SFB 823

Cycling on the extensive and intensive margin: The role of paths and prices

Manuel Frondel, Colin Vance, Martin Wagner

Discussion Paper

Nr. 44/2016

SFB 823

Cycling on the Extensive and Intensive Margin: The Role of Paths and Prices

Manuel Frondel*, Colin Vance†, and Martin Wagner ‡

August 2016

Abstract: Drawing on a panel of German survey data spanning 1997-2013, this paper identifies the correlates of non-recreational bicycling, focusing specifically on the roles of bicycle paths and fuel prices. Our approach conceptualizes ridership as a two stage decision process comprising the discrete choice of whether to use the bike (i.e. the intensive margin) and the continuous choice of how far to ride (i.e. the extensive margin). To the extent that these two choices are related and, moreover, potentially influenced by factors unobservable to the researcher, we explore alternative estimators using two-stage censored regression techniques to assess whether the results are subject to biases from sample selectivity. A key finding is that while higher fuel costs are associated with an increased probability of undertaking non-recreational bike trips, this effect is predicated on residence in an urbanized region. We also find evidence for a positive association with the extent of bike paths, both in increasing the probability of non-recreational bike travel as well as the distance traveled.

JEL classification: D13, Q41, Q58.

Keywords: bicycle paths, fuel prices, non-recreational cycling

*RWI – Leibniz Institute for Economic Research, Ruhr University Bochum

†Corresponding author, RWI – Leibniz Institute for Economic Research, Jacobs University Bremen; Hohenzollernstr. 1–3, 45128 Essen, Germany. E-mail: vance@rwi-essen.de.

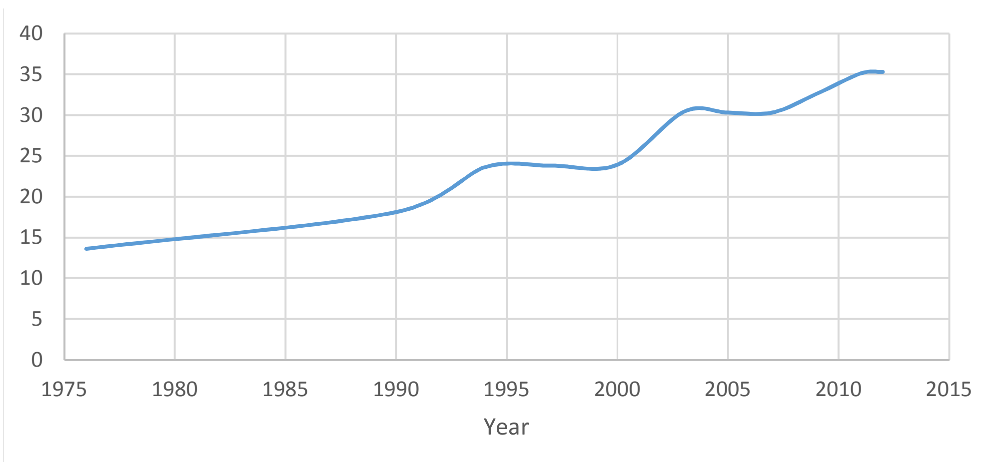
‡Faculty of Statistics, Technical University Dortmund, Institute for Advanced Studies, Vienna, and Bank of Slovenia, Ljubljana

Acknowledgements: We are grateful for invaluable comments and suggestions by Christoph M. SCHMIDT. This work has been partly supported 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". We also gratefully acknowledge financial support by the German Federal Ministry of Education and Research (BMBF) under grant 01LA1113A.

1 Introduction

The promotion of bicycling is widely recognized to advance multiple goals toward sustainable transportation policy. Beyond reducing air pollution, noise, congestion, and other negative externalities associated with the automobile, bicycling contributes to health, increases mobility, and affords shelter from volatile fuel prices. In recognition of these benefits, the German government recently released a traffic plan that aims to increase the share of bicycle trips from 10 to 15% by 2020 (BMVI, 2016), a particularly ambitious objective given the 47% increase in total bicycle mileage already achieved between 2000 and 2012 (Figure 1).

Figure 1: Total Bicycle Travel in billions of Kilometers in Germany



Source: (BMVI, 2015).

Reaching the new target raises the question of what policy-levers can be availed to encourage bicycle usage. A sizable literature has emerged to address this topic, comprehensively summarized in a review of bicycle commuting studies by Heinen et al. (2010). A recurrent theme is that public policy can play an effective role in promoting bicycling, especially insofar as it shifts the relative costs of alternative transport modes in favor of cycling.

Two broad strands of literature have emerged in this vein, one of which examines the role of non-monetary costs, such as those related to safety, physical effort, time, the enjoyment derived from the trip, and other factors that attract or repel people from using the bicycle (Rietveld and Daniel, 2004; Handy et al., 2010; Ritter and Vance, 2011). The associated policy interventions evaluated in this literature include traffic speed and volume, bicycle infrastructure, integration with public transit (Rodrıguez and Joo, 2004; Moudon et al., 2005; Parkin et al., 2008; Winters et al., 2011), as well as communication campaigns that propagate information on the benefits of cycling (Lanzendorf and Busch-Geertsema, 2014) and harness social network effects (Goetzke and Rave, 2011).

