SFB 823

Heterogeneous rebound effects: Comparing estimates from discrete-continuous models

Manuel Frondel, Fernanda Martinez Flores, Colin Vance

Discussion Paper

Nr. 2/2016

Heterogeneous Rebound Effects: Comparing Estimates from Discrete-Continuous Models

Manuel Frondel[1,2], Fernanda Martinez Flores[1,2], and Colin Vance[1,3]

[1]Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
[2]Ruhr University Bochum (RUB)
[3]Jacobs University Bremen

January 2016

Abstract: Discrete-continuous models have become a common technique for addressing selectivity biases in data sets with endogenously partitioned observational units. Alternative two-stage approaches have been suggested by LEE (1983), DUBIN and MCFADDEN (1984), and DAHL (2002), all of which capture the first-stage discrete choice by the multinomial logit model, while the second-stage outcome equation is estimated using OLS. The nonlinearity introduced by the selection bias correction implies that the second-stage coefficients cannot be interpreted as marginal effects. Instead, the marginal effects are obtained using the estimates from both the first and second stages, a step that has been widely neglected in the applied literature. After deriving formulae for the marginal effects obtained from these selection correction approaches, we estimate a joint model of automobile ownership and distance driven to quantify the rebound effect, the behaviorally induced increase in driving that results from higher fuel economy. Our example illustrates that the pattern of rebound effects varies substantially depending on the method of selection bias correction.

JEL Classifications: D12, Q21, Q41

Keywords: Discrete-continuous models; marginal effects; car use.

Correspondence: Manuel Frondel, RWI, Hohenzollernstr. 1-3, D-45128 Essen, Germany, Email: frondel@rwi-essen.de.

Acknowledgements: We are grateful for invaluable comments and suggestions by Christoph M. SCHMIDT and Reinhard MADLENER. This work has been supported by the NRW Ministry of Innovation, Science, and Research (BMBF) within the framework of the project "Rebound effects in NRW" and by the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG), within the framework of Project A3, "Dynamic Technology Modeling".

1 Introduction

Discrete-continuous models have become a common technique for addressing selectivity biases in data sets with endogenously partitioned observational units. In the transportation literature, these models have often been used to analyze the inter-related choices of vehicle ownership and vehicle utilization (e. g. WEST, 2004; BENTO et al. 2005, 2009; FRONDEL, VANCE, 2009). The application of such approaches also extends to other contexts including consumption expenditures (KLEE, 2008), energy use (MANSUR, MENDELSOHN, MORRISON, 2008; MILLS, SCHLEICH, 2014), and agricultural production (BLANC et al., 2008).

When selection covers a large number of discrete choices, LEE (1983), DUBIN and MCFADDEN (1984), and DAHL (2002) provide three alternative methods for handling the selection process, all of which are based on a two-step estimation method. In the first stage, a multinomial logit model is employed to derive selection correction terms that are included as explanatory variables in the second-stage specification. The coefficients from this second-stage model, however, cannot be directly interpreted as marginal effects. Applied analyses often either sidestep this issue by limiting the interpretation to the sign and significance of the coefficients, or they incorrectly interpret the coefficients as marginal effects. This note derives the formulae of the marginal effects resulting from each of those three discrete-continuous models and illustrates that the marginal effects can vary substantially depending on the method chosen.

Our application focuses on estimating the magnitude of the direct rebound effect, the behaviorally induced offset in the reduction of energy consumption following an improvement in energy efficiency that lowers the per-kilometer costs of driving. Recent literature on the determinants of car mileage has demonstrated that estimates of the role of fuel costs may be subject to considerable heterogeneity. WADUD et al. (2010), for example, include interaction terms in a random-effects model to allow for differential fuel price elasticities according to the household's income, geographic location, and other socioeconomic attributes. Similarly, MATIASKE, MENGES, and SPIESS (2012) capture heterogeneity in the fuel efficiency elasticity via a quadratic specification that allows the magnitude of the elasticity to vary with the level of fuel efficiency. More recently, LIN and PRINCE (2013) examine whether the magnitude of the fuel price elasticity varies with the degree of fuel price volatility.

The present study contributes to this line of inquiry by exploring how rebound effects vary according to the number of cars owned. Focusing on petrol car drivers, we draw on travel survey data from Germany to estimate a joint model of automobile ownership and distance driven, from which observation-specific rebound effects are estimated and presented graphically. Our analysis suggests that while the mean rebound effects, estimated at about 45-50% for single-car households, hardly differ between single- and multiple-car households for the DUBIN-MCFADDEN and DAHL approaches, the observation-specific estimates are characterized by substantial heterogeneity within these two household classifications.

2 Selection Bias Correction Approaches

We begin by presenting the selection bias correction approaches proposed by LEE (1983), DUBIN and MCFADDEN (1984), and DAHL (2002), thereby following closely BOURGUIGNON, FOURNIER and GURGAND (2007) and adopting their unified framework. These authors provide for a detailed discussion of the advantages and shortcomings of these approaches and model the first-stage decision by the following latent-variable equation:

$$y_j^* = \gamma_j^T \mathbf{z} + \eta_j, \qquad j = 1, ..., M, \tag{1}$$

where T indicates the transposition of a vector and j designates one out of M exclusive choice decisions that are based on utilities y_j^*. In our example, the first-stage decision refers to car ownership and we consider $M = 3$ alternatives, with the base case $j = 1$ standing for households that do not possess a car, $j = 2$ denoting single-car households, and $j = 3$ multi-car households. Vector $\mathbf{z}$ represents the maximum set of explanatory variables for all alternatives.

