

Cycling Mode Choice Amongst US Commuters: The Role of Climate and Topography

Justin Tyndall

jtyndall@hawaii.edu

University of Hawai'i Economic Research Organization
and University of Hawai'i Department of Economics
2424 Maile Way, Saunders Hall 540, Honolulu, HI, USA, 96822

August, 2020

Abstract

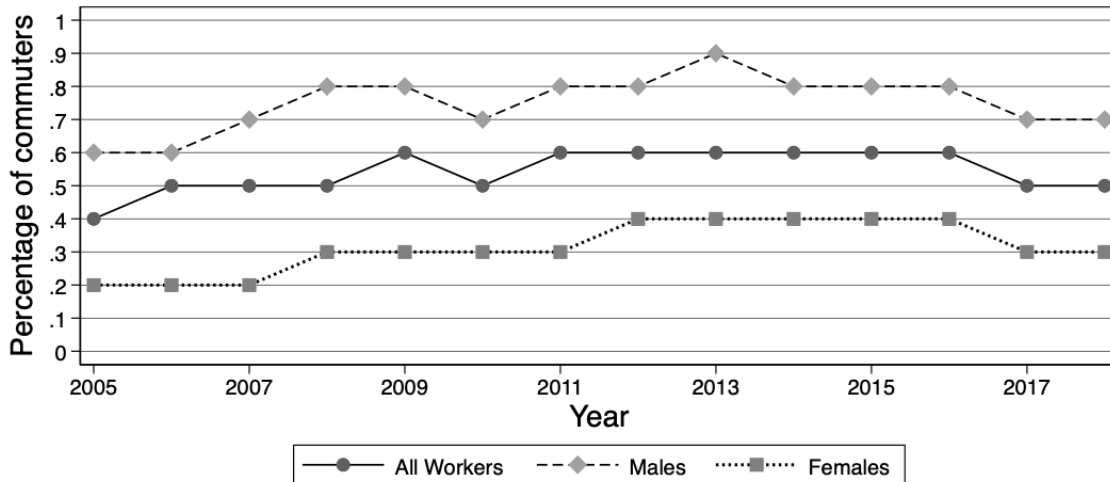
Cycling to work is uncommon in most areas of the US, but relatively common in a particular set of metros and neighbourhoods. Explanations for this spatial heterogeneity often focus on differences in local geography, with some areas being allegedly more suitable for cycling. I estimate the role of topography and climate in determining the share of a metro's workers who cycle to work and the probability a particular worker chooses to cycle. I combine a US wide data set of commute flows with detailed elevation and climate data. I find that climate and topography play essentially no role in explaining cycling mode share across metros. Across workers, the hilliness of a commuter's route is found to be statistically irrelevant to cycling mode choice.

Cycling, Transportation, Commuting, Mode Choice, Urban Planning

1 Introduction

Cycling has become a pillar of urban transportation planning. Promoting cycling as a form of commuting over private vehicle use can benefit the environment (Walsh et al., 2008), improve public health (De Hartog et al., 2010) and provide an affordable mobility option (Mattingly and Morrissey, 2014). Despite apparent advantages, only half of one percent of US workers commuted by bicycle in 2018, with the share staying relatively stable in recent years (Figure 1). This low average masks considerable heterogeneity across locations. For example, a commuter in metropolitan San Francisco is about 18 times as likely to cycle to work as a commuter in metropolitan Memphis. Within San Francisco, some neighbourhoods report 25% bicycle commuting shares, while others report 0%.¹ The cause of this enormous spatial heterogeneity in cycling uptake is unclear. This study will test whether the choice to cycle is substantially determined by local climate and topography. If cycling is a result of permanent geographic variables, it may be difficult for policy to shift workers to cycling in areas that lack suitable geography.

Figure 1: Share of US Commuters Who Cycle



Cycling has accounted for 0.5-0.6% of all commutes in recent years. Males choose cycling at roughly twice the rate of females. Data is from the American Community Survey, 1-Year Estimates.