Another strand of the literature has emphasized the importance of monetary costs, a central premise being that increases in the marginal costs of alternative modes, particularly for the automobile (Frondel and Vance, 2013), is among the most effective ways to increase bicycling (Pucher et al., 1999; Bergström and Magnusson, 2003; Pucher and Buehler, 2008). While there has been scant empirical work that tests this proposition using observed costs, studies by Noland and Kunreuther (1995) and Sardianou and Nioza (2015) both establish a correlation between perceived automobile cost and preferences for bicycling.

Using household level survey data from Germany, the present paper draws on elements from both these strands to identify policy tools for increasing cycling. Our approach conceptualizes ridership as a two stage decision process comprising the discrete choice of whether to use the bike (i.e. the intensive margin) and the continuous choice of how far to ride (i.e. the extensive margin). To the extent that these two choices are related and, moreover, potentially influenced by factors unobservable to the researcher, we explore alternative estimators using two-stage censored regression techniques to assess whether the results are subject to biases from sample selectivity. We are particularly interested in quantifying the roles of transport infrastructure and fuel costs as determinants of bicycle use.

A key finding is that while higher fuel costs are associated with an increased probability of undertaking non-recreational trips with the bike, this effect is predicated on residence in an urbanized region. We also find evidence for a positive association with the extent of bike paths, both in increasing the probability of non-recreational bike travel as well as the distance traveled. Robustness checks are undertaken that support a causative interpretation of these findings.

The remainder of the paper is structured as follows. The next section describes the data sources and their assembly for the quantitative analysis. Section 3 describes the econometric models, the explanatory variables included in the specification, and some technical details on the interpretation of the marginal effects. Section 4 catalogues the results, and Section 5 concludes the paper.

2 Data Assembly

The primary data source used in this research covers the 1999-2013 waves of the German Mobility Panel (MOP), a representative multi-year travel survey financed by the German Federal Ministry of Transport and Digital Infrastructure. Participating households are surveyed daily for a period of one week over each of three years, after which they exit the panel. The information collected includes individual attributes such as age, gender, and employment status, as well as mobility-related characteristics such as possession of a driver's license and ownership of a bicycle (Table 1). It also includes household attributes, such as income, car ownership, proximity to the nearest transit stop, residence in an urbanized neighborhood and other regional features. In addition, each adult household member fills out a trip log capturing relevant aspects of everyday travel conditions and travel behavior, including temperature, rainfall, distances traveled, modes used, activities undertaken, and activity durations.

Using the data from these logs, we derived a measure of the total weekly distance of non-recreational bicycle travel, which serves as the dependent variable. Non-recreational travel is defined as trips whose purpose is work, shopping or for completing tasks. In an effort to maintain a clear division with recreational travel, we exclude weekends from the analysis.

Table 1: Variable Definitions and Descriptive Statistics

Variable Name	Variable Definition	Mean	Std. Dev.
Lack of cars	Dummy: 1 if the number of driver licenses is larger than the number of cars in the household	0.448	–
Lack of bikes	Dummy: 1 if the number household members is larger than the number of bikes in the household	0.162	–
Transit proximity	Walking distance in minutes to public transit stop	5.664	4.726
Rail transit	Dummy: 1 if this stop is serviced by rail transit	0.102	–
Bike path extent	Total Length of bike paths in km (in 100s)	1,202	1,284
Urban	Dummy: 1 if household situated in urbanized county	0.381	–
Petrol price	Petrol price in Euros per liter	1.140	0.280
Open space	Square km (in 1000s) of undeveloped land	0.753	0.603
County size	Areal extent of residence county in sq km (in 1000s)	0.864	0.605
# Rainy days	Number of rainy days in a week	2.390	1.481
Temperature	Temperature in degree Celsius	10.366	3.709
Female	Dummy: 1 if respondent is female	0.515	–
Degree	Dummy: 1 if respondent has a post-high school degree	0.413	–
Age	Age of respondent in years	48.725	15.144
# Kids	Number of kids in the household	0.492	0.819
License	Dummy: 1 if respondent owns a driving license	0.946	–
High income	Dummy: 1 if real monthly household income $\geq$ 3,000 €	0.356	–
Middle income	Dummy: 1 if real monthly household income $\geq$ 1,500 € and $<$3,000 €	0.540	–
Full time employed	Dummy: 1 if respondent is full time employed	0.408	–
Year trend	Year of observation	2005.9	4.435

The data was pared along several additional dimensions. First, we limit the sample to adults over 17 years of age who do not report having a mobility-constraining handicap.