The second-stage equation of interest, capturing in our case the decision on the distance traveled with all household vehicles, is given by

$$y = \boldsymbol{\beta}^T \mathbf{x} + u, \tag{2}$$

where $\mathbf{x}$ is another parameter vector that is assumed to differ from $\mathbf{z}$ in at least one variable for the model to be non-parametrically identified. For disturbance u, it is assumed that $E(u|\mathbf{x}, \mathbf{z}) = 0$ and $Var(u|\mathbf{x}, \mathbf{z}) = \sigma^2$.

Without any loss of generality, it can be assumed that out of the M choices, $j = k$ will be selected and, hence, y_k will be observed, as

$$y_k^* > \max_{j \neq 1} y_j^*. \tag{3}$$

Defining $\varepsilon_k := \max_{j \neq k}(y_j^* - y_k^*) = \max_{j \neq k}(\gamma_j^T \mathbf{z} + \eta_j - \gamma_k^T \mathbf{z} - \eta_k)$, condition (3) equals

$$\varepsilon_k < 0. \tag{4}$$

Assuming that the disturbances η_j are independent and identically distributed, which implies the well-known IIA hypothesis of the independence from irrelevant alternatives, and, furthermore, that the η_j's follow a Gumbel or Type I extreme value distribution, LUCE (1959) developed the multinomial logit model:[1]

$$P_k := P(\varepsilon_k < 0|\mathbf{z}) = \frac{\exp(\gamma_k^T \mathbf{z})}{\sum_m \exp(\gamma_m^T \mathbf{z})}, \tag{5}$$

where P_k denotes the probability for observing alternative k and parameter estimates can be obtained using maximum likelihood methods (MCFADDEN, 1974).

In their seminal article, DUBIN and MCFADDEN (1984), henceforth DMF, conceptualize the idea that both the demand for durable goods, such as cars, and their use may not be the result of independent consumer decisions, but may depend on common factors, such as household size and income. That is, in formal terms, the disturbance of the outcome equation, u, may be correlated with the disturbances η_j of the choice equations. As a consequence, least squares estimates of the parameters $\boldsymbol{\beta}$ of the outcome equation would not be consistent.

Among other methods, such as the instrumental variable approach, DMF propose the Conditional Expectation Correction Method to obtain consistent estimators and the challenge is to consistently estimate $\boldsymbol{\beta}$ by taking account of the – generally non-vanishing – conditional mean of u:

$$y = \boldsymbol{\beta}^T \mathbf{x} + E(u|\varepsilon_k < 0, \Gamma) + w, \tag{6}$$

where Γ is defined by $\Gamma := (\gamma_1^T \mathbf{z}, ..., \gamma_M^T \mathbf{z})^T$ and w is a residual that is mean-independent of the regressors.

In extending the two-step selection bias correction method introduced by HECKMAN (1979), which is adequate when selection is just among two choices, several approaches have been proposed in the literature that differ in the concrete specification of the conditional mean $E(u|\varepsilon_k < 0, \Gamma)$, which, ultimately, is a function $\mu(P_1, ..., P_M)$ of the multinomial logit probabilities P_j. In other words, the bias correction approaches differ in the way the function $\mu(P_1, ..., P_M)$ is specified and they also distinguish in the restrictions that are imposed on $\mu(P_1, ..., P_M)$. These restrictions are of two types: restrictions on the covariance matrix of the error terms and linearity assumptions.

First, DMF invoked a linearity assumption for the disturbances η_j of the choice

[1]For the Gumbel distribution, the cumulative distribution and density functions read $G(\eta) = \exp(-e^{-\eta})$ and $g(\eta) = \exp(-\eta - e^{-\eta})$, respectively.

equations, so that the outcome equation is given by[2]

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma \frac{\sqrt{6}}{\pi} \left[\sum_{j \neq k} r_j \frac{P_j \ln(P_j)}{1 - P_j} - r_k \ln(P_k) \right] + w, \tag{7}$$

where r_j is a correlation coefficient between u and η_j and $\lambda_{DMFj} := \sigma \frac{\sqrt{6}}{\pi} r_j$ is a set of M coefficients to be estimated.[3]

Second, in a widely quoted, earlier article, LEE (1983) proposed a generalization of the HECKMAN method that allows for any parameterized error term, rather than normally distributed errors, and in contrast to the DMF approach involves only a single correction term:

$$y = \boldsymbol{\beta}^T \mathbf{x} - \sigma \rho_k \frac{\phi(\Phi^{-1}(P_k))}{P_k} + w, \tag{9}$$

where Φ is the cumulative distribution function of the standard normal distribution, ϕ is the respective density, correlation parameter ρ_k is a scalar, and $\lambda_{Lee} := \sigma \rho_k$ is the coefficient of the selection term to be estimated. In this approach, the single correction term includes only the probability P_k to be selected on the observed outcome k.[4] Similar to the DMF approach, consistent estimates of $\boldsymbol{\beta}$ can be obtained by first estimating $\widehat{P_k}$ on the basis of a multinomial logit model and then using these probability estimates to calculate an estimate of the selection correction regressor $\phi(\Phi^{-1}(P_k))/P_k$. Adding this regressor to the outcome equation (2) allows for employing OLS to get unbiased estimates of $\boldsymbol{\beta}$, provided that the restrictions underlying this approach hold true.

