al. (2008)	U.S. Aggregate Time series (2001-2006)	OLS	0.03-0.08
Definition 4 $\eta_{p^{\mathrm{gas}}}$(VMT)	Frondel and Vance (2013)	Germany Household level Panel (1997-2009)	FE	0.46
	Frondel and Vance (2013)	Germany Household level Panel (1997-2009)	RE	0.70
	Bento et al. (2009)	U.S. Household level Cross-sectional (2001)	Structural	0.34
	Roth (2013)	U.S. Household level Cross-sectional (2001, 2009)	Structural	0.45
	Spiller (2013)	U.S. Household level Cross-sectional (2001, 2009)	Structural	0.62
	Gillingham (2013)	Vehicle level Cross-sectional (2001-2009)	IV	0.23

Notes: RE=Random Effect Model, FE=Fixed Effect Model, IV=Instrumental Variable Regression, SR: Short-Run, LR=Long-Run

Table 4.2 Summary Statistics

	1999	2001	2003	2005	2007	2009	2011
Family Size	2.80	2.75	2.70	2.69	2.69	2.67	2.67
	(1.43)	(1.43)	(1.42)	(1.40)	(1.42)	(1.43)	(1.42)
Age of Head	43.12	43.80	44.02	44.19	44.28	44.83	46.40
	(14.59)	(14.98)	(15.08)	(15.37)	(15.43)	(15.46)	(15.19)
Number of Adults	1.84	1.84	1.83	1.83	1.82	1.82	1.83
	(0.67)	(0.69)	(0.71)	(0.70)	(0.69)	(0.69)	(0.69)
Income ($2010)	77,692	79,305	75,193	77,797	75,448	76,835	74,941
	(90,872)	(92,831)	(105,891)	(125,221)	(74,474)	(113,592)	(82,400)
VMT (1000miles)	20.23	19.61	17.87	17.83	17.04	18.02	15.54
	(16.46)	(19.67)	(14.91)	(14.20)	(13.48)	(15.98)	(12.18)
Fuel Economy(MPG)	20.28	20.17	19.98	19.92	19.93	20.08	20.26
	(3.16)	(3.42)	(3.42)	(3.56)	(3.64)	(3.78)	(4.02)
Gasoline Price($/Gallon)	1.09	1.57	1.63	2.28	2.94	2.28	3.77
	(0.17)	(0.17)	(0.18)	(0.22)	(0.26)	(0.28)	(0.32)
Number of Vehicle Per HH	1.71	1.73	1.74	1.71	1.74	1.73	1.69
	(0.68)	(0.69)	(0.70)	(0.68)	(0.70)	(0.70)	(0.68)
Households with 1 Vehicle	42.1%	41.3%	40.9%	42.1%	40.7%	41.5%	43.0%
Households with 2 Vehicle	45.0%	44.5%	44.6%	45.0%	44.4%	43.8%	44.6%
Households with 3 Vehicle	12.9%	14.1%	14.5%	12.9%	15.0%	14.7%	12.4%
Observations	4,444	5,030	5,452	5,514	5,697	5,865	5,011

The standard deviations are reported in the parenthesis.

Table 4.3 The Estimates of Rebound Effect From Definition 1(Full Sample)

	OLS	RE	FE	IV	RE with IV	FE with IV
Dependent Var.			ln(VMT)			
ln(MPG)	0.543***	0.527***	0.556***	-0.429	-0.325	-1.049
	(0.088)	(0.060)	(0.114)	(0.641)	(0.696)	(0.807)
ln(Fuel Price)	-0.782***	-0.714***	-0.671***	-0.721***	-0.700***	-0.599***
	(0.093)	(0.063)	(0.106)	(0.097)	(0.069)	(0.101)
ln(Income)	0.096***	0.081***	0.046**	0.092***	0.079***	0.050***
	(0.019)	(0.013)	(0.020)	(0.018)	(0.014)	(0.018)
Age of Head	-0.011***	-0.011***		-0.010***	-0.012***	
	(0.001)	(0.001)		(0.001)	(0.001)	
Family Size	0.048***	0.039***	0.021	0.036***	0.034***	0.012
	(0.009)	(0.007)	(0.014)	(0.010)	(0.009)	(0.014)
Number of Adults	0.135***	0.074***	0.106***	0.105***	0.090***	0.109***
	(0.023)	(0.015)	(0.024)	(0.023)	(0.021)	(0.025)
Number of Vehicles	0.814***	0.710***	0.593***	0.566*	0.597**	0.224
	(0.200)	(0.158)	(0.188)	(0.292)	(0.249)	(0.278)
High School	-0.035	0.004		0.025	0.002	
	(0.041)	(0.030)		(0.038)	(0.032)	
Some College	0.064	0.088***		0.099**	0.090***	
	(0.041)	(0.030)		(0.040)	(0.031)	
Bachelor	0.047	0.057*		0.082*	0.068*	
	(0.045)	(0.033)		(0.043)	(0.035)	
Graduate	0.077*	0.090***		0.137**	0.112***	
	(0.047)	(0.035)		(0.053)	(0.040)	
Female Head	-0.014	-0.060**		0.018	-0.028	
	(0.034)	(0.024)		(0.036)	(0.031)	
Vehicle Class F.E.	X	X	X	X	X	X
CitySize F.E.	X	X		X	X	
Year F.E.	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X
$H_0 : \beta_1 = -\beta_2$[a]	3.60*	4.57	0.55	3.53	2.30	4.16**

[a] The test statistics for the hypothesis are reported in the last row.

Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.4 The Estimates of Rebound Effect from Other Definitions (Full Sample)

	Definition 2			Definition 3			Definition 4		
	Pooled OLS	RE	FE	Pooled OLS	RE	FE	Pooled OLS	RE	FE
Dependent Var.		ln(VMT)			ln(GasDemand)			ln(VMT)	
ln(MPG)				-0.467***	-0.464***	-0.411***			
				(0.068)	(0.046)	(0.083)			
ln(Fuel Price)				-0.734***	-0.675***	-0.614***	-0.755***	-0.699***	-0.668***
				(0.071)	(0.049)	(0.079)	(0.093)	(0.063)	(0.106)
ln(Operating Cost)	-0.635***	-0.605***	-0.607***						
	(0.065)	(0.043)	(0.079)						
ln(Income)	0.096***	0.081***	0.046**	0.088***	0.071***	0.045***	0.092***	0.079***	0.044**
	(0.019)	(0.013)	(0.020)	(0.014)	(0.010)	(0.015)	(0.019)	(0.013)	(0.020)
Age of Head	-0.011***	-0.011***		-0.011***	-0.011***		-0.012***	-0.011***	
	(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)	
Family Size	0.049***	0.039***	0.021	0.044***	0.039***	0.021*	0.043***	0.035***	0.018
	(0.009)	(0.007)	(0.014)	(0.007)	(0.006)	(0.011)	(0.009)	(0.007)	(0.014)
Number of Adults	0.133***	0.073***	0.106***	0.121***	0.068***	0.096***	0.147***	0.083***	0.112***
	(0.023)	(0.015)	(0.024)	(0.018)	(0.012)	(0.019)	(0.023)	(0.015)	(0.024)
Number of Vehicles	0.843***	0.721***	0.605***	0.813***	0.689***	0.555***	0.657***	0.633***	0.469**
	(0.202)	(0.160)	(0.186)	(0.201)	(0.155)	(0.176)	(0.202)	(0.164)	(0.211)
High School	-0.034	0.005		-0.031	0.003		-0.032	0.004	
	(0.041)	(0.030)		(0.031)	(0.023)		(0.041)	(0.030)	
Some College	0.065	0.088***		0.055*	0.075***		0.068*	0.092***	
	(0.041)	(0.030)		(0.032)	(0.024)		(0.041)	(0.030)	
Bachelor	0.047	0.057*		0.035	0.042		0.055	0.067**	
	(0.045)	(0.033)		(0.035)	(0.026)		(0.045)	(0.033)	
Graduate	0.074	0.088**		0.065*	0.077***		0.097**	0.106***	
	(0.047)	(0.035)		(0.037)	(0.028)		(0.047)	(0.035)	
Female Head	-0.016	-0.062**		-0.021	-0.065***		0.002	-0.046*	
	(0.034)	(0.024)		(0.027)	(0.019)		(0.034)	(0.024)	
Vehicle Class F.E.	X	X	X	X	X	X	X	X	X
CitySize F.E.	X	X		X	X		X	X	
Year F.E.	X	X	X	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X	X	X	X

Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.5 The Estimates of Rebound Effect (Single Vehicle Household)

	OLS	RE	FE	IV	RE with IV	FE with IV
Dependent Var.			ln(VMT)			
ln(MPG)	0.547***	0.483***	0.711***	-0.343	-0.334	-2.736**
	(0.135)	(0.095)	(0.221)	(0.844)	(0.876)	(1.342)
ln(Fuel Price)	-0.774***	-0.742***	-0.793***	-0.687***	-0.722***	-0.797***
	(0.165)	(0.120)	(0.220)	(0.165)	(0.116)	(0.218)
ln(Income)	0.090***	0.073***	0.068**	0.074***	0.070***	0.067**
	(0.030)	(0.019)	(0.035)	(0.027)	(0.019)	(0.034)
Age of Head	-0.013***	-0.014***		-0.013***	-0.014***	
	(0.001)	(0.001)		(0.001)	(0.001)	
Family Size	0.075***	0.048***	0.048	0.065***	0.041**	0.035
	(0.019)	(0.015)	(0.040)	(0.019)	(0.017)	(0.039)
Number of Adults	0.096**	0.062**	0.108*	0.092**	0.074**	0.067
	(0.043)	(0.030)	(0.058)	(0.039)	(0.035)	(0.059)
High School	-0.030	-0.031		0.004	-0.030	
	(0.069)	(0.049)		(0.062)	(0.048)	
Some College	0.078	0.059		0.087	0.061	
	(0.072)	(0.050)		(0.066)	(0.053)	
Bachelor	0.062	0.027		0.101	0.043	
	(0.080)	(0.055)		(0.071)	(0.057)	
Graduate	0.154*	0.099*		0.186**	0.133*	
	(0.083)	(0.059)		(0.089)	(0.072)	
Female Head	-0.016	-0.026		0.027	-0.001	
	(0.046)	(0.033)		(0.050)	(0.041)	
Vehicle Class F.E.	X	X	X	X	X	X
CitySize F.E.	X	X		X	X	
Year F.E.	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X
$H_0 : \beta_1 = -\beta_2$[a]	1.22	2.91*	0.07	1.60	1.41	6.61**

[a] The test statistics for the hypothesis are reported in the last row.

Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.6 The Estimates of Rebound Effect (Multi-Vehicle Households)

	OLS	RE	FE	IV	RE with IV	FE with IV
Dependent Var.			ln(VMT)			
ln(MPG)	0.520***	0.584***	0.576***	1.855***	1.271*	1.083
	(0.090)	(0.066)	(0.113)	(0.709)	(0.715)	(0.831)
ln(Fuel Price)	-0.809***	-0.666***	-0.572***	-0.770***	-0.710***	-0.538***
	(0.091)	(0.062)	(0.092)	(0.099)	(0.074)	(0.091)
ln(Income)	0.105***	0.087***	0.051***	0.140***	0.095***	0.058***
	(0.017)	(0.015)	(0.018)	(0.018)	(0.015)	(0.017)
Age of Head	-0.009***	-0.008***		-0.006***	-0.007***	
	(0.001)	(0.001)		(0.001)	(0.001)	
Family Size	0.047***	0.040***	0.012	0.048***	0.046***	0.022
	(0.009)	(0.008)	(0.014)	(0.010)	(0.009)	(0.013)
Number of Adults	0.140***	0.077***	0.079***	0.049*	0.065***	0.045**
	(0.026)	(0.015)	(0.021)	(0.027)	(0.021)	(0.022)
Number of Vehicles	-0.001	-0.048	-0.095	0.620	0.191	0.057
	(0.367)	(0.245)	(0.203)	(0.514)	(0.382)	(0.294)
High School	-0.021	0.007		-0.018	0.008	
	(0.040)	(0.032)		(0.041)	(0.030)	
Some College	0.063	0.085***		0.056	0.080***	
	(0.038)	(0.032)		(0.040)	(0.030)	
Bachelor	0.038	0.050		0.017	0.036	
	(0.043)	(0.036)		(0.044)	(0.035)	
Graduate	0.025	0.071*		-0.002	0.047	
	(0.046)	(0.039)		(0.052)	(0.038)	
Female Head	-0.023	-0.048*		-0.048	-0.058*	
	(0.038)	(0.029)		(0.040)	(0.030)	
Vehicle Class F.E.	X	X	X	X	X	X
CitySize F.E.	X	X		X	X	
Year F.E.	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X
$H_0 : \beta_1 = -\beta_2$	4.8**	0.83	0.00	2.62	0.67	0.46

