Transportation (2012) 39:409–432
DOI 10.1007/s11116-011-9355-8

Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes

Ralph Buehler · John Pucher

Published online: 6 July 2011

Abstract This article analyzes the variation in bike commuting in large American cities, with a focus on assessing the influence of bike paths and lanes, which have been the main approach to increasing cycling in the USA. To examine the role of cycling facilities, we used a newly assembled dataset on the length of bike lanes and paths in 2008 collected directly from 90 of the 100 largest U.S. cities. Pearson's correlation, bivariate quartile analysis, and two different types of regressions were used to measure the relationship between cycling levels and bikeways, as well as other explanatory and control variables. Ordinary Least Squares and Binary Logit Proportions regressions confirm that cities with a greater supply of bike paths and lanes have significantly higher bike commute rates—even when controlling for land use, climate, socioeconomic factors, gasoline prices, public transport supply, and cycling safety. Standard tests indicate that the models are a good fit, with R^2 ranging between 0.60 and 0.65. Computed coefficients have the expected signs for all variables in the various regression models, but not all are statistically significant. Estimated elasticities indicate that both off-street paths and on-street lanes have a similar positive association with bike commute rates in U.S. cities. Our results are consistent with previous research on the importance of separate cycling facilities and provide additional information about the potentially different role of paths vs. lanes. Our analysis also revealed that cities with safer cycling, lower auto ownership, more students, less sprawl, and higher gasoline prices had more cycling to work. By comparison, annual precipitation, the number of cold and hot days, and public transport supply were not statistically significant predictors of bike commuting in large cities.

R. Buehler (✉)
School of Public and International Affairs, Virginia Tech, Alexandria Center, 1021 Prince Street, Suite 200, Alexandria, VA 22314, USA
e-mail: ralphbu@vt.edu

J. Pucher
Bloustein School of Planning and Public Policy, Rutgers University, 33 Livingston Avenue, Room 363, New Brunswick, NJ 08901, USA
e-mail: pucher@rutgers.edu

Z 1

Keywords Bicycling · Urban transport · Infrastructure · Bike lanes · Bike paths · Sustainability

Introduction

The mounting body of evidence on the health benefits of cycling has led government agencies, public health organizations, and medical journals to advocate more cycling as a way to improve individual health as well as reduce air pollution, carbon emissions, noise, traffic dangers, and other harmful impacts of car use (British Medical Association 1992; Cavill et al. 2006; CEMT 2004; Dora and Phillips 2000; IOTF 2010; NACTO 2010; USDHHS 1996, 2008; USDOT 1994, 2004, 2010d). Cities around the world have been implementing a wide range of infrastructure, programs, and policies to encourage more cycling (Fietsberaad 2010; Heinen et al. 2010; Krizek et al. 2009; Pucher et al. 2010). Most American cities have focused on providing separate bicycling facilities such as off-street bike paths and on-street bike lanes (Alliance for Biking and Walking 2010; NACTO 2010; Pucher et al. 1999; USDOT 2010d). Past research suggests that separate cycling facilities are associated with higher cycling levels. There is contradictory evidence, however, on the impacts of different kinds of facilities. Some studies find that bike paths are associated with higher cycling levels, but that lanes are not. Other studies find that lanes are related to more cycling, but paths are not. Most prior research that distinguishes between paths and lanes focuses on only one city per study. Most comparative analysis of different cities is hampered by small sample size—usually fewer than 45 cities.

This article examines the link between cycling facilities and cycling levels by analyzing new data on bike lanes and paths in 90 of the 100 largest U.S. cities. The League of American Bicyclists and the Alliance for Biking and Walking collected the data for the authors directly from planners, transportation experts, and government officials in each city for the year 2008. The only comparable measure of bike lane supply available for all 90 cities was 'centerline miles' of roads with bike lanes. Data collected for bike paths combined off-road facilities exclusively for cycling as well as multi-use paths shared by cyclists, pedestrians, joggers, in-line skaters, and other non-motorized users. Our multiple regression analysis focuses on measuring the relationship of bike paths and lanes to cycling levels while controlling for cycling safety, socioeconomic factors, land-use, gasoline price, public transport supply, and climate.

Determinants of cycling: the role of off-street paths and on-street lanes

Several studies have estimated the relationship of bike paths and lanes to cycling levels. Results from aggregate cross-sectional studies indicate that there is a positive correlation between cycling levels and the supply of bike paths and lanes (Dill and Carr 2003; LeClerc 2002; Nelson and Allen 1997; Parkin et al. 2008). Based on a sample of 18 small and large U.S. cities, Nelson and Allen (1997) find that one additional mile of combined bike paths and lanes per 100,000 residents is associated with a 0.069% increase in commuters cycling to work. Based on a sample of 42 large U.S. cities, Dill and Carr (2003) find that each additional linear mile of bike lanes per square mile of city area is associated with an increase of roughly one percentage point in the share of bike commuters, even after controlling for days of rain, automobile ownership, and state spending on walking and cycling.

22

Analyzing data from the 1990 and 2000 U.S. Census, Barnes et al. (2006) find that increases in bike commute levels in Minneapolis and St. Paul were concentrated around newly constructed bike paths and lanes. Cleaveland and Douma (2009) apply the same methods in their case study analysis of six cities and report that the relationship of bike facilities and cycling levels is mediated by local circumstances, such as network connectivity, bike promotion programs, and location of bike facilities along commuting routes leading to downtown.