Second, we include only those households that own at least one bike, which covers 81% of the sample. Finally, we include only those households owning at least one car, as individuals in carless households are unlikely to be responsive to fuel prices if they do not have the option of substituting between bike and car travel. The resulting sample comprises 8,845 individuals. Of these, 3,821 participate in one of the survey years, 2,587 participate in two years, and 2,437 participate in all three survey years, resulting in a total sample size of 16,306 observations. To correct for the non-independence of repeat observations over multiple time points in the data, the regression disturbance terms are clustered at the level of the individual, so that the estimates of the standard errors are robust to this survey design feature.

We augmented the MOP with various external data sources to allow investigation of fuel prices and landscape pattern. Fuel prices are obtained from the web-site of Aral, one of Germany's largest gasoline retailers. Aral publishes nominal fuel prices by month dating back to 1999, thereby affording a tight temporal linkage with the MOP data, which is collected in the fall months of each year. We converted the fuel prices into real values using a consumer price index from Germany's Federal Ministry of Statistics.

Two landscape measures are used, one of which is derived from an Esri shapefile of bicycle paths in Germany. Using a Geographical Information System (GIS), we intersected this layer with another shapefile of German counties from the year 2005, at which time there were 439 counties having an average size of 814 square kilometers. The resulting intersected shapefile allows us to calculate the length of bike paths in each county. The other landscape metric, measuring the area of open space, was derived in a similar manner using Corine Land Cover satellite imagery obtained from the web site of the European Environmental Agency. The imagery distinguishes 26 land cover classes in raster format at a resolution of 100 $\times$ 100 meters, and is available for the years 2000 and 2006. We added up the area classified as forest, agricultural, wetlands, and water bodies within each county to obtain the square kilometers of open space. We assigned the 2000 value of open space to the years 1999 and 2001 through 2005, and the 2006 value to the years 2007 through 2013.

3 Methods

Roughly 72% of the individuals in the data do not use a bicycle over a given week and for whom the observation on weekly distance ridden is consequently not recorded. To accommodate the preponderance of such corner solutions, two-stage estimation procedures can be availed, such as the Heckit model proposed by Heckman (1979) and the two-part model (2PM), which was developed by Cragg (1971) as an extension to the Tobit model.

3.1 Estimators

The first stage of both models captures the extensive margin, ordering observations of the outcome variable y into two regimes defined by whether the individual uses the bike:

$$R = 1, \text{if} \quad R^* = \mathbf{x_1}^T \boldsymbol{\tau} + \epsilon_1 > 0 \quad \text{and} \quad R = 0, \text{if} \quad R^* \leq 0, \tag{1}$$

where R^* is a latent variable indicating the utility from bike use, R is an indicator for bike usage status, the $\mathbf{x_1}$ denote the determinants of this status, $\boldsymbol{\tau}$ is a vector of associated parameter estimates, and ϵ_1 is an error term having a standard normal distribution.

The estimates of $\boldsymbol{\tau}$ are obtained using a Probit model of the probability of a positive outcome, referred to as the selection equation:

$$P(y > 0|\mathbf{x_1}) = \Phi(\mathbf{x_1}^T \boldsymbol{\tau}). \tag{2}$$

The second stage of both the Heckit and 2PM captures the intensive margin by estimating an OLS regression of distance traveled, conditional on $R = 1$. This stage is based on the conditional expectation

$$E[y|R=1,\mathbf{x_2}] = E[y|y>0,\mathbf{x_2}] = \mathbf{x_2}^T\boldsymbol{\beta} + E(\epsilon_2|y>0,\mathbf{x_2}), \tag{3}$$

where y denotes the dependent variable, measured here as the kilometers of daily bike travel, and ϵ_2 is another error term, again assumed to be normally distributed.

The models are distinguished by the second stage specification of the explanatory variables. In the 2PM, where it is assumed that $E(\epsilon_2|y>0,\mathbf{x_2}) = \mathbf{0}$ and, hence,

$$E[y|y>0,\mathbf{x_2}] = \mathbf{x_2}^T\boldsymbol{\beta}, \tag{4}$$

whereas the unconditional expectation is given by: $E[y] = \Phi(\mathbf{x_1}^T\boldsymbol{\tau})\cdot\mathbf{x_2}^T\boldsymbol{\beta}$.

By contrast, the second stage OLS regression of the Heckit model includes the inverse MILLS ratio, $\lambda(\mathbf{x_1}^T\boldsymbol{\tau})) := \frac{\varphi(\mathbf{x_1}^T\boldsymbol{\tau})}{\Phi(\mathbf{x_1}^T\boldsymbol{\tau})}$, as an additional regressor to control for sample selectivity:

$$E[y|y>0] = \mathbf{x_2}^T\boldsymbol{\beta} + \beta_\lambda\cdot\lambda(\mathbf{x_1}^T\boldsymbol{\tau}), \tag{5}$$

where β_λ is called the sample-selection parameter and the inverse MILLS ratio is proportional to $E(\epsilon_2|y>0,\mathbf{x_2}) \neq 0$ when ϵ_2 is assumed to be normally distributed with constant variance: $\mathrm{Var}(\epsilon_2) = \sigma^2$.