t statistics are reported in parentheses and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.7 The Estimates of Rebound Effect from Other Definitions (Single Vehicle Household)

	Definition 2			Definition 3			Definition 4		
	Pooled OLS	RE	FE	Pooled OLS	RE	FE	Pooled OLS	RE	FE
Dependent Var.		ln(VMT)			ln(GasDemand)			ln(VMT)	
ln(MPG)				-0.475***	-0.507***	-0.333**			
				(0.103)	(0.072)	(0.157)			
ln(Fuel Price)				-0.724***	-0.698***	-0.698***	-0.749***	-0.729***	-0.797***
				(0.122)	(0.089)	(0.157)	(0.165)	(0.120)	(0.221)
ln(Operating Cost)	-0.623***	-0.578***	-0.746***						
	(0.109)	(0.075)	(0.156)						
ln(Income)	0.090***	0.072***	0.069**	0.075***	0.059***	0.059**	0.088***	0.071***	0.066*
	(0.030)	(0.019)	(0.035)	(0.021)	(0.014)	(0.025)	(0.030)	(0.019)	(0.034)
Age of Head	-0.013***	-0.014***		-0.013***	-0.013***		-0.014***	-0.014***	
	(0.001)	(0.001)		(0.001)	(0.001)		(0.001)	(0.001)	
Family Size	0.075***	0.049***	0.048	0.067***	0.047***	0.036	0.070***	0.044***	0.046
	(0.019)	(0.015)	(0.040)	(0.014)	(0.011)	(0.029)	(0.019)	(0.015)	(0.040)
Number of Adults	0.096**	0.060**	0.109*	0.080**	0.048**	0.085*	0.105**	0.069**	0.105*
	(0.043)	(0.030)	(0.058)	(0.033)	(0.023)	(0.044)	(0.043)	(0.030)	(0.058)
High School	-0.030	-0.030		-0.031	-0.026		-0.027	-0.030	
	(0.069)	(0.049)		(0.051)	(0.036)		(0.069)	(0.049)	
Some College	0.078	0.058		0.061	0.047		0.076	0.060	
	(0.073)	(0.050)		(0.054)	(0.038)		(0.073)	(0.050)	
Bachelor	0.061	0.026		0.032	0.002		0.067	0.037	
	(0.080)	(0.055)		(0.060)	(0.042)		(0.080)	(0.055)	
Graduate	0.150*	0.095		0.117*	0.068		0.178**	0.119**	
	(0.083)	(0.059)		(0.063)	(0.046)		(0.084)	(0.059)	
Female Head	-0.017	-0.027		-0.025	-0.040		0.002	-0.012	
	(0.047)	(0.033)		(0.035)	(0.025)		(0.046)	(0.033)	
Vehicle Class F.E.	X	X	X	X	X	X	X	X	X
CitySize F.E.	X	X		X	X		X	X	
Year F.E.	X	X	X	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X	X	X	X

Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.8 The Estimates of Rebound Effect from Other Definitions (Multi-Vehicle Household)

	Definition 2			Definition 3			Definition 4		
	Pooled OLS	RE	FE	Pooled OLS	RE	FE	Pooled OLS	RE	FE
Dependent Var.		ln(VMT)			ln(GasDemand)			ln(VMT)	
ln(MPG)	-0.635***	-0.605***	-0.607***						
	(0.065)	(0.043)	(0.079)						
ln(Fuel Price)	0.096***	0.081***	0.046**	0.088***	0.071***	0.045***	-0.755***	-0.699***	-0.668***
	(0.019)	(0.013)	(0.020)	(0.014)	(0.010)	(0.015)	(0.093)	(0.063)	(0.106)
ln(Operating Cost)	-0.011***	-0.011***		-0.011***	-0.011***				
	(0.001)	(0.001)		(0.001)	(0.001)				
ln(Income)	0.049***	0.039***	0.021	0.044***	0.039***	0.021*	0.092***	0.079***	0.044**
	(0.009)	(0.007)	(0.014)	(0.007)	(0.006)	(0.011)	(0.019)	(0.013)	(0.020)
Age of Head	0.133***	0.073***	0.106***	0.121***	0.068***	0.096***	-0.012***	-0.011***	
	(0.023)	(0.015)	(0.024)	(0.018)	(0.012)	(0.019)	(0.001)	(0.001)	
Family Size	0.843***	0.721***	0.605***	0.813***	0.689***	0.555***	0.043***	0.035***	0.018
	(0.202)	(0.160)	(0.186)	(0.201)	(0.155)	(0.176)	(0.009)	(0.007)	(0.014)
Number of Adults	-0.034	0.005		-0.031	0.003		0.147***	0.083***	0.112***
	(0.041)	(0.030)		(0.031)	(0.023)		(0.023)	(0.015)	(0.024)
Number of Vehicles	0.065	0.088***		0.055*	0.075***		0.657***	0.633***	0.469**
	(0.041)	(0.030)		(0.032)	(0.024)		(0.202)	(0.164)	(0.211)
High School	0.047	0.057*		0.035	0.042		-0.032	0.004	
	(0.045)	(0.033)		(0.035)	(0.026)		(0.041)	(0.030)	
Some College	0.074	0.088**		0.065*	0.077***		0.068*	0.092***	
	(0.047)	(0.035)		(0.037)	(0.028)		(0.041)	(0.030)	
Bachelor	-0.016	-0.062**		-0.021	-0.065***		0.055	0.067**	
	(0.034)	(0.024)		(0.027)	(0.019)		(0.045)	(0.033)	
Graduate				-0.734***	-0.675***	-0.614***	0.097**	0.106***	
				(0.071)	(0.049)	(0.079)	(0.047)	(0.035)	
Female Head				-0.467***	-0.464***	-0.411***	0.002	-0.046*	
				(0.068)	(0.046)	(0.083)	(0.034)	(0.024)	
Vehicle Class F.E.	X	X	X	X	X	X	X	X	X
CitySize F.E.	X	X		X	X		X	X	
Year F.E.	X	X	X	X	X	X	X	X	X
Vehicle Age F.E.	X	X	X	X	X	X	X	X	X