Disaggregate, individual-level studies report a preference for separate paths and lanes over cycling in traffic (Abraham et al. 2002; Akar and Clifon 2009; Broach et al. 2011; Dill 2009; Dill and Gliebe 2008; Howard and Burns 2001; Hunt and Abraham 2007; Krizek et al. 2007; Lusk et al. 2011; Menghini et al. 2010; Shafizadeh and Niemeier 1997). In a study of Calgary, Canada, Abraham et al. (2002) find that cycling along roads is perceived to be two to four times as onerous as cycling on a bike path in a park. Dill and Gliebe (2008) report that women and inexperienced cyclists in Portland, OR prefer riding on bicycle paths, lanes, and low traffic volume roads over cycling on busy streets.

Findings on the relative importance of paths compared to lanes are contradictory. Vernez-Moudon et al. (2005) report that household proximity to bike paths in Seattle, WA increases the likelihood to cycle by 20%, but they find no effect for bike lanes. Using a wide range of datasets and methods, Cervero et al. (2009), de Geus et al. (2008), and Dill and Voros (2007) report no positive correlation between bike lanes and cycling levels. By comparison, a Minneapolis, MN study by Krizek and Johnson (2006) finds an increased likelihood of cycling for individuals living within 400 m of a bike lane, but no significant impact of bike paths.

Controlling for other determinants of cycling, before-and-after studies show increased levels of cycling after the installation of bike lanes, but report mixed results for bike paths (City of Toronto 2001; City of Vancouver 1999; Cohen et al. 2008; Evenson et al. 2005). A revealed preference survey by Dill (2009) finds that cyclists in Portland are willing to increase trip distance and travel time to ride on bike paths compared to shorter, more direct routes that require cycling on roads with motor vehicle traffic. Furthermore, a revealed preference study by Aultman-Hall et al. (1998) finds that bike paths in Guelph, Ontario are more likely to be used by recreational cyclists than by commuters.

In short, many studies conclude that there is a significant relationship between cycling facilities and cycling levels, but the analyses cannot determine the direction of causation. Moreover, regression analysis of cycling levels is almost always cross-sectional, thus limiting inferences about changes over time. Measurements of cycling volumes before and after the installation of specific facilities provide the simplest kind of time-series evidence, but they almost never control for the range of other factors affecting cycling levels. Most individual-level studies focus on one or a few cities. Such disaggregate, individual level studies can help mitigate some of the problems of aggregate data analysis, but transferring the results to other cities may be difficult because of policy, land use, and cultural differences between cities. Moreover, single-city studies cannot control for the influence of factors such as climate and gasoline price, which do not vary much within any particular city. Aggregate studies usually have a much larger geographic range than disaggregate studies, but they rely on few observations, such as Nelson and Allen (1997) and Dill and Carr (2003), with samples of 18 and 42 cities, respectively. Thus, all studies of the impacts of cycling facilities have their limitations. Our own study is no exception, but it enables analysis of an extensive new dataset of 90 U.S. cities that permits differentiation between bike paths and bike lanes while controlling for a range of other variables.

23

Data sources and variables

Our regression analysis investigates the relationship between bike lanes and paths and cycling levels in 90 of the 100 largest U.S. cities as determined by population estimates of the 2008 American Community Survey (ACS) (USDOC 2009a). The ACS reports city data following jurisdictional and governmental boundaries (USDOC 2010). City governments provided information on the supply of bike paths and lanes within their official city boundaries. Unless indicated otherwise, data for the variables used in our analysis pertain to the area within the city government jurisdiction. Data for some variables, such as public transport service supply, are only available for the metropolitan statistical area (MSA), including the principal city, suburban areas, and smaller secondary cities. We explicitly indicate in our analysis when we used regional instead of local data. The dependent variable—cycling level—is measured at the city level in two different ways: (1) percentage of commuters by bicycle—bike mode share—which controls for the number of workers in each city; and (2) the number of bike commuters per 10,000 population, which controls for population size.

Data on cycling levels and bikeway facilities

Data on the share of workers regularly commuting by bicycle were derived from the American Community Survey (ACS) 2006–2008 three-year average sample. The specific question posed to survey respondents was: "How did you usually get to work last week?" Respondents were asked to indicate only the main mode if they used more than one. Pooling data from the ACS surveys for 2006, 2007, and 2008 increases sample size and improves the reliability of estimates. Ideally, we would have measured cycling rates for all trip purposes, but the ACS data only report information on commuting to work, and the ACS is the only source of comparable travel data for all cities. The 2001 and 2009 National Household Travel Surveys (NHTS) provide data for all trip purposes, but their sample sizes are less than 3% as large as the ACS surveys and do not permit statistically reliable estimates for individual cities.

Table 1 displays the top ten of the 90 cities in our sample based on three measures of bike commute levels. Large cities dominate the list of total bike commuters (last column), while cities in the Midwest, West, and Southwest have the highest share of bike commuters on a per capita basis (first two columns).

The League of American Bicyclists and the Alliance for Biking and Walking collected data for the authors on the supply of bike lanes and paths by directly contacting bike planners, transportation officials, and bicycling experts in the 100 largest cities. Data for 10 of the 100 cities were not available even after multiple attempts to obtain the information. In spite of the missing cities, the resulting database for 90 cities is the most current and extensive source of information on the extent of bikeway networks in large U.S. cities.

