Title
Econometric Models in Transportation

Author
Wong, Timothy Chong Ji

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Econometric Models in Transportation

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics – Transportation Economics

by

Timothy Chong Ji Wong

Dissertation Committee:
Professor David Brownstone, Chair
Associate Professor Jiawei Chen
Professor Emeritus Kenneth Small

2015

DEDICATION

To

my parents and sisters

in recognition of their continuous support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my committee chair, Professor David Brownstone, whose depth of understanding of economics and statistics has my utmost respect. Without his mentorship and guidance, this dissertation would not have been possible. I also thank him for supporting me financially through multiple research assistantships. His generosity and dedication to my development is invaluable.

I thank Professor Kenneth Small for his guidance and advice. His expertise in transportation economics contributed significantly to this work. Every interaction with Professor Small has brought me a wealth of knowledge. Without his incredible attention to detail, many errors and concerns in my research would have gone unnoticed. Above all, his humility is his greatest quality.

I thank Professor Jiawei Chen for his guidance. His teaching provided me with a springboard to explore many ideas in economics. He also provided avenues for me to improve my teaching abilities and his feedback on my research significantly improved this dissertation. His positive influence is invaluable.

I thank Professors Dale Poirier, Ivan Jeliazkov and Yaming Yu informing me that I am Bayesian. Their teachings transformed my perspective of statistics and reality. I also thank Professors Jan Brueckner and Linda Cohen for their invaluable teaching and advice. Their classes exposed me to a world of ideas in Transportation Economics, Urban Economics and Industrial Organization. These perspectives are reflected heavily in my dissertation.

I thank my peers from the incoming Ph.D. Economics class of 2010 at the University of California, Irvine. I owe a greater amount of my training in Economics to them than to any other source. As a collective, they created a positive atmosphere for learning without which the work here would never have been completed. They showed great patience in their willingness to teach me. They also provided remarkable moral and emotional support through many professional and personal challenges.

Finally, I thank my parents, sisters and extended family for their relentless support.

I recognize funding provided by the University of California, Irvine through the Department of Economics, Institute of Transportation Studies, School of Social Sciences and Center for Economics and Public Policy. I also recognize funding from the University of California Center for Energy and Environmental Economics, the University of California Multi-Campus Research Program in Sustainable Transport: Technology, Mobility and Infrastructure, and the University of California Center on Economic Competitiveness in Transportation.

CURRICULUM VITAE

TIMOTHY WONG

Education

Doctor of Philosophy in Economics – Transportation Economics — June 2015
University of California, Irvine

Master of Arts in Economics — June 2012
University of California, Irvine

Bachelor of Science in Economics — May 2010
University of Central Missouri

Publication

Lights, Camera, Legal Action! The Effectiveness of Red Light Cameras on Collisions in Los Angeles. Transportation Research Part A: Policy & Practice 69 (2014): 165-182

Professional Experience

Graduate Research Assistant — 2012–2015
Institute of Transportation Studies, University of California, Irvine

Teaching Assistant — 2010- 2014
School of Social Sciences, University of California, Irvine

Honors and Awards

UCCONNECT Doctoral Dissertation Grant — 2015
University of California Center on Economic Competitiveness in Transportation

Kenneth Small Award: Best Ph.D. Student in Transportation and Urban Economics — 2013
University of California, Irvine

Economics Summer Research Fellowship — 2013, 2014
University of California Irvine

Graduate Student Paper Contest, Third Place — 2013
Transportation Research Forum

Research Support, Borowski Student Policy Research Initiative — 2011-2012
Center for Economics and Public Policy, University of California, Irvine

Social Science Tuition Fellowship — 2010-2011
University of California, Irvine

Outstanding Undergraduate in Economics — 2008, 2010
University of Central Missouri

ABSTRACT OF THE DISSERTATION

Econometric Models in Transportation

By

Timothy Chong Ji Wong

Doctor of Philosophy in Economics – Transportation Economics

University of California, Irvine, 2015

Professor David Brownstone, Chair

The three chapters in this dissertation study and apply econometric models to answer questions in transportation economics. Chapter 1 and 2 analyze the Berry, Levinsohn and Pakes (BLP) discrete choice model for combined micro- and macro-level data. Chapter 1 considers the concerns of choice set aggregation and estimating consistent standard errors within the BLP Model. These concerns are studied within the context of a vehicle choice application with interest in estimating household valuation of fuel efficiency. Chapter 2 studies the numerical properties of the maximum likelihood approach to estimating this BLP model. Chapter 3 applies a Poisson-Log Normal panel data model to study the effect of red light cameras on collision counts in Los Angeles. The camera program suffered from weaknesses in enforcement that dampened the effectiveness of the program over time. The model considered here controls for this dampening effect.

Chapter 1 finds that choice set aggregation affects the point estimates obtained from the BLP model and causes standard errors to be too small. The use of inconsistent sequential standard errors also underestimates the magnitude of standard errors. These findings may partly explain

the disparity across existing estimates from choice models on the value households place on vehicle fuel efficiency.

Chapter 2 finds that the maximum likelihood estimation approach is able to find the global minimum regardless of choice of starting values, optimization routine used and tightness of convergence criteria. These findings highlight the benefits of estimating the BLP model on combined micro- and macro-level datasets using the maximum likelihood approach compared to using the nested fixed point approach and only macro level data where numerical stability is difficult to obtain.

Chapter 3 finds that controlling for the dampening effect from poor enforcement, the Los Angeles Automated Red Light Camera program decreased red light running related collisions but increased right-angle and injury collisions, as well as collisions overall.

Chapter 1

Choice Set Aggregation and Consistent Standard Errors in BLP Models for Micro- and Macro-level Data

1.1 Introduction

Multinomial choice models have become popular in demand estimation because, unlike systems of demand equations, the number of parameters to be estimated is not a function of the number of products, removing the obstacle of estimating markets with many differentiated products. One challenge of choice modeling in applications is determining the level of detail at which the choice set is defined. Modeling choices at their finest level can quickly cause the resulting choice set to grow so large that it exceeds the practical capabilities of estimation. In addition, it is common in micro-level datasets that household choices are not observed at the finest level possible. In such cases, researchers often aggregate choices to the level of detail that is observed. For example, when modeling a household's vehicle choice decisions, the exact choice set may contain vehicles at the make-model-trim level. Assume households decide between four vehicles: the Honda Civic LX, Honda Civic EX, Toyota Corolla G, and Toyota Corolla X; however, the researcher only observes the household's choice between two broad make-model groups, the Honda Civic and the Toyota Corolla. Hence, the researcher aggregates the exact choice set to the make-model level. In this case, the researcher creates an "average"

Honda Civic whose characteristics consist of the average attributes of the Civic LX and Civic EX, such as horsepower and curb weight, and likewise for the Corolla G and Corolla X.

Despite these practical considerations, aggregation of the choice set to broader levels removes useful variation from the choice set and misspecifies the choices faced by households, which can diminish researchers' ability to understand key choice determinants. This paper studies the effect of choice set aggregation on estimates derived from choice models, applying this practice to a common policy concern regarding vehicle choice. These effects are studied within the context of a Berry, Levinsohn and Pakes (henceforth, BLP) model for micro- and macro-level data estimated with a maximum likelihood approach. This model bears strong similarities to the models applied in Goolsbee and Petrin, 2004; Chintagunta and Dubé, 2005; Train and Winston, 2007; Langer, 2014 and Whitefoot et al., 2014. Specifically, I discuss two model specifications within this BLP framework that help overcome this concern of aggregation: an aggregation correction by McFadden, 1978, and a model for broad choice data by Brownstone and Li, 2014. McFadden's aggregation correction places a distributional assumption on the elements within each aggregated choice and uses the higher moments of the distribution in the utility specification. Brownstone and Li, 2014, define the choice probability of a broad group as the sum of the choice probabilities of the elements within that group, leveraging on the existence of aggregate market share data at the exact choice level for identification.

These results are particularly meaningful for evaluating the literature on vehicle choice. Multinomial choice models are commonly applied to vehicle choice data with interest in estimating households' willingness to pay for fuel efficiency. These estimates are important because they help measure the extent that households do not value energy saving investments like vehicle fuel efficiency as standard economic theory on rational behavior would predict.

This energy paradox has significant consequences. If households consistently undervalue energy saving investments, this implies that households are incurring excess environmental costs, even if proper mechanisms (e.g. carbon taxes,) are in place to control for environmental externalities. Therefore, implementing government policies that increase the energy efficiency of products, (e.g. energy labelling programs, minimum energy efficiency standards) may be in the best interest of both households and the environment. Despite the important implications of understanding the energy paradox, the existing literature on willingness to pay for fuel efficiency is inconclusive (Greene, 2010; Helfand and Wolverton, 2011). One potential source of the disparity in estimates could be the common practice of aggregating the exact vehicle choice set to broader groups based on vehicle class (Goldberg, 1998; Bhat and Sen, 2005; Bento et al., 2009; Jacobsen, 2013).

I find that aggregation affects both the point estimates and standard errors of parameters from the model. Estimates of household willingness to pay for vehicle fuel efficiency are twice as large when aggregation is modeled using McFadden's method, and four times smaller when the fully disaggregated alternatives are modeled using the Brownstone and Li approach. The Brownstone and Li approach better suits the current application because it is harder to justify the necessary distributional assumptions of McFadden's approach in the current setting. Perhaps more importantly, model estimates are more likely to appear statistically significant when choices are aggregated without correction because doing so ignores the measurement error it introduces to the choice attributes. Estimates of household willingness to pay for vehicle fuel efficiency are significant in the model that aggregates choices but not significant in the other two models. If these results extend to existing vehicle choice studies that aggregate choices, it suggests that current disparities on the energy paradox may in part stem from overconfidence in model estimates because of a failure to account for noise induced by choice set aggregation.

I am aware of one other study that investigates the concern of choice set aggregation. Spiller, 2012, studies aggregation within a discrete-continuous model of household vehicle fleet choice and utilization. She compares estimates from a model that aggregates vehicles to make-class categories to a model that defines vehicles at the make-model level, finding that demand for gasoline is more inelastic when choices are aggregated to broader levels.

Three recent papers define vehicles at the finest levels of detail possible when estimating household willingness to pay for fuel efficiency, though none of these papers employ choice models to household-level data in their work. Busse et al., 2013 employ a reduced form approach to study the effect of fuel prices on vehicle transaction prices, market shares and sales for new and used vehicles of different fuel economies while Allcott and Wozny, 2014 and Grigolon et al., 2014 employ choice models to macro-level data. Busse et al., 2014 do not find evidence that households undervalue fuel efficiency. Grigolon et al., 2014, find only modest undervaluation of fuel costs, while Allcott and Wozny, 2014, find a wide range of values varying from undervaluation to rational valuation, depending on various model assumptions.

These findings on choice set aggregation are also relevant to many other empirical questions that choice models are used to study, including the effect of mergers (Nevo, 2000), welfare gains from new products (Petrin, 2002, Goolsbee and Petrin, 2004), measuring market power (Nevo, 2001), and explaining the declining market share of dominant firms (Train and Winston, 2007).

The BLP model for micro- and macro-level data is often estimated sequentially. In the first stage, a choice model is fit where product specific constants are estimated. These constants are then used as dependent variables in an instrumental variables (IV) regression in the second stage. The standard errors from this two-step process are inconsistent because they ignore 1) the constraint that estimated shares equal macro-level market shares and 2) the correlations that exist between estimates across the two stages. Though the model has been applied to data for more than a decade, consistent standard errors for both stages of the model have never been formally derived. Previous studies that implement this model either correct only the standard errors of first stage parameters (BLP,2004; Chintagunta and Dubé, 2005), use inconsistent standard errors, or never formally state how they obtain the standard errors of model estimates.

Thus, in this paper, I also derive consistent standard errors for the entire model by recasting each sequence of the model jointly within a Generalized Method of Moments (GMM) framework. I examine the performance of both the inconsistent and consistent standard errors through a Monte Carlo study and the vehicle choice application. I show that the use of standard errors derived from sequential estimation can cause errors when hypothesis testing. Second stage standard errors from sequential estimation are downward biased. I also find that the first stage standard errors are too small. In application, the standard errors derived from the GMM framework are larger than the sequential standard errors, generally by a factor of 1.3 to 13. However, the GMM standard errors for fuel costs are larger by a staggering 14 to 43 times. These findings also relate to the disparate results across existing choice models on the energy paradox. Studies that draw inference using inconsistent standard errors (BLP, 2004; Train and Winston, 2007; Whitefoot et al, 2014) may portray estimates as more significant than they really are.

Both the concern of choice set aggregation and the inconsistency of the standard errors from sequential estimation are independent of the random coefficients framework typically included in the BLP model. Therefore, I simplify the common BLP model by removing the random coefficient specification. Including random coefficients will result in slightly larger standard errors due to sampling noise from simulated maximum likelihood estimation, and, if correctly specified, also result in richer substitution patterns across products.

This paper is structured as follows. In Section 2, I present the BLP model for micro- and macro-level data. In Section 3, I detail the maximum likelihood approach to estimating the model, including the inconsistent and consistent methods of obtaining standard errors. In Section 4, I demonstrate through a Monte Carlo experiment the inconsistency of standard errors obtained through sequential estimation and show that the consistent standard errors provide the appropriate coverage probabilities of the true value in the limit. In Section 5, I present the McFadden, 1978, correction for aggregation bias and the Brownstone and Li, 2014, model for broad choice data, and adapt these models to the BLP setting at hand. In Section 6, I apply the consistent standard error formula and aggregation corrections to a vehicle choice application and discuss the results, relating them to the existing literature on the energy paradox. Section 7 summarizes and concludes.

1.2 The BLP Model

The seminal BLP choice model (BLP, 1995) contributes an important innovation to the choice modeling literature by accommodating the endogeneity of product attributes, since the econometrician rarely observes the full set of attributes of products that induce households to make purchases. In BLP, 1995, the model is applied using only macro-level data. BLP, 2004, extends the model to applications that combine both micro- and macro-level data. In this paper, BLP addresses an important concern in micro-level choice modeling, the issue of unrepresentative sampling. To overcome this concern, they supplement their unrepresentative choice dataset with aggregate market share data that is believed to be more representative.

The BLP model for micro- and macro-level data used here is similar to BLP, 2004. The key difference is that BLP, 2004, incorporates household information into their model by constructing a moment that captures the covariance of product attributes and household attributes. The model here incorporates household information through a standard multinomial likelihood function, as is done in Train and Winston, 2007. The model is as follows:

Let $n = 1, \dots, N$ index households which can either purchase any of J products, $j = 1, \dots, J$ in the market or not purchase any product, characterized by selecting the "outside good", $j = 0$. The indirect utility of household n from the choice of product j, U_{nj} is assumed to follow the following linear specification:

$$U_{nj} = \delta_j + w_{nj}'\beta + \epsilon_{nj},$$

$$n = 1, \dots, N, \quad j = 0, 1, \dots, J,$$

where δ_j is a product specific constant that captures the "average" utility of product j. In other words, it is, the portion of utility from product j that is the same for all households. w_{nj} is a $(K_1 \times 1)$ vector of household attributes interacted with product attributes, β is its $(K_1 \times 1)$ vector of associated parameters, and ϵ_{nj} is an error term with mean zero that captures all remaining elements of utility provided by product j to household n. For the purpose of identification, average utility of the "outside good," δ_0, is normalized to zero. Households select the product that yields them the highest utility:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\, i \neq j \\ 0 & otherwise. \end{cases}$$

Assuming that ϵ_{nj} follows the type I extreme value distribution, the probability that household n, chooses product j, P_{nj} is:

$$P_{nj} = \frac{exp\left(\delta_j + w_{nj}{}'\beta\right)}{\sum_k exp(\delta_k + w_{nk}{}'\beta)}. \qquad (1.1)$$

In the "traditional" conditional logit model, $\delta = \{\delta_1, \dots, \delta_J\}$ and β, can be estimated by maximizing the following log-likelihood function:

$$L(y; \delta, \beta) = \sum_n \sum_j y_{nj}\, log\ (P_{nj}). \qquad (1.2)$$

An interesting feature of maximum likelihood (ML) estimation of the multinomial logit is how δ is identified. The first order condition for δ is that the log-likelihood function with respect to δ equals zero:

$$\frac{\partial L}{\partial \delta_j} = \sum_n \sum_i (y_{ni} - P_{ni})\, d_i^j = 0, \qquad \text{for } j = 1, \dots, J,$$

$$where\ d_i^j = \begin{cases} 1 & if\ i = j \\ 0 & otherwise. \end{cases}$$

Rearranging and dividing both sides by N, we find that

$$\frac{1}{N}\sum_n y_{nj} = \frac{1}{N}\sum_n P_{nj}, \qquad \text{for } j = 1, \dots, J.$$

The equation describes that the average utilities are estimated such that the predicted shares from the model, $\widehat{S_J} = \frac{1}{N}\sum_n P_{nj}$, match the in-sample shares, that is, the share of households in the sample who choose each of the products. (Train, 2009).

A key innovation of BLP, 2004, is that δ is estimated such that the predicted shares match aggregate market shares rather than in-sample shares. For $j = 1, \ldots, J$, δ_j is chosen such that

$$A_j = \widehat{S_J}$$

where A_j is the aggregate market share for product j. Berry (1994) shows that for any value of β, a unique δ exists such that the predicted shares match these aggregate market shares.

Matching predicted shares to aggregate market shares rather than in-sample shares improves estimation in the event that there is high sampling variance and hence, unrepresentative sample shares. In addition, by using aggregate market shares, average utilities can be estimated for products even if the in-sample shares for these products are zero.

Finally, it is assumed that the average utilities are a linear function of product attributes:

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$

where x_j is a $(K_2 \times 1)$ vector of exogenous product attributes, α_1 is a $(K_2 \times 1)$ vector of associated parameters, p_j is a $(K_3 \times 1)$ vector of product attributes that are endogenous with respect to average utility, α_2 is a $(K_3 \times 1)$ vector of associated parameters, and ξ_{1j} captures the average utility associated with attributes unobserved to the econometrician. Because p_j is endogenous with respect to average utility, it is correlated with unobserved attributes contained in ξ_{1j}, such that $E(\xi_{1j}|p_j) \neq 0$. There exists a set of instruments, z_j, that are correlated with p_j and uncorrelated with ξ_{1j}:

$$p_j = z_j{}'\gamma + \xi_{2j}$$

$$where\ E(\xi_{1j}|z_j) = 0.$$

For simplicity, I assume here that z_j is a $(K_3 \times 1)$ vector, (and therefore, γ is a $K_3 \times 1$ vector of associated parameters), that is, there are as many instruments as there are endogenous regressors, making the model just-identified. Although over-identification does not tremendously complicate estimation, with a just-identified model, optimal GMM methods are not necessary, which simplifies estimation.

In summary, the BLP model consists of the following equations:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\, i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + \mathrm{w_{nj}}'\beta + \epsilon_{nj}, \quad \epsilon_{nj} \sim \text{type I extreme value}$$

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$

$$p_j = z_j{}'\gamma + \xi_{2j}$$

$$A_j = \frac{1}{N}\sum_n P_{nj}$$

In many empirical settings, all observed attributes of households are categorical. For example, income is often only observed in categories of income ranges, and education in categories by highest level attained. When these types of data are used, and the number of household attributes is small compared to the number of households in the sample, it is common that some households in the sample are observed to have identical attributes. As long as these households are still distributed independently, it is possible to collapse households with identical attributes into representative "household types" that make repeated, independent choices. The repeated choices are simply an aggregation of the decisions made by each household in the sample, of that type. Let m denote "household type", where $m = 1, \ldots, M$. Let r_m denote the number of households belonging to type m. It must be the case that $\sum_m r_m = N$. Finally, let l_m index households within a given household type. Then,

$$y_{mj} = \sum\nolimits_{l_m=1}^{r_m} y_{nj}\, , \forall\, j = 0, 1, \dots, J\, .$$

It is easy to see that the aggregation of households to "household types" changes the outcome variable, because $\sum_j y_{mj} = r_m$ while $\sum_j y_{nj} = 1$. Nevertheless, aggregation leaves the likelihood function of the multinomial logit unchanged. There is a computational benefit from estimating at the "household type" level, rather than the household level since, as long as $r_m > 1$ for any m, then $M < N$.

To remain consistent with the Monte Carlo experiment and empirical applications that are conducted, the remainder of this paper assumes estimation with aggregation to "household types."

1.3 Estimation Procedure: A Maximum Likelihood Approach

Sequential estimation is conducted in two stages. The first stage involves estimating the average utilities, δ, and the parameters associated with the household-product interaction variables, β. The first stage estimates of δ are then used in the second stage estimation of the parameters associated with the product attributes, $\alpha = [\alpha_1\ \alpha_2]$.

δ and β are estimated through an iterative process. Conditional on some initial value of δ, maximize equation (1.2) with respect to β to obtain conditional maximum likelihood estimates, $\hat{\beta}$. Conditional on $\hat{\beta}$, estimate δ using the contraction-mapping algorithm developed in BLP, 1995. This algorithm is itself an iterative process which yields the estimate, $\hat{\delta}$, when the following equation is iterated on until convergence:

$$\delta_{j,t+1} = \delta_{j,t} + \ln(A_j) - \ln(\widehat{S}_j), \quad \forall\, j = 1, \dots, J$$

The algorithm enforces the constraint that the predicted shares equal the aggregate market shares. The maximum likelihood and contraction mapping processes are repeated iteratively until convergence.

The second stage of the sequential process estimates the parameters associated with the product attributes, α. To do this, standard IV estimation is applied, substituting the converged values, $\widehat{\delta}$, from the first stage, for the "true values" of the average utilities, δ. Let $X = [x_1'\ x_2'\ \dots\ x_J']'$. Let Z and P be similarly defined matrices containing the element, z_j and p_j respectively. Then the IV estimates are given by the familiar solution:

$$\hat{\alpha} = \left(\tilde{X}'Z(Z'Z)^{-1}Z'\tilde{X}\right)^{-1}\tilde{X}'Z\left(Z'Z\right)^{-1}Z'\hat{\delta}$$

where $\tilde{X} = [X\ \ P]$.

The standard errors for $\hat{\alpha}$ from IV estimation are downward biased. Standard IV estimation assumes that the dependent variable (in this case, δ) is known; however, in the BLP model, the dependent variable, $\hat{\delta}$, is an estimate. The standard errors for $\hat{\alpha}$ from IV estimation do not account for the variance inherent in $\hat{\delta}$ causing the standard errors of $\hat{\alpha}$ to be underestimated.

Additionally, in application, some researchers obtain the standard errors for $\hat{\beta}$ and $\hat{\delta}$ from the inverse of the Hessian of the logit likelihood function. However, $\hat{\beta}$ and $\hat{\delta}$ are not maximum likelihood estimates, since first stage estimation constrains predicted shares to match aggregate market shares, not in-sample shares. Unless the aggregate market shares equal in-sample shares, standard errors for $\hat{\beta}$ and $\hat{\delta}$ obtained in this manner are also inconsistent.

To obtain estimates of the correct standard errors for $\hat{\delta}, \hat{\beta}$ and $\hat{\alpha}$, recast each sequence of the estimation process as moments within a GMM framework. The GMM analogue to the sequential process just described involves three sample moments, one for each of the vectors of parameters δ, β and α. The first moment condition is formed from the first derivative of the logit log-likelihood function, with respect to β[1]:

$$G_1(\beta,\delta) = \frac{1}{N}\sum\nolimits_m H_{1m}(\beta,\delta).$$

$$= \frac{1}{N}\sum\nolimits_m \sum\nolimits_j y_{mj}(w_{mj} - \sum\nolimits_i P_{mi} w_{mi}) \,.$$

The second moment condition, that identifies $\hat{\delta}$, is formed from a vector that constrains the predicted shares of the products in the model to match the aggregate market shares when minimized[2]:

$$G_2(\beta,\delta) = \frac{1}{N}\sum\nolimits_m H_{2m}(\beta,\delta)$$

$$= \frac{1}{N}\sum\nolimits_m A - P_m.$$

where $A = [A_1\ A_2 \ldots A_J]'$ and $P_m = [P_{m1}\ P_{m2} \ldots P_{mJ}]'$

The third moment condition estimates $\hat{\alpha}$. It is formed from a vector that when minimized, stipulates that in expectation, the instruments for the attributes of the products are uncorrelated with the error term:

[1] To obtain the BLP, 2004 estimator of the model, replace this moment with one that interacts the average attributes of households with the attributes of vehicles they purchase, then averages over all vehicles in the choice set.
[2] Note that although households are aggregated to "household types" (as indicated by the summations across m, rather than n), expectations are still taken at the household level and not at the household-type level, that is, the first and second moments are averaged over N and not M.

$$G_3(\delta, \alpha) = \frac{1}{J}\sum\nolimits_j H_{3j}(\delta, \alpha)$$

$$= \frac{1}{J}\sum\nolimits_j z_j\left(\delta_j - x_j\alpha_1 - p_j\alpha_2\right).$$

One can also obtain point estimates of the BLP model by minimizing the objective function $Q(\theta) = G^{'}W_0G$ where $G = [G_1\ G_2\ G_3]^{'}$, $\theta = [\beta\ \delta\ \alpha]'$, and W_0 is a weight matrix. When the model is just-identified, this solution, $\hat{\theta}_{GMM}$, is equivalent to the solution from sequential ML estimation, $\{\hat{\beta}, \hat{\delta}, \hat{\alpha}\}$ because it satisfies the same estimation conditions.