In omitting this regressor, the 2PM imposes the assumption that $E(\epsilon_2|y>0,x_2) = 0$, which is a key consideration bearing on which model to use. When sample selection bias arises from the correlation of unobserved factors affecting both the binary and continuous outcomes, this assumption is invalid. In this case, the Heckit model may be more appropriate for corner-solution data (Wooldridge, 2010, p. 697). A statistically significant estimate on the coefficient of the inverse Mills ratio (IMR) serves as a standard test for selectivity bias.

However, a well-known impediment in estimating the Heckman model emerges when there is a high degree of collinearity between the independent variables and the IMR, resulting in high standard errors on the coefficient estimates and parameter instability. The incorporation of so-called exclusion restrictions - variables included in the first stage probit, but not the second stage OLS - ameliorates these problems by reducing multicollinearity among the predictors and the IMR in the outcome equation. In their absence, however, the consequences for the model estimates can be profound, with some studies suggesting that even when the Heckit is the true model, its relative inefficiency may be so severe as to justify the use of the 2PM (Leung and Yu, 1996; Manning et al., 1987; Hay et al., 1987). As a clear case for the superiority of one model over the other can often not be drawn (Vance and Ritter, 2014), we present both models to assess the robustness of the results.

3.2 Specification

Presuming that sample selectivity is deemed to be a source of bias that warrants the use of the Heckit, identification of the model requires the selection of at least one variable that uniquely determines the discrete choice of bicycle use, but not the continuous choice of distance traveled. In the present example, this selection can be informed by consideration of fixed costs, that is, costs that are incurred or avoided with the use of the bike, but not with distance traveled.

One source for such costs are the negotiations between family members over access to household transport modes, which will be more contentious when the availability of the mode is less than the demand for it. To capture these costs, we include two dummy variables, one indicating the existence of more adult members than available bikes, and the other indicating the existence of more licensed drivers than available cars (see the first two variables in Table 1). We expect the first dummy variable, which reflects bicycle scarcity, to be negatively associated with the probability of bike use, and the second variable, which reflects car scarcity, to be positively associated.

Two other identifying variables are included, the walking distance in minutes to the nearest public transit stop and a dummy variable indicating whether this stop is serviced by rail transit. As the former variable measures the fixed costs of accessing public transit, we expect it to be positively associated with bicycle use. The sign of the latter variable is ambiguous. To the extent that rail transit captures higher speed and comfort relative to other public modes like the bus and streetcar, we would expect it to have a negative association with bike travel. On the other hand, rail users can often take their bikes on board, which may encourage bike use through combining the two modes.

The remaining variables are included in both the discrete and continuous stages of the models. Of particular interest are the measures of landscape features and fuel prices. The landscape features include a dummy for residence in an urbanized county. We also include three continuous variables measuring the kilometers of bike paths as well as the area of undeveloped land and the areal extent of the household's county of residence, measured in square kilometers. Both bike paths and open space capture landscape features that are conducive to bicycling, and hence are expected to have a positive association. The sign of the urban dummy, which captures development and population density in the individual's county of residence, is ambiguous. On the one hand, urbanized counties will have a greater range of destinations that are reachable by bicycle relative to their suburban counterparts, but on the other, they will also have a larger selection of alternative modes, such as public transit and walking.

Beyond these direct effects, we also explore whether urban residency affects bicycle use indirectly via the fuel price. We would expect that the fuel price itself has a positive association with bicycle use by increasing the cost of car travel. It is plausible, moreover, that the strength of this association varies according to location. Bicycle travel in rural areas will be hindered by greater distances separating destinations for non-recreational trip purposes than in urban areas, where density is higher. We consequently interact the urban dummy with the fuel price to allow for differential effects, expecting the magnitude of the fuel price estimate to be stronger for individuals in urban areas.

The specification is completed with a suite of controls for temporal factors, person-level characteristics, and socioeconomic attributes of the household. The temporal controls include a count of the number of days in the week it rained, the average temperature in degrees Celsius over the week, and a year trend. Personal attributes are captured by dummies indicating gender, attainment of a post-high school degree, full time employment status, and ownership of a driver's license, as well as a continuous measure of age. Household attributes are captured by a count of the number of children in the household and two dummies indicating middle income and wealthy households, with lower income households as the base case.

4 Results

Table 2 presents estimates from the first-stage Probit model, while Table 3 presents the estimates from the outcome equations of the Heckit and 2PM models. To ease interpretation, the presentation focuses on the marginal effects, rather than the coefficient estimates.

4.1 The Extensive Margin: Bike Use

Turning first to estimates from the Probit model, three of the four identifying variables are statistically significant and have the expected signs. Individuals in households with more licensed drivers than cars have a probability that is 12 percentage points higher of using the bike sometime during the week for non-recreational travel, while those in households in which the number of adults exceeds the number of available bikes have a probability that is 15 percentage points lower. Both results suggest that the decision to use a bicycle is correlated with the availability of transportation options in the home. Although the proximity to the nearest public transit does not have a statistically significant association with the probability of bike use, the type of transit service does: individuals whose nearest transit stop is serviced by rail have a probability of bicycle use that is 3 percentage points higher than those serviced by alternative transit modes such as bus.