Cities use different methods for recording the extent of their facilities. To correct for that inconsistency and to ensure the comparability of data among cities, the League of American Bicyclists and the Alliance for Biking and Walking used a uniform definition of bike lanes: centerline miles of roads with bike lanes. In order to be included, bike lanes had to be clearly designated with pavement markings and signage. They exclude shared bus and bike lanes as well as 'sharrowed' lanes intended for joint use by motor vehicles and bicycles. Calculating centerline miles of bike lanes requires adding the length of all stretches of roadway with a bicycle lane. Centerline miles do not distinguish between streets with bike lanes on only one side, in only one direction, and streets with bike lanes

24

Table 1 Top ten of 90 of the 100 largest U.S. cities by daily bike commuting levels, 2006–2008

Rank	% of commuters by bike		Bike commuters per 10,000 population		Bike commuters in 1,000	
1	Portland, OR	4.7	Portland, OR	24.0	New York City, NY	24.0
2	Madison, WI	3.9	Madison, WI	22.2	Portland, OR	13.2
3	Minneapolis, MN	3.5	Minneapolis, MN	18.9	Chicago, IL	12.8
4	Boise, ID	3.4	Boise, ID	17.8	Los Angeles, CA	12.6
5	Seattle, WA	2.5	Seattle, WA	14.2	San Francisco, CA	10.7
6	San Francisco, CA	2.5	San Francisco, CA	13.5	Seattle, WA	8.1
7	Sacramento, CA	2.0	Washington, DC	9.9	Philadelphia, PA	7.5
8	Washington, DC	2.0	Sacramento, CA	8.9	Minneapolis, MN	6.8
9	Oakland, CA	1.9	Oakland, CA	8.8	Washington, DC	5.8
10	Tucson, AZ	1.8	Denver, CO	8.4	San Diego, CA	5.3

Source USDOC (2009a)

on both sides, serving both directions of travel. Thus, the centerline measure understates bicycle facility supply on roads with bike lanes in both directions relative to roads with bike lanes in only one direction. We had to accept that limitation of the centerline measure, since it is the only comparable statistic all 90 cities could compute.

Bike paths comprised both exclusive off-road facilities for cycling as well as multi-use paths intended for joint use by cyclists, pedestrians, joggers, in-line skaters, and other non-motorized users. In fact, most bike paths in American cities are such multi-use paths, while in Europe, they are often exclusively for cyclists, probably due to the much higher cycling volumes needed to justify completely separate paths only for cyclists (Alliance for Biking and Walking 2010; Fietsberaad 2010; USDOT 2010d).

Figure 1 plots the supply of bike paths per 100,000 population against bike lanes per 100,000 population for the 90 cities in our sample. Both variables were normalized by a natural logarithm transformation. There is only a weak bi-variate correlation (Pearson's $r = 0.2$) between bike path and lane supply; and it is not statistically significant at $P < 0.05$. Thus, it is not necessarily the case that cities with many bike paths have many bike lanes as well, nor that cities with few bike paths also have few bike lanes. The graphical analysis suggests that cities in the western United States have a larger supply of bike paths per capita than in other regions. That is confirmed by results of an Analysis of Variance (ANOVA) which indicate that cities in the West Census Region[1] have a larger supply of bike lanes than cities in the Midwest, South, or Northeast ($P < 0.05$). However, there was no statistically significant difference in the supply of bike paths across U.S. Census regions ($P < 0.05$). Bivariate correlations were either weak or not statistically significant ($P < 0.05$) between our main explanatory variables and the control variables we later introduce into our models. The Pearson correlation coefficients were statistically significant but weak between bike path and lane supply (combined) and cycling safety (−0.33), share of households without a car (−0.24), retail price of gasoline (+0.29), and annual precipitation (−0.37). Bivariate correlations were both weak and not statistically significant for the relationships between bike path and lane supply (combined) and share of

[1] The western Census region includes Alaska, Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

25

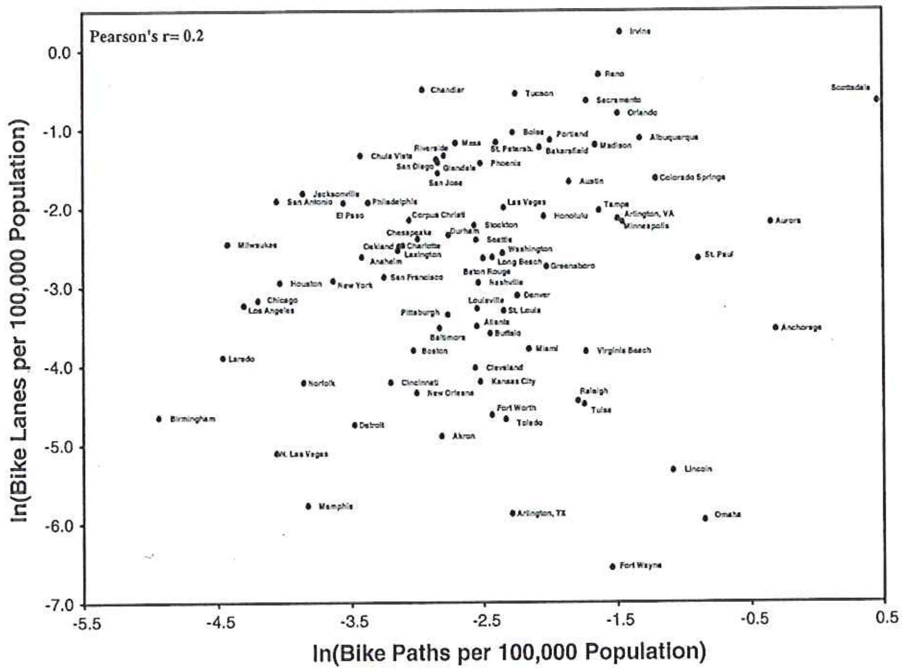

Fig. 1 Supply of bike paths and lanes in 90 of the 100 largest U.S. cities, 2008 (natural log. scale). *Source*: Data collected for the authors from each city by the League of American Bicyclists and the Alliance for Biking and Walking. *Notes*: Seven cities reported 0 miles of either bike lanes or paths. The natural logarithm of 0 is not defined. Thus, data for these cities are not displayed in Fig. 1. Please see footnote 3 for the treatment of those cities in the analysis. This *graph* was created with an Excel tool provided online by Wagner A. Kamakura

students in the population (+0.11), compact land use[2] (+0.06), public transport supply (+0.03), and annual number of hot (+0.02) and cold days (−0.07).