In line with derivations by Hansen, 1982, the variance of this GMM estimator, θ_{GMM}, is given by the following formula:

$$Var(\theta_{GMM}) = \left(M_0^{'}W_0M_0\right)^{-1}\left(M_0^{'}W_0S_0W_0M_0\right)\left(M_0^{'}W_0M_0\right)^{-1}, \tag{1.3}$$

where

$$M_0 = \begin{bmatrix} \frac{1}{N}\frac{\partial^2 L}{\partial \beta^2} & \frac{1}{N}\frac{\partial^2 L}{\partial \beta \partial \delta} & 0_{K_1 \times K_2} \\ \frac{1}{N}\sum\nolimits_m -\frac{\partial P_{mj}}{\partial \beta} & \frac{1}{N}\sum\nolimits_m -\frac{\partial P_{mj}}{\partial \delta} & 0_{(J-1)\times K_2} \\ 0_{K_2 \times K_1} & \frac{1}{\sqrt{NJ}}\sum\nolimits_j z_j{}' & \frac{1}{J}\sum\nolimits_j -z_j{}'x_j \end{bmatrix}$$

$$S_0 = \begin{bmatrix} \frac{1}{N}\sum\nolimits_m H_{1m}H_{1m}{}' & \frac{1}{N}\sum\nolimits_m H_{1m}H_{2m}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum\nolimits_m\sum\nolimits_j H_{1m}H_{3j}{}' \\ \frac{1}{N}\sum\nolimits_m H_{2m}H_{1m}{}' & \frac{1}{N}\sum\nolimits_m H_{2m}H_{2m}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum\nolimits_m\sum\nolimits_j H_{2m}H_{3j}{}' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum\nolimits_m\sum\nolimits_j H_{3j}H_{1nm}{}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum\nolimits_m\sum\nolimits_j H_{3j}H_{2m}{}' & \frac{1}{J}\sum\nolimits_j H_{3j}H_{3j}{}' \end{bmatrix}.$$

Since the current model is just-identified, it is efficient to set W_0 as the identity matrix. In over-identified cases $(K_3 > K_2)$, the two-step optimal GMM method provides more efficient estimates. Appendix A.1 provides a more detailed derivation and explanation of these standard errors.

Since $\frac{\partial H1}{\partial \alpha}, \frac{\partial H2}{\partial \alpha} = 0$, this GMM model is a sequential two-step estimator (Newey, 1984). Using the derivations by Murphy and Topel (1985), the following equation provides the correction for the downward bias present in the standard errors of IV estimates from the second stage of estimation:

$$Var(\hat{\alpha}) = s^2(X'Z(Z'Z)^{-1}Z'X) + G_{22}{}^{-1}\{G_{21}G_{11}^{-1}S_{11}G_{11}^{-1}G_{21} - G_{21}G_{11}^{-1}S_{12} - S_{21}G_{11}^{-1}G_{21}'\}G_{22}{}^{-1}$$

where $G_{11} = \begin{bmatrix} \frac{\partial H_1}{\partial \beta} & \frac{\partial H_1}{\partial \delta} \\ \frac{\partial H_2}{\partial \beta} & \frac{\partial H_2}{\partial \delta} \end{bmatrix}, G_{21} = \begin{bmatrix} \frac{\partial H_3}{\partial \beta} & \frac{\partial H_3}{\partial \delta} \end{bmatrix}, G_{22} = \begin{bmatrix} \frac{\partial H_3}{\partial \alpha} \end{bmatrix},\ S_{11} = \begin{bmatrix} \frac{1}{N}\sum_m H_{1m}H_{1m}{}' & \frac{1}{N}\sum_m H_{1m}H_{2m}{}' \\ \frac{1}{N}\sum_m H_{2m}H_{1m}{}' & \frac{1}{N}\sum_m H_{2m}H_{2m}{}' \end{bmatrix},$

The first term is the standard IV formula for standard errors. The second term is an upward correction to account for the first stage. This second term has three components. The first, containing the covariance matrix of the first stage parameters, $G_{11}^{-1}S_{11}G_{11}^{-1}$, accounts for the variance of δ, while the second and third components account for the correlation between the errors across the two stages.

1.3.1 Speeding up the Contraction Mapping Algorithm

The contraction-mapping algorithm can incur a high time cost particularly when tight stopping criteria are used, and the dimensions of the choice set are large. To speed up the contraction-mapping algorithm, I implement the Li, 2012 modification that augments the contraction mapping with an analytic Newton-Raphson algorithm. The modification produces a 280-fold improvement in the number of iterations necessary for convergence over the unmodified contraction mapping algorithm and a six-fold improvement in estimation time when using simulated data sets. The modification is as follows:

$$\delta_{j,t+1} = \delta_{j,t} + \mathrm{H}^{-1}[\ln(A_j) - \ln(\widehat{S}_j)]\ \forall j = 1,2\ \dots\ J,$$

where H is the matrix of first order partial derivatives of $[-\ln(\hat{S})]$, which can be shown to equal

$$H = \begin{bmatrix} 1 - \frac{\sum_n P_{n1}^2}{\sum_n P_{n1}} & -\frac{\sum_n P_{n1}P_{n2}}{\sum_n P_{n1}} & \cdots & -\frac{\sum_n P_{n1}P_{nJ}}{\sum_n P_{n1}} \\ -\frac{\sum_n P_{n2}P_{n1}}{\sum_n P_{n2}} & 1 - \frac{\sum_n P_{n2}^2}{\sum_n P_{n2}} & \cdots & -\frac{\sum_n P_{n2}P_{nJ}}{\sum_n P_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\sum_n P_{nJ}P_{n1}}{\sum_n P_{nJ}} & -\frac{\sum_n P_{nJ}P_{n2}}{\sum_n P_{nJ}} & \cdots & 1 - \frac{\sum_n P_{nJ}^2}{\sum_n P_{nJ}} \end{bmatrix}$$

One of the drawbacks of the Newton-Raphson algorithm is that in some situations, the method fails to converge. To avoid this problem from occurring in the empirical example, I only invoke the Newton modification when $\delta_{t+1} - \delta_t \leq 10^{-3}$. Using this ad-hoc rule, the contraction mapping algorithm does converge each time it is called in the empirical example, although the time savings are now only four-fold.

Recent work shows certain approaches to estimating the BLP models on macro-level data behave poorly. Nevo, 2000, Dubé et al., 2012, and Knittel and Metaxoglou, 2014, among others, note that the model is sensitive to starting values, requires very tight convergence criteria, can have multiple local minima, may falsely stop at points that are not even local minima, and is sensitive to the choice of optimization routine used. In light of these findings, in the following chapter of this dissertation, I study these concerns within the BLP model for combined micro- and macro-level data, finding that in these data settings, the ML approach does not suffer from such numerical concerns.

1.4 A Monte Carlo Experiment on the Consistency of Standard Errors

The following Monte Carlo experiment analyzes the properties of the standard errors from sequential ML estimation and the analytic GMM standard errors. The experiment proceeds as follows: Exogenous variables, $\{w_{mj}, x_j, z_j\}$ are generated for a large population of household types, N_{pop}, and a fixed number of products, J, For each generated household type, I create r_m households. For simplicity, an equal number of households are generated for each type, $r_m = r$, $\forall m$. More detailed information on the data generation process is available in Appendix A.2. Each iteration of the experiment then follows these steps:

1. New draws of the error terms, $\{\epsilon_{mj}, \xi_{1j}, \xi_{2j}\}$ are generated and with that, the endogenous variables, $\{y_{mj}, \delta_j, p_j\}$ are created as functions of the exogenous variables, error terms and the predetermined values of parameters, $\{\alpha, \beta, \gamma\}$.
2. Using the vector of selected products, $\{y_{mj}\}$, population market shares, A_j, are created.
3. A random subsample of household types, M, is selected from N_{pop}. The number of households in a given subsample, $N = \sum_m r_m = M \times r$
4. Estimation proceeds as outlined in the previous section; the model is estimated sequentially using the iterative ML and contraction mapping algorithm process in the first stage, and IV estimation in the second stage.
5. To obtain standard errors from the sequential process, the standard errors of β are calculated as the inverse of the negative Hessian matrix of the log-likelihood function, and the standard errors of α follow the standard IV variance formula. To obtain the corrected standard errors, I use the GMM variance formula in equation (1.3).

Table 1.1 displays the coverage probabilities of 80% confidence intervals constructed for β and α. These coverage probabilities were obtained from running 1000 Monte Carlo repetitions of the experiment for various sizes of N, where $J = 30, N_{pop} = 100{,}000$, and $r = 100$. The choice of $J = 30$ is sufficient to obtain asymptotic results in that dimension.

Table 1.1: Coverage probabilities of 80% confidence intervals for β and α, for $N = 2500$, 1000 and 60000, where $J = 30, N_{pop} = 100{,}000$, and $r = 100$.

Parameter	$N = 2{,}500$		$N = 10{,}000$		$N = 60{,}000$	
	Sequential	Joint	Sequential	Joint	Sequential	Joint
$\widehat{\beta_1}$	0.390	0.907	0.371	0.839	0.382	0.807
$\widehat{\beta_2}$	0.606	0.883	0.672	0.806	0.700	0.805
$\widehat{\alpha_0}$	0.789	0.813	0.791	0.796	0.810	0.810
$\widehat{\alpha_{11}}$	0.747	0.797	0.794	0.806	0.806	0.806
$\widehat{\alpha_{12}}$	0.597	0.858	0.746	0.805	0.781	0.797
$\widehat{\alpha_2}$	0.807	0.809	0.829	0.827	0.802	0.802

For each value of N, I report two sets of coverage probabilities, the "Sequential" column reports coverage probabilities calculated using standard errors from sequential estimation of the model. The "Joint" column reports coverage probabilities calculated using the analytic standard errors derived from equation (1.3).

As expected, Table 1.1 shows that the coverage probabilities for β from the sequential model converge to the wrong values. Confidence intervals for $\beta_1$and β_2 that should capture the true value of the parameter 80% of the time across repeated samples, only capture the true value 37-39% and 60-70% of the time respectively. While the sequential standard errors of β are too small in this study, and more severely so for β_1 than β_2, this result is specific to this Monte Carlo study. The sign and severity of the inconsistency is indeterminate across applications. As previously mentioned, the inconsistency of the sequential estimator stems from the fact that standard errors are derived from a likelihood function that does not contain the same information as was used to obtain the point estimates. Since the average utilities are estimated by matching predicted probabilities to aggregate shares, it is incorrect to obtain standard errors as if the predicted probabilities were matched to in-sample shares.

Confidence intervals for $\hat{\beta}$ constructed from the corrected standard errors perform much better. When $N = 2{,}500$, it can be inferred from the coverage probabilities that there is upward bias in the size of confidence intervals (and in turn, the standard errors) in small household samples. Nevertheless, as N increases to $60{,}000$, these coverage probabilities converge to 80%.

As theory predicts, the uncorrected standard errors of $\hat{\alpha}$ from sequential estimation are downward biased. This bias is more pronounced when $N = 2{,}500$, than when $N = 60{,}000$. In fact, when $N = 60{,}000$, the improvements from joint estimation are meager. The downward bias occurs because the standard IV standard errors do not account for the variance present in $\hat{\delta}$. Therefore, the larger the variances of the average utilities, $\hat{\delta}$, the larger the bias in the standard errors from sequential estimation. Since estimates of the average utilities are less precise when the household sample size is small, there are larger downward biases when N is small. As N increases, the variance of the estimates of $\hat{\delta}$ shrink, causing the bias of the uncorrected standard errors of $\hat{\alpha}$ to shrink as well. The standard errors from the joint GMM model, which do account for the variances of $\hat{\delta}$, have the coverage probabilities of approximately 80% across all three values of N considered.

Figure 1.1 displays box plots of the coverage probabilities of the 80% confidence intervals for the 29 δ_j's from the same Monte Carlo experiment. The left side presents the coverage probabilities using standard errors from the sequential model, while the right side presents the coverage probabilities using the corrected standard errors from the joint model.

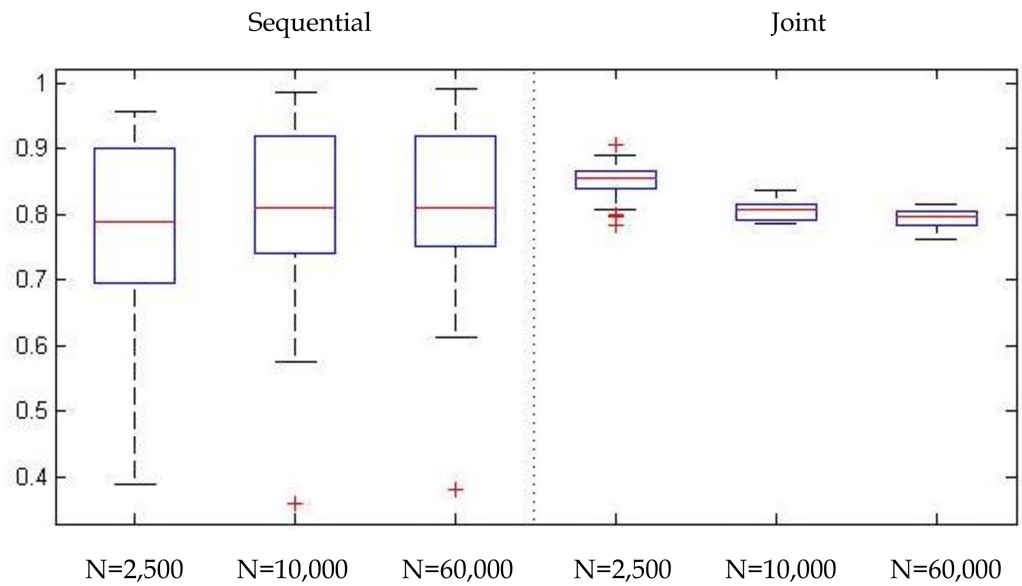


Figure 1: Box plots of the coverage probabilities of 80% confidence for the "average" utilities estimated when $M = 25, 100$ and 600, where $J = 30, N_{pop} = 100{,}000$, and $r = 100$.

The three box plots on the left side of the figure show there is large variation in the coverage probabilities of the $\delta_j's$, though this variation narrows as N increases. The inability to obtain coverage probabilities of the $\delta_j's$ that center around 0.8 reflect the inconsistency of using maximum likelihood derivations to calculate the standard errors of estimates that are not obtained from maximum likelihood estimation.

The three box plots on the right side of the figure suggest that there is an upward bias in the standard errors for many of the δ_j's in small samples. When $N = 2{,}500$, even the first quartile value is greater than 80%, and the distribution shows several outliers. Nevertheless, this bias diminishes quickly, as sample sizes increases. As shown in the box plots for $N = 10{,}000$ and $N = 60{,}000$, the coverage probabilities for the 80% confidence intervals for all of the δ_j's center tightly around 80%.

To increase efficiency of the second stage estimates, Goolsbee and Petrin, 2004, and Chintagunta and Dubé, 2005, weigh the second stage observations by the inverse of the variance of δ but use the sequential approach to calculate the standard errors.[3] Running a Monte Carlo study on this method, I find that the second stage standard errors behave similarly to the sequential standard errors in Table 1.1 at best. In many Monte Carlo runs, second stage standard errors were more downward biased when weights were used.

1.5 Methods of Aggregation

In this section, I introduce two methods that address the concerns related to aggregation of products to broad levels. The first model, introduced in McFadden, 1978, proposes that the covariance matrix of the attributes within each broad group and the number of products within each broad group be included in the likelihood function of the choice model. The second model is a model for broad choice data, introduced in Brownstone and Li, 2014. In their model, equation (1.2) is defined in terms of the broad choice sets from which household choices are observed and the broad choice probabilities are defined as the sum of the probabilities of the exact choices contained within each broad choice. Before introducing the two models, it is necessary to formally define broad choice data.

[3] Goolsbee and Petrin, 2004, do not incorporate macro-level share data in their model, hence their standard errors of δ obtained from the Hessian of the log-likelihood function are consistent. Chintagunta and Dubé, 2005, obtain standard errors of δ through a parametric bootstrap.

Define C as the exact choice set that contains all products, $j = 1, 2, \dots, J$. C is decomposed into B groups, denoted $C_b, b = 1, 2, \dots, B$. Each product, j, belongs to only one choice group such that, $C = \bigcup_{b=1}^{B} C_b$ and $\bigcap_{b=1}^{B} C_j = \emptyset$. Individuals' exact choices, y_{nj}, are not observed. Instead, what is observed are individuals' choices among the broad choice groups, Y_{nb}:

$$Y_{nb} = \begin{cases} 1 \; if \; y_{nj} \in C_b \\ 0 \; otherwise. \end{cases}$$

1.5.1 McFadden's Method for Aggregation

The empirical application concerned in McFadden, 1978, is the modeling of household choice of residential location. Here, the broad choice groups are communities where households are known to reside while the exact choice set contains the dwellings within these communities.

Let w_{nj} denote the observed attributes of household n interacted with the attributes of dwelling j. When the number of dwellings within a community is large, and w_{nj} behaves as if it is independently identically normally distributed with mean, w_{nj}^*, and variance Ω_{nb}, then, McFadden, 1978, shows that for the conditional logit with linear utility specification, the probability that household n chooses community b converges to:

$$\bar{P}_{nb} = \frac{\exp(\delta_b + w_{nb}^{*}{}' \beta + \frac{1}{2} \beta' \Omega_{nb} \beta + \log(D_b))}{\sum_k \exp(\delta_k + w_{nk}^{*}{}' \beta + \frac{1}{2} \beta' \Omega_{nk} \beta + \log(D_k))} \quad (1.4)$$

where D_b is the number of dwellings in community b.

The presence of the term, $\frac{1}{2}\beta'\Omega_{nb}\beta$, in (1.4) comes from the fact that the sample mean and sample sum of squared errors are sufficient statistics for the normal distribution with unknown mean and variance. To account for the distribution of characteristics of products within group

b, it is necessary to condition on both quantities. The intuition here is that community attributes with larger variances should have a greater impact on the probability that the community is selected. A simpler approach to incorporate Ω_{nb}, is to relax the constraint that its associated parameter, β, is equal to the parameter on w_{nb}^*. This approach yields consistent estimates without the complexity of non-linear constraints.

Lerman, 1977, explains that the $\log(D_b)$ term is a measure of community size. "Other conditions being equal, a large tract (i.e., one with a large number of housing units) would have a higher probability of being selected than a very small one, since the number of disaggregate opportunities is greater in the former than the latter." Here, the coefficient associated with $\log(D_b)$ is assumed to be one, because it is assumed that the logit model applies to each product in the exact choice set. Should this assumption not hold, then the coefficient on $\log(D_b)$ will differ from one.

The simplest way to apply McFadden's method to address aggregation concerns within the BLP setting, is to apply the likelihood function as in equation (1.2), to household choice of communities, replacing the choice probability defined in equation (1.1) with equation (1.4).

One limitation of this method is that it does not controls for the aggregation present in the product specific constants, δ_b. Recall that in the BLP model, the product specific constants are a function of product attributes. Using McFadden's method in this model, it is assumed that $\delta_b$is a linear combination of x_b^* and p_b^*, the means of x_j and p_j. To apply McFadden's method to account for the aggregation in x_b^* and p_b^*, one would have to include the covariance matrix of x_b and p_b in P_{nb}. Since P_{nb} would then be a function of both β and α, the sequential approach to estimating the BLP model would no longer be valid. Because of the estimation complexities this introduces, in this paper, I do not control for aggregation in x_b^* and p_b^*. Additionally, while McFadden's method applies easily to the conditional logit and nested logit, it is not immediately intuitive how the method can be applied to frameworks with more flexible substitution patterns, such as the mixed logit.

Nevertheless, the BLP model with McFadden's method has its advantages. The model is relatively easy to compute, particularly if the non-linear constraints are ignored, and unlike the proceeding model, does not require aggregate market share data at the exact choice level for identification.

1.5.2 A BLP Model for Broad Choice Data

When the researcher observes individuals' choices among broad choice groups, but macro-level market share data is available at the exact choice level, Brownstone and Li propose estimating household choice at the broad group level, defining the probability of choosing a broad group as the sum of the probabilities of the exact choices contained within the group. This involves replacing the likelihood function in equation (1.2) with the following:

$$L(y; \delta, \beta) = \sum_n \sum_b Y_{nb} \, log \ (\tilde{P}_{nb}) \tag{1.5}$$

where $\tilde{P}_{nb} = \sum_{j \in C_b} P_{nj}$ and P_{nj} is defined as in equation (1.1).

Estimation of this model follows the maximum likelihood approach detailed in Section 3. The contraction mapping algorithm is used to estimate δ_j, while β is estimated by maximizing equation (1.5).

Brownstone and Li, 2014, show that equation (1.5) is not globally concave and generally, has less concavity than equation (1.2). They also show that the parameters of the model, estimated on exact choice data, are better identified and have smaller variances than estimates from the model for broad choice, given the added uncertainty in the model for broad choice, stemming from having only partial information on household decisions.

There are numerous advantages to estimating the broad choice model over McFadden's method. It avoids aggregation altogether and preserves the sequential estimation process. Also, the model for broad choice does not require that asymptotic distributional assumptions be placed on the variables of the exact choice set, X_{nd}, within each broad group as is the case with McFadden's method. There may not always be an intuitive way to partition the exact choice set into groups that are all large to best approximate the asymptotic normality assumption required for consistency when using McFadden's method. Additionally, the model for broad choice produces estimates of product specific constants for each product in the exact choice set, rather than product specific constants at the broad group level. This allows for more observations in the second stage regression, providing more variation and power to the estimates in that stage. This addresses a concern of these types of models, raised in BLP, 2004. Because these models are typically estimated on single cross-sections of data, there are often not enough observations in the second stage to obtain precise estimates of β.

However, the model for broad choice is poorly identified without macro-level market share data at the exact choice level. Additionally, the model for broad choice can be computationally burdensome to estimate because of the higher dimensions of δ. Finally, given that there are many more first stage parameters to be estimated in this model, larger household sample sizes are required to obtain significant results. Because these first stage parameters are used in the second stage, this can also affect the efficiency of the second stage estimates as well.

1.6 A Vehicle Choice Application

Standard economic theory dictates that households should invest in an energy saving opportunity if the upfront cost of investment is lower than the present value of future savings from decreased energy bills. However, analysts of the energy industry have long debated the existence of an energy paradox, that households undervalue energy saving investments, hence underinvest in such technologies and over-consume energy. Explanations for this paradox include market failures such as imperfect information, principle-agent issues, credit constraints, and non-rational behavior such as loss aversion and hyperbolic discounting (Gillingham and Palmer, 2014). If the paradox exists, resulting market inefficiencies mean that many households in the state are spending more on fuel than they would if market failures did not exist. Additionally, the paradox suggests that private vehicle travel incurs excess environmental costs, even if proper mechanisms (e.g. carbon taxes) are in place to control for environmental externalities from vehicle emissions. This would mean there is a role for government intervention to correct the market failures through instruments that encourage increased household investments in fuel efficiency technologies such as rebates, taxes, fuel efficiency information programs and mandated fuel efficiency standards.

The energy paradox remains contested, in part because existing estimates of household valuations of energy saving investments are varied and inconclusive. Some of this disagreement comes from studies on vehicle choice (Greene, 2010). In this section, I apply the standard error corrections and methods for aggregation from the previous sections to a vehicle choice application in effort to explain some of this lack of consensus on whether the energy paradox exists.

I model the choices of households in the United States who purchase new model year 2008 vehicles between October 2007 and September 2008. As in Train and Winston, 2007, this model omits households that buy used vehicles or do not make any vehicle purchase over the sample period. Train and Winston, 2007, contend that "preferences among new car buyers can be estimated more accurately by estimating directly on a sample of new car buyers," and has the "practical advantage that it can include explanatory variables whose distributions are not known for the general population."

Data are obtained from Brownstone and Bunch, 2013. They compile data on households and their purchase decisions from the 2009 National Household Transportation Survey (NHTS). The NHTS sample is not a simple random sample. Households in 20 regions were oversampled because metropolitan transportation planning organizations in those regions sponsored larger samples for their own use. In addition, a single interview was conducted for each household between April 2008 and May 2009. Households who were interviewed earlier are more likely to have purchased model year 2008 vehicles after their NHTS interview, and this purchase is not reflected in the sample.

Multinomial choice models still yield consistent estimates of model parameters despite the use of stratified samples as long as all household heterogeneity is fully captured in the model specification through w_{nj} (Manski and Lerman, 1977). However, to obtain population averaged estimates from this model, one must place weights that correct sample stratification.

If household heterogeneity is not fully specified in the model, then estimates of α will not yield the average effect of vehicle attributes on utility. For example, assume that wealthy households are less sensitive to vehicle price than the rest of the population and that this heterogeneity is not accounted for in the model. Then, if wealthy households are over-represented in the sample, the estimates of the average effect of household disutility from vehicle price will be too small.

Alternatively, if the researcher has information about the distribution of characteristics in the population, he could place weights on the observations such that the weighted observations are representative of the population. One can incorporate such weights in estimation through the Weighted Exogenous Sampling Maximum Likelihood Estimator (WESML) by Manski and Lerman, 1977. WESML was developed to address choice-based sampling concerns (where the sample is stratified based on the observed choices, y_{nj}) but also applies when one is interested in estimating "average effects" without a full specification of household heterogeneity.

The results here assume that w_{nj} sufficiently captures household heterogeneity such that estimates from the model are consistent for α. To test the robustness of this assumption, I also estimate the BLP model for broad choice data using sampling weights. These results are presented in Appendix A.3.

Vehicle attributes are provided by the Volpe Center and supplemented with data from Polk, the American Fleet Magazine, and the National Automobile Dealers Association. Vehicle price data are adjusted adding the gas guzzler tax for some vehicles and subtracting estimated purchase subsidies for hybrid vehicles. Vehicle attribute data are available at the trim level, however, NHTS household choices are only observed at the Make/Model/Fuel-type level.

Macro-level market share information at the Make/Model/Fuel-type/trim[4] level is also obtained from the Volpe Center. They collect information on production volumes which

[4] "Make" refers to the manufacturer of the vehicle (e.g. Ford, General Motors, Toyota.) "Model" refers to the product name (e.g. Focus, Chevrolet Impala, Prius.) "Fuel-type" refers to the power source to move the vehicle (gasoline, natural gas, gasoline-electric hybrid). "Trim" denotes different configurations of

represent all model year 2008 vehicles that are produced (and eventually sold.) These production volumes are adjusted to omit fleet vehicle purchases.

There are 10,500 NHTS households in the dataset who purchase at least one new model year 2008 vehicle during the sample period. All household characteristic variables are categorical in nature. Because of this I aggregate to 4157 unique "household types," with between one and forty-one households within each type. There are 235 broad groups of vehicles that households choose from, and 1120 vehicles in the exact choice set.

Table 1.2 provides some descriptive statistics about the NHTS household sample as well as a comparison of NHTS in-sample vehicle shares to the Volpe aggregate market shares. Table 1.3 summarizes the utility specification that is used in the models.