Third, DAHL (2002) has recently proposed to restrict the set of probabilities in cor-

[2]DMF's linearity assumption for $E(u|\varepsilon_k < 0, \Gamma)$, expressed in terms of the η_j's, rather than ε_k, represents a linear combination of the standardized disturbances, $(\eta_j - E(\eta_j)/\sqrt{Var(\eta_j)}$:

$$E(u|\eta_1, ..., \eta_M) = \sigma \frac{\sqrt{6}}{\pi} \sum_{j=1,...,M} r_j(\eta_j - E(\eta_j)),$$

as $Var(\eta_j) = \pi^2/6$ for all $j = 1, ..., M$.

[3]By additionally imposing the restriction $r_1 + ... + r_M = 0$ on (7), DMF reduced the number of coefficients c_j to be estimated from M to $M - 1$, yielding the following outcome equation:

$$y = \boldsymbol{\beta}^T \mathbf{x} + \sigma \frac{\sqrt{6}}{\pi} \sum_{j \neq k} r_j \left[\frac{P_j \ln(P_j)}{1 - P_j} + \ln(P_k) \right] + w, \tag{8}$$

BOURGUIGNON, FOURNIER and GURGAND (2007) criticize the imposition of restriction $r_1 + ... + r_M = 0$ as unnecessary, as the correction function $\mu(P_1, ..., P_M)$ is non-linear, rather than linear in the probabilities. In their Monte Carlo experiments, these authors find that this restriction is a source of bias when incorrectly imposed.

[4]BOURGUIGNON, FOURNIER and GURGAND (2007) emphasize that the parsimony of this specification comes at the cost of fairly restrictive assumptions. LEE imposed the distributional assumption that the joint distribution of u and $\Phi^{-1}(P_k)$ does not depend on Γ, as well as the linearity assumption $E(u|\varepsilon_k, \Gamma) = \sigma \rho_k \Phi^{-1}(F_{\varepsilon_k}(\varepsilon_k|\Gamma)$, where $F_{\varepsilon_k}(.|\Gamma)$ denotes the cumulative distribution function of ε_k.

rection function $\mu(P_1, ..., P_M)$ to a subset $S \subset \{1, ..., M-1\}$, thereby invoking the sufficiency assumption

$$f(u, \varepsilon_k|\Gamma) = f(u, \varepsilon_k|P_1, ..., P_{M-1}) = f(u, \varepsilon_k|P_i, i \in S) \tag{10}$$

that this subset S exhausts all the relevant information. A special case proposed by DAHL (2002) is to invoke the hypothesis that $S = \{k\}$, that is, that the probability P_k to be selected on the observed outcome k is the only information needed for estimation and function $\mu(P_k)$ is approximated by series expansions, such as a polynomial:

$$y = \boldsymbol{\beta}^T \mathbf{x} + \mu(P_k) + w. \tag{11}$$

This procedure drastically reduces the dimension of the correction function $\mu(P_1, ..., P_M)$ and, hence, the number of corresponding coefficients λ_{Dahlj} to be estimated.

Marginal Effects

As in the case of HECKMAN's sample selection model, the coefficients of variables that appear in both the choice and outcome equations do not lend themselves to direct interpretation. Rather, the marginal effects of such variables must be calculated using a nonlinear function of the underlying model parameters to correct for the selectivity effect.

Assuming that the first-stage variable z_{kl} is also included in vector $\mathbf{x}$ of the second-stage regression, so that $x_l = z_{kl}$, for the DMF approach, the marginal effect of z_{kl} is given by the partial derivative of Equation (7):

$$\begin{aligned}\frac{\partial y}{\partial z_{kl}} &= \beta_l + \sigma\frac{\sqrt{6}}{\pi}\left[\sum_{j\neq k} r_j \frac{\partial P_j}{\partial z_{kl}} \frac{[\ln P_j + P_j/P_j](1-P_j) + P_j \ln P_j}{(1-P_j)^2} - r_k \frac{\frac{\partial P_k}{\partial z_{kl}}}{P_k}\right] \\ &= \beta_l + \sigma\frac{\sqrt{6}}{\pi}\left[\sum_{j\neq k} r_j \frac{\partial P_j}{\partial z_{kl}} \left[\frac{(\ln P_j + 1)}{1-P_j} + \frac{P_j \ln P_j}{(1-P_j)^2}\right] - r_k \frac{\frac{\partial P_k}{\partial z_{kl}}}{P_k}\right],\end{aligned} \tag{12}$$

where

$$\frac{\partial P_k}{\partial z_{kl}} = \frac{\gamma_{kl}\exp(\boldsymbol{\gamma}_k^T \mathbf{z})\left[\sum_m \exp(\boldsymbol{\gamma}_m^T \mathbf{z})\right] - \exp(\boldsymbol{\gamma}_k^T \mathbf{z})\sum_m \gamma_{km}\exp(\boldsymbol{\gamma}_m^T \mathbf{z})}{\left[\sum_m \exp(\boldsymbol{\gamma}_m^T \mathbf{z})\right]^2} \tag{13}$$

and $\frac{\partial P_j}{\partial z_{kl}}$ is calculated accordingly. If variable x_l instead merely emerges from the second-stage regression, but not from the first-stage multinomial logit model, the marginal effect (12) simplifies to $\frac{\partial y}{\partial x_l} = \beta_l$, a result that holds for the approaches of LEE and DAHL as well.