Control variables

In estimating the relationship between bikeways and bike commuting, our multiple regression analysis controls for other determinants of cycling commonly cited in the literature. We describe these variables and their measurement in detail below. Summary statistics of our main explanatory and control variables are presented in Table 2.

Cycling safety is an important determinant of cycling levels. The causation probably goes in both directions. Several studies confirm that increased cycling safety encourages more people to cycle (Alliance for Biking and Walking 2010; Fietsberaad 2006, 2010; Jacobsen et al. 2009a; Pucher and Buehler 2008; USDOT 2010d). Conversely, the concept of 'safety in numbers' proposes that, as more people cycle, it becomes safer because more cyclists are more visible to motorists, and an increasing percentage of motorists are also cyclists, which probably makes them more considerate of cyclists when driving. As cycling grows, it is increasingly viewed as normal, gains legitimacy as a means of travel, and

[2] Compact land use is measured by a composite 'sprawl index' with lower values for sprawled development and higher values for compact development as explained in detail later in the text.

26

Table 2 Descriptive statistics for variables in the analysis

Variable	Mean	Median	SD	Cases	Description & measurement	Source
Bike share of commuters	0.8	0.6	0.9	90	Percent of workers regularly commuting by bike	American Community Survey 2006–2008 averages (USDOC 2009a)
Bike commuters per capita	4.1	2.7	4.6	90	Daily total number of workers regularly commuting by bike per 10,000 population	
Bike lane supply	13.5	7.1	19.3	90	Miles of bike lanes in city per 100,000 population	Data collected from each city individually; Population data are ACS 2006–2008 averages (USDOC 2009a)
Bike path supply	12.9	7.9	20.0	90	Miles of bike and shared-use paths in city per 100,000 population	
Cycling safety	6.7	5.8	4.2	90	State level data: three year average of bicyclist fatality rate per 10,000 bike commuters	NHTSA 2006–2008, averages (USDOT 2010a) ACS 2006–2008 averages (USDOC 2009a)
College students	8.4	7.8	2.9	90	Percent of total population enrolled in college or university	ACS 2006–2008 averages (USDOC 2009a)
Car access	13.4	9.9	9.7	90	Percent of households without a motorized vehicle	ACS 2006–2008 averages (USDOC 2009a)
Sprawl index	101.3	101.8	23.2	90	Regional index combining 22 variables measuring residential density, mix of land uses, strength of downtowns, and connectivity of street network. (Note: higher scores = less sprawl)	Ewing et al. (2002)
Public transport supply	18,915	16,831	10,564	90	Regional annual vehicle miles of public transport supply per 1,000 inhabitants	National Transit Database (USDOT 2008)
Gasoline price	263.5	262.1	8.7	90	Average state retail price of gasoline (in cents) (2006–2008)	U.S. Department of Energy (2010a, b)
Hot weather	52.5	34.5	46.0	90	30 year average of annual number of days above 90°F	National Climatic Data Center (NCDC) (2010)
Cold weather	61.9	51.5	54.7	90	30 year average of annual number of days below 32°F	National Climatic Data Center (NCDC) (2010)
Annual precipitation	31.9	34.7	15.6	90	30 year average of annual inches of precipitation	National Climatic Data Center (NCDC) (2010)

Sources Alliance for Biking and Walking (2010), Ewing et al. (2002), National Climatic Data Center (2010), USDOC (2009a, b), USDOE (2010a), USDOT (2008, 2010a)

27

generates more public and political support for more and better cycling facilities. Regardless of which explanation is correct, several studies find significant time-series as well as cross-sectional evidence of 'safety in numbers' (Elvik 2009; Jacobsen 2003; Robinson 2005).

In our analysis, we measured safety as cyclist fatalities per 10,000 bike commuters at the state level. The National Highway Safety Administration (NHTSA) reports annual fatalities for states but not for cities. Reliable cyclist fatality data are not available at the city level. Cyclist fatalities are rare events, so cities with little cycling have few fatalities and do not collect such data systematically. Thus, the fatality rates used in our analysis refer to cycling safety in the overall state and not the city itself. In addition to that geographic discrepancy, the fatality rate is only a rough approximation of actual cycling safety. Cyclist fatalities result from all trip purposes and not just the trip to work, but the measure of exposure in the denominator of the fatality rate includes only bike commuters. As mentioned earlier, the only nationally comparable source of travel data for all trip purposes is the NHTS. Because the NHTS sample size is less than 3% as large as the ACS sample, it cannot be disaggregated to the state or city level with statistical reliability to calculate total bike trips for all trip purposes. Thus, the fatality rate we calculated is only a very rough approximation, but it helps capture the sharp differences in cycling safety across states: ranging from less than 2 fatalities per 10,000 bike commuters in Alaska, Colorado, Minnesota, and Oregon to over 20 in Alabama (Alliance for Biking and Walking 2010).