Standard errors are reported in parentheses and clustered by household, and * denotes significance at the 10%, ** at the 5%, and *** at the 1%, respectively.

Table 4.9 The Estimates of Rebound Effect by Income Decile

Decile	Definition1			Definition2			Definition3			Definition4		
	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE	OLS	RE	FE
1	0.423	0.438	0.459	0.518	0.514	0.536	0.959	0.842	0.853	1.051	0.915	0.981
	(0.092)	(0.062)	(0.114)	(0.072)	(0.048)	(0.082)	(0.088)	(0.060)	(0.100)	(0.118)	(0.078)	(0.138)
2	0.489	0.484	0.508	0.588	0.563	0.572	0.806	0.760	0.716	0.814	0.781	0.774
	(0.089)	(0.060)	(0.114)	(0.068)	(0.045)	(0.081)	(0.078)	(0.056)	(0.087)	(0.101)	(0.072)	(0.118)
3	0.525	0.516	0.535	0.634	0.603	0.601	0.734	0.697	0.674	0.731	0.706	0.721
	(0.089)	(0.060)	(0.115)	(0.068)	(0.045)	(0.083)	(0.076)	(0.052)	(0.084)	(0.100)	(0.067)	(0.112)
4	0.538	0.525	0.551	0.651	0.611	0.616	0.724	0.675	0.638	0.725	0.682	0.694
	(0.089)	(0.060)	(0.116)	(0.067)	(0.045)	(0.081)	(0.078)	(0.053)	(0.086)	(0.102)	(0.068)	(0.117)
5	0.537	0.526	0.553	0.645	0.61	0.608	0.709	0.674	0.619	0.718	0.692	0.675
	(0.088)	(0.060)	(0.113)	(0.065)	(0.044)	(0.079)	(0.078)	(0.053)	(0.084)	(0.103)	(0.068)	(0.114)
6	0.547	0.537	0.569	0.657	0.621	0.624	0.696	0.650	0.592	0.711	0.667	0.642
	(0.088)	(0.060)	(0.114)	(0.066)	(0.044)	(0.080)	(0.074)	(0.052)	(0.081)	(0.096)	(0.066)	(0.108)
7	0.535	0.540	0.572	0.642	0.624	0.622	0.732	0.649	0.592	0.758	0.670	0.642
	(0.089)	(0.060)	(0.116)	(0.066)	(0.044)	(0.082)	(0.078)	(0.053)	(0.087)	(0.102)	(0.068)	(0.116)
8	0.557	0.551	0.593	0.665	0.635	0.639	0.661	0.619	0.530	0.669	0.639	0.568
	(0.088)	(0.060)	(0.114)	(0.066)	(0.044)	(0.080)	(0.074)	(0.051)	(0.080)	(0.096)	(0.065)	(0.107)
9	0.554	0.568	0.607	0.659	0.652	0.646	0.659	0.556	0.467	0.689	0.579	0.510
	(0.090)	(0.061)	(0.114)	(0.067)	(0.045)	(0.079)	(0.078)	(0.053)	(0.083)	(0.102)	(0.067)	(0.110)
10	0.553	0.558	0.618	0.653	0.637	0.648	0.671	0.595	0.454	0.723	0.640	0.484
	(0.088)	(0.061)	(0.113)	(0.067)	(0.046)	(0.079)	(0.078)	(0.055)	(0.085)	(0.100)	(0.070)	(0.114)
$H_0 : \beta_1^1 = \beta_1^{10}$	9.61***	22.06***	17.78***	7.18***	14.92***	6.18**	15.80***	26.64***	30.21***	10.87***	18.77***	24.18***
$H_0 : \beta_1^2 = \beta_1^9$	6.03**	24.64***	14.38***	4.78**	17.94***	5.35**	8.04***	33.47***	22.54***	3.42*	19.79***	14.71***
$H_0 : \beta_1^3 = \beta_1^8$	3.04*	7.79***	9.04***	2.01	4.41**	2.42	2.85*	7.01***	10.85***	1.30	3.30*	7.66***

Standard errors are reported in parentheses and are clustered by household. The all estimates of the rebound effect are significant at 1 % level, we drop *** for the limit of space.

CHAPTER 5. GENERAL CONCLUSIONS

This dissertation consists of three independent studies that focus on welfare measurement, causation of addition emission, and effectiveness of policy in environmental economics. The first study proposes a new framework to solve the failure of convergent validity assumption in combining RP and SP. The second and third studies estimate the factor and magnitude of interesting variables to influence the choice of vehicle and its usage using the panel study of income dynamics (PSID) data.

The first study examines the method to combine revealed preference (RP) and stated preference (SP) data. Even though combining RP and SP data is recently very common in nonmarket valuation literature, several studies have reported the critical assumption, convergent validity, is often rejected. Therefore, prior studies have chosen using either combining two data sources or one source, according to the result of the convergent validity assumption. The goal for this study is to propose an alternative framework that allows for possible divergence among individuals in terms of the consistency between their RP and SP responses. The empirical results suggest that somewhat less than half the sample exhibits inconsistent preferences. Welfare estimates in our proposed latent model are significantly different with a conventional combining model and single class model.