---

standard equipment and amenities for a given vehicle make and model, such as manual or automatic transmission, fabric of leather seats, and number of engine cylinders (e.g. Honda Civic DX, Honda Civic LX.)

Table 1.2: Descriptive Statistics of the NHTS sample and market shares

NHTS Socioeconomic Attribute Variables	Sample Value (%)	
Percent retired with no children	34.17	
Percent whose children is under the age of 15	26.89	
Percent living in urban areas	68.07	
Percent of household respondents with college degree	48.12	
Average gasoline price at time of vehicle purchase ($)	3.46	
Household Income Distribution†:		
Less than $25,000	5.98	
$25,000 - $75,000	35.36	
$75,000 - $100,000	16.62	
Greater than $100,000	35.28	
Income Missing	6.76	
Household Size Distribution		
1	10.96	
2	49.31	
3	17.14	
4+	22.58	
Market share of MY2008 vehicle purchases by Manufacturer (NHTS household in-sample shares vs. Volpe aggregate market shares)	Share (%) NHTS	Volpe
General Motors	21.76	20.76
Toyota	18.82	18.98
Honda	15.53	13.45
Ford	13.87	12.05
Other Japanese	8.87	7.58
Chrysler	8.53	9.79
European	6.46	7.59
Korean	4.30	5.06

†Although five household income categories are observed, I use only four in the empirical application. I combine the lowest two categories into one for purposes of identification as I find the results for the two categories are very similar.

Table 1.3: Vehicle attributes, x_j, and vehicle-household attribute interactions, w_{nj}, included in the estimated model

x_j	w_{nj}
Price	(Price) × (75,000<Income<100,000)
Horsepower/Curb Weight	(Price) × (Income>100,000)
Hybrid	(Price) × (Income Missing)
Curb Weight	(Prestige) × (Urban)
Wagon	(Prestige) × (Income>100,000)
Mid-Large Car	(Performance Car) × (Income>100,000)
Performance Car	(Japan) × (Urban)
Small-Medium Pickup	(Van) × (Children under 15)
Large Pickup	(Large SUV) × (Children under 15)
Small-Medium SUV	(Small SUV) × (Children under 15)
Large SUV	(Korea) × (Rural)
	(Seats≥5) × (Household Size≥4)
	(Mid-Large Car) × (Retired)
	(Prestige) × (Retired)
	(Import) × (College)
	(Prestige) × (Japan) × (College)
	(Prestige) × (Europe) × (College)
	(Prestige) × (Japan) × (Urban)
	(Performance Car) × (College)
	Fuel Operating Cost (cents per mile)
	(Fuel Operating Cost) × (College)

Note: Fuel operating cost is the product of gallons per mile and fuel price (in cents per mile)
"Korea," "Japan," and "Europe" are dummy variables that equal 1 if the vehicle is made in that region and 0 otherwise.
"Prestige" is a dummy variable that equals 1 if the vehicle is classified as a "prestige brand" by the American Fleet Magazine.
The following vehicle classes were adopted from the American Fleet Magazine: Mid-Large Car, Performance Car, Small-Medium Pickup, Large Pickup, Small-Medium SUV and Large SUV.

I consider three specifications of the BLP model for micro- and macro-level data: a model that aggregates to the Make-Model-Fuel type level (BLP with aggregated choices), the BLP with McFadden's method for aggregation, where data is also grouped at the Make-Model-Fuel type level but the number of choices and co-variance matrix of attributes within each group are utilized in estimation, and the Brownstone and Li, 2014, BLP model for broad choice data.

First, I present the results of the standard error corrections for the BLP model with aggregated choices in Table 1.4. I focus this discussion on the price and fuel operating cost estimates. Results for the full model are listed in Appendix A.4. Table 1.4 shows that the uncorrected standard errors from sequential estimation are smaller than the corrected standard errors. These findings are consistent with the findings from the Monte Carlo study in Table 1.1. Hence, a likely explanation for this is that the former are biased downward. Most corrected standard errors are larger by a factor of about 2 to 3. The corrected standard error for fuel operating cost, however, is 18 times larger. This is a concern given the importance of this variables in constructing estimates of how much households value fuel efficiency improvements.

Table 1.4: BLP with Aggregated Choices: The sequential and GMM standard errors for select parameters.

Variable	BLP with Aggregated Choices					
	Estimated Parameter	Uncorrected Standard Error		Corrected Standard Error		Ratio of Corrected to Uncorrected Standard Errors
(Price) × (75,000<Income<100,000)	0.065	0.004	***	0.014	***	3.067
(Price) × (Income>100,000)	0.102	0.004	***	0.015	***	3.556
(Price) × (Income Missing)	0.094	0.005	***	0.015	***	3.140
Fuel Operating Cost (cents per mile)	-2.877	0.053	***	0.953	***	18.064
(Fuel Operating Cost) × (College)	-0.061	0.009	***	0.020	***	2.225
Price	-0.116	0.019	***	0.026	***	1.368

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table 1.5 presents select point estimates and standard errors, both corrected and uncorrected, for the BLP model for broad choice data. Full results are in Appendix A.4. The ratio of corrected to uncorrected standard errors display similar behavior as in Table 1.4. Again, the greatest correction occurs for the fuel operating cost variable, and the ratios are larger across the board compared to Table 1.4.

Table 1.5: BLP with Broad Choice Data: Sequential and GMM standard errors for select parameters.

Variable	BLP Model for Broad Choice Data			
	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error	Ratio of Corrected to Uncorrected Standard Errors
(Price) × (75,000<Income<100,000)	0.038	0.006 ***	0.052	9.379
(Price) × (Income>100,000)	0.123	0.008 ***	0.100	13.343
(Price) × (Income Missing)	0.079	0.006 ***	0.056	9.383
Fuel Operating Cost (cents per mile)	-0.599	0.048 ***	2.044	42.908
(Fuel Operating Cost) × (College)	-0.057	0.013 ***	0.076	5.792
Price	-0.098	0.008 ***	0.097	12.686

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table 1.6 presents the point estimates and corrected standard errors for the price, fuel cost, horsepower, and curb weight variables from the BLP model with Aggregated Choices, BLP model with McFadden's method for aggregation, and the BLP for broad choice data.[5] These variables vary by trim and hence are aggregated in both the BLP model with aggregated choices and the BLP Model with McFadden's method. Recall that the BLP model with McFadden's method controls for aggregation in the {Price × Income} and {Fuel Operating Cost} variables but not for the aggregation in variables held constant across households (i.e. price, horsepower/curb weight, curb weight.) In the BLP model for broad choice, there is no aggregation of vehicle attributes.

There are interesting comparisons to be made across the three models regarding the significance of estimates. Table 1.6 shows that more variables appear statistically significant with

[5] For results from the full model specifications, see Appendix A.4.

aggregation than without. When using McFadden's method, only the coefficients on curb weight and fuel operating cost are significant at the 1% level. By controlling for aggregation in the first stage, all coefficients on price variables and the coefficient on Fuel Operating Cost × College lose significance. In the BLP model for broad choice data, all variables reported in Table 1.6 lose significance, as is the case for all but two variables in the entire specification (see Table A4).

There are two explanations for why variables are less significant when aggregation is either accounted for or avoided altogether. First, by treating the aggregated choice set as if it is the exact choice set, the BLP model with aggregated choices ignores the uncertainty from the fact that choices are only partially observed, and aggregated vehicle attributes contain measurement error. Hence, the model produces smaller standard errors than the two models that do account for aggregation. These findings are consistent with comparisons between the "peakedness" of the likelihood function of the exact choice model and the broad choice model, presented in Brownstone and Li, 2014. Second, with respect to the BLP model for broad choice data, because of the high dimensionality of δ, estimates of these parameters are less precise. The large standard errors associated with these estimates inflate the standard errors of the second stage coefficients making them less significant. Though the BLP model for broad data allows for more observations in the second stage, unfortunately, the gains in precision from these added observations are swamped by the lack of precision from the increase in the dimension of $\hat{\delta}$.

Table 1.6: Select estimates across BLP models with various methods for aggregation.

Variable	BLP with Aggregated Choices			BLP with McFadden's Method			BLP for Broad Choice Data	
	Estimated Parameter	Corrected Standard Error		Estimated Parameter	Corrected Standard Error		Estimated Parameter	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.065	0.014	***	0.001	0.067		0.038	0.052
(Price) × (Income>100,000)	0.102	0.015	***	0.004	0.056		0.123	0.100
(Price) × (Income Missing)	0.094	0.015	***	0.011	0.080		0.079	0.056
Fuel Operating Cost (cents/mile)	-2.877	0.953	***	-2.946	0.263	***	-0.599	2.044
(Fuel Operating Cost) × (College)	-0.061	0.020	***	-0.027	0.466		-0.057	0.076
Price	-0.116	0.026	***	-0.064	0.120		-0.098	0.097
Horsepower / Curb weight	158.582	53.803	***	144.232	93.200		20.737	111.690
Curb Weight	7.569	2.511	***	7.084	1.907	***	0.002	0.006

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Given that there are more alternatives in the BLP model for broad choice data and more variables in the BLP model with McFadden's method, I cannot speak of the differences in magnitudes of coefficients across the three models. However, comparisons can be made across utility independent quantities such as the estimates of willingness to pay for improvements in fuel operating costs.

Table 1.7: Willingness to pay estimates across the three model specifications

Willingness to pay for a 1 cent/mile improvement in fuel efficiency (thousands)†	Estimated Parameter	Uncorrected Standard Error		Corrected Standard Error‡		Ratio of Corrected to Uncorrected Std. Errors	Implied Discount Rate
BLP Model with Aggregated Choices	24.695	4.090	***	10.128	**	2.477	-23.675
BLP Model with McFadden's Method	46.083	14.663	***	83.105		5.667	-28.132
BLP Model for Broad Choice Data	6.123	0.683	***	22.706		33.234	-10.785

*Note: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*
† willingness to pay for a 1 cent/mile reduction in fuel operating costs for households with no college education and income below $75,000 (in thousands of dollars).
‡ calculated using the delta method:

$$Var(williness\ to\ pay) = Var\left(\frac{\beta_{fuelop}}{\alpha_{price}}\right) = \frac{\beta_{fuelop}^2}{\alpha_{price}^4}\sigma_{price}^2 + \frac{1}{\alpha_{price}^2}\sigma_{fuelop}^2 - \frac{2\beta_{fuelop}}{\alpha_{price}^3}\rho_{fuelop,price}\sigma_{price}\sigma_{fuelop},$$

$$\sigma_{price}^2 = var(\alpha_{price}), \sigma_{fuelop}^2 = var\ (\beta_{fuelop}), \rho_{fuelop,price} = corr(\beta_{fuelop}, \alpha_{price})$$

Table 1.7 presents the implied willingness to pay estimates for a 1 cent/mile improvement in fuel operating cost, in thousands of dollars, for households with no college education and incomes less than $75,000. These quantities are constructed using the estimates in Table 1.6. This 1 cent/mile improvement in fuel cost is a 7.4% improvement over the average fuel operating cost of households in the sample. The final column of Table 1.7 provides the implied discount rate assuming vehicles are held for 14 years, with an annual mileage of 18,778, assumptions used in Greene, 2010. The negative discount rates in this column indicate that across all three models, households overvalue future fuel savings compared to present day investments in vehicle fuel efficiency.

The BLP model with McFadden's method has the largest willingness to pay estimates while the BLP model with broad choice data has the smallest estimates, though only the BLP model with aggregated choices presents an estimate that is significant at any of the conventional levels. An interesting point to draw from Table 1.7 is that the use of uncorrected standard errors leads one to believe that all three estimates of willingness to pay are significant, whereas with corrected standard errors only one of them (BLP with aggregated choices) is, and only at the 5% level. For the BLP model for broad choice data, the difference is driven largely by the fact that using the uncorrected standard errors grossly understates the uncertainty in the fuel operating cost coefficients. For the BLP model with McFadden's method, the difference is because the uncorrected standard errors understate the uncertainty in the price coefficients. The impreciseness of these estimates suggests that when aggregation is accounted for, this dataset does not speak loudly enough to provide conclusive evidence on the existence of the energy paradox within this model specification.

One will also notice from Table 1.6 that the estimates associated with horsepower and curb weight vary more drastically across models than estimates associated with other variables. This means that many utility invariant quantities formed from these variables such as willingness to pay for horsepower and willingness to pay for curb weight will differ across models as is the case with vehicle fuel efficiency. This suggests that choice set aggregation also affects the estimates associated with attributes other than fuel efficiency. In addition, fuel efficiency, horsepower, and curb weight are highly correlated variables. It is likely that the cost of aggregation increases in the face of collinearity because aggregation further obfuscates the ability to disentangle these competing effects. This is reflected in the instability of the estimates of these three variables across the three models.

The Make/Model/Fuel-type groups contain between one and fifty-five trims, with 47.7% of the groups containing 3 trims or less. Thus the assumption that attributes within each broad group follow a normal distribution may not hold for all Make-Model groups. This means that the estimates from the BLP model with McFadden's method for aggregation may be less plausible in this empirical application. The BLP model with McFadden's method may perform better if choices are aggregated to a higher level (for example, the Make/Class level) such that there are many more vehicle trims within each broad group. The distributional assumptions may be better approximated in such a setting.

There are a number of possible reasons why willingness to pay for fuel efficiency is not estimated with sufficient precision in the BLP Model for Broad Choice Data to provide conclusions about the existence of an energy paradox. Given the larger choice set in this model, an even larger number of households is required for asymptotic assumptions to apply. A larger household sample may also be required to obtain significant estimates. It is also possible that there is a lack of sufficient variability in vehicle fuel efficiency within this dataset to pin down household valuation of this attribute. In addition, vehicle fuel efficiency is highly correlated with horsepower and curb weight, making it difficult to disentangle the value households place on each of these attributes. Finally, it is possible that the model specification is simply too rich to be identified by the current data because of the uncertainty from only broadly observing choices. If pared down to a smaller model, it may perform better.

1.7 Conclusion

In this paper, I examine the implications of choice set aggregation on parameter estimates in multinomial choice models, a practice common when household choices are not fully observed and when modeling choices at the most detailed level renders the model too large for estimation in standard computing environments. I discuss two models that account for choice set aggregation. The first is a method for aggregation by McFadden, 1978, that places distributional assumptions on the elements within each aggregated alternative and uses the higher order moments of the distribution in the utility specification. The second is a model for broad choice by Brownstone and Li, 2014, that defines the choice probability of a broad group of products as the sum of the probabilities of products within that group, circumventing the need for aggregation. Applying these models, and a model that aggregates choices, to vehicle choice data reveals that aggregation affects the point estimates of the model. In addition, standard errors are smaller when the presence of aggregation is ignored, because the measurement error from aggregation does not get accounted for in estimation.

Choice set aggregation is studied within a model that employs the innovations from a series of Berry, Levinsohn, and Pakes papers. One can incorporate these innovations into choice models by estimating the model sequentially. However, standard errors from this estimation procedure are inconsistent. I derive consistent standard errors for the model by recasting the model within a GMM framework, and find, both in a Monte Carlo study and a vehicle choice application, that for all parameters of interest, the GMM standard errors are larger than the inconsistent standard errors derived from sequential estimation. For most variables, the GMM standard errors are larger by a factor of 1.3 to 5, though the largest correction inflates the standard errors of a variable by a factor of 18.

If these findings on choice set aggregation and inconsistent sequential standard errors extend to existing vehicle choice studies that aggregate the choice set, than this may help explain some of the variation in existing estimates of households' valuations of fuel efficiency from choice models which have been a source of conflict on the presence of an energy paradox. First, the variation in estimates across studies may be driven in part by the practice of choice set aggregation. Second, the use of inconsistent standard errors from the sequential estimation process may lead researchers to be overconfident in the significance of model estimates. Finally, models that aggregate the choice set ignore the measurement error that aggregation induces also causing standard errors of their estimates to be too small. Addressing these concerns may help increase coherency across studies on how much value households place on fuel efficiency.

The findings from this paper should also serve as a call to search for better data and identification strategies in vehicle choice models so that future research can provide more precise estimates on the matter. For example, there is greater variation in vehicle fuel efficiency in the present decade than the last because of higher fuel prices, more hybrid vehicle models and tighter fuel efficiency standards. Also, data on used-vehicle purchases will introduce attractive variation in vehicle attributes since the mix of vehicle attributes has changed quite rapidly in the last decade. Exploiting this variation may provide tighter estimates of household valuation for fuel efficiency and a clearer resolution to the energy paradox.

More generally, these findings send a cautionary message to choice model practitioners on the importance of giving due consideration to how choice sets are defined. Aggregating choices without accounting for the variation across aggregated alternatives may lead the researcher to flawed and invalid conclusions from model estimates. Given the popularity of multinomial choice models across a variety of fields, including transportation, industrial organization and marketing, such practices may have widespread consequences. This paper brings attention to two existing models that provide methods to address aggregation in hope that practitioners will adopt these methods in future work.

Chapter Two

The Numerical Performance of the BLP Model for Combined Micro- and Macro-level Data

2.1. Introduction

Discrete choice models have become popular in demand estimation because of their ability to model rich substitution patterns between large arrays of products while avoiding the curse of dimensionality that plagues systems of demand equations. In addition, innovations by Berry, Levinsohn and Pakes, 1995, henceforth BLP, that allow discrete choice models to accommodate controls for the endogeneity of product attributes have added to the attractiveness of these models. Though first developed for settings where only macro-level data are available, the BLP model has since been extended to settings where both micro- and macro-level data are combined in estimation (BLP,2004, Goolsbee and Petrin, 2004, Train and Winston, 2007). Variations of the BLP model have been applied to study a host of empirical questions such as the effect of mergers (Nevo, 2000a), welfare gains from new products (Petrin, 2002) and measuring market power (Nevo, 2001).

Estimating the BLP model can be challenging. This is because estimation requires inverting a nonlinear system of market share equations. When only macro-level data is available, the most common estimation approach involves nesting a contraction mapping algorithm that estimates the market share equations within the parameter search of a Generalized Method of Moments (GMM) objective function (Berry, 1994, BLP, 1995). By nesting the algorithm, the dimensionality of the objective function is significantly reduced. Using this approach, estimation involves both an "outer loop" that searches over the parameters of the GMM objective function, and a nested "inner loop" to the contraction mapping algorithm. Henceforth, this method is referred to as the nested fixed point (NFP) approach.

The GMM objective function of the NFP approach is highly non-linear and optimization of this function has proven challenging. Numerous studies note concerns with the effectiveness of using the NFP approach to obtain reliable estimates of BLP model parameters. One of the earliest papers to highlight problems with the NFP approach is Nevo, 2000b where it is noted that for certain values of parameters, the objective function can be undefined. He writes that "poor starting values, different scaling of variables and the non-linear objective would cause this to happen."

Dubé et al., 2012, investigate the sensitivity of BLP model estimates to the tightness of convergence tolerance levels of the "inner loop" and the "outer loop" in NFP estimation. Through a Monte Carlo study, they find that unless both loops are set to very tight convergence tolerance levels, errors from the "inner loop" propagate to the "outer loop" causing estimates to converge to a common wrong solution that is not even a local minimum. In addition, they also test the sensitivity of model estimates to starting points under varying convergence tolerance levels. They find that estimates converge 95-100% of the time when convergence tolerance levels of both loops are set tight (10^{-14} for the inner loop and 10^{-6} for the outer loop), but only 54-76% of the time when both loops are set loose (10^{-4} for the inner loop and 10^{-2} for the outer loop.) In light of these findings, they propose recasting the NFP objective function as a mathematical program with equilibrium constraints (MPEC). This estimation method has better numerical properties and, in certain settings, speed advantages over the NFP approach.

Knittel and Metaxoglu, 2014, study how estimates from the NFP approach vary depending on the optimization algorithm used. They estimate the BLP model on two widely used data sets: an automobile data set from BLP, 1995, and a breakfast cereal data set from Nevo, 2000b. They consider eleven different optimization algorithms, and find that estimates are in fact sensitive to the optimization algorithm in use. For example, within the automobile dataset, the own-price elasticities for the product with the highest market share ranges from -3.48 to -0.93 with a mean of 2.77 and a standard deviation of 0.53. The average change in profits due to a hypothetical merger for automobiles ranges from $485 million to $2548 million with a mean of $787 million and a standard deviation of $316 million. The authors also find that local minima with comparable objective function values can have very large differences in economic implications. In a hypothetical merger exercise, two local minima in the automobile data set with comparable objective function values reported changes in profits and consumer welfare that differ from the true values by a factor of 2 .5 and 2 respectively.

Skrainka, 2011, studies the small sample properties of the BLP model for macro data. He finds that for sample sizes that are commonly observed in empirical work, parameter estimates and elasticities from the BLP model are considerably biased, and even when very large empirical settings are considered (50 markets and 100 products), results still show finite sample bias. He hypothesizes that part of this small sample bias is related to the fact that GMM estimators have poor finite sample properties. He also shows that the bias is a function of the quality of instruments used. The most commonly applied instruments from BLP, 1995, cause larger biases than instruments derived from supply equations.

While the numerical performance of the NFP approach for BLP models for macro-level data has received much attention, the same cannot be said of estimation methods for BLP models that combine micro- and macro-level data in estimation. In this paper, I help fill this gap by investigating the numerical performance of the maximum likelihood (ML) approach to estimating the BLP model that combines micro- and macro-level data. The ML approach was first developed in Goolsbee and Petrin, 2004 but also received early attention in Chintagunta and Dubé, 2005, and Train and Winston, 2007. I test the sensitivity of estimates from the ML approach to different starting values, tightness of convergence criteria and minimization using two different optimization routines.

The BLP model for combined micro- and macro-level data is a multinomial logit model for micro data with two departures. The first is the imposition of a constraint that matches predicted shares from the model to macro-level market share data rather than in-sample shares. The second is the nesting of an instrumental variables framework within the choice model to account for the endogeneity of product attributes. The ML approach is a sequential estimation procedure. In the first stage, the parameters of the logit choice model are estimated while imposing the macro-level market share constraint. The second stage involves estimating an instrumental variables regression to account for the endogeneity of product characteristics. When the model is just identified, the ML approach is equivalent to estimating a Generalized Method of Moments (GMM) model where the first order condition of the logit likelihood function is the first moment function of the model, the market share constraint is the second, while the instrumental variables framework forms the third.

Despite being a GMM estimator, the ML approach is named as such to differentiate it from another GMM estimator for the same model developed in BLP, 2004, that incorporates consumer demographics into the model through a moment that captures the covariance of product attributes and household attributes, rather than through a likelihood function.

One advantage of micro-level data is that it allows the researcher to observe consumers' demographics and their product choices, rather than just population market shares and product characteristics. The presence of such data allows the researcher to directly model the effect of consumer demographics on product choices without depending as heavily on a random coefficient specification to account for consumer heterogeneity. Thus, a priori, one may expect models with micro-level data to have better numerical properties than models that rely solely on macro-level data. However, the ML approach still uses some similar estimation techniques as the NFP approach, such as the contraction mapping algorithm, hence may be susceptible to the same estimation concerns that plague the NFP approach highlighted here.

I find that the ML approach to estimating the BLP model for combined micro- and macro-level data is insensitive to the concerns of starting values, tightness of convergence criteria and choice of minimization routine that have been shown to plague the BLP model for macro data. The finding that the BLP model estimates well even with loose convergence criteria suggests that researchers need not incur the large time costs associated with tight convergence tolerances because reliable estimates of model parameters can be obtained with looser tolerances. More broadly, the findings in this paper suggest that there are significant computational advantages from having micro-level data on consumer choice as such data allows econometricians to estimate demand models with objective functions that are more numerically stable.

An important distinction should be made between existing research on the numerical properties of the NFP approach and the work contained in this paper. All existing work on this topic that I am aware of include a random coefficient (i.e. mixed logit) utility specification in the BLP model. Here, I only consider fixed coefficients. The random coefficient framework could make the numerical properties of the models studied here worse because it introduces the burden of simulation methods in estimation which often adds noise to the estimation process.

This paper is structured as follows. In Section 2, I present the random utility framework that underlies BLP models. In Section 3, I present the methods for estimating the BLP model when both micro- and macro-level data are available and when only macro data is available. In Section 4, I present a Monte Carlo study that tests the sensitivity of model estimates to the various numerical concerns. Section 5 concludes.

2.2. The BLP Model

Let $n = 1, \dots, N$ index consumers who can either purchase any of J products, $j = 1, \dots, J$ in the market or not purchase any product, characterized by selecting the "outside good", $j = 0$. The indirect utility of consumer n from the choice of product j, U_{nj} is assumed to follow the following linear specification[6]:

$$U_{nj} = \delta_j + w_{nj}'\beta + \epsilon_{nj},$$

$$n = 1, \dots, N, \quad j = 0, 1, \dots, J,$$

where δ_j is the "average" utility of product j, that is, the portion of utility from product j that is the same for all consumers. w_{nj} is a $(K_1 \times 1)$ vector of consumer characteristics interacted with product characteristics, β is its $(K_1 \times 1)$ vector of associated parameters, and ϵ_{nj} is an error term

[6] As mentioned in the introduction, BLP models most commonly assume a random coefficient framework, that is $U_{nj} = \delta_j + w_{nj}'\beta_n + \epsilon_{nj}$ where $\beta_n \sim N(\tilde{\beta}, \Omega)$. In this paper, β is assumed to be fixed.

with mean zero that captures all remaining elements of utility provided by product j to consumer n. For the purpose of identification, average utility of the "outside good," δ_0, is normalized to zero. Consumers select the product that yields them the highest utility:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\, i \neq j \\ 0 & otherwise. \end{cases}$$

Assume that ϵ_{nj} follows the type I extreme value distribution. Then, the probability that consumer n chooses product j is given by the following logit probability:

$$P_{nj} = \frac{exp\left(\delta_j + w_{nj}{}'\beta\right)}{\sum_k exp(\delta_k + w_{nk}{}'\beta)}.$$

In the conditional logit model, $\delta = \{\delta_1, \ldots, \delta_J\}$ and β, can be estimated by maximizing the following log-likelihood function:

$$L(y; \delta, \beta) = \sum_n \sum_j y_{nj}\, log\ (P_{nj}). \qquad (2.1)$$

An interesting feature of maximum likelihood (ML) estimation of the conditional logit is how δ is estimated. As with any ML estimator, set the first derivative of the log-likelihood function with respect to δ equal to zero:

$$\frac{\partial L}{\partial \delta_j} = \sum_n \sum_i (y_{ni} - P_{ni})\, d_i^j = 0, \qquad \text{for } j = 1, \ldots, J,$$

$$where\ d_i^j = \begin{cases} 1 & if\ i = j \\ 0 & otherwise. \end{cases}$$

Rearranging and dividing both sides by N yields:

$$\frac{1}{N}\sum_n y_{nj} = \frac{1}{N}\sum_n P_{nj}, \qquad \text{for } j = 1, \ldots, J.$$

As the equation describes, the average utilities are estimated such that the predicted shares from the model, $\widehat{S_J} = \frac{1}{N}\sum_n P_{nj}$, match the in-sample shares, that is, the share of consumers in the sample who choose each of the products. (Train, 2009).