For LEE's approach, using the derivative $[\phi(x)]' = -2x\phi(x)$ of density function $\phi(x) = 1/\sqrt{2\pi}\exp(-x^2)$ and $\Phi'(x) = \phi(x)$, the derivative of the cumulative standard normal distribution $\Phi(x)$, the marginal effect of a variable $x_k = z_{kl}$ derives from Equation (9):

$$\begin{aligned}\frac{\partial y}{\partial z_{kl}} &= \beta_l + \sigma\rho_k \frac{\phi(\Phi^{-1}(P_k))}{(P_k)^2}\frac{\partial P_k}{\partial z_{kl}} + 2\frac{\sigma\rho_k}{P_k}\Phi^{-1}(P_k)\phi(\Phi^{-1}(P_k))[\Phi^{-1}(P_k)]'\frac{\partial P_k}{\partial z_{kl}} \\ &= \beta_l + \sigma\rho_k \frac{\partial P_k}{\partial z_{kl}}\left[\frac{\phi(\Phi^{-1}(P_k))}{(P_k)^2} + 2\frac{\Phi^{-1}(P_k)}{P_k}\right], \end{aligned} \tag{14}$$

as due to the inverse function rule, $[\Phi^{-1}(P_k)]' = 1/[\Phi]'(\Phi^{-1}(P_k)) = 1/\phi(\Phi^{-1}(P_k))$, and, hence, the terms $\phi(\Phi^{-1}(P_k))$ and $[\Phi^{-1}(P_k)]'$ neutralize each other.

For DAHL's approach, the marginal effect depends on the functional form of $\mu(P_k)$ and its derivative $\frac{d\mu}{dP_k}$:

$$\frac{\partial y}{\partial z_{kl}} = \beta_l + \frac{d\mu}{dP_k}\frac{\partial P_k}{\partial z_{kl}}. \tag{15}$$

Finally, we note that for all three approaches, the calculated marginal effects can be interpreted as elasticities provided that both the dependent variable and the explanatory variable of interest are logged, such as in our rebound example, where $y = \ln s$ denotes logged monthly kilometers driven and $z_{kl} = \ln p$ designates logged real fuel prices (Table 1).

3 Data

The data used in this research is mainly drawn from the German Mobility Panel (MOP 2015) and covers fourteen years, spanning 2000 through 2014, and 3,564 households, yielding a total of 6,631 observations. Travel survey information, which is recorded at the level of the automobile, is used to derive the dependent and explanatory variables (see FRONDEL, PETERS, VANCE, 2008, for more details on this survey). To abstract from complexities associated with households who own a mix of diesel and petrol cars, which comprise about 5% of the sample, we focus here exclusively on petrol car owners. In other work, FRONDEL and VANCE (2014) have shown that the responsiveness to fuel prices does not differ significantly between petrol and diesel drivers.

With respect to the incidence of car ownership, the dependent variable of the first-stage multinomial logit model, 19% of the sample households do not possess a car, whereas the remainder either own just one car (56%) or have multiple cars (25%). Hence, we conceive the first-stage model to reflect the discrete choice between three exclusive alternatives: owning either no car, or just one car, or two and more cars. The dependent variable of the second-stage equation is given by the total monthly distance driven in kilometers (Table 1).

Tabelle 1: Variable Definitions and Descriptive Statistics

Variable	Variable Definition	Mean	Std. Dev.
s	Monthly kilometers driven	984.96	937.81
p	Real petrol price in Euros per liter	1.41	0.13
income	Real income in €	2,359.78	1,016.78
# cars	Number of cars owned by the household	1.06	0.66
insurance class	Car insurance cost class (1-12)	6.12	2.79
% age$<$ 09	Share of household members younger than 9	0.03	0.10
% age $10-17$	Share of household members between 10-17	0.05	0.13
% employed	Share of full-time employed household members	0.40	0.40
% retired	Share of retired household members	0.39	0.46
size$=$ 1	Dummy: 1 if household has 1 member	0.33	–
size$=$ 2	Dummy: 1 if household has 2 members	0.41	–
size$=$ 3	Dummy: 1 if household has 3 members	0.13	–
size$=$ 4	Dummy: 1 if household has 4 members	0.10	–
size$>$ 4	Dummy: 1 if household has more than 4 members	0.02	–
big city	Dummy: 1 if household resides in a big city	0.50	–
shop	Dummy: 1 if there is a shop for basic needs at walking distance	0.86	–
rail stop	Dummy: 1 if there is a rail stop at walking distance	0.14	–

The suite of control variables that are hypothesized to influence both car ownership and the extent of motorized travel encompass, among others, the fuel price, the demographic composition of the household, its income, as well as measures characterizing the surrounding landscape pattern and public transport infrastructure. Non-parametric model identification requires that at least one variable is included in the first-stage selection equation that determines car ownership, but not in the second-stage equation on car use (HECKMAN, 1979). A candidate variable is an insurance index reflecting part of the fixed costs of owning a car that might affect the decision of purchasing a car, but are unlikely to impact on distance traveled. This index, which serves as a proxy for insurance costs, is taken from the German Insurance Association and ranges between 1 and 12, with increasing index values pointing towards higher insurance costs. It indicates the average insurance cost per car at the zip-code level.