The second study examine the interesting link between obesity and gasoline consumption. Even though there are several studies to explore the relationship between obesity and gasoline consumption, the literature relies on aggregate data or focuses on either the choice of fuel economy or usage. These approaches can lead to wrong conclusions, due to omitted variable bias from unobserved household characteristics and partial effects in either fuel economy or VMT. With unique household level data contained with the vehicle information, gasoline consumption, body mass index (BMI), and other household characteristics, we estimate the impact of obesity through both reduced- and structural-form models. To solve omitted variable bias and endogeneity problems, we adopt the fixed effects model and instrumental variable approaches as the reduced-form approach, and joint discrete and continuous model as the structural-form approach. Our findings show that the comprehensive impact of obesity and overweight on gasoline consumptions is little or ambiguous in contrast to the results of prior studies considering either driving or vehicle choices.

The third study visits the very classical and controversial issue in energy economics. There are numerous studies to estimate the rebound effect in vehicle usage or gasoline consumption. The range of estimates of the rebound effect in the literature is very wide, 0.03 to 0.9. The effectiveness of the Corporate Average Fuel Economy (CAFE) standards, a primary policy to reduce gasoline consumption in the U.S., largely depends on the size of the rebound effect. The difficulty of estimating the rebound effect is how to deal with the endogeneity of fuel economy. There are several approaches to solve endogeneous variables such as instrumental variable regression. However, even though good instruments are difficult to find and using panel data is a possibly good strategy, there is no study using U.S. panel data. This study adopts the Panel Study of Income Dynamics (PSID) - not used in the literature. Our results show that 1% increase in fuel prices or fuel economy (MPG) leads to a 0.41% to 0.67% increase in driving miles. Moreover, this study shows that the elasticity of vehicle miles traveled (VMT) with respect to both fuel economy and fuel price to be statistically significant. In addition, we find that low income households are more responsive to changes in gasoline prices, but less sensitive to changes in fuel economy.

APPENDIX A. ADDITIONAL MATERIAL FOR CHAPTER 2

Table A.1 Generated Data Experiments - Model 2 Parameter Estimates

	TRUE	TRUE	N=200			N=500			N=1000		
Parameter	s	values	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%
	0.10	-1.20	-1.15	-1.30	-1.02	-1.14	-1.25	-1.04	-1.15	-1.26	-1.05
	0.25	-1.20	-1.12	-1.25	-0.98	-1.11	-1.23	-0.99	-1.11	-1.27	-1.00
β^{RP}	0.50	-1.20	-1.15	-1.37	-0.96	-1.14	-1.31	-1.02	-1.14	-1.32	-1.00
	0.75	-1.20	-1.33	-1.60	-1.11	-1.32	-1.53	-1.17	-1.33	-1.52	-1.18
	0.90	-1.20	-1.61	-1.91	-1.39	-1.61	-1.79	-1.44	-1.61	-1.79	-1.44
	0.10	-0.60	-0.61	-0.66	-0.57	-0.61	-0.66	-0.58	-0.61	-0.64	-0.58
	0.25	-0.60	-0.63	-0.71	-0.58	-0.63	-0.68	-0.59	-0.63	-0.66	-0.59
β^{SP}	0.50	-0.60	-0.66	-0.73	-0.61	-0.66	-0.73	-0.61	-0.66	-0.71	-0.62
	0.75	-0.60	-0.72	-0.80	-0.65	-0.72	-0.78	-0.67	-0.72	-0.77	-0.67
	0.90	-0.60	-0.77	-0.83	-0.71	-0.76	-0.82	-0.73	-0.76	-0.80	-0.73
	0.10	-0.70	-0.77	-0.86	-0.70	-0.77	-0.84	-0.72	-0.77	-0.83	-0.72
	0.25	-0.70	-0.89	-1.06	-0.79	-0.89	-1.02	-0.80	-0.89	-0.97	-0.81
γ^{RP}	0.50	-0.70	-1.17	-1.37	-0.99	-1.17	-1.37	-1.01	-1.17	-1.32	-1.01
	0.75	-0.70	-1.68	-2.02	-1.33	-1.65	-1.92	-1.41	-1.66	-1.88	-1.45
	0.90	-0.70	-2.26	-2.70	-1.83	-2.26	-2.55	-1.99	-2.24	-2.46	-2.01
	0.10	-0.50	-0.55	-0.63	-0.49	-0.55	-0.60	-0.51	-0.55	-0.58	-0.52
	0.25	-0.50	-0.63	-0.72	-0.55	-0.63	-0.71	-0.58	-0.62	-0.67	-0.58
γ^{SP}	0.50	-0.50	-0.77	-0.91	-0.69	-0.77	-0.86	-0.69	-0.77	-0.84	-0.71
	0.75	-0.50	-0.95	-1.09	-0.85	-0.95	-1.02	-0.86	-0.95	-1.03	-0.89
	0.90	-0.50	-1.09	-1.19	-1.01	-1.09	-1.16	-1.03	-1.09	-1.13	-1.05
	0.10	-1.80	-1.61	-1.77	-1.47	-1.61	-1.73	-1.49	-1.61	-1.67	-1.52
	0.25	-1.80	-1.38	-1.52	-1.22	-1.38	-1.48	-1.26	-1.37	-1.45	-1.28
ρ^{RP}	0.50	-1.80	-1.10	-1.24	-0.93	-1.09	-1.19	-0.95	-1.08	-1.17	-0.99
	0.75	-1.80	-0.91	-1.00	-0.79	-0.89	-0.98	-0.80	-0.89	-0.97	-0.82
	0.90	-1.80	-0.83	-0.90	-0.74	-0.82	-0.88	-0.75	-0.82	-0.87	-0.78
	0.10	-0.40	-0.39	-0.43	-0.35	-0.39	-0.41	-0.37	-0.39	-0.40	-0.37
	0.25	-0.40	-0.37	-0.41	-0.33	-0.37	-0.40	-0.34	-0.37	-0.39	-0.34
ρ^{SP}	0.50	-0.40	-0.34	-0.39	-0.30	-0.34	-0.37	-0.31	-0.34	-0.36	-0.32
	0.75	-0.40	-0.33	-0.38	-0.28	-0.33	-0.35	-0.30	-0.33	-0.35	-0.31
	0.90	-0.40	-0.32	-0.36	-0.29	-0.32	-0.35	-0.30	-0.32	-0.34	-0.31