A key innovation of BLP, 2004, is that δ is estimated such that the predicted shares match macro-level market shares rather than in-sample shares. For $j = 1, \dots, J$, δ_j is chosen such that

$$A_j = \widehat{S_J}$$

where A_j is the macro-level market share for product j. Berry, 1994, shows that for any value of β, a unique δ exists such that the predicted shares match these macro-level market shares.

Matching predicted shares to macro-level market shares rather than in-sample shares improves estimation in the event that there is high sampling variance and hence, unrepresentative sample shares. In addition, by using macro-level market shares, average utilities can be estimated for products even if the in-sample shares for these products are zero.

Finally, it is assumed that the average utilities are a linear function of product attributes:

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$

where x_j is a $(K_2 \times 1)$ vector of exogenous product attributes, α_1 is a $(K_2 \times 1)$ vector of associated parameters, p_j is a $(K_3 \times 1)$ vector of product attributes that are endogenous with respect to average utility, α_2 is a $(K_3 \times 1)$ vector of associated parameters, and ξ_{1j} captures the average utility associated with attributes unobserved to the econometrician. Because p_j is endogenous with respect to average utility, it is correlated with unobserved attributes contained in ξ_{1j}, such that $E(\xi_{1j}|p_j) \neq 0$. There exists a set of instruments, z_j, that are correlated with p_j and uncorrelated with ξ_{1j}:

$$p_j = z_j{}'\gamma + \xi_{2j}$$

$$where\ E(\xi_{1j}|z_j) = 0.$$

For simplicity, I assume here that z_j is a $(K_3 \times 1)$ vector, (and therefore, γ is a $K_3 \times 1$ vector of associated parameters), that is, there are as many instruments as there are endogenous regressors, making the model just-identified. Although over identification does not tremendously complicate estimation, with a just-identified model, optimal GMM methods are not necessary, which simplifies estimation.

In summary, the BLP model consists of the following equations:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\ i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + \mathrm{w_{nj}}'\beta + \epsilon_{\mathrm{nj}}, \quad \epsilon_{\mathrm{nj}} \sim \text{type I extreme value}$$

$$A_j = \frac{1}{N}\sum_n P_{nj}$$

$$\delta_j = x_j{}'\alpha_1 + p_j{}'\alpha_2 + \xi_{1j},$$

$$p_j = z_j{}'\gamma + \xi_{2j}$$

2.3. Estimation Procedures

In this section, I detail three estimation methods for the BLP model. In the first subsection, I detail the ML approach which is a sequential estimation procedure to estimating the BLP model on a combination of micro- and macro-level data. In the second and third subsections, I describe the NFP and MPEC approaches to estimating the BLP model. Both methods are used when the model is applied solely to macro-level data. These approaches are equivalent to each other though not equivalent to the ML approach. In fact, they should be less efficient than the ML approach since they use less data in estimation. Finally, in the fourth subsection, I detail a Newton-Raphson augmentation to the contraction mapping algorithm from Li, 2012 that I implement to decrease the algorithm's time to convergence.

2.3.1. The Maximum Likelihood Approach

In the presence of both micro-level data on consumer demographics and product choice, and macro-level data on product market shares, BLP, 2004, show that the model can be estimated sequentially. The estimation approach detailed here is similar to BLP, 2004. The key difference is that BLP, 2004, incorporates household information into their model by constructing a moment that captures the covariance of product attributes and household attributes. The model here incorporates household information through a logit likelihood function, as is done in Train and Winston, 2007.

Sequential estimation can be conducted in two stages. The first stage involves estimating the average utilities, δ, and the consumer-alternative interaction parameters, β. The first stage estimates of δ are then used in the second stage estimation of the parameters associated with the alternative attributes, $\alpha = [\alpha_1\ \alpha_2]$.

Since δ and β are dependent on each other, they are estimated through an iterative process. Conditional on some initial value of δ, maximize equation (2.1) with respect to β to obtain conditional maximum likelihood estimates, $\hat{\beta}$. Conditional on $\hat{\beta}, \delta$ is then estimated using the contraction-mapping algorithm developed in BLP, 1995. This algorithm is itself an iterative process which yields the estimate, $\hat{\delta}$, when the following equation is iterated on until convergence:

$$\delta_{j,t+1} = \delta_{j,t} + \ln\left(A_j\right) - \ln(\widehat{S}_j), \quad \forall\, j = 1, \dots, J$$

The contraction mapping algorithm enforces the constraint that the predicted shares equal the macro market shares. With these new estimates, $\widehat{\delta}$, the maximum likelihood process can be estimated again to update $\hat{\beta}$. The maximum likelihood and contraction mapping processes are repeated iteratively until convergence.

The second stage of the sequential process estimates the parameters, α, associated with product attributes, $\{p_j, x_j\}$. To do this, two stage least squares (2SLS) estimation is used, substituting the converged values, $\widehat{\delta}$, from the first stage, for the true values of the average utilities, δ. The 2SLS estimates are given by the familiar instrumental variables solution:

$$\hat{\alpha} = \left(\tilde{X}'Z(Z'Z)^{-1}Z'\tilde{X}\right)^{-1}\tilde{X}'Z(Z'Z)^{-1}Z'\hat{\delta} \tag{2.2}$$

where $\tilde{X} = [X \;\; P]$.

To obtain estimates of the standard errors for $\hat{\delta}, \hat{\beta}$ and $\hat{\alpha}$, recast each sequence of the estimation process as moments within a GMM framework. The GMM analogue to the sequential process just described involves three sample moments, one for each of the vectors of parameters δ, β and α. The first moment is the first derivative of the logit log-likelihood function with respect to β:

$$G_1(\beta, \delta) = \frac{1}{N}\sum_{\mathrm{n}} \mathrm{H}_{1\mathrm{n}}(\beta, \delta).$$

$$= \frac{1}{N}\sum_{\mathrm{n}}\sum_{\mathrm{j}} \mathrm{y}_{\mathrm{nj}}(\mathrm{w}_{\mathrm{nj}} - \sum_{\mathrm{i}} \mathrm{P}_{\mathrm{ni}}\mathrm{w}_{\mathrm{ni}}) .$$

The second moment, that identifies $\hat{\delta}$, constrains the predicted shares of the alternatives in the model to match the macro market shares:

$$G_2(\beta, \delta) = \frac{1}{N}\sum_{n} H_{2n}(\beta, \delta)$$

$$= \frac{1}{N}\sum_{n}\sum_{j} A - P_n.$$

where $A = [A_1\ A_2\ ...\ A_J]'$ and $P_n = [P_{n1}\ P_{n2}\ ...\ P_{nJ}]'$

The third moment estimates $\hat{\alpha}$. This 2SLS moment condition stipulates that in expectation, the instruments, z_j, are uncorrelated with the error term:

$$G_3(\delta, \alpha) = \frac{1}{J}\sum_{j} H_{3j}(\delta, \alpha) \tag{2.3}$$

$$= \frac{1}{J}\sum_{j} z_j(\delta - x_j\alpha).$$

The sequential ML approach is equivalent to minimizing the objective function $Q(\theta) = G'W_0G$ where $G = [G_1\ G_2\ G_3]'$ and W_0 is a weight matrix of appropriate size, when the model is just-identified because it satisfies the same estimation conditions.

The variance of this GMM estimator, θ_{GMM} is as follows:

$$Var(\theta_{GMM}) = (M_0' W_0 M_0)^{-1}(M_0' W_0 S_0 W_0 M_0)(M_0' W_0 M_0)^{-1},$$

where

$$M_0 = \begin{bmatrix} \frac{1}{N}\frac{\partial^2 L}{\partial \beta^2} & \frac{1}{N}\frac{\partial^2 L}{\partial \beta \partial \delta} & 0_{K_1 \times K_2} \\ \frac{1}{N}\sum_n -\frac{\partial P_{nj}}{\partial \beta} & \frac{1}{N}\sum_n -\frac{\partial P_{nj}}{\partial \delta} & 0_{(J-1)\times K_2} \\ 0_{K_2 \times K_1} & \frac{1}{\sqrt{NJ}}\sum_j z_j{}' & \frac{1}{J}\sum_j -z_j{}' x_j \end{bmatrix}$$

$$S_0 = \begin{bmatrix} \frac{1}{N}\sum_n H_{1n}H_{1n}{}' & \frac{1}{N}\sum_n H_{1n}H_{2n}{}' & \frac{1}{\sqrt{NJ}}\sum_n \sum_j H_{1n}H_{3j}{}' \\ \frac{1}{N}\sum_n H_{2n}H_{1n}{}' & \frac{1}{N}\sum_n H_{2n}H_{2n}{}' & \frac{1}{\sqrt{NJ}}\sum_n \sum_j H_{2n}H_{3j}{}' \\ \frac{1}{\sqrt{NJ}}\sum_n \sum_j H_{3j}H_{1n}{}' & \frac{1}{\sqrt{NJ}}\sum_n \sum_j H_{3j}H_{2n}{}' & \frac{1}{J}\sum_j H_{3j}H_{3j}{}' \end{bmatrix}.$$

Since the current model is just-identified, it is efficient to set W_0 equal to the identity matrix.

2.3.2 The Nested Fixed Point Approach

When the econometrician possesses only macro-level data, then he does not observe the products individual consumers purchase or the characteristics of the consumers that makes the purchase. However, he does observe product market shares at the aggregate level, A_j, product attributes, $\{x_j, p_j z_j\}$ and possibly also the distribution of consumer characteristics in the population such as income and education. In the absence of micro data, the most common approach to estimating the model is the nested fixed point (NFP) approach.

The NFP method to estimating the BLP model, involves concentrating δ out of the parameter search so that optimization is restricted to β and α. This is done by nesting the contraction mapping algorithm within the 2SLS moment condition such that equation (2.3) becomes:

$$G_3(\beta, \alpha) = \frac{1}{J}\sum_j H_{3j}(\delta(\beta), \alpha)$$

$$= \frac{1}{J}\sum_j z_j\big(\delta(\beta) - x_j\alpha\big).$$

For each value of β and α, the contraction mapping algorithm is used to estimate δ. These estimates of δ are then used to evaluate $G_3(\beta, \alpha)$. By doing this, the first and second moments are no longer explicitly involved in estimation. The second moment is made binding through the contraction mapping algorithm while the first moment enters implicitly, through the calculation of the logit choice probabilities in the contraction mapping algorithm.

When consumer characteristics are unobserved, as is often the case in macro data settings, one can account for consumer heterogeneity by sampling consumer attributes like income and education repeatedly from the known distribution of consumer characteristics, then averaging the choice probabilities across samples, that is:

$$P_j = \frac{1}{R}\sum_{r=1}^{R}\frac{\exp\big(\delta_j + w_{j(r)}\beta\big)}{\sum_k \exp(\delta_k + w_{k(r)}\beta)}$$

When the model is over-identified as is typically the case in empirical work, α can also be concentrated out of the objective function by expressing it as in equation (2.2). The IV moment then becomes:

$$G_3(\beta) = \frac{1}{J}\sum_j H_{3j}(\delta(\beta), \alpha(\delta(\beta)))$$

$$= \frac{1}{J}\sum_j z_j\big(\delta(\beta) - x_j\alpha(\delta(\beta))\big)$$

and the search is conducted solely over β. The attractiveness of the NFP method is that it reduces a $(K_1 + K_2 + J)$ -dimensional problem to one with only K_1 parameters.

The NFP method should yield the same model estimates as the ML approach, although the NFP method makes use of macro-level data only, while the ML approach incorporates micro data as well. Hence one would expect the NFP method to be less efficient than the ML approach.

2.3.3 Mathematical Programming with Equilibrium Constraints

More recently, Dubé et al. (2012) has propose an alternative estimation procedure to the NFP method. They recast the NFP GMM objective function as a mathematical program with equilibrium constraints (MPEC). This constrained optimization formula is

$$\min_{\alpha,\beta} \frac{1}{J}\sum_j z_j\big(\delta(\beta) - x_j\alpha\big)$$

$$\text{subject to } A = \frac{1}{N}\sum_n P_n$$

MPEC removes the need for the contraction mapping algorithm and has a speed advantage, because the iterative nature of the contraction mapping algorithm can slow down estimation significantly. The speed advantage of MPEC holds when the econometrician observes many markets and the number of products considered is not too large (<500 products). In addition, Dubé et al.,2012 show that MPEC is much more numerically stable than the NFP approach. MPEC has become a popular alternative to the NFP approach. Recent papers that have employed this approach include Skrainka, 2011, Gillen et al, 2014, and Gordon and Hartmann, 2013.

2.3.4 Speeding up the Contraction Mapping Algorithm

As previously mentioned, the contraction-mapping algorithm can incur a high time cost because it can take many iterations to reach convergence, especially when tight convergence criteria are used. To speed up the contraction-mapping algorithm in the Monte Carlo study here, I implement a simple modification using an analytic Newton-Raphson algorithm (Li, 2012). This modification produces a 280-fold improvement in number of iterations over the unmodified contraction mapping algorithm, and a six-fold improvement in estimation time. The modification is as follows:

$$\delta_{t+1} = \delta_t + \mathrm{H}^{-1}[\ln(M_j) - \ln(\hat{S})]$$

where H is the matrix of first order partial derivatives of $[\ln(M_j) - \ln(\hat{S})]$, which can be shown to equal

$$H = \begin{bmatrix} 1 - \frac{\sum_n P_{n1}^2}{\sum_n P_{n1}} & -\frac{\sum_n P_{n1}P_{n2}}{\sum_n P_{n1}} & \cdots & -\frac{\sum_n P_{n1}P_{nJ}}{\sum_n P_{n1}} \\ -\frac{\sum_n P_{n2}P_{n1}}{\sum_n P_{n2}} & 1 - \frac{\sum_n P_{n2}^2}{\sum_n P_{n2}} & \cdots & -\frac{\sum_n P_{n2}P_{nJ}}{\sum_n P_{n2}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\sum_n P_{nJ}P_{n1}}{\sum_n P_{nJ}} & -\frac{\sum_n P_{nJ}P_{n2}}{\sum_n P_{nJ}} & \cdots & 1 - \frac{\sum_n P_{nJ}^2}{\sum_n P_{nJ}} \end{bmatrix}$$

2.4. A Monte Carlo Study: Starting Values and Convergence Criteria

To compare the sensitivity of the ML algorithms to different starting values, convergence criteria and optimization routines, I conduct a Monte Carlo study conducted as follows.

I generate a single data set from the following model:

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\ i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + w_{mj1}\beta_1 + w_{mj2}\beta_2 + \epsilon_{nj}, \quad where\ n \in m$$

$$\delta_j = \alpha_0 + x_{j1}\alpha_{11} + x_{j2}\alpha_{12} + p_j\alpha_2 + \xi_{1j},$$

$$p_j = z_j\gamma + \xi_{2j}$$

$$(\xi_{1j}, \xi_{2j}) \sim N(0_2, \Omega), \quad \Omega = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

where

$$J = 30$$

$$N_{pop} = 1{,}000{,}000$$

x_{j1} is a $J \times 1$ vector of binary values, drawn from a Bernoulli distribution with success probability of 0.5,

x_{j2} is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2,

v_{m1} and v_{m2} are $N_{pop} \times 1$ vectors of continuous values drawn from a standard normal distribution bounded between -2 and 2,

w_{mj1} is an $N_{pop}J \times 1$vector of interactions between v_{m1} and x_{j1} ,

w_{mj2} is an $N_{pop}J \times 1$vector of interactions between v_{m2} and x_{j2} , and

z_j is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2.

Finally,

$$\beta_1 = 0.8,\ \ \beta_2 = -0.7, \alpha_0 = \ 0,\ \ \alpha_{11} = -.5,\ \ \alpha_{12} = 1,\ \ \alpha_2 = -0.5, \text{and}\ \ \gamma = 1.3.$$

I use the choices generated from this dataset to create macro-level market shares, then randomly select a subsample of consumers, $N = 20{,}000$, from N_{pop} to serve as the Monte Carlo sample.

I generate 100 different sets of starting values and estimate the model 100 times using a different set of starting values each time. I provide the analytic gradient in estimation and use the KNITRO minimization package in MATLAB (Byrd, Hribar, and Nocedal, 1999; Byrd, Nocedal and Waltz, 1999) to conduct the parameter search. To confirm that the gradients were programmed correctly, they are checked against their numerical equivalents, and are found to always be equal to the ninth or tenth decimal. KNITRO is the optimization package used in Dubé et al., 2012 and is also a package considered in Knittel and Metaxoglou, 2014. It has quickly become the choice optimization routine for estimation of BLP models.

As in Dubé et al., 2012, I conduct the Monte Carlo study considering various convergence tolerance levels. For the ML algorithm, three convergence criteria have to be set, one each on the contraction mapping algorithm and the likelihood function, and an "outer" loop that iterates between the two. I consider a "tight" algorithm, where the contraction mapping tolerance is 10^{-14} while both the likelihood tolerance and outer loop tolerance are 10^{-6}, and a "loose" algorithm where the contraction mapping tolerance is 10^{-4} and the likelihood and outer loop tolerances are 10^{-2}.

In addition, I also run the Monte Carlo study using the NFP approach on quasi-macro data. In this setting, I estimate the model using the same simulated data as used in the Monte Carlo study for the ML algorithms, but ignoring the data on consumer choices, y_{nj}. This assumes the case where the researcher observes individual consumer characteristics but does not observe consumers' choices from the product space. This is in contrast to the more common macro-level data scenario where the researcher only knows the distribution of consumer characteristics within the population and has to sample from this distribution to simulate consumer characteristics.

For the NFP approach, I consider a "tight" algorithm where the inner loop tolerance is set at 10^{-14} and the outer loop tolerance is set at 10^{-6}, a "loose-both" algorithm where the inner loop tolerance is 10^{-4} and the outer loop tolerance is 10^{-2} and a "loose-inner" algorithm where the inner loop tolerance is 10^{-4} and the outer loop tolerance is 10^{-6}. The results from the Monte Carlo runs on the NFP and ML approaches are displayed in Table 2.1.

Table 2.1: Three NFP and two ML implementations: Varying starting values for one generated data set

	NFP loose both	NFP loose-inner	NFP tight	ML loose	ML tight	Truth
Fraction that converge to a minimum	0.89	0.89	0.88	1.00	1.00	
Mean Estimate						
β_1	0.4660	0.4660	0.4660	0.7028	0.7028	0.8000
α_1	0.3921	0.1926	0.4103	0.4252	0.4252	0.5000
Std. Dev. of Estimate						
β_1	2.1462	2.1462	2.1462	0.63×10^{-4}	0.17×10^{-7}	
α_1	0.1127	0.0975	0.0677	0.12×10^{-6}	0.31×10^{-8}	
Lowest objective function	1.7×10^{-31}	3.2×10^{-31}	2.0×10^{-30}	9.8×10^{-11}	2.4×10^{-17}	
Average computation time	4.77	4.88	17.54	3.62	79.69	

The second row in Table 2.1 shows the fraction of runs that converge to a local minimum. The three NFP routines converge to a minimum about 90% of the time. The routines converge to a saddle point in all other runs. Both ML routines always converge to a local minimum. The third row of Table 2.1 shows the mean estimates of β_1 and $\alpha_1$across the 100 Monte Carlo runs while the fourth row shows the empirical standard deviation of the estimates across the 100 Monte Carlo runs. Rows three and four show that for the NFP approach, the tightness of the tolerance levels matter for the estimates of α_1. The estimates of α_1 are more sensitive to starting values under "NFP loose both" and "NFP loose-inner". Of these three methods, "NFP tight" presents the least variable estimates that are also, on average, closest to the truth. Comparatively, even when loose tolerances are used for the ML algorithm, estimates are much closer to the truth, with small standard deviations across the Monte Carlo runs.

The fifth row shows the lowest objective function value attained across all 100 runs. The NFP routines achieve smaller "lowest values" than the ML routines even though the NFP routine utilizes less data. This is because the NFP routine has fewer convergence criteria. The NFP routine nests one algorithm within the other, thus only requires two convergence criteria. The ML routine iterates between algorithms and thus requires three convergence criteria. This extra loop in the ML routines causes a loss in precision compared to the NFP routines. Nevertheless, the objective function of the ML method is less sensitive to starting values, and finds the minimum more often than the NFP approach.

Finally, the sixth row shows average computation time across the Monte Carlo runs. There is a clear trade-off between computation time and precision. The "ML tight" routine is much slower than the other four estimation routines, and though the variance in estimates of α and β are smaller under “ML tight,” they are already relatively precise when using “ML loose” suggesting that the gains from using tighter convergence tolerances are small. This is in contrast to the findings of Dubé et al., 2012 who show that for the NFP approach, it is crucial to use tight convergence tolerances to obtain consistent model estimates.

I also study the sensitivity of model estimates to optimization algorithms. I re-run the Monte Carlo study in Table 2.1 using the “fminunc” search routine provided by MATLAB. I provide the routine with analytic first and second derivatives for minimization. This routine uses a trust-region method to obtaining the solution to the optimization problem. Interestingly, when using “fminunc” rather than the KNITRO for optimization, the results in Table 2.1 remain unchanged. This result is promising for the ML approach as it shows that the approach is insensitive to starting values and convergence tolerance levels under either optimization routine. However, for the NFP approach, this finding is inconsistent with that of Knittel and Metagoxlu, 2014 who find that the choice of optimization package affects the point estimates that are yielded as the solution. In addition, Dubé et al., 2012 also note that the “fminunc” solver often fails to converge to a local minimum when minimizing the NFP objective function.

I conjecture that what is driving the poor results from the NFP approaches in Table 2.1 is not the poor numerical properties of the NFP objective function, but rather, a lack of model identification from having only a single cross-section of data. This explanation also gains support from two other facts. First, I find that the numerical properties of the NFP approach do not improve when I increase the number of consumers, N, and choices, J, in the generated data set. Asymptotic theory on the BLP model (Berry et al, 2004b) suggests that this should be the case, though evidence from Skrainka, 2011, suggests that very large increases are required before asymptotic behavior becomes evident. Second, the results in Table 2.1 for the NFP approach are much worse than the findings reported in Dubé et al., 2012 who run a similar study using a mixed logit model. Dubé et. al, 2012 allow for multiple cross sections of data, which is probably necessary for the NFP approach to estimate well. Extending this Monte Carlo study to a panel data setting may yield the NFP approach more promising results.

2.5. Conclusion

Recent studies have cast doubt on the estimation properties of the Nested Fixed Point (NFP) approach typically used to estimate the BLP model for macro-level data. Dubé et. al, 2012 demonstrate that the method lacks robustness to starting values and tightness of convergence criteria. Knittel and Mexagolu, 2014 show that the estimates obtained using the NFP approach is sensitive to the choice of optimization routine for minimization. In light of these findings, in this paper, I study whether the concerns of starting values, tightness of convergence criteria and choice of optimization algorithms also apply to the maximum likelihood (ML) approach to estimating the BLP model for combined micro- and macro-level data.

The Monte Carlo study conducted in this paper yields positive results regarding the properties of the ML approach to estimating this model. The ML approach is robust to starting values and provides consistent estimates even when loose convergence criteria are used. The ML approach is also insensitive to the KNITRO optimization routine and the "fminunc" optimization routine in MATLAB. The insensitivity to the ML approach to the tightness of convergence criteria mean that researchers can obtain consistent parameter estimates without incurring the large time costs associated with tight convergence tolerances.

More broadly, the findings in this paper suggest that there is great value in micro-level choice datasets. One solution to the numerous challenges faced when trying to obtain reliable estimates from macro-level data choice models is to supplement macro-level datasets with micro-level data on consumer demographics and product choice. Though obtaining micro-level data is often more expensive, such datasets circumvent numerous estimation challenges that plague BLP models that rely solely on macro-level data. In addition, this study shows that with micro-level data, one may obtain consistent estimates of model parameters from just a single cross section of data. Evidence here and in other papers (Skrainka, 2011) suggest that with macro data, it is likely that one needs many cross sections for the model to estimate well. Researchers and policymakers should take these considerations into account when deciding the amount of resources they want to allocate to data collection.

Chapter 3

Lights, Camera, Legal Action! The Effectiveness of Red Light Cameras on Collisions in Los Angeles

3.1 Introduction

In 2009, road collisions at intersections killed 7,043 people in the United States, of which 676 were attributed to drivers failing to stop at red lights (Federal Highway Administration, 2010a,b). Tangible economic costs of collisions alone are estimated at about 313.6 billion dollars a year, in 2013 prices, primarily because of losses due to injury, death, property damage, travel delay and insurance administration. If one utilizes value of statistical life estimates rather than just productivity loss estimates, these costs rise by over 70% (Small and Verhoef, 2007). Such large figures warrant that road safety policies be studied carefully in order that resources are best used to reduce collisions and their associated costs.

Red light cameras (henceforth cameras) have been introduced in cities across the United States since the 1990s to increase enforcement of traffic laws in the hope of increasing safety. These camera systems are triggered to capture an image of the intersection if a vehicle crosses the stopping line after a specified time, once the light has turned red. License plate information from the photographs is then used to issue a ticket to the vehicle owner. Evidence on the effectiveness of cameras in increasing safety at intersections remains inconclusive. While past research indicates that cameras decrease the number of right-angle collisions, some findings suggest that the cameras also increase rear-end collisions significantly, because road users brake more suddenly in the presence of these cameras (Erke, 2009).