Two socioeconomic variables we included were share of students in the population and percent of households without a car. Previous studies find that individuals in households with more cars are less likely to ride a bicycle, while students are more likely to cycle (Dill and Carr 2003; Heinen et al. 2010; Pucher and Buehler 2006). We did not include per-capita income because of its high correlation with car ownership (Pearson's $r = 0.6$). The most important impact of income on cycling levels is via car ownership (Dill and Voros 2007; Heinen et al. 2010; Stinson and Bhat 2003). Moreover, the two most recent national travel surveys for the United States, the 2001 and 2009 NHTS, reveal no statistically significant difference in cycling levels among income groups, but a large and statistically significant difference by car ownership levels (Buehler et al. 2011; Pucher et al. 2011a; USDOT 2010b, c).

Previous studies have shown that cycling levels are higher in dense, mixed-use developments with short trip distances and proximity of households to destinations such as offices, stores, and restaurants (Baltes 1997; Ewing and Cervero 2001, 2010; Guo et al. 2007; Handy 1996; Litman 2007a; Moudon et al. 2005; Parkin et al. 2008; Pucher and Buehler 2006; Zahran et al. 2008). Moreover, studies find that a grid-pattern road network increases levels of cycling because short blocks and frequent intersections provide easier bike access and more flexible bicycle route choice to most destinations (Ewing and Cervero 2010).

In our study, we approximate the influence of the built environment by using the composite sprawl index that was developed by Ewing et al. (2002). The sprawl index combines 22 different variables measuring various aspects of urban form, mix of land uses, density, and street network connectivity. Of the cities included in our study, the metropolitan areas with the worst sprawl ratings (lowest numerical values) were Riverside-San Bernardino, CA (14.2), Greensboro, NC (46.8), Raleigh, NC (54.2), and Atlanta, GA (57.7). The metropolitan areas with the best sprawl ratings (highest numerical values) were: New York City, NY (177.8), San Francisco, CA (146.8), and Honolulu, HI (140.2). Although the sprawl index refers to the metropolitan area as a whole, it is also useful for comparing land-use characteristics of the central cities included in our study. For example,

28

the index specifically considers several measures of downtown strength and overall compactness of the urban area. There is no comprehensive land-use index that provides comparable information for central cities only. Thus, we had to assume that the relative differences in land use among metropolitan areas as a whole reflect the relative differences among their central cities.

Public transport may also influence cycling levels. Some studies show that coordinating cycling with public transport can encouraging more cycling as well as more public transport use (Brons et al. 2009; Givoni and Rietveld 2007; Hegger 2007; Martens 2004, 2007; TRB 2005; USDOT 1998). Other studies, mainly from Europe, suggest that public transport may compete with bicycling for short trip distances in cities with good public transport supply (Fietsberaad 2010; Heinen et al. 2010; Pucher and Buehler 2007; Schwanen 2002). Our study includes a variable measuring public transport vehicle miles per capita from the National Transit Database (NTD) for the year 2008 (USDOT 2008). Data were only available at the metropolitan level, since service areas of public transport agencies almost always extend beyond central city boundaries into the suburbs (USDOT 2008).

Few studies specifically examine the impact of gasoline prices and taxes on cycling levels (Pucher and Buehler 2006; Rashad 2009). However, many studies find that higher gasoline prices lead to less driving (Buehler 2010; DeJong and Gunn 2001; Epsey 1998; Hanly et al. 2002; Litman 2007b). In our study we use average gasoline prices by state for the years 2006–2008, as reported by the Energy Information Administration (EIA) (US-DOE 2010a). Comparable data on gasoline prices in each of the 90 cities in our study were not available for the years 2006–2008. The state data are only proxies for the unavailable city data, but at least they capture major differences in state gasoline tax rates, fuel distribution costs, and state standards for fuel composition, all of which help determine the final retail price of gasoline (USDOE 2010a, b). The state rates do not, however, reflect variation within states in gasoline taxes and prices.

Previous research shows that climate and topography can affect cycling levels. Several studies find that cycling is deterred by rain as well as by very cold or hot weather (Baltes 1997; Bergström and Magnusson 2003; Dill and Carr 2003; Gatersleben and Appleton 2007; Heinen et al. 2010; Nankervis 1999; Stinson and Bhat 2003; Winters et al. 2007). Our analysis includes three variables measuring weather and climate: (1) average annual number of days that reach temperatures of over 90°F; (2) average number of days below 32°F; and (3) annual precipitation levels. We used 30 year average data for each city provided by the National Climatic Data Center (2010).

Almost all studies find that flat topography facilitates cycling, and that cyclists choose routes that avoid steep gradients (Hunt and Abraham 2007; Menghini et al. 2010; Rietveld and Daniel 2004; Timperio et al. 2006; Vandenbulcke et al. 2011). Topography uninterrupted by harbors, bays, and rivers also favors cycling by enabling more direct routes (Pucher et al. 2011c). However, standardized indices of topography do not yet exist for the cities in our sample. Thus, we were not able to control for the influence of topography on cycling levels.

Similarly, it was not possible to include variables measuring the extent and quality of the many other policies and programs that might potentially affect cycling levels (Heinen et al. 2010; Krizek et al. 2009; Pucher et al. 2010). These measures include, for example, bike parking, bike racks on buses, bike sharing programs, cycling training courses, media campaigns, and educational events (APBP 2002; Brons et al. 2009; Fietsberaad 2010; Givoni and Rietveld 2007; Hegger 2007; Hunt and Abraham 2007; Martens 2007; Netherlands Ministry of Transport 2009; Noland and Kunreuther 1995; Taylor and Mahmassani 1996; TRB 2005; Wardman et al. 2007). Comparable data for these programs are not available for most of the 90 cities.