Table A.2 Generated Data Experiments - Model 3 Parameter Estimates

	TRUE	TRUE	N=200			N=500			N=1000		
Parameter	s	values	Mean	5%	95%	Mean	5%	95%	Mean	5%	95%
	0.10	0.40	0.33	0.24	0.42	0.33	0.26	0.39	0.33	0.25	0.39
	0.25	0.40	0.38	0.28	0.46	0.38	0.29	0.46	0.38	0.30	0.44
ω	0.50	0.40	0.45	0.36	0.52	0.45	0.38	0.53	0.45	0.38	0.51
	0.75	0.40	0.47	0.40	0.55	0.48	0.43	0.55	0.48	0.43	0.53
	0.90	0.40	0.45	0.39	0.53	0.45	0.41	0.49	0.45	0.42	0.49
	0.10	-2.00	-1.81	-2.60	-1.36	-1.79	-2.46	-1.34	-1.79	-2.43	-1.36
	0.25	-2.00	-1.61	-2.23	-1.23	-1.59	-2.20	-1.23	-1.58	-2.16	-1.25
β	0.50	-2.00	-1.44	-1.94	-1.14	-1.42	-1.80	-1.15	-1.42	-1.77	-1.14
	0.75	-2.00	-1.50	-1.80	-1.19	-1.47	-1.69	-1.26	-1.48	-1.71	-1.28
	0.90	-2.00	-1.71	-1.95	-1.41	-1.69	-1.86	-1.49	-1.69	-1.84	-1.53
	0.10	-3.00	-0.96	-1.13	-0.80	-0.97	-1.15	-0.80	-0.97	-1.14	-0.79
	0.25	-3.00	-1.09	-1.27	-0.89	-1.09	-1.28	-0.91	-1.08	-1.29	-0.90
γ	0.50	-3.00	-1.36	-1.57	-1.16	-1.36	-1.60	-1.14	-1.35	-1.59	-1.14
	0.75	-3.00	-1.83	-2.22	-1.51	-1.81	-2.07	-1.55	-1.82	-2.05	-1.61
	0.90	-3.00	-2.36	-2.77	-1.98	-2.36	-2.62	-2.08	-2.35	-2.54	-2.11
	0.10	-0.80	-1.40	-1.68	-1.17	-1.40	-1.62	-1.22	-1.39	-1.61	-1.23
	0.25	-0.80	-1.17	-1.38	-0.97	-1.16	-1.34	-1.00	-1.16	-1.33	-1.03
ρ	0.50	-0.80	-0.91	-1.05	-0.75	-0.90	-1.03	-0.78	-0.90	-1.01	-0.79
	0.75	-0.80	-0.79	-0.89	-0.67	-0.78	-0.85	-0.68	-0.78	-0.84	-0.70
	0.90	-0.80	-0.77	-0.85	-0.69	-0.77	-0.82	-0.70	-0.77	-0.82	-0.70