This paper studies the City of Los Angeles Automated Photo Red Light Enforcement Program which was in effect from April 2006 to July 2011 and estimates the effect these cameras had on collisions. Leveraging on the city's size and abundance of available potential control intersections, intersections with cameras (henceforth treated intersections) are compared against nearby intersections without cameras, matched on observable characteristics. The spatial correlation that arises from this design is incorporated into the study through the use of a random coefficient specification within a Poisson regression model. Similar methods are adopted by Hallmark et al., 2010, in their study of cameras in Davenport, Iowa. The estimation design in this paper also attempts to capture the potential spillover effects that cameras may have on neighboring intersections by considering three sets of control groups with varying distances from the treated intersections.

The Los Angeles camera program suffered from legal setbacks that dampened the effectiveness of the program over time. Because the Los Angeles Superior Courts began rejecting citations issued from cameras, citations issued later in the program were more frequently ignored. In addition, the installation of cameras at intersections were accompanied by increases in amber and all-red phase signal timings at the same intersections, complicating the ability to uncover the effectiveness of cameras alone on collisions. However, by assuming that the effect of cameras on collisions decayed over time, due to enforcement weaknesses, while the effect of traffic light phase timings did not, it is possible to separate the effect of cameras on collisions from the effect of traffic light phase timings on collisions. This is done by estimating the yearly effect of the program on collisions for each of the five years of the program. The assumption just stated suggests that any reversal in trends of the yearly effects must be assigned to the decay in the effectiveness of camera enforcement over time.

Controlling for the decay in enforcement, this study finds that cameras decrease red-light-running collisions. Nevertheless, cameras also increase right-angle, rear-end, and injury-related collisions. As a result, cameras cause a net increase in collisions overall. The paper is organized as follows: Section 2 presents a summary of existing findings and estimation concerns on the effectiveness of cameras. Section 3 provides an overview of camera programs in Los Angeles. Sections 4 and 5 provide details on the data used and estimation strategy employed. In section 6, the results from estimation are presented and discussed, while section 7 summarizes and concludes the study.

3.2 Literature Review

There is a well-developed existing literature on the effect of cameras on road safety. Retting et al., 2003, and Aeron-Thomas and Hess, 2009, provide summaries of the existing work on the issue. In addition, Høye, 2013, conducts a meta-analysis of the camera literature, which provides a comprehensive index of academic papers that address the effect of cameras on collisions and other safety indicators.

Past research indicate that cameras reduce red-light-running violations. Studies of Oxnard, California (Retting et al., 1999a), Fairfax County, Virginia (Retting et al., 1999b), Salem, Oregon (Ross and Sperley, 2011) and Virginia Beach, Virginia (Martinez and Porter, 2006) find that cameras decrease violations when cameras were installed. Porter et al., 2012, find that when the same cameras in Virginia Beach were turned off, red light running rose dramatically, by almost 300% in the subsequent month, and by 400%, one year later.

The effectiveness of cameras at reducing collisions, however, is more ambiguous. Retting and Kyrychenko, 2002, and Hallmark et al., 2010, find that cameras reduce the overall number of collisions. Retting and Kyrychenko, 2002, find that cameras decrease injury related collisions. Hu et al., 2011, study a cross section of 99 cities and find that cameras decrease fatal collisions. Haque et al., 2010, find that cameras in Singapore decrease motorcycle collisions. Other studies find less encouraging results in support of cameras. Some find that, at best, cameras have no effect on collisions overall (Chin and Quddus, 2002; Burkey and Obeng, 2004; Garber et al., 2007). Many studies also find that cameras increase rear-end collisions (Burkey and Obeng, 2004; Council et al., 2005a; Garber et al., 2007; Shin and Washington, 2007). These increases suggest that, in the presence of cameras, drivers brake more abruptly to avoid citations, causing such collisions to occur.

Høye, 2013, provides some rationale for the lack of cohesive results. In her meta-analysis, she highlights two primary concerns when studying camera programs that cause heterogeneity in results: selection bias and spillover effects. Studies that do not control for selection bias find reductions in right angle collisions that are twice as large as studies that do not. She also finds evidence that publication bias accounts for heterogeneity in results across studies. Overall, she finds that cameras increase all collisions by 6%, decrease right-angle collisions by 13%, and increase rear end collisions by 40%. Right-angle injury collisions decrease by 33% and rear end injury collisions increase by 19%

Selection bias arises because the placement of cameras at intersections is, in most cases, non-random. Cameras are generally placed at intersections where more collisions occur. Unobservable characteristics of these intersections that make them particularly dangerous (for example, high traffic volume) are hence correlated with the presence of cameras. Unless this correlation is addressed, any estimate of the effect of cameras on collisions will also capture some of the effect of these unobservable characteristics, creating bias (Rubin, 1974).

Spillover effects arise because the presence of cameras at certain signalized intersections within a city may cause drivers to behave differently at signalized intersections without cameras. The traditional thought is that the spillover effect is positive (Retting et al., 1999a,b; Shin and Washington, 2007). Cameras help drivers develop better driving habits, making them less likely to run red lights at all intersections. However, it is also conceivable that the spillover effect is negative. A driver who wants to shorten travel time could run more red lights at intersections without cameras knowing he is more likely to receive a citation for such an offence at an intersection with cameras.

These concerns have been addressed previously in different ways. One way to control for selection bias is by conditioning on an extensive set of variables that affect safety at intersections. This strategy is employed by Burkey and Obeng, 2004, Chin and Quddus, 2002, and Haque et al. 2010, among others. Generally, the set of control variables includes data on the number of lanes at intersections, weather conditions, length of amber lights, speed limits and traffic volume. By controlling for these variables, the estimate of the effect of the cameras is free of selection bias if one contends that conditional on these variables, placement of cameras is random. One shortcoming of this estimation design is that it does not account for spillover effects unless non-camera intersections are chosen from outside the potential spillover effect region.

In the absence of extensive and detailed data on variables affecting safety at intersections, an alternative strategy to overcome selection bias is to use control groups to serve as comparisons to treated intersections. This strategy controls for all characteristics (both observed and unobserved) that are identical across the intersections in the control group and the intersections with cameras. In the Hallmark et al., 2010, study of Davenport, Iowa, the authors select control intersections that have similar traffic volumes, collision frequencies and roadway types. When studying the camera program in Oxnard, California, Retting and Kyrychenko, 2002, consider the effect cameras have on all signalized intersections within the city, not just those that received cameras. They use as a control group, the neighboring cities of San Bernardino, Santa Barbara and Bakersfield, California. Hence, they obtain estimates that capture both the effect of the cameras on treated intersections and spillover effects as well.

3.3 An Overview of Los Angeles Red Light Camera Programs

In November, 2005, the City of Los Angeles, in cooperation with the Los Angeles Police Department (LAPD) approved a contract with a private vendor to install and provide camera enforcement services at 32 intersections across the city (KABC-TV News, 2005). Figure 3.1 displays the locations of the 32 intersections across the city.

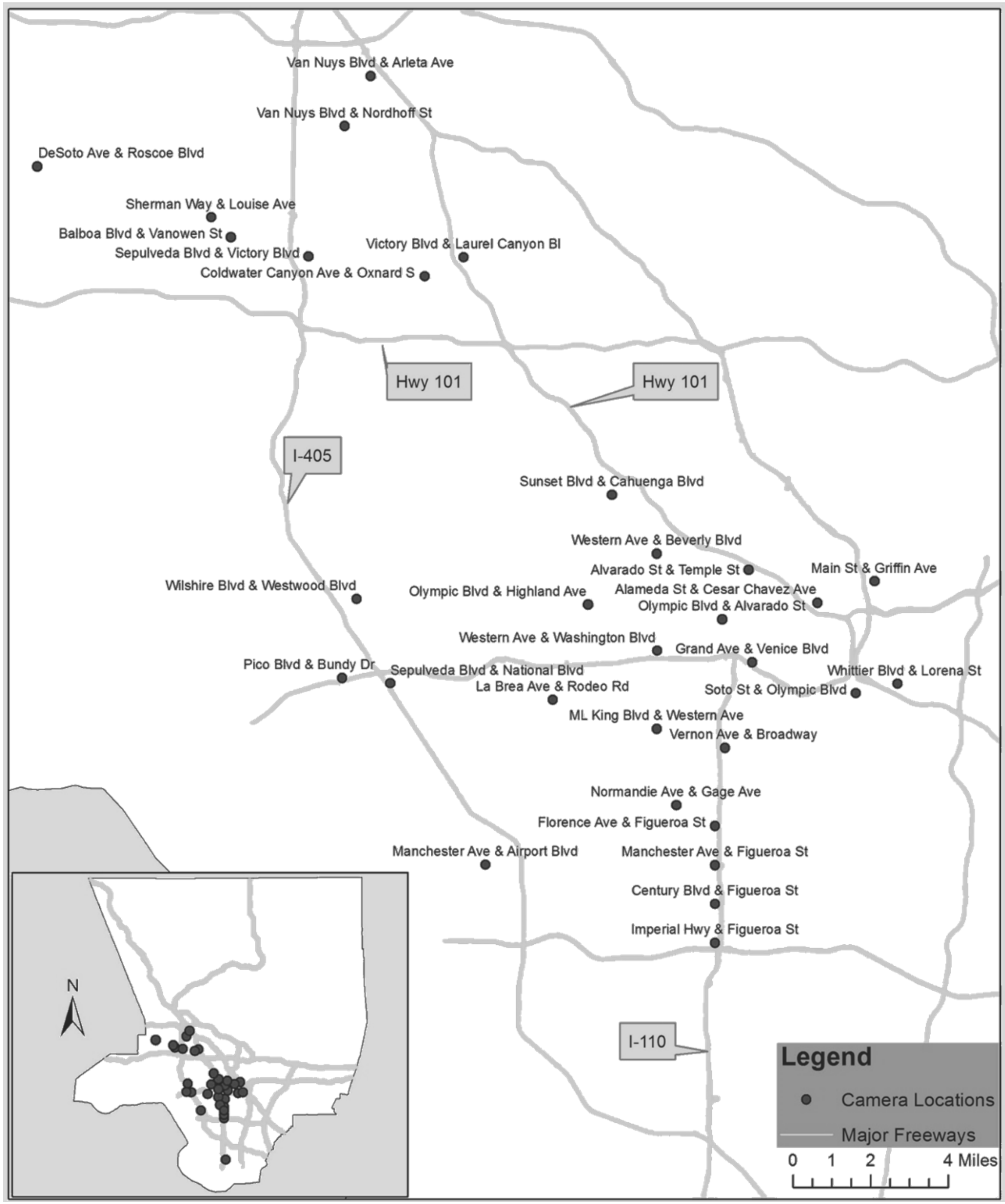


Figure 3.1: Location of Red Light Cameras under the City fo Los Angeles Automated Photo Red Light Enforcement Program, 2006-2011. Adapted from Compton, 2011.

Selection of the locations of these cameras was non-random. Selection was made based on collision records from 2003 and 2004 with an equal number of cameras being placed under each of the four LAPD Bureaus' jurisdictions (Compton, 2011). Additionally, intersections where installation of cameras would be expensive and more challenging were omitted as were intersections adjacent to state highways that required state approval of any changes (Greuel, 2010). Cameras were installed at staggered times between April 4th, 2006 and December 12th, 2007. Once a camera and related signs were posted, the LAPD began a 30-day warning period during which violators were issued warnings instead of tickets.

On July 27th, 2011, the Los Angeles City Council voted unanimously to discontinue the camera program. This decision was mainly because of escalating costs. Since the program began, only 60% of citations issued had been paid leaving the city council to cover an annual $1-1.5 million in program costs. The low percentage of citation payment was due to a judicial interpretation of the California Vehicle Code. According to this interpretation of the vehicle code, moving citations had to be issued to the driver of the vehicle; however, the citations from the cameras were instead issued to the registered owner of the vehicle. Consequently, the Los Angeles Superior Court was unwilling to enforce camera citations and made the decision not to notify the Department of Motor Vehicles about the unpaid camera citations. Because of this, no holds were placed on drivers' licenses and vehicle registration renewals if citations were ignored. (Bloomekatz et al., 2011).

Other reasons for the program's discontinuation include new state legislation that reduced the maximum fines for "rolling right turns," disputes about the implementation of the program since cameras were not chosen solely in the interest of public safety, and the possibility that the concurrent changes in all-red and amber light phase times were the cause of any safety improvements, and not the cameras themselves (Greuel, 2010).

3.4 Data

Data were obtained from numerous sources. Compton, 2011, obtained information about the camera program, including where and when the cameras were installed, from the LAPD. These same data are used here. Characteristics of intersections, such as number of lanes and intersection type, were obtained from Google Maps.

Collision records were obtained from the California Highway Patrol Statewide Integrated Traffic Records System (CHP SWITRS). This database was also the source of collision records for the Retting and Kyrychenko, 2002, study of cameras in Oxnard, California. The CHP SWITRS records every police reported collision in California from 2001 onwards. The system was queried for all collisions within Los Angeles from January 2006 to December 2010. All 252,406 collisions obtained from CHP SWITRS were checked for spelling errors and cross checked with corresponding streets on Google Maps. Similar data clean up processes are conducted in the Burkey and Obeng, 2004, study of Greensboro, North Carolina. This data are then used to create intersection-month observations. Collisions are aggregated by month for each intersection. This creates the outcome variable of interest, the number of collisions at a particular intersection in a particular month.

From this data set, I consider five categories of collisions where the effect of cameras is worth investigating: all collisions, right-angle collisions, rear-end collisions, red light collisions and injury collisions. "All collisions" encompass every collision in the dataset for the time period of study. Three variables supplied within the SWITRS dataset are used to define the remaining four categories considered. First, the dataset contains a *type of collision* variable which defines collisions into seven types based on the point of contact between vehicles. Two of the seven types are used: right-angle collisions and rear-end collisions. Second, the primary *collision factor violation* variable categorizes collisions by the traffic violation that police determine caused the collision. This variable is used to obtain a fourth category of collisions for this study: red-light-running related collisions. Note that this category is not entirely objective since police discretion is involved in deciding which collisions are caused by red light running. Finally, the *collision severity* variable defines each collision into five collision types, ranging from property damage only (no injury) to fatal. This variable is used to obtain injury collisions. Injury collisions encompass four of the five collisions defined in the *collision severity* variable: fatal, severe injury, visible injury and complaint of pain. 44.95% of collisions are property damage only, 44.21% are complaint of pain collisions, 9.43% are visible injury collisions, 1.15% are severe injury collisions and 0.27% are fatal collisions. The distribution of collisions across the five considered categories are summarized in Table 3.1 and illustrated by two Venn diagrams in Figure 3.2. Right-angle and rear-end collisions are mutually exclusive; however they are not mutually exclusive of red light and injury collisions. Similarly, red light and injury collisions are not mutually exclusive of each other either.

Table 3.1: Three-way (relative) frequency table of collisions by category: type of collision, red light collisions and injury collisions

Type of Collision	Non- Red Light Collision		Red Light Collision		**Total**
	Non-Injury	Injury	Non-Injury	Injury	
Other	21067 *(16.64%)*	24417 *(19.28%)*	594 *(0.47%)*	2444 *(1.93%)*	48522 *(38.32%)*
Rear-end	12521 *(9.89%)*	18605 *(14.69%)*	80 *(0.06%)*	137 *(0.11%)*	31343 *(24.75%)*
Right-angle	6767 *(5.34%)*	24806 *(19.59%)*	2544 *(2.01%)*	12647 *(9.99%)*	46764 *(36.93%)*
Total	40355 *(31.87%)*	67828 *(53.56%)*	3218 *(2.54%)*	15228 *(12.03%)*	126629 *(100.00%)*

The average number of months observed before a treated intersection receives a camera is 16.46 months, while the average number of months observed after a treated intersection receives a camera is 42.53 months. A longer "before" period, where all intersections are untreated, cannot be used because the city operated a camera program with a different private vendor prior to this program which expired in June 2005. It is not known which intersections received cameras under the earlier program and when these cameras were removed, though news reports about the defunct cameras began circulating around November 2005 (KABC-TV News 2005), suggesting the possibility of contamination from the previous program had dwindled by the end of 2005.

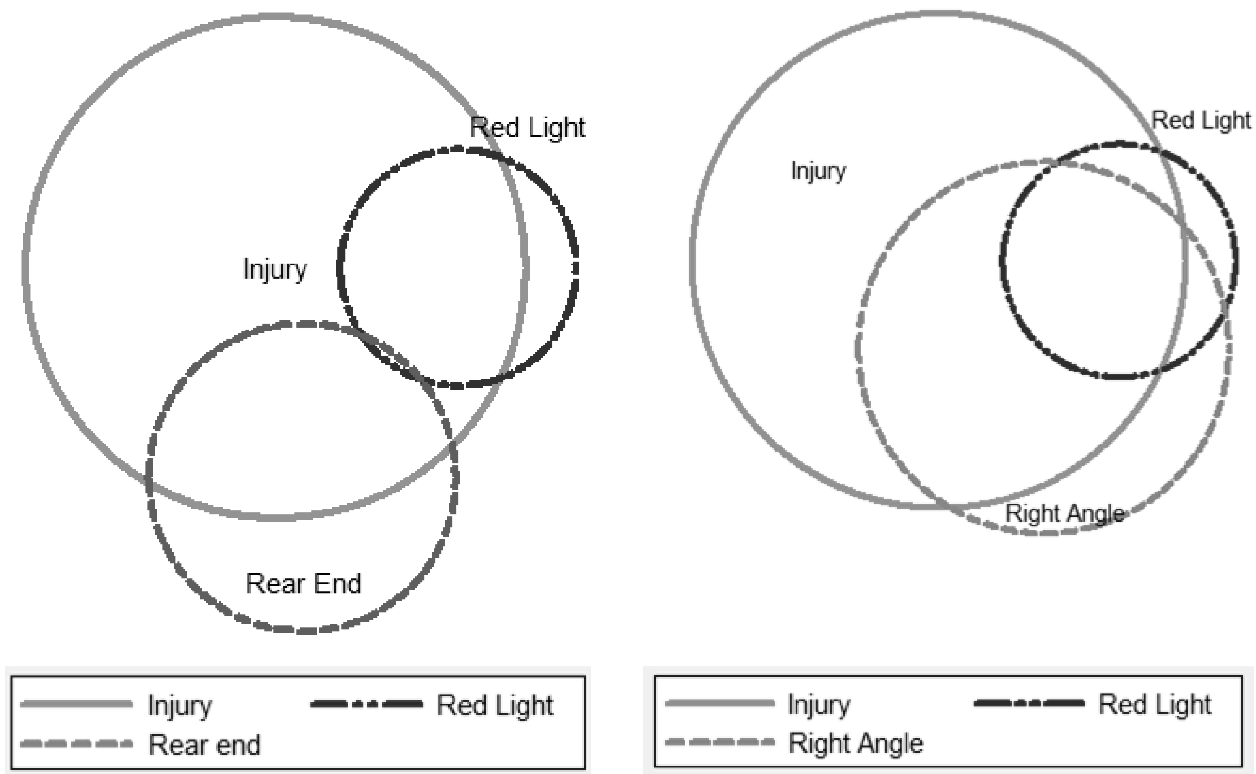


Figure 3.2: Proportional Venn Diagram showing the relationships between four of the categories of collisions studied

**Rear-end Collisions and Right-Angle Collisions are mutually exclusive, hence there is no overlap between their respective sets.*

Additionally, Compton, 2011, provides evidence that the previous program did not contaminate driver behavior in the current program.

3.5 Estimation

The estimation design employed here considers concerns of selection bias and spillover effects highlighted in existing studies of camera programs.

To address concerns of selection bias, I make use of control intersections, similar to the study by Hallmark et al., 2010. While the strategy of conditioning on an extensive set of observable intersection characteristics is an attractive approach to addressing the problem, these data are not available for intersections in Los Angeles. However, the camera program in Los Angeles is well suited for the use of control intersections as comparisons to treated intersections. Because the city is very large, there are many untreated intersections that are potential controls. For each treated intersection, I select four nearby intersections, one each to the north, south, east and west of the treated intersection as controls for that treated intersection. [7] Therefore, each control intersection shares one common street with the treated intersection.

These control intersections are also selected by matching on three observable characteristics: type of intersection (i.e. four-way), presence of a traffic light, and number of lanes. If no intersection with a common street fulfills the matching criteria, it is replaced with the closest intersection that does match the observable characteristics. The use of four nearby intersections serve as good counterfactuals to the treated intersections; not only are they similar to treated

[7] Four treated intersections on Figueroa Street (Imperial Highway, Florence Ave, Century Boulevard, and Manchester) and two treated intersections in the San Fernando Valley (Balboa Boulevard/Vanowen Boulevard and Sherman Way/Louise Ave) are located too closely to have their own set of controls. The solution employed was to merge these intersections into two separate groups that share the same controls.

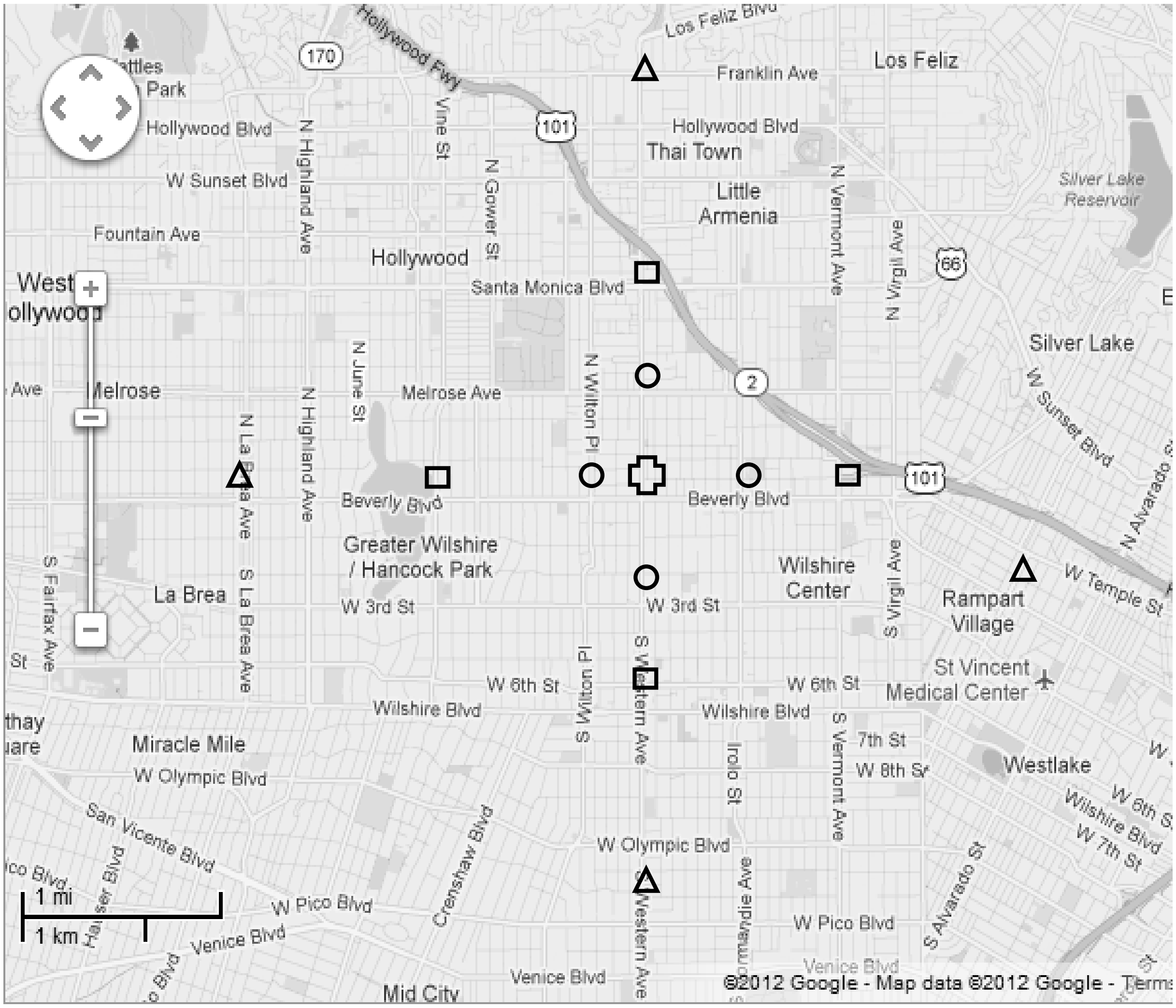
intersections on matched observable characteristics, they are also similar on many unobservable characteristics, such as traffic volume, driver characteristics and weather patterns. However, this design raises concerns of spatial correlation.

In modeling, it is necessary to assume that observations are independent of each other. This is equivalent to saying that conditional on observed factors, the remaining unobserved effects (or errors) are random, or uncorrelated across observations. However, in this study, since control intersections are selected nearby to treated intersections, observations are located in clusters rather than randomly distributed across the city. Therefore, it is likely that errors are non-random, due to spatial correlation. The proper specification of a model should incorporate this concern. Details on how this study addresses spatial correlation is described later.

To address potential spillover effects, three rings of nearby intersections are considered: 0.5 mile, 1 mile and 2 miles away from the treated intersections respectively.[8] Figure 3.3 shows an example of the three sets of control intersections for a treated intersection. The spillover effect should be strongest at the 0.5-mile ring and weakest at the 2-mile ring. The tradeoff is that the 0.5-mile ring is often better matched to the treated intersection on observables such as number of lanes than the 2-mile ring, hence, better controls for the effect of unobservable characteristics like traffic volume and weather, on collisions. Table 3.2 presents the mean collisions per month at intersections before and after cameras were installed for both the treated and treated intersections 0.5-mile away. Table B1 in the appendix presents additional summary statistics of the data in use.

---

[8] Most signalized intersections in the city are located about 0.5-miles apart; hence it is not feasible to consider rings of intersections at closer distances than 0.5 and 1-mile from the treated intersection.