4 Empirical Results

Table 2 reports the first-stage multinomial logit estimation results using car ownership as the dependent variable, with the reference group being households without cars. In the interest of brevity, it suffices to note that all of the statistically significant results have signs that are consistent with intuition, including the negative sign of the insurance cost index, which serves to identify the model.

Tabelle 2: First-Stage Multinomial Logit Estimation Results on the Number of Vehicles of a Household

	Single-Car Households			Multiple-Car Households		
	Coeff.s		Std. Errors	Coeff.s		Std. Errors
$\ln(p)$	−0.563		(0.483)	0.064		(0.648)
insurance class	−0.066	**	(0.020)	−0.059	*	(0.027)
% age< 09	−0.979		(0.805)	−3.145	***	(0.865)
% age 10 − 17	−0.239		(0.607)	−2.643	***	(0.674)
% employed	−0.012		(0.201)	0.320		(0.292)
% retired	0.119		(0.188)	−0.716	*	(0.290)
size= 2	0.741	***	(0.127)	3.363	***	(0.321)
size= 3	0.558		(0.292)	4.265	***	(0.414)
size= 4	1.058	*	(0.435)	5.011	***	(0.541)
size> 4	2.448	*	(1.041)	6.494	***	(1.091)
big city	−0.772	***	(0.121)	−1.203	***	(0.161)
shop	−1.267	***	(0.194)	−2.061	***	(0.225)
rail stop	−0.264	*	(0.132)	−0.613	**	(0.194)
$\ln(income)$	2.268	***	(0.176)	4.129	***	(0.247)
Constant	−13.963	***	(1.279)	−31.230	***	(1.851)
Pseudo R^2				0.299		

Note: Standard errors are clustered at the household level. Number of observations: 6,631
* denotes $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, respectively.

Tables A1 and A2 of the appendix present the second-stage results from the models of distance driven for the single- and multiple-car households, distinguished by the three sample selection correction models of LEE, DMF, and DAHL. By and large, this distinction appears to have little bearing on the magnitude of the coefficient estimates.[5] With respect to the selectivity parameters, only in the DMF model is one of the parameters statistically significant, and only for the single-car case. On this basis, we would conclude that sample selectivity does not appear to be a major issue with this data, which likely accounts for the similarity of the coefficients.

To illustrate the interpretation of the marginal effects, we focus on the fuel price elasticity. Subject to certain assumptions, this elasticity can be used to gauge the magnitude of the rebound effect (see e. g. FRONDEL, RITTER, VANCE, 2012), measuring the extent to which motorists increase driving in response to decreases in the per-kilometer costs of driving through improvements in fuel economy. Table 3 presents the mean

[5] We have cross-checked our estimation results using the selmlog Stata command written by Marc GURGAND and Martin FOURNIER and offered at the following internet site: http://www.parisschoolofeconomics.com/gurgand-marc/selmlog/selmlog13.html.

estimates of the rebound effect associated with each of the three models, calculated by averaging over the observation-specific estimates.

For the single-car case, these mean estimates, which equal the negative of the fuel price elasticities, range between 0.45 and 0.5, and are perfectly in line with those of our former studies that focus on single-car households (FRONDEL, VANCE, 2013a). Larger discrepancies between the coefficients and rebound effects are seen for the multiple-car case, which may be attributed to the considerably higher standard errors of the estimates. The multiple-car case also reveals larger discrepancies across the estimates, with an unreliably large rebound of 1.42 estimated for the Lee model that reflects the low estimation precision given by the large standard error of 0.923.