APPENDIX B. ADDITIONAL MATERIAL FOR CHAPTER 3

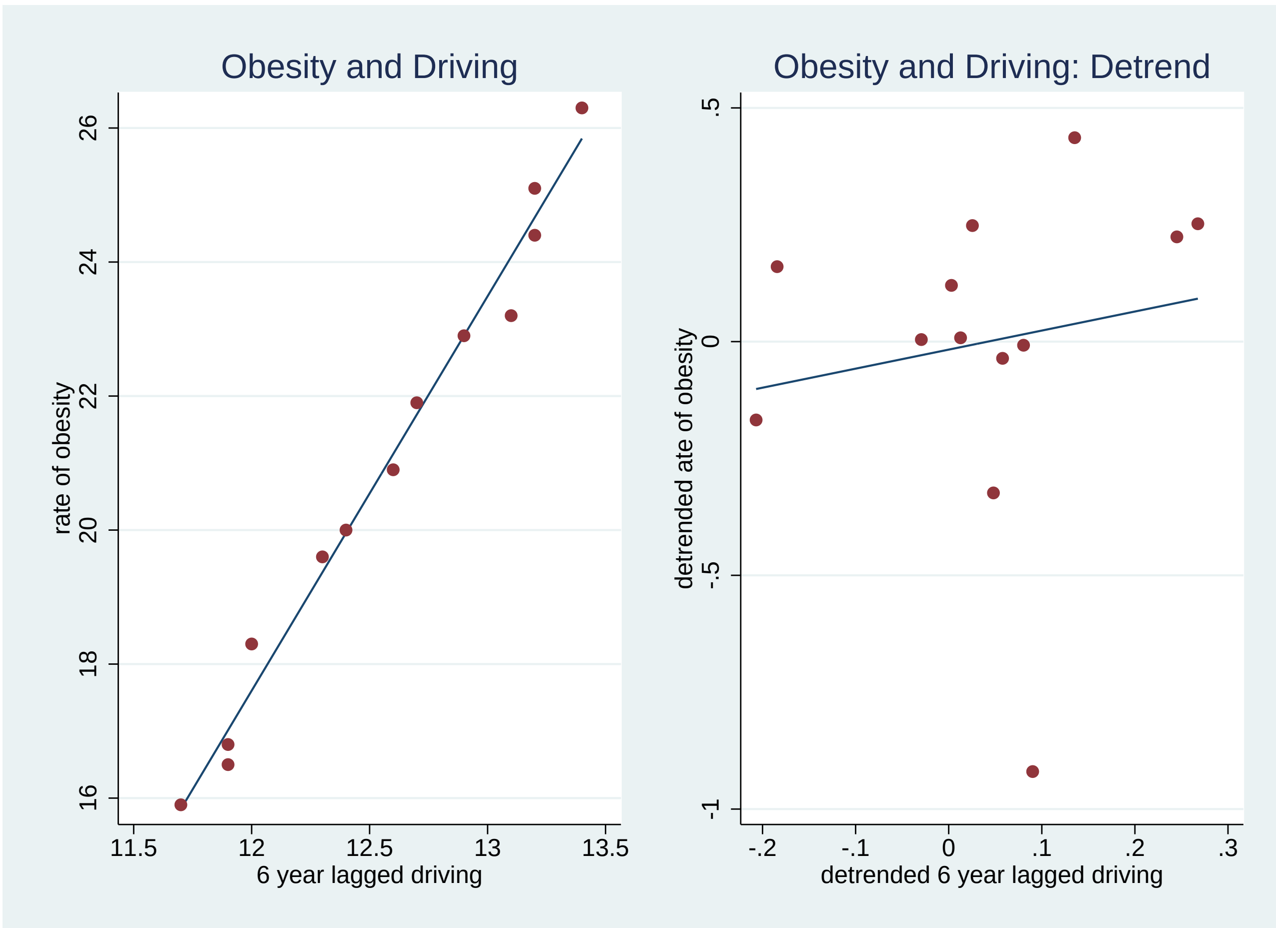

Figure B.1 Spurious regression: Rate of Obesity and Driving

Table B.1 Full Estimates of Structural Model

	1999		2001		2003		2005		2007		2009	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
α parameter												
Vehicle Age #1	2.24	0.02	2.16	0.03	2.20	0.03	2.23	0.03	1.95	0.04	1.92	0.04
Vehicle Age #2	2.28	0.02	2.18	0.03	2.23	0.04	2.20	0.02	1.95	0.04	1.93	0.04
Vehicle Age #3	2.27	0.02	2.18	0.02	2.22	0.03	2.21	0.02	1.97	0.04	1.96	0.04
Vehicle Age #4	2.25	0.02	2.16	0.02	2.21	0.04	2.21	0.02	1.95	0.04	1.95	0.03
Vehicle Age #5	2.22	0.02	2.13	0.02	2.19	0.04	2.19	0.03	1.94	0.04	1.95	0.04
Obesity	0.18	0.03	0.44	0.02	0.36	0.01	0.04	0.02	0.04	0.02	0.29	0.01
Obesity*(WB/100)	-0.09	0.02	-0.32	0.01	-0.25	0.01	0.06	0.01	-0.03	0.01	-0.20	0.01
Overweight	-0.23	0.02	-0.35	0.03	-0.28	0.03	-0.19	0.01	-0.26	0.02	-0.28	0.02
Overweight*(WB/100)	0.21	0.01	0.32	0.02	0.26	0.02	0.20	0.01	0.27	0.01	0.27	0.01
HP/WT	0.15	0.03	-0.20	0.02	0.41	0.03	-0.41	0.02	0.86	0.03	-0.09	0.03
(HP/WT)*(Head Age/100)	0.01	0.06	-0.31	0.02	-1.12	0.05	0.95	0.02	0.30	0.04	0.17	0.03
# of Adults	0.04	0.01	0.06	0.01	-0.01	0.02	0.01	0.01	0.07	0.02	0.12	0.02
β	-0.62	0.06	-0.82	0.08	-0.78	0.06	-1.20	0.08	-1.19	0.06	-1.05	0.04
λ	-7.29	0.13	-6.86	0.13	-7.01	0.08	-7.23	0.08	-6.60	0.06	-7.14	0.09
τ Parameter												
Midsize	-1.81	0.06	-0.49	0.02	-0.44	0.04	0.84	0.03	-0.55	0.04	0.38	0.03
Fullsize	-0.16	0.04	-0.77	0.04	-0.39	0.03	-0.94	0.02	-0.79	0.04	0.23	0.02
Luxury Car	-0.67	0.03	-0.81	0.03	-1.77	0.06	-1.21	0.02	-1.90	0.06	-0.85	0.02
Small SUV	0.24	0.04	0.43	0.03	-0.54	0.02	-0.90	0.05	0.37	0.03	-0.06	0.03
Large SUV	0.74	0.03	-0.23	0.03	-0.33	0.04	0.52	0.06	0.42	0.03	1.18	0.05
Small Truck	-0.32	0.03	-0.81	0.07	-1.93	0.03	-0.37	0.03	-1.94	0.04	0.56	0.03
Large Truck	-0.07	0.03	-0.60	0.06	-0.48	0.05	0.56	0.03	-0.21	0.02	1.13	0.03
Minivan	-0.67	0.03	1.59	0.04	-0.25	0.03	-0.21	0.03	0.24	0.04	0.49	0.03
European	-0.54	0.06	0.78	0.03	1.27	0.04	1.37	0.03	0.09	0.03	-0.37	0.05
Asian	0.08	0.04	-0.88	0.04	1.30	0.05	1.91	0.06	-0.11	0.03	-0.49	0.03
Vehicle Age #1	-1.23	0.03	0.63	0.05	0.76	0.06	0.29	0.03	0.05	0.03	-0.52	0.02
Vehicle Age #2	0.68	0.04	-0.62	0.02	0.48	0.07	-0.29	0.03	0.11	0.06	-1.16	0.05
Vehicle Age #3	-0.55	0.03	-0.31	0.02	-0.18	0.02	0.63	0.03	-0.59	0.05	-0.70	0.02
Vehicle Age #4	0.76	0.03	-0.72	0.02	0.29	0.03	-1.00	0.05	-0.26	0.02	0.03	0.03
WT/100	-2.91	0.03	-4.08	0.02	-4.11	0.03	-3.72	0.03	-3.88	0.05	-4.46	0.04
WB/100	0.74	0.08	-0.22	0.03	-0.87	0.03	0.20	0.05	0.04	0.03	-0.70	0.03
HP/WT	-1.39	0.07	0.56	0.06	0.07	0.05	0.65	0.02	-1.00	0.02	-0.05	0.03
φ parameter												
(Head Age/100)	0.12	0.03	-1.28	0.06	-1.75	0.07	0.04	0.02	-0.74	0.06	-1.22	0.07
College Degree of Head	0.50	0.03	-1.23	0.04	1.15	0.03	-0.38	0.03	1.06	0.03	-1.01	0.05
MSA $<$ 250k	-0.13	0.03	0.44	0.02	0.58	0.02	-0.58	0.03	0.59	0.02	0.10	0.03
250k $\leq$ MSA $<$ 500k	-0.74	0.05	-0.35	0.03	1.02	0.02	0.90	0.03	0.14	0.03	-0.76	0.03
500k $\leq$ MSA $<$ 1m	0.08	0.04	-0.29	0.03	0.02	0.02	-0.77	0.03	0.10	0.03	1.30	0.05
1m $\leq$ MSA $<$ 3m	-0.52	0.03	-0.11	0.03	-1.22	0.05	0.63	0.02	-0.49	0.02	-0.69	0.04
MSA geq 3m	-1.21	0.04	-1.39	0.03	0.75	0.03	0.28	0.03	0.37	0.03	-0.68	0.03
μ	7.83	0.10	7.85	0.12	7.72	0.07	7.97	0.09	7.60	0.10	7.91	0.08
σ	-1.34	0.05	-1.58	0.08	-1.56	0.06	-1.65	0.07	-1.54	0.07	-1.45	0.07