Intersection with red light camera
0.5-mile control intersection
1-mile control intersection
2-mile control intersection

Figure 3.3: Control intersections for Red Light Camera at Beverly Boulevard and Western Avenue

Table 3.2: Mean collisions per month at treated and 0.5-mile control intersections before and after treatment

Type of Collision		Treated Intersections		Control Intersections		Difference in Means	
		Before	After	Before	After	Before	After
All	Mean	1.0778	1.1159	0.8266	0.8077	0.2513 ***	0.3082 ***
	Std. Dev.	0.0511	0.0299	0.0233	0.0138	0.0562	0.0329
Right-angle	Mean	0.3949	0.3798	0.2853	0.2821	0.1096 **	0.0977 ***
	Std. Dev.	0.0307	0.0169	0.0130	0.0078	0.0333	0.0186
Rear-end	Mean	0.2840	0.3279	0.2147	0.2098	0.0694 **	0.1181 ***
	Std. Dev.	0.0248	0.0156	0.0114	0.0064	0.0273	0.0169
Red Light	Mean	0.1673	0.0953	0.1158	0.1039	0.0514 **	0.0085
	Std. Dev.	0.0188	0.0080	0.0080	0.0046	0.0205	0.0093
Injury	Mean	0.6637	0.6892	0.5049	0.5071	0.1588 ***	0.1821 ***
	Std. Dev.	0.0377	0.0237	0.0169	0.0107	0.0412	0.0260

All control intersections selected are four-way intersections with traffic lights and at least three lanes on each street in the intersection. The minimum number of lanes for any of the treated intersections is also three. For the 0.5-mile control ring, there is only one instance where a control intersection selected does not share a common street with the treated intersection, or 0.008% of all intersections. For the 1-mile ring, there are 14 such instances (11.6%), and for the 2-mile ring, 25 such instances (20.7%). Hence, the quality of matches is significantly higher at the 0.5-mile ring compared to the 2-mile ring. The absolute differences in the number of lanes at the treated and control intersections are presented in Table 3.3. These absolute differences are calculated for all control intersections that have a street in common with the treated intersection. Differences in number of lanes are taken first for the street in common between the two intersections and then for the streets that are uncommon between the two intersections. For example, in Figure 3.3, the treated intersection is located at Beverly Boulevard and Western Avenue. The 0.5-mile control intersection, west of the treated intersection, is located at Beverly Boulevard and Wilton Place. The differences in number of lanes in this case is the difference in the number of lanes on Beverly Boulevard at its intersections with Western Avenue and Wilton

Place (common street) and the difference in the number of lanes at Western Avenue and Wilton Place at their respective intersections with Beverly Boulevard (uncommon street).

Table 3.3: Absolute difference in number of lanes at treated and control intersections

Control Ring		Average absolute difference in number of lanes	Standard Error	Min	Max	Zero Difference (%)
0.5-mile	Common Street	0.7983	0.9074	0	4	42.86
	Uncommon Street	1.3193	1.1194	0	6	26.05
1-mile	Common Street	0.7905	0.8737	0	3	44.76
	Uncommon Street	1.2667	1.1458	0	6	28.57
2-mile	Common Street	0.8022	0.7919	0	3	41.76
	Uncommon Street	1.2308	1.0962	0	5	28.57

The intersection-month observations created from the SWITRS data are used to create an intersection-by-month panel dataset. Since the outcome variable, the number of collisions at a particular intersection in a particular month, is count in nature, it is assumed to follow a Poisson distribution. A limitation of the Poisson distribution is that it is equidispersed (its variance is always equal to its mean) and cannot model error correlations across observations. The latter is necessary to model spatial correlation that arises from the use of control intersections in this study. To overcome these shortcomings, the Poisson distribution is "mixed" with a log-normal distribution to form the Poisson-log-normal model. This mixing distribution relaxes the assumption of equidispersion[9] and allows for error correlations across observations. Hallmark et al., 2010, employ a similar model in their study of cameras. The model, first developed by Hausman et al., 1984, is as follows:

Let y_{it} be the number of collisions at intersections $i = 1, \dots, n$ across months, $t = 1, \dots, T$, X be an $nT \times k$ matrix of covariates that predict y, and x_{it} be a $1 \times k$ row from the matrix X. Let β be a $k \times 1$ vector of coefficients corresponding to the k covariates in X. Let Wbe an $nT \times g$ matrix of

[9] For a derivation of the model's overdispersion property, see Appendix B.1.

covariate values, and w_{it} be a $1 \times g$ row from the matrix W. The g covariates in W have corresponding coefficients, b_i where b_i is a $g \times 1$ vector.

Then,

$$y_{it}|\beta, b_i \sim Poisson\ (\mu_{it})$$

where μ_{it} is the conditional mean,

$$\mu_{it} = E(y_{it}|\beta, b_i) = \exp(x_{it}\beta + w_{it}b_i)$$

$$b_i \sim N(\eta, D).$$

Note that since $b_i \sim N(\eta, D), \exp(\mathrm{b_i}) \sim \log \mathrm{Normal}$. This is how a log-normal distribution is "mixed" with a Poisson distribution. Since the coefficients b_i are stochastic, they are called "random coefficients."

Assuming observations are independent and identically distributed,

$$f(y_i|\beta, b_i) = \prod_{t=1}^{T} p(y_{it}|\beta, b_i) \tag{3.1}$$

where p is the Poisson mass function and $y_i = \{y_{it}\}, t = 1, \dots, T$.

The joint density of y_i and b_i is

$$f(y_i, b_i|\beta, \eta, D) = f(y_i|\beta, b_i)\phi(b_i, |\eta, D) \tag{3.2}$$

where $\phi(b_i, |\eta, D)$ is a g −variate normal density function with mean η and variance matrix D.

The likelihood function is the product of the individual likelihood contributions, where b_i is marginalized over its distribution, hence

$$L(y|\beta,\eta,D) = \prod_{i=1}^{n} L_i(y_i|\beta,\eta,D)$$
$$= \prod_{i=1}^{n} \int f(y_i, b_i|\beta,\eta,D)\, db_i \,.$$

If the matrix, W, is empty or alternatively, if D is a zero matrix, then the model reduces to a simple Poisson regression without random coefficients. The $b_i{}'s$ allow for heterogeneous effects of covariates on the outcome variable. In addition, the $b_i' s$ can be interpreted as representing error components that create correlations across the intersections (Brownstone and Train, 1999).

In the first application of this model, μ_{it} is specified as follows:

$$\mu_{it} = \exp(treatment_{it}\, \beta_{treatment} + \gamma_t \beta_{time} + group_{ih} b_i).$$

The treatment variable, $treatment_{it}$ is an indicator variable that takes on the value 1 if there is a camera at intersection i in month t, and 0 if that intersection does not have a camera in month t:

$$treatment_{it} = \begin{cases} 1 & if\ intersection\ i\ has\ a\ camera\ in\ month\ t \\ 0 & otherwise. \end{cases}$$

γ_t is a $1 \times T$ row vector of month dummy variables, equal to 1 for an observation in month t and 0 otherwise. These variables capture the effects common to all intersections in each month.

$group_{ih}$ is a $1 \times g$ row vector of group specific dummies. Each treated intersection and its associated control intersections are placed in a group, h. In all, there are a total of g groups of intersections, $h = 1,2, \dots, g$. For each observation, only the $h-th$ entry in $group_{ih}$ takes on the value of 1, corresponding to the group, h, that the observation (intersection i) belongs to. The remaining columns take on the value of 0:

$$\{group_{ih}\} = \begin{cases} 1 & if\ intersection\ \ i\ is\ in\ group\ h \\ 0 & otherwise. \end{cases}$$

The associated coefficients for up_{ih} , b_i, are random coefficients. As previously mentioned, random coefficients can be interpreted as error components. Following this interpretation, the use of the $group$ variable allows the model to have an additional error component that is group specific, in addition to the independent and identically distributed error term that affects all observations. Additionally, the random coefficient allows correlation across groups as well. This captures the spatial correlation concerns previously noted. These correlations manifest in the covariance matrix of the random coefficients, D.

To be more precise,

$$\begin{aligned} Cov\left(w'_{it}b_i\, , w'_{js}b_j\right) &= E[(w'_{it}b_i - w'_{it}\eta)(\, w'_{js}b_j - w'_{js}\eta)] \\ &= w'_{it}\, D\, w_{js} \end{aligned}$$

since D is the covariance of b_i and b_j.

Denoting D_{lh} as the element in row l and column h of matrix D, the covariance between two intersections **within** the same group, g, is simply, D_{gg}, and the covariance between two intersections in **different** groups, g and h, is, D_{gh}.

Additionally, the group specific dummies allow each intersection to have its own intercept, thereby controlling for all intersection-specific effects that are constant over time. For example, if there are no changes to the number of lanes and speed limit at an intersection over the time period studied, then the coefficient b_i captures the total effect of these factors on collisions. This removes the need for data on intersection characteristics insofar as these characteristics are constant over time.

The parameters β, η, and D, are estimated with Bayesian inference as outlined in Chib et al., 1995, using the following priors:

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N\left(0_g, I_g \times 100\right), \;\; D^{-1} \sim Wishart\left(g + 4, I_g\right).$$

An Expectation-Maximization routine (Dempster et al., 1977) is first used to obtain suitable starting values which are then fed into a Gibbs sampler to obtain the joint posterior distribution of the estimates, β, η, and D. Within the Gibbs sampler, the data augmentation technique (Tanner and Wong 1987) is used to estimate the latent $b_i's$ (Chib et al., 1995).

While the parameters β, η and D are informative, of greater interest is the effect of cameras on the outcome variable, number of collisions. This "treatment effect" is the difference in the expected number of collisions at treated intersections, with and without the cameras (and light timing changes):

$$E_{it}(\, Treatment\ Effect_{\ it}) = E_{it}(y_{it} | \beta, \eta, D, treatment_{it} = 1) - E_{it}(y_{it} | \beta, \eta, D, treatment_{it} = 0)$$

The aggregate marginal effect of treatment on collisions can be obtained by taking the expectation of the aforementioned quantity over the distribution of the estimated parameters, β, η, D:

$$E_{\beta,\eta,D} E_{it}[Treatment\ Effect_{it}\,] \quad \forall i \;\in \{i\colon \max\{treatment_{it}\} = 1\}.$$

3.6 Results

For the sake of parsimony, this section presents results using the 0.5-mile ring as controls since this ring best addresses the problem of selection bias. Results using the 1 and 2-mile rings are provided in Appendix B.[10]

As described in Section 4, five non-mutually exclusive categories of collisions are considered: all, right-angle, rear-end, red light and injury collisions. Table 3.4 presents the marginal effects of the program on collisions in Los Angeles from 2006 to 2010. Because amber light and all-red phase times were changed concurrently with the installation of cameras, this marginal effect captures both the effects of the cameras and the light timing changes. Thus, all results from this specification have to be interpreted as the aggregate effect of both changes on collisions.

Table 3.4: The marginal effect of treatment on all, right-angle, rear-end, red light and injury collisions from 2006 to 2010 using 0.5-mile control ring

Type of Collisions	Marginal Effect	95% Probability Interval		% Change
All	0.1866	0.0790	0.2929	16.72
Right-angle	0.0905	0.0419	0.1393	23.82
Rear-end	0.1130	0.0682	0.1598	34.46
Red Light	-0.0116	-0.0306	0.0085	-12.12
Injury	0.1532	0.0869	0.2166	22.23

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000 iterations. The following non-informative priors are used for all regressions

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N(0_g, I_g \times 100), \quad D^{-1} \sim Wishart(g+4, I_g).$$

The marginal effects presented here, and in all subsequent tables, have posterior distributions that are Poisson-Log Normal. This distribution is not necessarily symmetric. Therefore, probability intervals are presented because they present a more informative measure of uncertainty than standard deviations.

[10] See Figures B2.1 & B2.2, Tables B2.1 & B2.2

The estimates in Table 3.4 show that when "all collisions" are considered, treatment causes a significant increase of 0.18 collisions a month. This translates to a 17% increase in collisions. The model also estimates increases in right-angle and rear-end collisions of 24% and 34% respectively. The marginal effect of treatment on red light collisions is negative but insignificant. Injury collisions increase by 22%.

Results in Table 3.4 show that the program increased collisions overall. However, because of the concurrent changes to light timing phases and the lack of legal enforcement of the program, it does not tell us the specific effect cameras had on collisions. To estimate this effect, we need a model that captures the potential decay in enforcement over time.[11] To do this, I permit the treatment variable to differ by year. [12] This is done by generating one treatment variable for each of the five years in the sample, 2006-2010. The five variables are obtained by multiplying the previously defined treatment variable, $treatment_{it}$, by year dummy variables.

The results from this specification are presented in Table 3.5. The percent changes in collisions due to treatment are also summarized in Figures 3.4-3.9. The estimates for 2006 across all five types of collisions are imprecise because only four intersections receive cameras that year, hence these results are not discussed here. Nevertheless, the remaining results from this specification are insightful. Similar patterns emerge across time for four of the categories of collisions. For "all collisions", right-angle, rear-end and injury collisions, the treatment effect is positive for all five years. This effect first exhibits a rising trend but this trend reverses in the later years in the sample. For red light camera related collisions, the treatment effect is negative for the first three

[11] Appendix B.2 present results for an alternative specification that also controls for decay in enforcement: a re-estimation of the single treatment variable model for the years 2006-2008.

[12] The choice to separate treatment out by year rather than half year or quarter is made purely for the sake of parsimony. The estimation time cost involved lends favor to smaller models.

years in the sample, then becomes positive for the final two. Unfortunately, the estimates for rear-end and red light camera related collisions are imprecise since these collisions are the most infrequent among the five types of collisions.

Table 3.5: The marginal effect of treatment on all, right-angle, rear-end, red light and injury collisions from 2006 to 2010 by year for 0.5-mile control ring

Type of Collision	Year	Marginal Effect	95% Probability Interval		% Change
All	2006	0.4489	-0.0465	0.9821	40.22
	2007	0.0892	-0.0530	0.2307	7.99
	2008	0.1756	0.0434	0.3065	15.73
	2009	0.3016	0.1753	0.4253	27.03
	2010	0.1440	0.0287	0.2624	12.90
Right-angle	2006	0.0442	-0.1993	0.3249	11.65
	2007	0.1000	0.0220	0.1831	26.34
	2008	0.0972	0.0298	0.1627	25.60
	2009	0.1075	0.0438	0.1768	28.30
	2010	0.0687	0.0108	0.1303	18.10
Rear-end	2006	-0.0315	-0.3001	0.2573	-9.61
	2007	0.0808	-0.0145	0.1744	24.64
	2008	0.0890	0.0117	0.1689	27.14
	2009	0.0638	-0.0152	0.1414	19.45
	2010	0.0624	-0.0074	0.1337	19.04
Red Light	2006	-0.0543	-0.1642	0.1031	-57.00
	2007	-0.0114	-0.0452	0.0310	-11.97
	2008	-0.0636	-0.0863	-0.0372	-66.71
	2009	0.0302	-0.0042	0.0685	31.71
	2010	0.0036	-0.0245	0.0342	3.75
Injury	2006	-0.0915	-0.3456	0.1944	-13.27
	2007	0.1082	0.0144	0.2069	15.70
	2008	0.1308	0.0443	0.2172	18.98
	2009	0.2107	0.1230	0.3044	30.57
	2010	0.1681	0.0855	0.2558	24.39

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000. The following non-informative priors are used for all regression

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N\left(0_g, I_g \times 100\right), \quad D^{-1} \sim Wishart\left(g + 4, I_g\right).$$

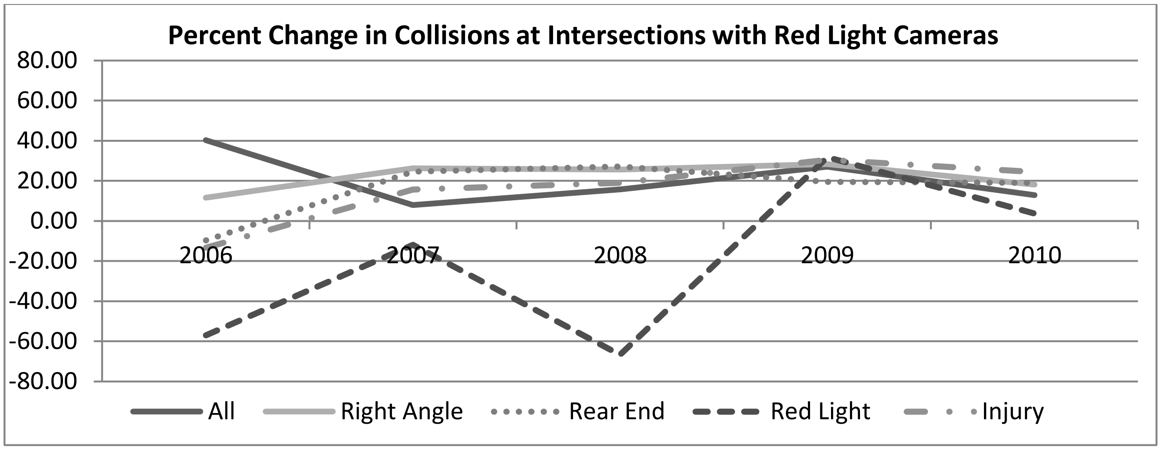


Figure 3.4: The marginal effect of treatment on collisions by year using the 0.5-mile control ring

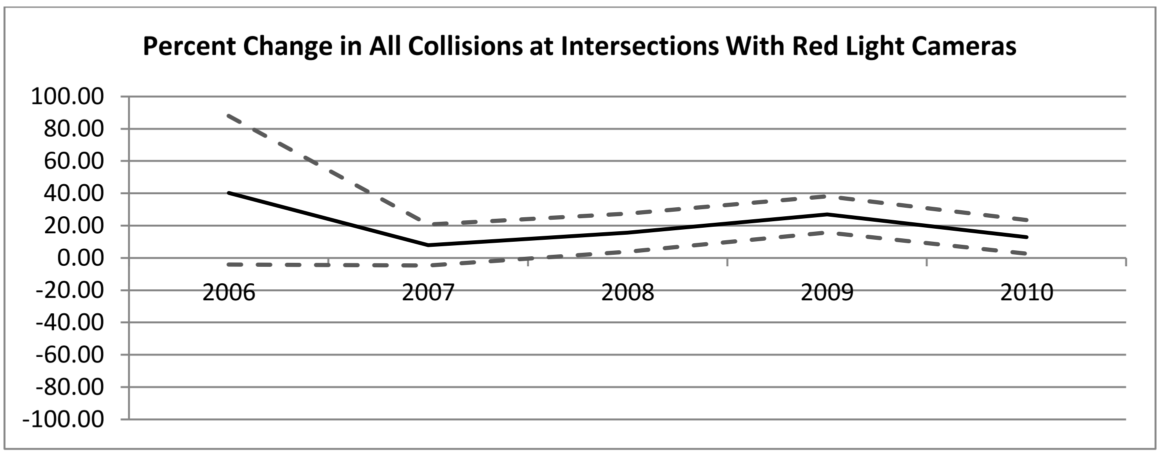


Figure 3.5: Point and 95% interval estimates of the marginal effect of treatment on all collisions by year, using the 0.5-mile control ring

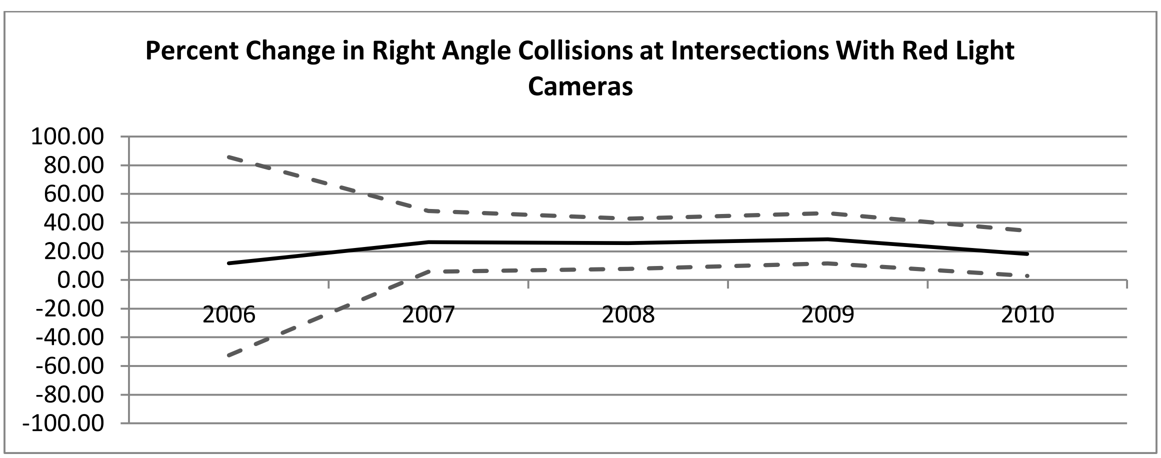


Figure 3.6: Point and 95% interval estimates of the marginal effect of treatment on right-angle collisions by year, using the 0.5-mile control ring

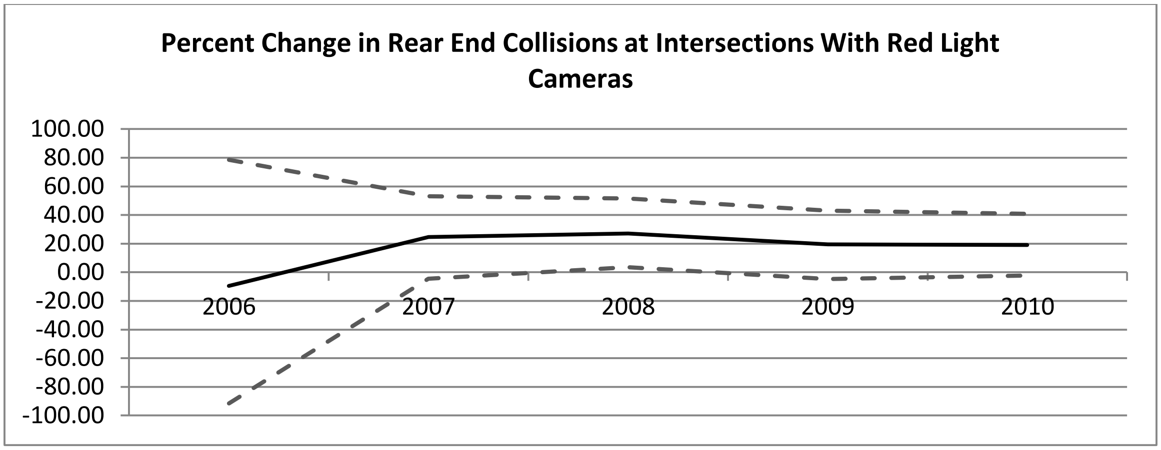


Figure 3.7: Point and 95% interval estimates of the marginal effect of treatment on rear-end collisions by year, using the 0.5-mile control ring

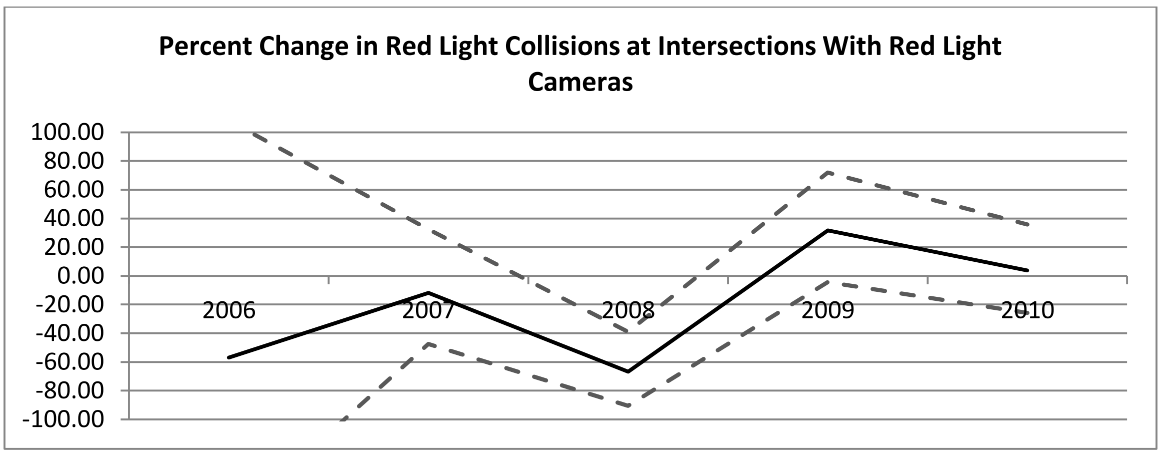


Figure 3.8: Point and 95% interval estimates of the marginal effect of treatment on red light collisions by year, using the 0.5-mile control ring

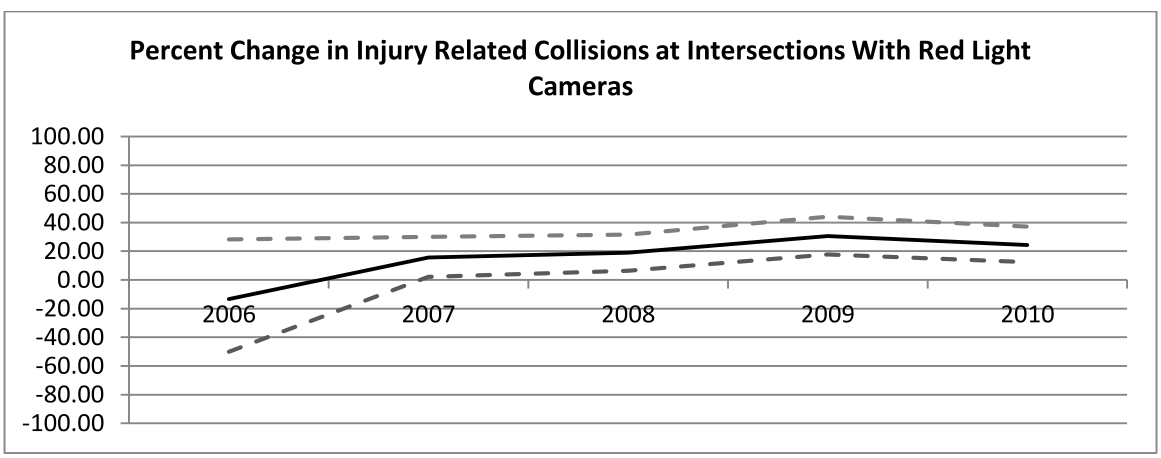


Figure 3.9: Point and 95% interval estimates of the marginal effect of treatment on injury related collisions by year, using the 0.5-mile control ring

The initial general trends observed across all five categories of collisions could be a result of increased awareness of the program as it grew in size. The reversals in trends that are observed across all five types of collisions around 2009 and 2010 allow us to make causal statements about the effectiveness of cameras alone on collisions. If we are willing to contend that the treatment variable captures only the effects of light timing changes and cameras, and that the effect on collisions from the changes to all-red and amber light timings remains constant over time, then we can attribute these trend reversals to the decay in effectiveness of legal enforcement on collisions. This assumption allows us to infer that cameras, when fully enforced, decrease red light collisions but increase "all" and right-angle collisions. In addition, there is weaker evidence that cameras increase rear-end and injury collisions.