Table 2: First-Stage Probit Estimation Results.

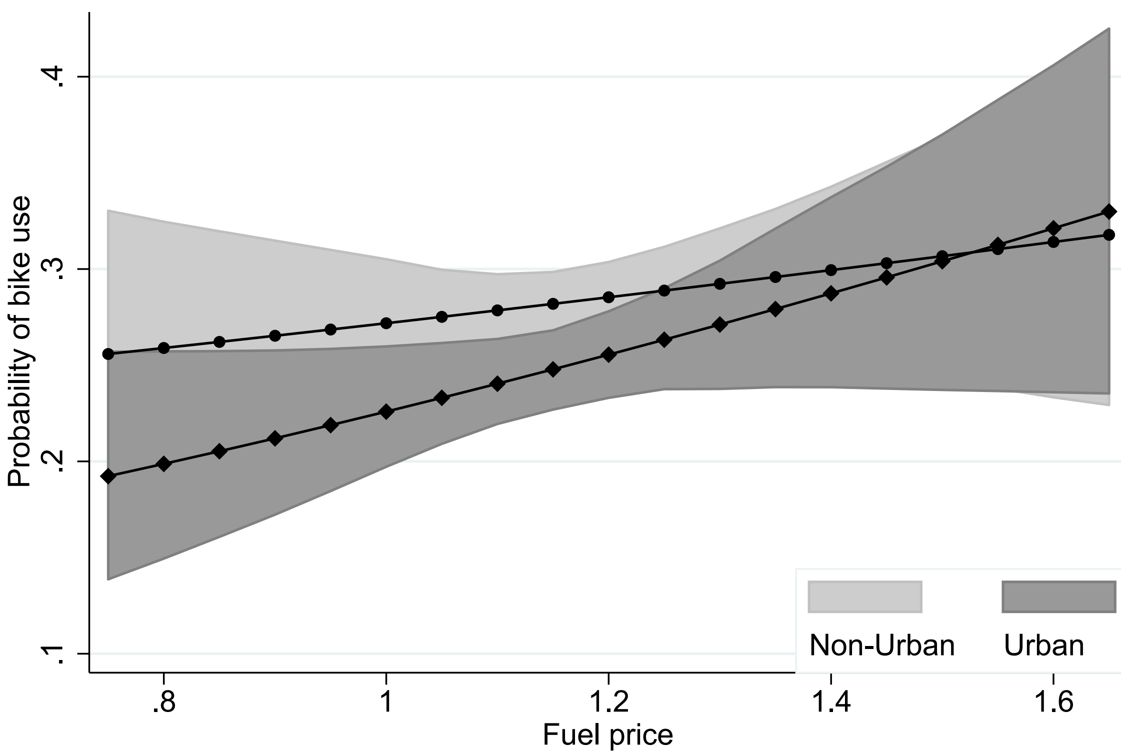
Variable	Marginal Effects	Std. Errors
Lack of cars	** 0.119	(0.009)
Lack of bikes	** -0.149	(0.009)
Rail transit	* 0.032	(0.014)
Transit proximity	-0.001	(0.001)
Bike path extent	0.034	(0.004)
Urban	** -0.035	(0.011)
Petrol price	0.096	(0.060)
Petrol price × *Urban*	* 0.144	(0.061)
Open space	0.040	(0.050)
County size	-0.047	(0.048)
# Rainy days	-0.001	(0.002)
Temperature	** 0.011	(0.001)
Female	-0.005	(0.009)
Degree	** 0.058	(0.009)
Age	** 0.002	(0.000)
# Kids	** 0.019	(0.005)
License	** -0,143	(0.021)
High income	** -0,063	(0,015)
Middle income	* -0,031	(0,014)
Full time employed	** -0,086	(0,010)
Year trend	-0.005	(0.004)
Number of Observations:		16,306

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively.

Further evidence for the role of the physical environment is seen for the estimates of bike path extent and the dummy indicating that the household resides in an urbanized county. As expected, bike paths are positively correlated with the probability of bicycle use. Each 100 kilometer increase in the bike path extent is associated an 3.4 percentage point increase in the probability that the bike is used for non-recreational travel. Conversely, residence in an urbanized county has a negative association, resulting in a probability of bike use that is 3.5 percentage points lower than for those who reside in rural counties.

The influence of urban location is also seen to work through the fuel price. Among those in rural counties, the effect of the fuel price is not statistically different from zero. This is contrasted by a highly significant slope effect estimated among those in urbanized counties, for whom a 1 Euro increase in the fuel price is associated with a 14.4 percentage point increase in the probability of bicycle use. Further insights into the differences between urban and rural counties can be gleaned from Figure 1, which shows the predicted probabilities of bike use and the associated 95% confidence intervals.