APPENDIX C. ADDITIONAL MATERIAL FOR CHAPTER 4

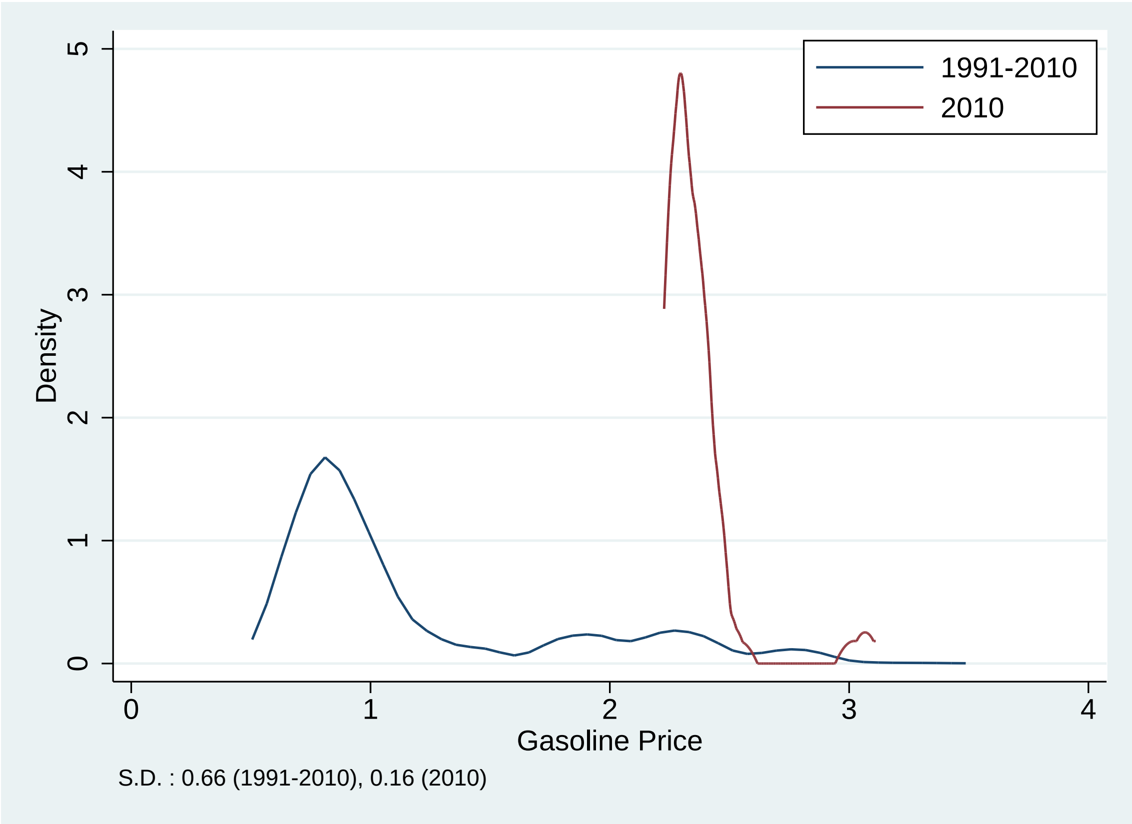

Figure C.1 Density of Gasoline Price 1991-2010 vs 2010

Data Sources

- Panel Study of Income Dynamics (PSID): http://psidonline.isr.umich.edu
  - Household characteristics such as income, age of head, family size and others.
  - The PSID publicly provides state-level address and manufacturer, make, broad type, and model year on vehicle each household owns.
  - In this paper, we use MSA-level address and specific vehicle models under conditions of a restricted use contract.
- EPA fuel-economy: http://www.fueleconomy.gov/feg/download.shtml
- WARDS Automotive Year Book: http://wardsauto.com/subscriptions/auto-yearbook
- American Chamber of Commerce Researchers Association's Regional Cost of Living Index (ACCCRA) http://www.coli.org/
- EIA Unleaded Regular Gasoline Price, U.S. City Average Retail Prices http://www.eia.gov/totalenergy/data/monthly/#prices

Data Construction

- Based on address reported in PSID data, we merge the household characteristics with gasoline prices (unleaded regular prices including taxed) in ACCRA cost of living index.
- Unfortunately, ACCRA gasoline prices does not cover MSAs, for the missing observations, we impute them using EIA state-level gasoline prices.
- Using the specific model shown in PSID, we merge MPG from EPA. However, EPA fuel economy data include from 1984. For prior models, we apply WARDS Automotive Yearbook.
- We impute missing observations by regression technique.