There is also evidence suggesting that light timing changes increase red light related collisions. In 2009, not only does the trend of decreases in red light collisions end, the sign of the treatment effect switches from negative to positive. If we assume that the decay due to legal enforcement weaknesses is complete by 2009, then the sign change in 2009 must be because the light timings changes increase such collisions, since the treatment effect for 2009 no longer captures any effect of cameras on collisions.

One explanation for the observed increases in right-angle and rear-end collisions is that the presence of cameras induce drivers to do one of two things when the traffic light signal transitions from green to red, either they brake more dramatically to avoid entering the intersection and being captured on camera, or they accelerate in attempt to make it through the intersection before being caught on camera.

Because the camera program was preceded by an earlier program, a shorter period, before which the cameras were installed, is used than would have been preferred. Therefore, results presented here are susceptible to spillovers from this earlier program if drivers were not aware of the earlier program's termination. Under certain assumptions, we can postulate how such spillovers would bias the results of this study. If we assume that cameras have the same effect on collisions in both programs, then the spillovers from the existing program would cause downward bias in the magnitude of the effect of cameras but not the sign of the effect, that is, the results presented here are smaller in magnitude than they should be. If we do not make that assumption, then the signs and magnitudes of the bias are indeterminate. However, as mentioned in Section 4, I believe the concerns related to the spillover bias from the earlier program are small because news reports informed drivers about termination of the program. Compton, 2011, finds evidence in support of this fact as her study of this program is robust to concerns of spillovers from the earlier program.

3.6.1. Spillover Effects

Comparisons across rings do not tell a clear story about spillover effects. Table 3.6 presents the marginal effects of treatment across all three control rings for the single treatment variable specification. For all five types of collisions, the effect of treatment becomes less positive as distance of control ring from treated intersection increases. For red light related collisions, nearby intersections are more dangerous than distant intersections after treatment suggesting a negative spillover effect. For the other four categories, untreated intersections are safer the closer they are to treated intersections.

Table 3.6: The marginal effect of treatment on all, right-angle, rear-end, red light and injury collisions from 2006 to 2010 using 0.5-mile, 1-mile and 2-mile control rings

Type of Collision	Control Intersections	Marginal Effect of Treatment	95% Probability Interval		% Change
All	0.5-mile	0.1866	0.0790	0.2929	16.72
	1-mile	0.1416	0.0192	0.2607	12.69
	2-mile	0.1248	-0.0010	0.2494	11.19
Right-angle	0.5-mile	0.0905	0.0419	0.1393	23.82
	1-mile	0.0748	0.0136	0.1320	19.71
	2-mile	0.0682	0.0045	0.1294	17.96
Rear-end	0.5-mile	0.1130	0.0682	0.1598	34.46
	1-mile	0.1038	0.0471	0.1550	31.66
	2-mile	0.1042	0.0489	0.1608	31.78
Red Light	0.5-mile	-0.0116	-0.0306	0.0085	-12.12
	1-mile	-0.0030	-0.0233	0.0178	-3.12
	2-mile	0.0022	-0.0195	0.0242	2.30
Injury	0.5-mile	0.1532	0.0869	0.2166	22.23
	1-mile†	-	-	-	-
	2-mile†	-	-	-	-

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000 iterations. The following non-informative priors are used for all regressions

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N(0_j, I_j \times 100), \quad D^{-1} \sim Wishart(j+4, I_j).$$

†Results omitted because model EM algorithm failed to converge

There is no clear story about spillover effects for the treatment-by-year specification either. While comparisons across rings for "all collisions" suggest the same relationship as observed in Table 3.6, comparisons across rings for other types of collisions are not insightful. Relevant results for these comparisons are available in Tables 3.5, B2.1 and B2.2.

This inconclusive evidence on spillover effects is consistent with Høye, 2013, whose study raised doubts on the existence and extent of spillover effects. It is possible that the lack of conclusive evidence about the nature of spillover effects from cameras is because the positive and negative spillover effects are cancelling each other out and all the model is capturing is noise.

While quantitative results from estimation vary across rings, qualitatively, results across all three sets of control rings tell the same story, suggesting that even if spillover effects are of concern, they cast doubt on the magnitudes but not the signs of the estimated effects.

3.6.2 Sensitivity Analysis

Model sensitivity analysis is conducted to ensure that the results of the model do not hinge on particular assumptions. First, results for all three specifications are robust to choice of the prior distribution used in Bayesian estimation. Marginal effects remain unchanged with the use of both stronger and weaker priors. A second concern is that the dependent variable, number of collisions, used for all regressions is based on counts of collisions up to 200 feet away from intersections. This choice of distance is arbitrary. However, results do not change when the models are applied to collisions 100 feet and 300 feet away from intersections. Finally, when potential spillover effects from a concurrent camera program managed by the Los Angeles County Metropolitan Transit Authority are accounted for, by omitting intersections located in close proximity to LACMTA cameras, results do not vary significantly either.

3.7 Conclusion

The findings here suggest that cameras increase collisions overall, as well as right-angle and injury collisions. There is weaker evidence that cameras increase rear-end collisions as well but decrease red-light running related collisions. These findings show that camera surveillance of city intersections may not improve collisions unilaterally. It appears that drivers may brake more dramatically causing rear-end collisions, or accelerate to try to get through the intersection before the camera is triggered, causing right angle collisions. At best, we can hope that it reduces the more costly collisions at the expense of minor collisions.

Additionally, this study suggests there may be a disconnect between the effect of cameras on violations and collisions. Cameras have been shown to decrease violations but I find they increase collisions overall. This suggests that the types of violations we are issuing may not be fully justified in terms of the likelihood that they translate into collisions. Policymakers may want to consider a review of safety laws so that only actions that threat safety are defined as violations. That being said, the negative findings here should not be used as conclusive evidence against the case for cameras.

One caveat is that this study provides conclusions only about the average effect that cameras have on collisions. It may be the case that cameras are more effective at certain intersections than others, depending on the design of the intersection. For example, Burkey, 2005, finds that cameras are less effective at intersections with high traffic volumes while Høye, 2013, finds that cameras are most effective when warning signs are posted at some but not all camera intersections. Unfortunately, this study does not have the data to determine the conditions under which cameras may be most effective. This should be an undertaking of future research.

A second caveat is that the design of this study focuses on the effects cameras have on driver behavior and does not provide an exhaustive economic cost-benefit analysis of the program. As just mentioned, it is plausible that cameras increase minor injury collisions but significantly decrease severe injury and fatal collisions, which are 2.5 and 50 times more costly than minor injuries respectively (Council et al., 2005b). It is not possible to ascertain this effect within this model because the frequency with which fatal collisions occur is too low.[13] Fatal collisions make up only 0.87% of injury collisions; hence it is unlikely that fatal collisions are driving the observed increases in injury related collisions. A possible strategy to study the economic costs and benefits of this program is to weigh each collision by its average estimated cost based on collision severity estimates, and apply the same methodology of this paper in a censored data model with random coefficients.

The answer to these questions will provide a more definitive answer on whether or not camera programs are worth continuing. Until then, when investing in camera programs, policy makers should be aware of the potential unintended consequences that could arise from implementation of the technology.

[13] There are only 26 fatal collisions in the 0.5-mile control ring sample or 0.29% of observations, 21 fatal collisions in the 1-mile control ring sample (0.25%), and 17 fatal collisions in the 2-mile control ring sample (0.20%).

References

Aeron-Thomas, A. S., Hess, S., 2005. Red-light cameras for the prevention of road traffic crashes" Cochrane Database of Systematic Reviews, Issue 2, Art. No. CD003862 pub2. John Wiley & Sons Ltd. Hoboken, NJ.

Allcot, H., and Wozny, N., 2014. Gasoline Prices, Fuel Economy, and the Energy Paradox. The Review of Economics and Statistics. 96(5): 779-795.

Allcott, H., and Greenstone, M., 2012. Is There an Energy Efficiency Gap? Journal of Economic Perspectives. 26(1): 3-28.

Bento, A. M., Goulder, L. H., Jacobsen, M. R., von Haefen, R. H., 2009. Distributional and Efficiency Impacts of Increased US Gasoline Taxes. American Economic Review. 99(3):667 – 699.

Berry, S. T., 1994. Estimating Discrete-Choice Models of Product Differentiation. The RAND Journal of Economics. 25(2), 242-262.

Berry, S.T., Levinsohn, J., Pakes, A., 1995. Automobile Prices in Market Equilibrium. Econometrica. 63(4), 841-890.

Berry, S.T., Levinsohn, J., Pakes, A., 2004. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. Journal of Political Economy. 112(1), 68-105.

Bhat, C., Sen, S., 2006. Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-continuous Extreme Value (MDCEV) Model. Transportation Research Part B: Methodological 40:35-53.

Bloomekatz, A., Sewell A., Mather, K., 2011. Who knew L.A.'s red-light camera fines were 'voluntary'? Los Angeles Times, 27 July. Accessed November 11, 2012. http://articles.latimes.com/2011/jul/27/local/la-me-0727-red-light-cameras-20110727.

Brownstone, D., Bunch, D. 2013. A Household New Vehicle Purchase Model to Support Analysis of the Impact of CAFÉ Standards. VOLPE Report.

Brownstone, D., Li, P., 2014. A Model for Broad Choice Data. Department of Economics Working Paper, UC Irvine.

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics. 89: 109-129.

Burkey, M. L., 2005. "A response to unfounded criticisms of Burkey and Obeng (2004) made by the IIHS." Department of Economics and Transportation/Logistics, NCA&T State University, Greensboro, NC.

Burkey, M. L., and Obeng, K., 2002. A detailed investigation of crash risk reduction resulting from red light cameras in small urban areas. Greensboro, N.C.: Transportation Institute, North Carolina Agricultural and Technical State University, Urban Transit Institute, Transportation Institute.

Busse, M.R., Knittel, C.R., Zettelmeyer, F., 2013. Are Consumers Myopic? Evidence from New and Used Car Purchases. American Economic Review. 103(1): 220 – 256.

Byrd R.H., Nocedal, J., Waltz, R. A., 1999. KNITRO: An Integrated Package for Nonlinear Optimization. In G di Pillo and M. Roma (Eds.), Large-Scale Nonlinear Optimization. Springer. Pp 35-59.

Byrd, R. H., Hribar, M. E., Nocedal, J., 1999. An Interior Point Method for Large Scale Nonlinear Programming. SIAM Journal on Optimization. 9(4), 877-990.

Chantagunta, P., Dubé, J., 2005. Storekeeping Unit Brand Choice Model that combines Household Panel Data and Store Data. Journal of Marketing Research 42(3), 368-379.

Chib, S., Greenberg E., and Winkelmann, R., 1998. Posterior simulation and Bayes factors in panel count data models. Journal of Econometrics. 86 (1):33-54.

Chin, H. C., Quddus, M. A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis Prevention. 35 (2): 253-259.

Compton, A., 2011. The Safety Impact of Red Light Cameras on Collisions in Los Angeles. University of California, Irvine Economics Honors Thesis.

Connell, R., 2010. Red-light cameras prompt revenue questions even as LAPD report says program is helping curb accidents. Los Angeles Times, 26 March. Accessed November 28, 2012. http://latimesblogs.latimes.com/lanow/2010/03/red-light-cameras-prompt-revenue-questions-even-as-lapd-report-says-program-is-helping-curb-accidents.html.

Council, F. M., Persaud, B., Eccles, K., Lyon, C., Griffith, M. C., 2005a. Safety Evaluation of Red-Light Cameras. Report FHWA-HRT-05-048.

Council, F. M., Zaloshnja E., Miller, T., Persaud, B., 2005b. Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries. Report FHWA-HRT-05-051.

Dempster, A. P., Laird, N.M., Rubin, D. B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 39 (1): 1-38.

Dubé, J., Fox, J.T., Su, C., 2011. Improving the Numerical Performance of BLP Static Dynamic Discrete Choice Random Coefficients Demand Estimation. Econometrica 80(5), 2231-2267.

Erke, A., 2009. Red light for red-light cameras? A meta-analysis of the effects of red-light cameras on crashes. Accident Analysis and Prevention. 41 (5): 897-905.

Federal Highway Administration. 2010a. Synthesis and Evaluation of Red Light Running Automated Enforcement Programs in the United States. Publication No. FHWA-IF-00-004.

Federal Highway Administration. 2010b. Red-Light Running. US Department of Transportation FHA Resources. Accessed November 28, 2012. http://safety.fhwa.dot.gov/intersection/redlight/data/.

Garber, N., J., 2005. An evaluation of red light camera (photo-red) enforcement programs in Virginia: a report in response to a request by Virginia's Secretary of Transportation. Charlottesville, VA: Virginia Transportation Research Council.

Gillen, B.J., Shum, M., Moon, H.R., 2014. Demand Estimation with High-Dimensional Product Characteristics. In Ivan Jeliazkov , Dale J. Poirier (Eds.), Bayesian Model Comparison (Advances in Econometrics, Volume 34) Emerald Group Publishing Limited, pp. 01 – 32.

Gillingham, K., Palmer, K., 2014. Bridging the Energy Efficiency Gap: Policy Insight from Economic Theory and Empirical Evidence. Review of Environmental Economics and Policy. 8(1): 18-38.

Goldberg, P.K., 1998. The Effects of the Corporate Average Fuel Efficiency Standards in the U.S.. The Journal of Industrial Economics, XLVI(1):1-33.

Goolsbee, A., Petrin, A., 2004. The Household Gains from Direct Broadcast Satellites and the Competition with Cable TV. Econometrica. 72(2), 351-381.

Gordon, B. R., Hartmann, W.R., 2013. Advertising Effects in Presidential Elections. Marketing Science. 32(1), 19-35.

Greene, D., 2010. How Households Value Fuel Economy: A Literature Review. Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Report EPA-420-R-10-008.

Greuel, W., 2010. Audit of the Photo Red Light Program. City of Los Angeles Office of the Controller.

Grigolon, L., Reymaert, M., Verboven, F., 2014. Consumer Valuation of Fuel Costs and the Effectiveness of Tax Policy: Evidence from the European Car Market. Working Paper.

Hallmark, S., Orellana, M., McDonald, T., Fitzsimmons, E., Matulac, D., 2010. Red Light Running in Iowa: Automated Enforcement Program Evaluation with Bayesian Analysis. Transportation Research Record: Journal of the Transportation Research Board. 2182:48-54.

Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica. 50(4): 1029-1054.

Haque, M. M., Ching H. C., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis and Prevention. 42(1):203-212.

Hausman, J., Hall, B. H., and Griliches, Z., 1984. " Econometric Models for Count Data with an Application to the Patents-R&D Relationship." Econometrica. 52 (4):909-938.

Helfand, G., Wolverton, A., 2011. Evaluating the Household Response to Fuel Economy: A Review of the Literature. International Review of Environmental and Resource Economics 5, 103-146.

Høye, A., 2013. Still red light for red light cameras? An update. Accident Analysis and Prevention. 55 (2013) 77-89.

Hu, W., McCartt, A.T., Teoh., E.R., 2011. Effect of red light camera enforcement on fatal crashes in large US cities. Journal of Safety Research. 42(4):277-282.

Jacobsen, M., 2013. Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity. American Economic Journal: Economic Policy. 5(2):148-187.

KABC-TV News. 2005. L.A. to Double Number of Red Light Cameras. 1 November. Accessed November 30, 2012. http://abclocal.go.com/kabc/story?section=news/local&id=3592318.

Knittel, C., Metaxoglu, K., 2014. Estimation of Random Coefficient Demand Models: Two Empiricists' Perspective. Review of Economics and Statistics. 96(1), 34-59.

Lerman, S. R., 1977. Location, Housing, Automobile Ownership, and Mode to Work: A Joint Choice Model. Transportation Research Board Record. 610:6-11.

Li, P. 2012. Speeding up the Contraction Mapping Algorithm. In Essays on missing data models and MCMC estimation. Ph.D. Dissertation. UC Irvine Department of Economics.

Manski, C. F., Lerman, S. R., 1977. The Estimation of Choice Probabilities from Choice Based Samples. Econometrica. 45(8): 1977-1988.

Martinez, K.L.H., and Porter, B. E., 2006. Characterizing red light runners following implementation of a photo enforcement program. Accident Analysis and Prevention. 38 (5): 862-870.

McFadden, D., 1978. Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, and J.Weibull (Eds.), Spatial Interaction Theory and Planning Models. North-Holland, Amsterdam: 75-96.

Murphy, K. M., Topel, R. H., 2002. Estimation and Inference in Two-Step Econometrics Models. Journal of Business and Statistics. 20:1 88-97.

Nevo, A., 2000a. Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. Rand Journal of Economics. 31, 395-421.

Nevo, A., 2000b. A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. Journal of Economics and Management Strategy. 9(4), 513-548.

Nevo, A., 2001. Measuring Market Power in the Ready-to-Eat Cereal Industry. Econometrica. 69(2), 307-342.

Newey, W. K., 1984. A Methods of Moments Interpretation of Sequential Estimators. Economics Letters. 14, 201–206.

Petrin, A., 2002. Quantifying the Benefits of New Products: The Case of the Minivan. Journal of Political Economy. 110, 705-729.

Porter, B. E., Johnson, K. L., Bland, J. F., 2013. Turning off the cameras: Red light running characteristics and rates after photo enforcement legislation expired. Accident Analysis and Prevention. 50: 1104-1111.

Retting, R. A., and Kyrychenko, S. Y., 2002. Reductions in Injury Crashes Associated With Red Light Camera Enforcement in Oxnard, California. American Journal of Public Health. 92 (11): 1822-1825.

Retting, R. A., Ferguson, S. A., Hakkert, A. S., 2003. Effects of Red Light Cameras on Violations and Crashes: A Review of the International Literature. Traffic Injury Prevention. 4 (1): 17-23.

Retting, R. A., Williams, A. F., Farmer C. M., and Feldman, A. F., 1999a. Evaluation of red light camera enforcement in Oxnard, California. Accident Analysis and Prevention. 31 (3): 169-174.

Retting, R. A., Williams, A. F., Farmer C. M., and Feldman, A. F., 1999b. Evaluation of Red Light Camera Enforcement in Fairfax, Virginia." ITE Journal 69: 30–34.

Ross, J., Sperley, M., 2011. Red Light Running Camera Assessment Final Report SPR 304-521. Oregon Department of Transportation Research Section 72 (8): 1-56.

Rubin, D. B., 1974. Estimating the Causal Effects of Treatments in Randomized and Non-Randomized Studies. Journal of Educational Psychology. 66 (5): 688-701.

Shin, K., Washington, S., 2007. The impact of red light cameras on safety in Arizona. Accident Analysis and Prevention. 39 (6): 1212-1221.

Skrainka, 2011. A Large Scale Study of the Small Sample Performance of Random Coefficient Models of Demand. Working Paper.

Small, K. A., Verhoef, E. T., 2007. The Economics of Urban Transportation. Routledge: New York. 101-103.

Spiller, E., 2012. Household Vehicle Bundle Choice and Gasoline Demand: A Discrete-Continuous Approach. Working Paper.

Tanner, M. A., and Wong, W. H., 1987. The Calculation of Posterior Distributions by Data Augmentation. Journal of the American Statistical Association. 82 (398): 528-540.

Train, K. E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Train, K. E., Winston, C., 2007. Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers. International Economic Review. 48(4), 1469-1496.

Appendix A

A.1. Derivation of the corrected variance-covariance matrix of the sequential estimator

The variance of any GMM estimator is given by the following formula:

$$Var(\theta_{GMM}) = (M_0' W_0 M_0)^{-1}(M_0' W_0 S_0 W_0 M_0)(M_0' W_0 M_0)^{-1},$$

where

$$M_0 = \begin{bmatrix} \frac{1}{N}\frac{\partial^2 L}{\partial \beta^2} & \frac{1}{N}\frac{\partial^2 L}{\partial \beta \partial \delta} & 0_{K_1 \times K_2} \\ \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \beta} & \frac{1}{N}\sum_m -\frac{\partial P_{mj}}{\partial \delta} & 0_{(J-1)\times K_2} \\ 0_{K_2 \times K_1} & \frac{1}{\sqrt{NJ}}\sum_j z_j' & \frac{1}{J}\sum_j -z_j' x_j \end{bmatrix}$$

where

$$\frac{\partial^2 L}{\partial \beta^2} = -\sum_m \sum_j p_{mj}(w_{mj} - \overline{w}_m)(w_{mj} - \overline{w}_m)'$$

$$\frac{\partial^2 L}{\partial \beta \partial \delta} = -\sum_n \sum_j p_{nj}(w_{nj} - \overline{w}_n)\left(d_j^i - p_{ni}\right)' \forall i = 1, \ldots, J-1;$$

$$\frac{\partial P_{mj}}{\partial \beta} = P_{mj}(w_{mj} - \overline{w}_m)$$

$$\frac{\partial P_{mj}}{\partial \delta_i} = \begin{cases} P_{mj}(1 - P_{mj}) & if\ j = i \\ P_{mj}P_{m-j} & otherwise \end{cases}$$

where $\overline{w}_m = \sum_l p_{ml} w_{ml}$,

$$S_0 = \begin{bmatrix} \frac{1}{N}\sum_m H_{1m}H_{1m}' & \frac{1}{N}\sum_m H_{1m}H_{2m}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{1m}H_{3j}' \\ \frac{1}{N}\sum_m H_{2m}H_{1m}' & \frac{1}{N}\sum_m H_{2m}H_{2m}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{2m}H_{3j}' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{3j}H_{1nm}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j H_{3j}H_{2m}' & \frac{1}{J}\sum_j H_{3j}H_{3j}' \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{N}\{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}\{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}' & \frac{1}{N}\left(\sum_m M_j - P_{mj}\right)\{\sum_j y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j \{\boldsymbol{z_j(\delta - x_j\alpha)}\}\{\sum_m \mathbf{y_{mj}(w_{mj}} - \sum_i \mathbf{P_{mi}w_{mi}})\}' \\ \frac{1}{N}\{\sum_m y_{mj}(w_{mj} - \sum_i P_{mi}w_{mi})\}\left(\sum_m M_j - P_{mj}\right)' & \frac{1}{N}\left(\sum_m M_j - P_{mj}\right)\left(\sum_m M_j - P_{mj}\right)' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j \{\boldsymbol{z_j(\delta - x_j\alpha)}\}\left(\sum_m \boldsymbol{M_j - P_{mj}}\right)' \\ \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j \{\sum_m \mathbf{y_{mj}(w_{mj}} - \sum_i \mathbf{P_{mi}w_{mi}})\}\{\boldsymbol{z_j(\delta - x_j\alpha)}\}' & \frac{\mathbf{1}}{\sqrt{\boldsymbol{NJ}}}\sum_m \sum_j \left(\sum_m \boldsymbol{M_j - P_{mj}}\right)\{\boldsymbol{z_j(\delta - x_j\alpha)}\}' & \frac{1}{J - K_2}\sum_j z_j(\delta - x_j\alpha)(\delta - x_j\alpha)' z_j' \end{bmatrix}$$

If the model is just identified (as is assumed in this application), then it is sufficient to set $W_0 = I$. In an over identified setting, the two-step optimal GMM procedure can be used to generate the appropriate estimate of W_0.

The expectations for the first two moments, $G_1$and G_2, are taken across households while the expectation for the third moment, G_3, is taken across alternatives. This means that the estimates, β and δ, obtained from minimization of $G_1$and G_2, are consistent and asymptotically efficient as $N \to \infty$ while the estimate of α obtained from G_3 is consistent and asymptotically efficient as $J \to \infty$. Berry, Linton and Pakes, 2002, note that for asymptotic normality to hold, N must increase at rate J^2.

Because of this, there is a peculiarity to the estimation of S_0 that is worth noting. $S_0$is the sum of the outer product of the contributions to the moment function. While the cross product of H_{1m} and H_{2m} is simply $H_{1m}H_{2m}'$, calculating the other cross products is not straightforward as there is a dimensionality mismatch between H_{1m} and H_{3j}, and, H_{2m} and H_{3j} because there are N observations in H_{1m} and H_{2m} and only J observations in H_{3j}.

The solution to this is to divide each of the N moment contributions in H_{1m} and H_{2m} by $\sqrt{N}$ and to divide each of the J moment contributions in H_{3m} by $\sqrt{J}$. Then, each of the N moment contributions in $H_{1m}$and H_{2m} is multiplied with each of the J moment contributions in H_{3j}, and these quantities are summed. This produces the formulas in entries (3,1), (3,2), (1,3) and (2,3) of S_0. Dividing by $\sqrt{N}$ and $\sqrt{J}$ ensures that these quantities converge to finite values as J and $N \to \infty$.

A.2. Data Generating Process for the Monte Carlo Study

$$y_{mj} = \sum_{l_m=1}^{r_m} y_{nj} \; \forall n \in m$$

$$y_{nj} = \begin{cases} 1 & if\ U_{nj} > U_{ni} \quad \forall\ i \neq j \\ 0 & otherwise. \end{cases}$$

$$U_{nj} = \delta_j + w_{mj1}\beta_1 + w_{mj2}\beta_2 + \epsilon_{nj}, \quad where\ n \in m$$

$$\delta_j = \alpha_0 + x_{j1}\alpha_{11} + x_{j2}\alpha_{12} + p_j\alpha_2 + \xi_{1j},$$

$$p_j = z_j\gamma + \xi_{2j}$$

$$(\xi_{1j}, \xi_{2j}) \sim N(0_2, \Omega), \quad \Omega = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

where

x_{j1} is a $J \times 1$ vector of binary values, drawn from a Bernoulli distribution with success probability of 0.5,

x_{j2} is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2,

v_{m1} and v_{m2} are $M \times 1$ vectors of continuous values drawn from a standard normal distribution bounded between -2 and 2,

w_{mj1} is an $MJ \times 1$vector of interactions between v_{m1} and x_{j1} ,

w_{mj2} is an $MJ \times 1$vector of interactions between v_{m2} and x_{j2} , and

z_j is a $J \times 1$ vector of continuous values drawn from a standard normal distribution bounded between -2 and 2.