Figure 2: Fuel Prices and the Probability of Bike Use



The figure illustrates that notwithstanding the higher slope coefficient estimated for those in urban counties, the predicted probabilities of bike use between the two groups are statistically indistinguishable over the entire range of fuel prices. Among urban residents, the mean probability ranges from a low of just under 20% at a fuel price of 70 Euro cents per liter to a high of about 33% at a fuel price of 1.65 Euros per liter, albeit subject to large confidence intervals at the extremes. Taken together, the results suggest that while urban residents have a lower probability of using the bicycle, they show greater sensitivity in switching to this mode with an increase in fuel prices.

Among the socio-demographic control variables, the attributes of post high school education, age, and the presence of children in the household all increase the probability of non-recreation bike travel. Conversely, holders of driver's licenses, full-time employees, and those living in high-income households all have a lower probability. These two effects, which have been documented in other studies, comport with the idea that those having a higher opportunity cost of their time are less inclined to use the bike for non-recreational travel.

4.2 The Intensive Margin: Distance Traveled

The second-stage specifications of the Heckit and 2PM, presented in Table 3, are distinguished by the inclusion of the inverse Mills ratio (IMR) in the Heckit model to avoid bias from sample selectivity. The high degree of statistical precision on the coefficient estimate of the IMR would indicate that its inclusion is warranted, with the negative sign of β_λ suggesting that, on net, unobservable factors that increase the probability of bicycle use decrease the distance driven. One such factor could be the density of retail outlets in the immediate vicinity of the household, which could encourage non-recreational bike use but lower the distance traveled.

Nevertheless, as Leung and Yu caution, the statistical significance of β_λ is not reliable when the inverse Mills ratio is highly correlated with the explanatory variables. To gauge the extent to which collinearity afflicts the results of the Heckit, they recommend using the condition number, a diagnostic tool suggested by Belsley et al. (1980b). This measure, which indicates how close a data matrix is to being singular, is computed from the eigenvalues of the moment matrix. A higher condition number indicates a greater likelihood of collinearity problems, whereby Belsley et al. (1980a) suggest a maximum threshold of 30 on the basis of Monte Carlo experiments. The condition number estimated with the present data is 43.7, indicating that multi-collinearity may in fact undermine the stability of the results from the Heckit.

Table 3: Two-Part (2PM) and Heckit Models of Distance Traveled

	2PM		Heckit	
Variable	Marginal Effects	Std. Errors	Marginal Effects	Std. Errors
Female	** -2.211	(0.412)	** -6.300	(0.994)
Degree	** 2.092	(0.408)	** 2.827	(0.948)
# Rainy days	* -0.247	(0.092)	* -0.669	(0.236)
Temperature	** 0.267	(0.046)	* 0.147	(0.104)
Bike path extent	** 1.103	(0.181)	* 1.197	(0.414)
Urban	-0.693	(0.573)	-0.176	(2.947)
Petrol price	3.323	(2.402)	2.876	(6.187)
Petrol price $\times$ *Urban*	3.671	(2.290)	0.997	(6.215)
Open space	1.385	(1.909)	2.024	(4.754)
County size	-2.053	(1.856)	-3.581	(4.616)
# Kids	0.395	(0.224)	-0.149	(0.558)
Age	** 0.041	(0.014)	0.029	(0.034)
License	** -5.405	(1.069)	** -7.118	(1.854)
High income	** -1.933	(0.627)	** -2.766	(1.696)
Middle income	* -1.795	(0.608)	* -3.979	(1.520)
Full time employed	* -0,913	(0.453)	* 3.282	(1.239)
Year trend	-0.092	(0.148)	0.097	(0.383)
Inverse Mills Ratio	–	–	** -13.013	(2.397)
Number of Obs.	4,615		4,615	

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively.

While the conflicting evidence of these diagnostics complicates the choice of the preferred model, the difficulty would to some extent be rendered moot were the estimates of the Heckit and 2PM found to be similar. In this regard, the results present a largely confirmatory picture. With respect to the fuel price, for example, the estimates from both the 2PM and the Heckit model are statistically insignificant, contrasted by statistically significant estimates of bike path extent. The estimate from the Heckman model suggests that a 100 kilometer increase in path length is associated with an increase in the distance traveled of about 1.2 kilometers, with a comparable estimate implied by the 2PM. Other cases of statistically significant estimates across both the Heckit and 2PM are seen for the controls for females and rainfall, both of which have negative estimates.

4.3 Robustness Check

We have thus far refrained from ascribing a causative interpretation to the results, recognizing the possibility of endogeneity bias arising from omitted variables and reverse causation. The threat of such bias is particularly high for variables that are clearly subject to household choices, such as residency in a neighborhood with close proximity to public transit and the ownership of transport assets like bicycles and cars. Correcting for endogeneity using instrumental variables in this instance is difficult given the multiple sources from which the endogeneity may originate: there are several explanatory variables for which the assumption of zero correlation with the error term could be called into question. We instead gauge the robustness of the results through a strategy that, first, relies on a more parsimonious specification and, second, focuses on two policy-relevant variables, fuel prices and bike path extent, both of which are argued to be exogenous.