29

Table 3 Bike commute levels by quartile of independent variables and bivariate Pearson's correlations for the 90 largest U.S. cities

	Share of bike commuters by quartile of independent variable				Difference fourth minus first quartile	Bivariate correlation with share of bike commuters
	First quartile	Second quartile	Third quartile	Fourth quartile		
Bike lanes per 100,000 pop.	0.4	0.7	0.9	1.3	+0.9**	0.5**
Bike paths per 100,000 pop.	0.5	0.8	0.8	1.2	+0.7**	0.3**
Bike paths and lanes per 100,000 pop.	0.5	0.6	0.7	1.5	+1.0**	0.5**
Cyclist fatality rate	1.5	0.6	0.6	0.4	−1.1**	−0.5**
% College students	0.4	0.6	1.1	1.3	+0.8**	0.5**
% Households without car	0.8	0.5	1.1	1.0	+0.2*	0.1
Sprawl index	0.5	0.8	0.9	1.1	+0.6**	0.2*
Transit revenue miles per capita	0.6	0.6	1.0	1.1	+0.5*	0.1
Gas price	0.4	0.7	0.8	1.5	+1.1**	0.5**
Days above 90°F	1.4	0.6	0.8	0.6	−0.8**	−0.3**
Days below 32°F	0.9	0.8	0.5	1.1	+0.2	0.1
Annual inches of precipitation	0.8	1.1	0.7	0.5	−0.5**	−0.2**

	Bike commuters per 10,000 population by quartile of independent variable				Difference fourth minus first quartile	Correlation with bike commuters per 10,000 population
	First quartile	Second quartile	Third quartile	Fourth quartile		
Bike lanes per 100,000 pop.	1.7	3.3	4.7	6.6	+4.8**	0.5**
Bike paths per 100,000 pop.	2.4	3.7	3.9	6.2	+4.0**	0.3**
Bike paths and lanes per 100,000 pop.	2.5	2.7	3.5	7.8	+5.3**	0.5**
Cyclist fatality rate	7.6	2.7	3.0	1.9	−4.2**	−0.5**
% College students	2.0	2.7	5.3	6.4	+4.4**	0.5**
% Households without car	3.9	2.2	5.6	4.8	+0.9*	0.1
Sprawl index	2.4	4.1	4.3	5.4	+3.0**	0.2*
Transit revenue miles per capita	3.0	2.7	5.2	5.7	+2.7**	0.1
Gas price	1.8	3.6	4.0	7.4	+5.6**	0.5**
Days above 90°F	7.1	3.0	3.9	2.7	−3.7**	−0.3**
Days below 32°F	4.5	4.0	2.4	5.5	+1.0	0.1
Annual inches of precipitation	4.0	5.7	3.6	2.1	−1.9**	−0.2**

** Significant at the 95% level

* Significant at the 90% level

Bivariate relationships

Bicycling to work is positively correlated with both bike paths and bike lanes (see Table 3, last column). Estimates of the correlation coefficients between bike commuting and bike lanes are slightly larger than for bike paths, but the magnitude of the coefficients is not significantly different at $P < 0.05$. Our grouping of cities into quartiles of bike path and lane supply shows that bike commuting in cities with the most bike lanes per 100,000 population (4th quartile) are three to four times higher than in cities with the fewest bike lanes (1st quartile). The difference between quartiles is less pronounced for bike paths—with slightly more than twice as much bike commuting in the 4th compared to the 1st quartile. The table also displays the combined relationship of bicycle paths and lanes on bike commuting. There is three to four times as much bike commuting in cities with the most paths and lanes (4th quartile) as in cities with the least bike path and lane supply (1st quartile).

The correlation coefficients for the control variables suggest the same directions of relationships as previous studies we reviewed, but not all coefficients are statistically significant. City cycling levels and state bike fatality rates have a statistically significant negative correlation. The actual relationship might be stronger, but the state data are obviously an imperfect proxy for city cycling safety. Cities with a higher percentage of students have higher levels of bike commuting. A higher share of households without a car is associated with more bike commuting, but the bivariate correlation is not statistically significant. Bicycle commuting levels are higher in central cities of more compact metropolitan areas. Cities with more public transport supply per capita have higher cycling levels, but the correlation coefficient is not statistically significant. State gasoline retail prices and city cycling levels have a statistically significant positive correlation—consistent with the theory that higher costs of driving encourage cycling. As found by earlier studies, extreme weather conditions deter cycling. Our dataset shows that cycling levels are lower in cities with more days per year with temperatures of 90°F or higher and more annual precipitation. We found no statistically significant relationship between the number of cold days per year and bike commuting.

Multiple regression analysis

The quartile and correlation analysis presented above investigate the relationship between bike commuting and each independent variable, one at a time. The multiple regressions presented below examine the relationship of cycling levels and bike paths and lanes while controlling for safety, socioeconomics, land use, public transport supply, gasoline price, and climate.

We estimated two sets of models. The first model is a log–log Ordinary Least Square (OLS) regression with the natural log of bike commuters per 10,000 population as dependent variable. The second model is a Binary Logit Proportions Model with the share of bike commuters in each city as dependent variable. In both types of models the independent variables are expressed as natural log to assure a more normal distribution of otherwise skewed explanatory variables.