Finally, $\beta_1 = 0.8$, $\beta_2 = -0.7$, $\alpha_0 = 0$, $\alpha_{11} = -0.5$, $\alpha_{12} = 1$, $\alpha_2 = 0.5$, and $\gamma = 1.3$.

A.3. Results from the BLP Model with Sampling Weights

The household sample obtained from the 2009 National Household Transportation Survey is not a representative sample. Households in 20 regions were oversampled because metropolitan transportation planning organizations in those regions sponsored larger samples for their own use. In addition, a single interview was conducted for each household between April 2008 and May 2009. Households who were interviewed earlier are more likely to have purchased model year 2008 vehicles after their NHTS interview, and these purchases are not reflected in the sample.

Estimates from the choice model will not be consistent under this sampling scheme if household heterogeneity along the stratified dimensions are not fully accounted for in the model specification. The main results of this paper assume that heterogeneity is fully specified. To test this assumption, I estimate the BLP model for broad choice data incorporating sampling weights in estimation that allow the data to approximate a simple random sample.

Unfortunately, in estimation, a few variables no longer converge once weights are used. These variables are:

College X Prestige X Japan

College X Prestige X Europe

Fuel Operating Cost

College X Fuel Operating Cost

Because the "fuel operating cost" variables no longer converge, to capture the effect of fuel efficiency on utility, "gallons per miles" is included in the model. Since the estimates from the weighted and unweighted specifications are no longer directly comparable, in Table C1, I present the implied discount rates from both models for the average household:

Table A1: Implied Discount Rates from household willingness to pay for fuel efficiency

	Willingness to Pay (thousands)*	Implied Discount Rate†
BLP Broad Choice without Weights (I)	-0.263	Undefined
BLP Broad Choice without Weights (II)	8.213	-13.809
BLP Broad Choice with Weights	0.526	-19.710

Notes:
**For the BLP Broad Choice without Weights, willingness to pay is for a 1 cent/mile improvement in fuel operating cost. For BLP Broad Choice with Weights, willingness to pay is for a gallon per mile improvement in fuel efficiency. BLP Broad Choice without Weights (I) takes the weighted average of willingness to pay across income and college groups. BLP Broad Choice without Weights (II) takes the weighted average of willingness to pay across income and college groups assuming that the willingness to pay for households with income greater than $100,000 is zero.*
† Implied interest rates are calculated assuming households drive vehicles 18,778 miles a year and hold vehicles for 14 years.

BLP Broad Choice without weights (I) yields an undefined interest rate because the average willingness to pay for a one cent/mile reduction in fuel operating cost across the sample is negative. This is driven by the fact that the coefficient on vehicle price, used to construct the willingness to pay measure, is positive for the households in the highest income category. BLP Broad Choice without weights (II) forces the negative willingness to pay values of high income households to equal zero. This results in an average willingness to pay of $8,213.95 and an implied discount rate of -14.8%.

The BLP Broad Choice with weights estimates that households are willing to pay $525.58 for a one gallon per mile improvement in fuel efficiency. This implies a discount rate of -19.71%. These findings provide some evidence that the model is robust to the use of sampling weights. All three methods in the table generate implied interest rates that suggest an overvaluation of fuel efficiency. When a zero-bound constraint is imposed in aggregating the willingness to pay estimates from the model without weights, it yields an implied interest rate that is qualitatively similar to the implied interest from the model estimated with weights (-14% vs -19%).

A.4. Full Results from the Three Vehicle Choice Specifications

Table A2: BLP model with aggregated choices: parameter estimates

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.065	0.004 ***	0.0136 ***
(Price) × (Income>100,000)	0.102	0.004 ***	0.0153 ***
(Price) × (Income Missing)	0.094	0.005 ***	0.0153 ***
(Prestige) × (Urban)	0.609	0.095 ***	0.2165 ***
(Prestige) × (Income>100,000)	0.167	0.091 *	0.1652
(Performance Car) × (Income>100,000)	0.242	0.103 **	0.1427 *
(Japan) × (Urban)	0.367	0.048 ***	0.1040 ***
(Van) × (Children under 15)	0.849	0.084 ***	0.1238 ***
(Large SUV) × (Children under 15)	0.157	0.056 ***	0.0776 **
(Small SUV) × (Children under 15)	0.503	0.112 ***	0.1683 ***
(Korea) × (Rural)	-0.614	0.105 ***	0.1408 ***
(Seats≥5) × (Household Size≥4)	0.406	0.032 ***	0.1026 ***
(Mid-Large Car) × (Retired)	0.953	0.050 ***	0.0977 ***
(Prestige) × (Retired)	0.575	0.073 ***	0.1396 ***
(Import) × (College)	0.398	0.047 ***	0.0694 ***
(Prestige) × (Japan) × (College)	-0.127	0.177	0.2941
(Prestige) × (Europe) × (College)	-0.598	0.139 ***	0.2156 ***
(Prestige) × (Japan) × (Urban)	-0.180	0.154	0.3329
(Performance Car) × (College)	0.622	0.099 ***	0.1671 ***
Fuel Operating Cost (cents per mile)	-2.877	0.053 ***	0.9533 ***
(Fuel Operating Cost) × (College)	-0.061	0.009 ***	0.0202 ***
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.116	0.019 ***	0.026 ***
Horsepower/Curb weight	158.582	24.622 ***	53.803 ***
Hybrid	-13.222	0.815 ***	4.102 ***
Curb weight	7.569	0.309 ***	2.511 ***
Wagon	0.892	1.248	1.384
Mid-Large Car	-0.685	0.540	0.548
Performance Car	-0.354	0.850	0.889
Small-Medium Pickup	6.196	1.087 ***	2.101 ***
Large Pickup	5.231	1.155 ***	1.454 ***
Small-Mid SUV	9.292	1.668 ***	3.198 ***
Large SUV	2.275	0.472 ***	0.866 ***

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table A3: BLP model with McFadden's method: parameter estimates

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.001	0.002	0.067
(Price) × (Income>100,000)	0.004	0.001 ***	0.056
(Price) × (Income Missing)	0.011	0.004 ***	0.080
(Prestige) × (Urban)	0.620	0.094 ***	2.735
(Prestige) × (Income>100,000)	1.195	0.072 ***	2.257
(Performance Car) × (Income>100,000)	0.712	0.112 ***	1.121
(Japan) × (Urban)	0.363	0.048 ***	1.625
(Van) × (Children under 15)	0.946	0.084 ***	5.639
(Large SUV) × (Children under 15)	0.251	0.055 ***	2.339
(Small SUV) × (Children under 15)	0.679	0.111 ***	4.999
(Korea) × (Rural)	-0.597	0.105 ***	3.655
(Seats≥5) × (Household Size≥4)	0.427	0.035 ***	0.222 *
(Mid-Large Car) × (Retired)	0.891	0.049 ***	3.914
(Prestige) × (Retired)	0.429	0.072 ***	1.096
(Import) × (College)	0.437	0.047 ***	1.960
(Prestige) × (Japan) × (College)	-0.096	0.176	9.624
(Prestige) × (Europe) × (College)	-0.607	0.139 ***	11.408
(Prestige) × (Japan) × (Urban)	-0.195	0.154	2.158
(Performance Car) × (College)	0.691	0.113 ***	1.676
Fuel Operating Cost (cents per mile)	-2.946	0.056 ***	0.263 ***
(Fuel Operating Cost) × (College)	-0.027	0.009 ***	0.466
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.064	0.020 ***	0.120
Horsepower/Curb weight	144.232	26.475 ***	93.200
Hybrid	-12.288	0.880 ***	3.364 ***
Curb weight	7.084	0.398 ***	1.907 ***
Wagon	-0.054	1.342	1.403
Mid-Large Car	-0.572	0.615	4.454
Performance Car	-0.517	0.928	2.251
Small-Medium Pickup	6.817	1.201 ***	2.220
Large Pickup	4.777	1.352 ***	3.047
Small-Mid SUV	2.973	0.604 ***	0.885 ***
Large SUV	2.676	1.054 **	4.440

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Table A4: BLP model with broad choice data: parameter estimates

w_{nj}	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
(Price) × (75,000<Income<100,000)	0.038	0.006 ***	0.052
(Price) × (Income>100,000)	0.123	0.008 ***	0.100
(Price) × (Income Missing)	0.079	0.006 ***	0.056
(Prestige) × (Urban)	-1.253	0.137 ***	0.698 *
(Prestige) × (Income>100,000)	-1.591	0.167 ***	1.381
(Performance Car) × (Income>100,000)	1.518	0.212 ***	2.159
(Japan) × (Urban)	-0.453	0.075 ***	0.662
(Van) × (Children under 15)	0.788	0.144 ***	0.934
(Large SUV) × (Children under 15)	0.707	0.166 ***	1.160
(Small SUV) × (Children under 15)	0.352	0.084 ***	0.584
(Korea) × (Rural)	0.341	0.157 **	0.917
(Seats≥5) × (Household Size≥4)	-1.664	0.351 ***	1.063
(Mid-Large Car) × (Retired)	0.108	0.080	0.287
(Prestige) × (Retired)	-0.415	0.102 ***	0.339
(Import) × (College)	0.362	0.074 ***	0.518
(Prestige) × (Japan) × (College)	-0.294	0.194	1.338
(Prestige) × (Europe) × (College)	-1.747	0.296 ***	2.043
(Prestige) × (Japan) × (Urban)	-0.420	0.177 **	1.156
(Performance Car) × (College)	2.815	0.327 ***	4.622
Fuel Operating Cost (cents per mile)	-0.599	0.048 ***	2.044
(Fuel Operating Cost) × (College)	-0.057	0.013 ***	0.076
x_j	Estimated Parameter	Uncorrected Standard Error	Corrected Standard Error
Price	-0.098	0.008 ***	0.097
Horsepower/Curb weight	20.737	8.960 **	111.690
Hybrid	-2.193	0.727 ***	9.051
Curb weight	0.002	0.000 ***	0.006
Wagon	0.048	0.454	1.209
Mid-Large Car	2.258	0.317 ***	2.185
Performance Car	-3.280	0.392 ***	5.066
Small-Medium Pickup	-1.179	0.464 **	3.228
Large Pickup	-0.777	0.392 **	3.035
Small-Mid SUV	2.888	0.304 ***	1.030 ***
Large SUV	2.697	0.496 ***	1.760

*Notes: * denotes significance at the 10% level. ** denotes significance at the 5% level. *** denotes significance at the 1% level.*

Appendix B

B.1. Dispersion in the Poisson-Log Normal Model

Let Σ_i be the covariance matrix of $(y_i b_i)$:

$$\Sigma_i = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}_{(g+1)\times(g+1)}$$

where $\Sigma_{11}$is the unconditional variance of y_i, Σ_{22} is the $g \times g$ unconditional covariance matrix of the $b_i's$ and $\Sigma_{12} = \Sigma_{21}{}'$ is the covariance between y_i and b_i.

Note that y_i is not equidispersed in equation (3.2) as it is in equation (3.1); unlike in the standard Poisson model, the variance of y_i is not equal to its mean because according to the partition inverse theorem:

$$\begin{aligned} \Sigma_{11} &= \Sigma_{11|2} + \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21} \\ &= E(\mu_{it}) + \Sigma_{12}D^{-1}\Sigma_{21}\ . \end{aligned}$$

B.2. Estimating the model using the 2006 - 2008 time period

An additional specifications of the model is considered to address the issue of decay in enforcement of the program over time. This is a re-estimation of the model for the period of 2006 to 2008. Media reports highlighting fine collection issues first surfaced around March 2010 (Connell, 2010) suggesting that people started ignoring citations in 2009 and early 2010. Hence, the omission of observations from 2009 and 2010 controls for the lack of legal bite that plagued the program. Table B3 presents the results from this specification for all three control rings.

Results for the 0.5-mile ring in Table B3 tell a similar story as those in Table 3.4, though magnitudes of estimates are larger with the exception of rear-end and injury collisions. The treatment effect for "all collisions" rises to 21% (compared to 17% in Table 3.4), right-angle collisions to 35% (24% in Table 3.4), and rear-end collisions to 32% (34% in Table 3.4). Injury collisions, however, fall to 16% (22% in Table 3.4). In addition, this specification estimates a significant decrease in red light collisions of 54% as opposed to a non-significant decrease of 12% in Table 3.4. The difference in results across specifications for right-angle and red light collisions suggest that the decay dampened the effect of increases in cameras on right-angle collisions and reduced the positive effect that cameras have on red light related collisions.

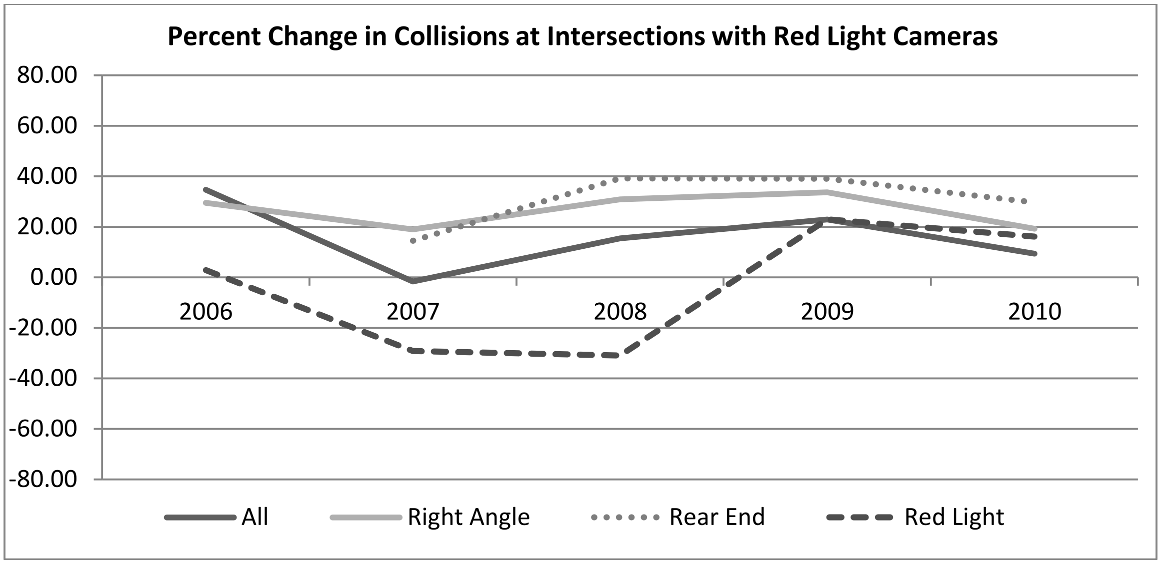


Figure B1.1: The marginal effect of treatment on collisions by year using 1-mile control ring
Estimate of marginal effect on rear-end collisions in 2006 omitted because it is an implausible outlier

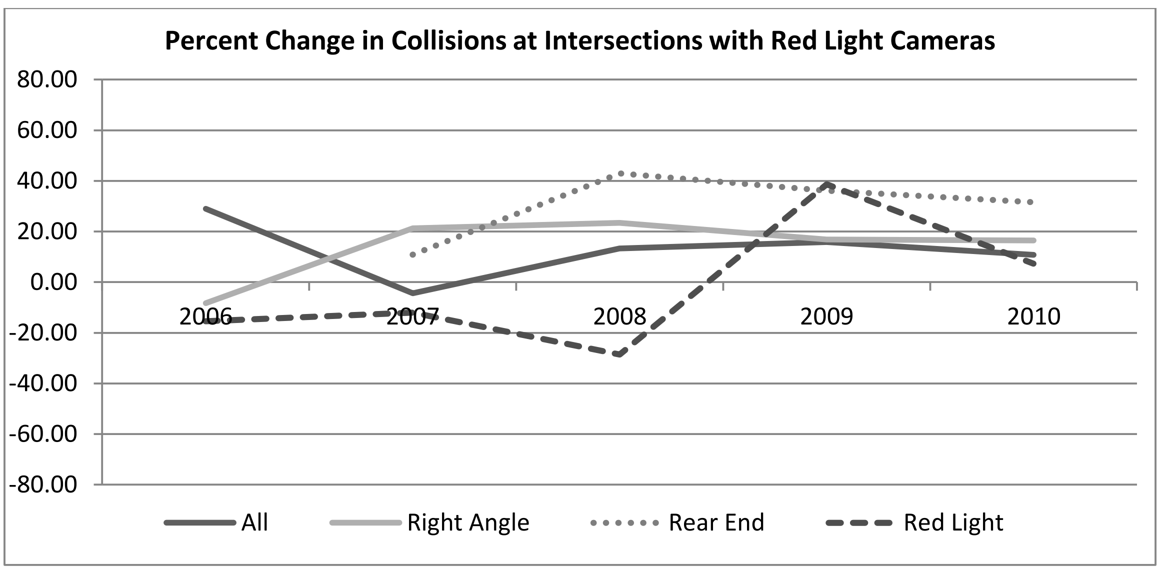


Figure B1.2: The marginal effect of treatment on collisions by year using 2-mile control ring†
Estimate of marginal effect on rear-end collisions in 2006 omitted because it is an implausible outlier

Table B1: Descriptive statistics of intersection-month collision counts, before and after treatment for the five types of collisions, by treatment status

Type of Collision			Treated intersections		0.5-mile controls		1-mile controls		2-mile controls	
			Before	After	Before	After	Before	After	Before	After
All	Number of Collisions	0	199	485	833	2516	859	3446	862	2688
		1	158	505	560	1626	535	2020	499	1391
		2	104	255	265	712	210	906	222	626
		3	35	109	84	139	97	327	90	208
		4	11	36	21	68	43	103	30	64
		5	4	13	6	24	13	32	4	26
		6	2	3	1	3	1	4	1	4
		7	1	0	0	1	0	1	1	3
		8	0	0	0	1	0	1	0	1
		Mean	1.0778	1.1159	0.8266	0.8077	0.8476	0.7741	0.7993	0.7396
		Std Dev	0.0511	0.0299	0.0233	0.0138	0.0257	0.0138	0.0246	0.0142
Right Angle	Number of Collisions	0	360	977	1340	3975	1347	3978	1311	3965
		1	116	336	363	998	336	923	333	866
		2	30	81	59	191	63	155	54	156
		3	5	12	8	21	10	21	10	18
		4	3	0	0	4	1	5	1	5
		5	0	0	0	1	1	0	0	1
		6	0	0	0	0	0	0	0	0
		7	0	0	0	0	0	0	0	0
		8	0	0	0	0	0	0	0	0
		Mean	0.3949	0.3798	0.2853	0.2820	0.2850	0.2589	0.2785	0.2508
		Std Dev	0.0307	0.0169	0.0130	0.0077	0.0137	0.0075	0.0135	0.0075
Rear End	Number of Collisions	0	391	1020	1440	4215	1421	4144	1393	4168
		1	103	320	285	872	271	790	259	735
		2	19	58	40	93	56	131	48	97
		3	0	7	5	9	9	15	9	10
		4	0	1	0	1	1	2	0	1
		5	1	0	0	0	0	0	0	0
		6	0	0	0	0	0	0	0	0
		7	0	0	0	0	0	0	0	0
		8	0	0	0	0	0	0	0	0
		Mean	0.2840	0.3279	0.2147	0.2098	0.2355	0.2174	0.2235	0.1921
		Std Dev	0.0248	0.0156	0.0114	0.0064	0.0127	0.0069	0.0124	0.0064

Red Light	Number of Collisions	0	439	1275	1575	4692	1569	4617	1515	4588
		1	64	128	185	463	181	440	180	392
		2	11	3	10	30	7	24	14	26
		3	0	0	0	4	1	1	0	3
		4	0	0	0	1	0	0	0	2
		5	0	0	0	0	0	0	0	0
		6	0	0	0	0	0	0	0	0
		7	0	0	0	0	0	0	0	0
		8	0	0	0	0	0	0	0	0
		Mean	0.1673	0.0953	0.1158	0.1038	0.1126	0.0966	0.1217	0.0919
		Std Dev	0.0188	0.0080	0.0080	0.0046	0.0080	0.0043	0.0085	0.0045
Injury	Number of Collisions	0	304	708	1198	3133	1217	3180	1220	3210
		1	170	441	542	1368	509	1282	484	1227
		2	62	156	164	415	146	364	148	358
		3	18	40	26	82	36	76	32	79
		4	3	15	6	22	13	12	4	17
		5	1	1	1	3	1	3	0	1
		6	1	0	0	0	1	0	0	0
		7	0	0	0	0	0	0	0	0
		8	0	0	0	0	0	0	0	0
		Mean	0.6637	0.6892	0.5049	0.5070	0.5055	0.4679	0.4724	0.4605
		Std Dev	0.0377	0.0237	0.0168	0.0107	0.0179	0.0103	0.0168	0.0104
Number of Observations			514	1406	1770	5190	1758	5082	1709	5011

Table B2.1: The marginal effect of treatment on all, right-angle, rear-end, and red light collisions from 2006 to 2010 by year using 1-mile control ring†

Type of Collision	Year	Marginal Effect	95% Probability Interval		% Change
All	2006	0.3867	-0.1219	0.9104	34.65
	2007	-0.0189	-0.1756	0.1431	-1.69
	2008	0.1718	0.0381	0.3131	15.39
	2009	0.2547	0.1223	0.3833	22.82
	2010	0.1043	-0.0216	0.2322	9.34
Right-angle	2006	0.1120	-0.1799	0.4391	29.49
	2007	0.0718	-0.0228	0.1626	18.89
	2008	0.1170	0.0374	0.1999	30.80
	2009	0.1275	0.0485	0.2060	33.58
	2010	0.0731	-0.0036	0.1432	19.24
Rear-end	2006	0.5844	0.3055	0.9119	178.24
	2007	0.0473	-0.0194	0.1203	14.42
	2008	0.1285	0.0641	0.1939	39.18
	2009	0.1278	0.0668	0.1938	38.99
	2010	0.0972	0.0405	0.1586	29.63
Red Light	2006	0.0027	-0.1088	0.1555	2.81
	2007	-0.0278	-0.0648	0.0120	-29.19
	2008	-0.0295	-0.0540	-0.0042	-30.97
	2009	0.0218	-0.0150	0.0606	22.93
	2010	0.0153	-0.0164	0.0508	16.09

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000. The following non-informative priors are used for all regression

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N(0_g, I_g \times 100), \quad D^{-1} \sim Wishart(g+4, I_g).$$

†Injury collision results not provided because model EM algorithm failed to converge

Table B2.2: The marginal effect of treatment on all, right-angle, rear-end, and red light collisions from 2006 to 2010 by year using 2-mile control ring†

Type of Collision	Year	Marginal Effect	95% Probability Interval		% Change
All	2006	0.3236	-0.1699	0.8963	28.99
	2007	-0.0484	-0.2034	0.1113	-4.34
	2008	0.1479	-0.0026	0.2974	13.25
	2009	0.1771	0.0335	0.3176	15.87
	2010	0.1197	-0.0208	0.2511	10.73
Right-angle	2006	-0.0315	-0.3001	0.2573	-8.30
	2007	0.0808	-0.0145	0.1744	21.27
	2008	0.0890	0.0117	0.1689	23.43
	2009	0.0638	-0.0152	0.1414	16.79
	2010	0.0624	-0.0074	0.1337	16.44
Rear-end	2006	0.5517	0.2653	0.8850	168.26
	2007	0.0355	-0.0385	0.1120	10.83
	2008	0.1407	0.0652	0.2168	42.91
	2009	0.1183	0.0503	0.1950	36.07
	2010	0.1034	0.0403	0.1695	31.54
Red Light	2006	-0.0147	-0.1262	0.1381	-15.41
	2007	-0.0115	-0.0465	0.0303	-12.09
	2008	-0.0273	-0.0512	0.0001	-28.60
	2009	0.0369	-0.0015	0.0776	38.71
	2010	0.0069	-0.0221	0.0397	7.27

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000. The following non-informative priors are used for all regression

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N\left(0_g, I_g \times 100\right), \quad D^{-1} \sim Wishart\left(g+4, I_g\right).$$

†Injury collision results not provided because model EM algorithm failed to converge

Table B3: Marginal effect of treatment on all, right-angle, rear-end, red light and injury collisions from 2006 to 2008 using 0.5-mile, 1-mile and 2-mile control rings

Type of Collision	Control Intersections	Marginal Effect of Treatment	95% Probability Interval		% Change
All	0.5-mile	0.2367	0.1637	0.3099	20.57
	1-mile	0.1488	0.0555	0.2415	12.94
	2-mile	0.0873	0.0058	0.1656	7.59
Right-angle	0.5-mile	0.1400	0.0956	0.1848	34.76
	1-mile	0.1151	0.0735	0.1577	28.56
	2-mile	0.1385	0.0959	0.1844	34.38
Rear-end	0.5-mile	0.1122	0.0743	0.1503	31.97
	1-mile	0.1155	0.0773	0.1537	32.91
	2-mile	0.1353	0.0959	0.1752	38.54
Red Light	0.5-mile	-0.0421	-0.0579	-0.0259	-53.70
	1-mile	0.1158	0.0778	0.1554	147.77
	2-mile	-0.0215	-0.0370	-0.0036	-27.40
Injury	0.5-mile	0.1138	0.0650	0.1651	16.49
	1-mile	0.1643	0.1120	0.2190	23.82
	2-mile	0.1823	0.1306	0.2373	26.42

Estimates based on 10,000 Markov chain Monte Carlo draws with a burn-in of 1000. The following non-informative priors are used for all regression

$$\beta \sim N(0_k, I_k \times 100), \quad \eta \sim N\left(0_j, I_j \times 100\right), \quad D^{-1} \sim Wishart\left(j+4, I_j\right).$$