To this end, we follow the strategy used by Deaton and Stone (2014) and re-estimate the model with a specification that is purged of variables deemed to pose a source of endogeneity bias. These include the dummies indicating the prevalence of bicycles and cars in the household, rail services at the nearest transit stop, and whether the individual holds a driving license. We also exclude the continuous measure of walking minutes to the nearest public transit stop. As this re-specification eliminates all of the variables needed for the identification of the Heckit model, we now focus exclusively on the estimates from the 2PM.

Whether a causative interpretation applies to the remaining variables rests on the argument that none of these variables are themselves outcomes in the notational experiment at hand (Angrist and Pischke, 2008, p. 64). The exogeneity of fuel prices and the demographic attributes would seem to fulfill this requirement, but whether the same applies to the urban agglomeration dummy and the extent of bike paths is perhaps more questionable, since households may settle in such areas based on their transport preferences. While not being able to completely rule out this possibility, we find support in an argument made by Bento et al. (2005), who reason that reliance on city-wide variables that are measured at a sufficiently large geographic scale are not subject to bias from the endogenous location choices of households. The measures of bike path extent and urban agglomeration, being both based on county-level designations, meet this criterion.

The results of this robustness check, presented in Table 4, confirm most of the results from the 2PM presented in Table 3, but with some deviations. The most notable is that the fuel price now registers as a statistically significant determinant of distance traveled among people living in an urban area, with each Euro increase associated with a 2.06 kilometer increase in cycling per week for non-recreational travel. Beyond this, many of the estimates are of a lower magnitude. The influence of bicycle paths, for example, retains its statistical significance, but its magnitude is reduced substantially, by about 45% relative to the estimate in Table 3. Likewise, the gender indicator in the more parsimonious model of Table 4 suggests a considerably weaker negative effect among females.

5 Conclusion

Focusing on non-recreational travel, this paper has drawn on German household survey data from 1997 to 2013 to estimate the correlates of bicycle use. By employing both the Heckit and the Two-Part Model, our empirical approach distinguished between the discrete choice of whether to use the bike for non-recreational purposes – the extensive margin – from the continuous choice of distance traveled – the intensive margin. Against the backdrop of Germany's goal to increase bicycling, a suite of policy-relevant variables has been included in the analysis that, collectively, has to date received scant empirical scrutiny: the walking time in minutes to the nearest transit stop, a dummy indicating whether this stop is serviced by rail, the price for petrol fuel, and the extent of bike paths in the household's county of residence.

Table 4: Two-Part Model of Bicycle Use and Distance Traveled, Parsimonious Specification

	Probit		2PM	
	Marginal	Std.	Marginal	Std.
Variable	Effects	Errors	Effects	Errors
Female	-0.004	(0.007)	** -0.580	(0.159)
Degree	** 0.050	(0.007)	** 0.836	(0.168)
# Rainy days	0.000	(0.002)	* -0.067	(0.036)
Temperature	** 0.010	(0.001)	** 0.141	(0.019)
Length of bike paths	** 0.037	(0.003)	* 0.608	(0.091)
Urban	** -0.043	(0.009)	* -0.524	(0.230)
Petrol price	0.099	(0.061)	1.689	(0.976)
Petrol price $\times$ Urban	* 0.151	(0.060)	* 2.057	(0.908)
Open space	0.003	(0.038)	-0.006	(0.774)
County size	-0,015	(0.036)	-0.339	(0.753)
# Kids	** 0.023	(0.004)	** 0.277	(0.093)
Age	** 0.002	(0.000)	** 0.021	(0.006)
High income	** -0.049	(0.012)	** -0.759	(0.250)
Middle income	-0.010	(0.012)	* -0.433	(0.246)
Full time employed	** -0.103	(0.008)	** -1.038	(0.171)
Year trend	-0,006	(0.004)	-0,072	(0.060)
Number of Obs.	16,306		4,615	

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively.

Three main findings emerge. First, the proximity of rail transit is associated with an increased probability of bicycle use, perhaps owing to the possibility – common to rail systems in Germany – for riders to take their bicycles along on trains. This result, however, is subject to the caveat that it may suffer from endogeneity bias arising from residential choice decisions. The second main finding is that bike paths, which we argue are an exogenous variable given the scale of their measurement, have positive impacts on both the probability of bicycle use and the distance traveled. Last, we find a positive impact of fuel prices on the probability of bicycle use, but one that only holds among households located in highly urbanized counties. Among such households, we additionally find evidence from the two-part model that higher fuel prices increase the distance traveled by bicycle for non-recreational purposes.