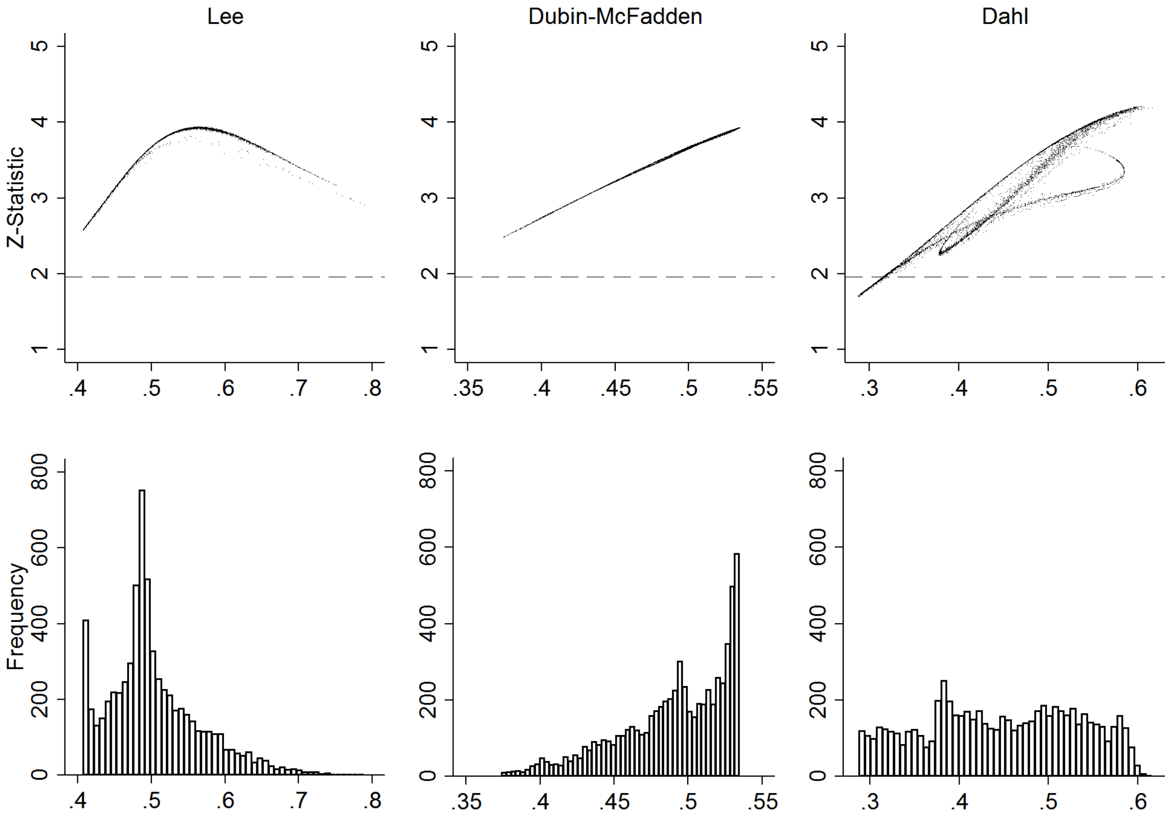
Tabelle 3: Mean Rebound Effects

	Lee		DMF		Dahl	
	Rebound	Std. Errors	Rebound	Std. Errors	Rebound	Std. Errors
Single-Car Households	0.504	(0.145)	0.490	(0.138)	0.447	(0.152)
Multiple-Car Households	1.420	(0.923)	0.314	(0.193)	0.468	(0.221)

The mean rebound effects presented in Table 3 obscure substantial heterogeneity over the individual estimates. The degree of this heterogeneity can be gleaned by plotting the observation-specific effects, presented for the single-car case in Figure 1. The top panel of Figure 1 presents this plot for the whole range of observations. For all three models, the majority of points falls outside the absolute 1.96 threshold on the vertical axis that indicates significance at the 5% level, with magnitudes along the horizontal axis ranging between 0.3 and 0.8. This suggests that between 30 and 80% of potential fuel reductions from an efficiency improvement is lost to increased driving. The histograms in the second panel facilitate a more transparent view of the distribution of rebound effects by showing the frequency corresponding to each value. The frequencies resulting from the DAHL model are relatively uniform, while having peaks between 0.4 and 0.55 for the LEE and DMF models.

For the multiple-car case presented in Figure 2, the estimates of the DMF model are not statistically significant over the entire range. The LEE model has a limited range of statistically significant estimates varying between 0.1 and 0.8; otherwise the values are of an implausibly large magnitude and estimated with a low degree of precision. In contrast, the majority of estimates from the DAHL model is statistically significant, varying between 0.2 and 0.8.