The log–log specification for the first set of models has two advantages. First, it normalizes the skewed independent and dependent variables, thus helping to meet assumptions of the OLS regression. Second, it allows interpreting the regression coefficients directly as elasticities or percentage changes in bike commuting, which makes the results more intuitive and easier to understand.[3]

[3] Seven cities reported 0 miles of bike lanes or bike paths. These cities would have been lost in our models, because the natural logarithm of 0 is not defined. Thus, we followed the common procedure of transforming the bike lane and path per 100,000 population variable by adding 1, which yields a log value of 0 for the 7 cities. We also estimated the models without this transformation, with only 83 cities. Significance, sign, and magnitude of coefficients and goodness of fit were very similar to the results of the models presented in this paper.

31

Table 4 Multiple regression analysis of bike commuters per 10,000 population and bike commute share (continues on next page)

	OLS regression of ln(bike commuters per 10,000 population)						Binary logit proportions model for share of bike commuters[a]	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Elasticity at mean
ln (bike lanes per 100,000 population)	**0.361**	**0.310**	**0.305**	**0.299**	**0.314**	**0.311**	**0.404**	**0.250**
	(5.85)**	(3.78)**	(3.76)**	(3.69)**	(3.78)**	(5.14)**	(5.65)**	(6.19)**
ln (bike paths per 100,000 population)	**0.267**	**0.245**	**0.302**	**0.181**	**0.251**	**0.230**	**0.147**	**0.091**
	(2.75)**	(2.88)**	(3.55)**	(2.26)**	(2.98)**	(2.90)**	(2.09)**	(2.08)**
ln (fatality rate per 10,000 bike commuters)		**−0.366**		**−0.397**	**−0.412**	**−0.277**	**−0.514**	**−0.320**
		(2.30)**		(2.45)*	(2.51)**	(1.97)*	(4.35)**	(4.35)**
ln (percent of students in population)		**0.859**	**0.904**	**0.863**	**0.808**	**0.879**	**0.544**	**0.340**
		(3.70)**	(3.74)**	(3.49)**	(3.53)**	(4.39)**	(2.52)**	(2.50)**
ln (percent of households without car)		**0.339**	**0.370**		**0.378**	**0.300**	**0.499**	**0.310**
		(2.55)**	(2.77)**		(2.80)**	(2.72)**	(3.57)**	(3.52)**
ln (sprawl index)		**0.362**	**0.436**	**0.426**		**0.353**	**0.340**	**0.210**
		(2.29)**	(2.55)*	(2.84)**		(2.13)**	(2.46)**	(2.33)**
ln (transit revenue miles of service per capita)		−0.106	−0.064	0.028	−0.070		−0.266	−0.140
		(0.58)	(0.33)	(0.17)	(0.37)		(1.63)	(1.39)
ln (state gas retail price)		**5.161**	**6.655**	**5.752**	**4.544**	**5.166**	**4.905**	**3.000**
		(1.76)*	(2.17)**	(1.92)*	(1.65)*	(2.14)**	(2.18)**	(2.19)**
ln (annual number of days above 90°F)		0.025	−0.049	0.005	0.022		0.01	−0.010
		(0.28)	(0.59)	(0.05)	(0.25)		(0.14)	(0.14)
ln (annual number of days below 32°F)		−0.048	−0.025	−0.029	−0.048		−0.026	−0.020
		(1.60)	(0.77)	(1.00)	(1.55)		(0.93)	(0.09)
ln (annual inches of precipitation)		0.105	−0.032	0.212	0.106		0.233	0.140
		(0.58)	(0.19)	(1.20)	(0.57)		(1.50)	(1.52)

32

Table 4 continued

	OLS regression of ln(bike commuters per 10,000 population)						Binary logit proportions model for share of bike commuters[a]	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Elasticity at mean
Constant	−0.265	−31.843	−41.186	−36.061	−27.026	−32.639	−34.669	
	(1.24)	(1.94)*	(2.40)**	(2.16)**	(1.69)*	(2.39)**	(2.74)**	
Observations	90	90	90	90	90	90	90	
Adjusted R^2	0.33	0.65	0.62	0.63	0.63	0.64	Pseudo LL (Intercept): −9.048	
F-statistic	27.44	18.14	16.31	17.00	18.37	26.00	Pseudo LL(Full): −3.399	
	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**	Pseudo R^2(McFadden): 0.62	

Note coefficients of statistically significant variables shown in bold print

Absolute value of robust *t*/*z* statistics in parentheses

[a] Logistic regression estimated via STATA GLM (generalized linear models) with logit link function, binomial distribution, and robust standard errors

* Significant at 10%

** Significant at 5%

33

Models 1 through 6 in Table 4 present the results of the OLS regression with the natural log of bike commuters per 10,000 population as dependent variable. Model 1 only includes bike path and lane supply as explanatory variables without controlling for other factors. Results confirm the positive correlation between cycling levels and bike path and lane supply from the bivariate analysis. Path and lane supply alone account for 33% of the variability in bike commuting (Adj. $R^2 = 0.33$). However, this model is underspecified and likely suffers from omitted variables bias, since theoretically relevant control variables are missing. Model 2 includes control variables for cycling safety, socioeconomics, land use, public transport supply, gasoline prices, and climate. The independent variables of Model 2 have joint significance at the 99% level ($F = 18.1$) and account for 65% of the variability in bike commuters per capita (Adj. $R^2 = 0.65$).

Coefficients are consistent with relationships reported in most other studies, but not all estimators are statistically significant. Both bike lanes and bike paths per 100,000 population are significant predictors for bike commuting. A 10% greater supply of bike lanes is associated with a 3.1% greater number of bike commuters per 10,000 population. Similarly, a 10% greater supply of bike paths is associated with a 2.5% higher level of bike commuting. As in our previous correlation analysis, a t-test comparison shows that the coefficients for bike lanes and paths are not significantly different from each other at the 95% confidence level.