Taken together, these results suggest that policy-makers can avail a combination of measures to encourage bicycle use, some of which may have synergistic effects. Higher fuel costs, for example, not only directly promote a shift to bicycle use, but can also generate revenue for complementary measures that make bicycling more enjoyable and convenient, for example through the construction of bike lanes or dedicated paths that separate automobiles from bicycle traffic. Additional synergies that facilitate the coupling of bicycle travel with public transit use also hold promise for mitigating transport-related externalities.

References

Angrist, J. D., Pischke, J.-S., 2008. Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Belsley, D. A., Kuh, E., Welsch, R. E., 1980a. Detecting and assessing collinearity. Wiley Series in Probability and Statistics. Regression diagnostics: Identifying influential data and sources of collinearity, 85–191.

Belsley, D. A., Kuh, E., Welsch, R. E., 1980b. Regression diagnostics. Wiley & Sons, New York.

Bento, A. M., Cropper, M. L., Mobarak, A. M., Vinha, K., 2005. The effects of urban spatial structure on travel demand in the united states. Review of Economics and Statistics 87 (3), 466–478.

Bergström, A., Magnusson, R., 2003. Potential of transferring car trips to bicycle during winter. Transportation Research Part A: Policy and Practice 37 (8), 649–666.

Bundesministerium für Verkehr und digitale Infrastruktur (BMVI), 2015. Verkehr in Zahlen 2015/2016. `http://www.bmvi.de/SharedDocs/DE/Artikel/K/verkehr-in-zahlen_2015.html`.

Bundesministerium für Verkehr und digitale Infrastruktur (BMVI), 2016. Verkehr und Mobilität. `https://www.bmvi.de/SharedDocs/DE/Artikel/G/fahrradverkehr.html`.

Cragg, J. G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica, 829–844.

Deaton, A., Stone, A. A., 2014. Evaluative and hedonic wellbeing among those with and without children at home. Proceedings of the National Academy of Sciences 111 (4), 1328–1333.

Frondel, M., Vance, C., 2013. Re-Identifying the Rebound: What About Asymmetry? Energy Journal 34 (4), 43–54.

Goetzke, F., Rave, T., 2011. Bicycle use in Germany: Explaining differences between municipalities with social network effects. Urban Studies 48 (2), 427–437.

Handy, S. L., Xing, Y., Buehler, T. J., 2010. Factors associated with bicycle ownership and use: a study of six small US cities. Transportation 37 (6), 967–985.

Hay, J. W., Leu, R., Rohrer, P., 1987. Ordinary Least Squares and Sample-Selection Models of Health-Care Demand Monte Carlo Comparison. Journal of Business Economics & Statistics 5 (4), 499–506.

Heckman, J. J., 1979. Statistical models for discrete panel data. Department of Economics and Graduate School of Business, University of Chicago.

Heinen, E., Van Wee, B., Maat, K., 2010. Commuting by bicycle: an overview of the literature. Transport Reviews 30 (1), 59–96.

Lanzendorf, M., Busch-Geertsema, A., 2014. The cycling boom in large German cities - Empirical evidence for successful cycling campaigns. Transport Policy 36, 26–33.

Leung, S. F., Yu, S., 1996. On the choice between sample selection and two-part models. Journal of Econometrics 72 (1), 197–229.

Manning, W. G., Duan, N., Rogers, W. H., 1987. Monte Carlo evidence on the choice between sample selection and two-part models. Journal of Econometrics 35 (1), 59–82.

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., Weather, R. D., 2005. Cycling and the built environment, a US perspective. Transportation Research Part D: Transport and Environment 10 (3), 245–261.

Noland, R. B., Kunreuther, H., 1995. Short-run and long-run policies for increasing bicycle transportation for daily commuter trips. Transport Policy 2 (1), 67–79.

Parkin, J., Wardman, M., Page, M., 2008. Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation 35 (1), 93–109.

Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. Transport Reviews 28 (4), 495–528.

Pucher, J., Komanoff, C., Schimek, P., 1999. Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling. Transportation Research Part A: Policy and Practice 33 (7), 625–654.

Rietveld, P., Daniel, V., 2004. Determinants of bicycle use: do municipal policies matter? Transportation Research Part A: Policy and Practice 38 (7), 531–550.

Ritter, N., Vance, C., 2011. The determinants of bicycle helmet use: evidence from Germany. Accident Analysis & Prevention 43 (1), 95–100.

Rodrıguez, D. A., Joo, J., 2004. The relationship between non-motorized mode choice and the local physical environment. Transportation Research Part D: Transport and Environment 9 (2), 151–173.

Sardianou, E., Nioza, E., 2015. Who are the eco-bicyclists? Transportation Research Part D: Transport and Environment 34, 161–167.

Vance, C., Ritter, N., 2014. Is peace a missing value or a zero? On selection models in political science. Journal of Peace Research 51 (4), 528–540.

Winters, M., Davidson, G., Kao, D., Teschke, K., 2011. Motivators and deterrents of bicycling: comparing influences on decisions to ride. Transportation 38 (1), 153–168.

Wooldridge, J. M., 2010. Econometric analysis of cross section and panel data. MIT press.