Abbildung 1: Observation-specific Rebound Effects of One-Car Households

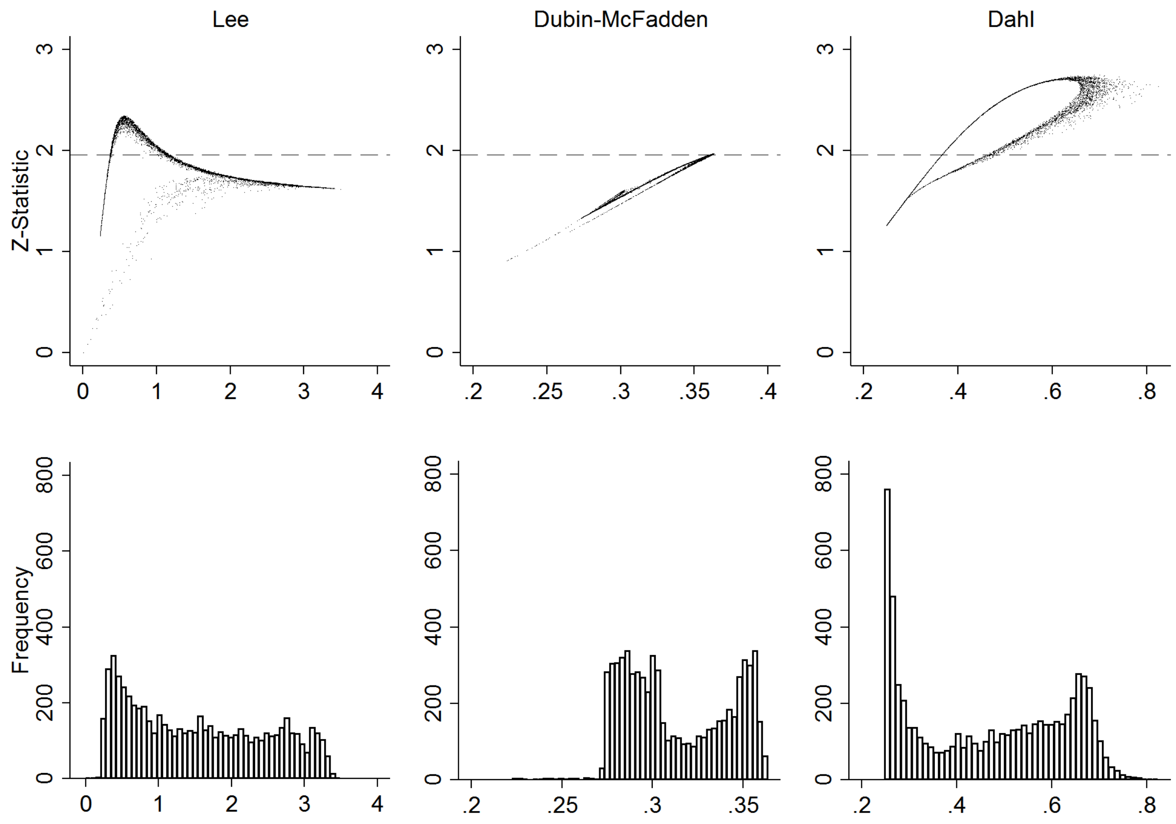


5 Conclusion

This note has demonstrated that the coefficients from discrete-continuous bias correction models cannot be interpreted as marginal effects when the variable of interest appears in both the selection and outcome equation, an issue that has largely been neglected in the applied literature. After deriving the marginal effects corresponding to the selection bias correction approaches of LEE (1983), DUBIN and MCFADDEN (1984), and DAHL (2002), we have presented an empirical example that employs each of these approaches to analyze the correlates of car ownership and car use using German survey data.

Our example illustrates substantial heterogeneity in the spread of price elasticities, interpreted here as rebound effects, for each of the approaches. We find rebound effects varying between 0.2 and 0.8, suggesting that upwards of 80% of the emissions reduction from an efficiency improvement is lost to increased driving. Hence, the re-

Abbildung 2: Observation-specific Rebound Effects of Multiple-Car Households



bound effect may be substantial and is not an exaggeration (FRONDEL, VANCE, 2013b). For single-car households, specifically, the mean rebound estimates range between 0.45 and 0.5 and are perfectly in line with those of our former studies (FRONDEL, VANCE, 2013a), indicating that sample selectivity does not appear to be a major issue with this data.

Appendix

Table A1: Second-Stage Bias Correction Results for Single-Car Households.

	Lee			DMF			Dahl		
	Coeff.s		Std. Errors	Coeff.s		Std. Errors	Coeff.s		Std. Errors
$\ln(p)$	−0.489	***	(0.137)	−0.443	**	(0.139)	−0.471	***	(0.137)
% age$<$ 09	−0.145		(0.149)	−0.192		(0.172)	−0.115		(0.149)
% age $10-17$	−0.232		(0.138)	−0.313		(0.163)	−0.214		(0.140)
% employed	0.245	***	(0.063)	0.263	***	(0.064)	0.253	***	(0.063)
% retired	−0.187	**	(0.061)	−0.221	***	(0.064)	−0.193	**	(0.061)
size$=2$	0.234	***	(0.038)	0.351	***	(0.092)	0.250	***	(0.039)
size$=3$	0.298	***	(0.068)	0.480	***	(0.145)	0.318	***	(0.069)
size$=4$	0.376	***	(0.076)	0.551	***	(0.160)	0.391	***	(0.076)
size>4	0.547	***	(0.095)	0.703	***	(0.180)	0.559	***	(0.097)
big city	−0.043		(0.031)	−0.040		(0.050)	−0.031		(0.031)
shop	−0.142	***	(0.038)	−0.145	*	(0.065)	−0.126	**	(0.039)
rail stop	−0.036		(0.040)	−0.045		(0.044)	−0.034		(0.039)
$\ln(income)$	0.172	***	(0.042)	0.225		(0.125)	0.168	***	(0.042)
constant	5.545	***	(0.314)	5.334	***	(0.948)	5.569	***	(0.307)
λ_{Lee}	0.076		(0.060)	–		–	–		–
λ_{DMF0}	–		–	0.492		(0.300)	–		–
λ_{DMF1}	–		–	0.127		(0.082)	–		–
λ_{DMF2}	–		–	0.729	*	(0.330)	–		–
λ_{Dahl1}	–		–	–		–	−0.588		(0.475)
λ_{Dahl2}	–		–	–		–	0.688		(0.425)
R^2	0.132			0.134			0.133		

Note: Standard errors are clustered at the household level. Number of observations: 3,696.
* denotes $p<0.05$, ** $p<0.01$, *** $p<0.001$, respectively.

Table A2: Second-Stage Bias Correction Results for Multiple-Car Households.