Cycling safety is statistically significant as well. A 10% higher cyclist fatality rate per 10,000 commuter cyclists is associated with 3.7% fewer bike commuters per 10,000 population. A 10% higher share of students in the population is associated with 8.6% more bike commuting. A 1% increase in the retail price of gasoline is associated with a 5.2% increase in cycling levels. The cross-price elasticity of bike commuting with respect to gasoline price may seem high, but it is in line with other models estimating the relationship between gasoline prices and cycling levels (Pucher and Buehler 2006; Rashad 2009). The coefficients for public transport supply and the climate variables—number of days per year with temperatures of 90°F or higher, 32°F or lower, and precipitation—are not statistically significant.

Models 3 through 6 present regression results for reduced models, excluding explanatory variables to control for potential multicollinearity and endogeneity. For example, prior research suggests that bike paths and lanes contribute to lower cycling fatality rates (CEMT 2004; Fietsberaad 2010; Lusk et al. 2011; Pucher and Buehler 2008; Reynolds et al. 2009). Possible multicollinearity due to the inclusion of both cyclist fatality rate and bikeway supply variables in our model may siphon off strength from the bike path and lane coefficients. In our dataset of 90 cities, bivariate Pearson's correlations between the fatality rate and the supply of bike paths and lanes are below 0.3, and tests for multicollinearity do not indicate any serious problem.[4,5] Endogeneity is a second potential problem arising from the inclusion of the cyclist fatality rate variable, since 'safety in numbers' suggests that cycling safety increases with higher cycling levels (Jacobsen 2003; Jacobsen et al. 2009b). Model 3 excludes the cyclist fatality rate variable in order to test for the possible distorting influence of any multicollinearity and endogeneity problems caused by its inclusion in the model. The Model 3 estimate of the coefficient for bike path supply is only slightly larger (+0.05) than in Model 2—possibly related to greater safety of off-street

[4] Variance Inflation Factor (VIF) yields scores for individual variables below 2.7 and a score of 1.9 for the overall equation. Tolerance values are all above 0.4.

[5] A possible reason for this low correlation may be that state cyclist fatality rates are imperfect proxies for city fatality rates.

34

facilities (Lusk et al. 2011). *T*-tests show that the estimated coefficients for bike lanes, bike paths, and all other variables in Model 3 are not statistically different from Model 2.

Including car access and the sprawl index as explanatory variables may also introduce bias into Model 2. Some studies suggest that individuals who cycle more are less likely to own an automobile (Dill and Voros 2007; Parkin et al. 2008; Stinson and Bhat 2003). Similarly, studies show that individuals who prefer to cycle more may choose to live in more compact communities (Heinen et al. 2010; Krizek et al. 2009). Inclusion of these two variables might cause simultaneous equations bias, since cycling levels may also affect the choice to own a car or to live in a compact community. Moreover, car access and sprawl may themselves be negatively correlated with each other, since studies show that individuals living in compact urban areas own fewer cars (Cervero 2003; Ewing et al. 2002, 2008). To test for the possible distorting effects caused by potential simultaneous equations bias and multicollinearity, Models 4 and 5 omit the car access and the sprawl index variables. Similar to our findings in the reduced Model 3, *t*-test comparisons show that the magnitude and significance of coefficients of the remaining variables in Models 4 and 5 do not change significantly from those estimated in Model 2, where all the variables were entered into the equation.

Finally, Model 6 presents results of a reduced model including only statistically significant variables. This model confirms results from Models 2 through 5, but probably suffers from omitted variables bias. In summary, goodness of fit measures and the direction, magnitude, and significance of the model, coefficients are very similar for Models 2 through 6. In all models, the coefficients for the key explanatory variables of interest—bike paths and bike lanes—remain significant, positive, and are not statistically different from each other at the 95% confidence level. Model 2 seems preferable, because it includes all theoretically relevant variables available for this study, and is thus less prone to omitted variable bias.

We also tested the robustness of our results by re-estimating Model 2 excluding cities with extreme values for the explanatory variables. Such outliers, for example, included cities with the most or least bikeway supply, the most extreme climates, highest and lowest car ownership levels, highest and lowest student share, highest and lowest gasoline prices, and most and least public transport supply. The coefficients estimated for Model 2 without the outliers were similar to our original estimates for the entire sample of 90 cities presented in Table 4.

To test further the robustness of our results, we estimated an additional equation, presented as Model 7 in Table 4, using the share of bike commuters in each city as the dependent variable. For this dependent variable, an OLS regression might estimate values beyond the range of actual possible values of the bike share of commuters (0–1.0). To address this issue, we followed Xing et al. (2010) by estimating a non-linear Binary Logit Proportions Model for bicycle mode share.[6] This estimation technique transforms the dependent variable into the 'log of odds' of the bike share of commuters and approximates a nonlinear Maximum Likelihood estimation (Xing et al. 2010). Transformation of the dependent variable and nonlinear estimation of the model assure that predicted mode shares lie between 0 and 1.0.

Model 7 displays the results of the Binary Logit Proportions regression. Standard test statistics suggest the model is a good fit. For example, McFadden's Pseudo R^2 is 0.62. All variable coefficients are consistent with the direction of relationships reported by most other studies. Similar to Models 1 through 6, the coefficients for bike paths and lanes are significant and positive, even after using this very different, non-linear estimation technique.

[6] For an alternative approach to estimating fractional response variables using a so-called 'quasi-likelihood estimation method,' see Papke and Wooldridge (1996).

35