	Lee		DMF		Dahl	
	Coeff.s	Std. Errors	Coeff.s	Std. Errors	Coeff.s	Std. Errors
$\ln(p)$	-0.224	(0.203)	-0.187	(0.277)	-0.248	(0.198)
% age$<$ 09	-0.761 **	(0.288)	-1.031	(0.648)	-0.717 *	(0.290)
% age $10-17$	-0.777 *	(0.319)	-1.055	(0.720)	-0.732 *	(0.306)
% employed	0.162	(0.093)	0.193	(0.129)	0.158	(0.090)
% retired	-0.421 **	(0.142)	-0.481	(0.279)	-0.403 **	(0.123)
size$=$ 2	0.679	(0.392)	0.811	(0.766)	0.351 *	(0.177)
size$=$ 3	1.071 *	(0.519)	1.315	(1.081)	0.727 **	(0.276)
size$=$ 4	1.184 *	(0.548)	1.472	(1.152)	0.836 **	(0.304)
size$>$ 4	1.350 *	(0.572)	1.667	(1.185)	1.001 **	(0.333)
big city	-0.184 **	(0.070)	-0.249	(0.133)	-0.173 **	(0.063)
shop	-0.221 *	(0.103)	-0.329	(0.233)	-0.203 *	(0.095)
rail stop	-0.057	(0.074)	-0.097	(0.127)	-0.053	(0.078)
$\ln(income)$	0.700 **	(0.267)	0.936	(0.560)	0.646 **	(0.214)
constant	0.762	(2.756)	-1.634	(5.999)	2.423	(1.549)
λ_{Lee}	-0.486	(0.333)	–	–	–	–
λ_{DMF0}	–	–	-0.628	(0.662)	–	–
λ_{DMF1}	–	–	-0.929	(1.018)	–	–
λ_{DMF2}	–	–	0.210	(0.289)	–	–
λ_{Dahl1}	–	–	–	–	-1.315	(0.695)
λ_{Dahl2}	–	–	–	–	0.514	(0.410)
R^2	0.132		0.132		0.133	

Note: Standard errors are clustered at the household level. Number of observations: 1,638.
* denotes $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, respectively.

References

BENTO, A., CROPPER, M., MUSHFIQ, A., VINHA, K. (2005) The effects of urban spatial structure on travel demand in the United States. *The Review of Economics and Statistics* 87(3), 466-478.

BENTO, A., GOULDER, L., JACOBSEN, M., VON HAEFEN, R. (2009) Distributional and efficiency impacts of increased US gasoline taxes. *American Economic Review* 99(3), 667-699.

BLANC, M., CAHUZAC, E., ELYAKIME, B., TAHAR, G. (2008). Demand for on-farm permanent hired labour on family holdings. *European Review of Agricultural Economics* 35(4), 493-518.

BOURGUIGNON, F., FOURNIER, M., GURGAND, M. (2007) Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons. *Journal of Economic Surveys* 21(1), 174-205.

DAHL, G. (2002) Mobility and the returns to education: testing a Roy Model with multiple markets. *Econometrica* 70(6), 2367-2420.

DUBIN, J., MCFADDEN, D. (1984) An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 52, 345-362.

FRONDEL, M., PETERS, J., VANCE, C. (2008) Identifying the rebound: Evidence from a German household panel. *Energy Journal* 29(4), 145-163.

FRONDEL, M., RITTER, N., VANCE, C. (2012) Heterogeneity in the rebound effect: Further evidence for Germany. *Energy Economics* 34(2), 461-467.

FRONDEL, M., VANCE, C. (2009) Do high oil prices matter? Evidence on the mobility behavior in German households. *Environmental and Resource Economics* 43(1), 81-94.

FRONDEL, M., VANCE, V. (2013a) The rebound is not an exaggeration. *Nature.* 494, 430.

FRONDEL, M., VANCE, V. (2013b) Re-Identifying the rebound: What about asymmetry?. *Energy Journal* 34(4), 43-54.

FRONDEL, M., VANCE, V. (2014) More pain at the diesel pump? An econometric comparison of diesel and petrol price elasticities. *Journal of Transport Economics and Policy*

48(3), 449-463.

HECKMAN, J. (1979) Sample selection bias as an specification error. *Econometrica* 47, 153-161.

KLEE, E. (2008). How people pay: Evidence from grocery store data. *Journal of Monetary Economics* 55(3), 526-541.

LIN, C. Y. C., PRINCE, L. (2013). Gasoline price volatility and the elasticity of demand for gasoline. *Energy Economics* 38, 111-117.

LEE, L. (1983) Generalized econometric models with selectivity. *Econometrica* 51, 507-512.

LUCE, R.D. (2005). Individual choice behavior: A theoretical analysis. *John Wiley and Sons*, New York.

MABIT, S.L., FOSGERAU, M. (2009) Controlling for sample selection in the estimation of the value of travel time. *The expanding sphere of travel behaviour research*. Emerald Group Publishing Limited, Howard House. 703-723.

MANSUR, E.T., MENDELSOHN, R., MORRISON, W. (2008). Climate change adaptation: A study of fuel choice and consumption in the US energy sector. *Journal of Environmental Economics and Management* 55(2), 175-193.

MATIASKE, W., MENGES, R., SPIESS, M. (2012). Modifying the rebound: It depends! Explaining mobility behavior on the basis of the German socio-economic panel. *Energy Policy* 41, 29-35.

MCFADDEN, D. (1974) Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*. Academic Press, New York. 105-142.

MILLS, B., SCHLEICH, J. (2014). Household transitions to energy efficient lighting. *Energy Economics* 46, 151-160.

MOP – German Mobility Panel (2015) http://www.ifv.kit.edu/english/26_MOP.php

WADUD, Z., GRAHAM, G. J., NOLAND, R. B. (2010) Gasoline demand with heterogeneity in household responses. *The Energy Journal* 31 (1), 47-74.

WEST, S. (2004) Distributional effects of alternative vehicle pollution control technologies. *Journal of Public Economics* 88(3), 735-757.