Across US metropolitan areas the distribution of cycling mode share is highly

¹Census Transportation Planning Products, 2012-2016.

skewed (Figure 2). Most metros have very low rates of cycling, while a few have relatively high rates. Among large US metros (population over 1 million), the metros with the highest rates of commuting by bicycle are Portland, Oregon (2.43%), San Francisco, California (2.00%) and San Jose, California (1.82%).² All of the top five metros in terms of cycling rate are located in the American West while the bottom five are all in the American South. Cycling rates for many metros in the South are remarkably low. For example, in the Memphis, Tennessee metro only one out of every 885 commuters (0.11%) choose to cycle.

Figure 2: Share of Commuters Who Cycle to Work, Distribution Across Metros



In 326 out of 382 metros fewer than 1% of commuters cycle to work. Data is from the Census Transportation Planning Products 2012-2016 commuting and workplace data. Corvallis, Oregon has the highest cycling rate of any metropolitan area at 7.9%.

Within existing literature, hilly terrain is frequently cited as a disincentive to cycling (Heinen et al., 2010; Parkin et al., 2008; Rietveld and Daniel, 2004; Rodriguez and Joo, 2004; Vandenbulcke et al., 2011). Several studies have analyzed the role of hills in setting cycling rates in Europe. Rietveld and Daniel (2004) looked at cycling across Dutch cities, Parkin et al. (2008) studied districts in England and Wales and Vandenbulcke et al. (2011) studied cyclists in Belgium. These studies all found that areas with steeper slopes had lower rates of cycling. The current paper will focus on

²Census Transportation Planning Products, 2012-2016.

the US, for which there is less empirical evidence on the role of hills in cycling mode choice. Determinants of cycling in the US may be substantially different than those in Europe, given that cycling in the US is relatively uncommon and the urban structure of US cities is unique to that of European cities.

Local climate is an other potential candidate to explain cycling rates. Extreme temperatures, rainfall and snow could potentially create barriers to cycling. Dill and Carr (2003) note that several US cities with high annual rainfall also have high rates of cycling, suggesting rainfall is not an important deterrent. The relationship between cycling rates and hourly weather were analysed for Montreal, Canada in Miranda-Moreno and Nosal (2011). The authors found substantial reductions in cycling when there was precipitation or when the temperature was extreme. Looking across Canadian cities, Winters et al. (2007) found cycling rates were inversely related to the number of days with precipitation and the number of days with sub-freezing temperatures.

The spatial distribution of cycling within metros, across neighbourhoods is also extremely heterogeneous. For example, 63% of metropolitan census tracts in the US report the share of cycling among commuters as zero.³ The remaining 37% of tracts also have a skewed distribution, with a few tracts having much higher levels of cycling than average.

The most commonly noted spatial variable that determines an individual's probability to cycle is distance to work, with cycling uncommon for long commutes (Cervero, 1996; Heinen et al., 2013). Neighbourhoods that are far from job centers will therefore have low cycling rates.

Urban form and the presence of infrastructure that enables cycling have been widely studied as potential causes of cycling. Studies of the built environment have shown that higher density, mixed-use neighbourhoods correlate with higher rates of resident cycling (Saelens et al., 2003; Xing et al., 2010). Some causal evidence from a neighbourhood relocation experiment in Australia is presented in Beenackers et al. (2012), suggesting a more cycling friendly neighbourhood causes residents to switch to cycling. Estimating the role of bicycle infrastructure is difficult because infrastructure is often directed to routes with latent demand for cycling. Bike lanes and other bicycle infrastructure have generally been found to be important to cycling uptake (Hunt and Abraham, 2007; Zhao, 2014). Bicycle amenities at the workplace such as

³American Community Survey, 2016, 5-year Estimates.

bicycle parking facilities, locker rooms and showers are sometimes found to be important (Abraham et al., 2002) and in other studies have been found to be largely irrelevant to cycling choice (Stinson and Bhat, 2004).

Proper infrastructure, such as bike lanes, can have a strong effect on the safety of cycling (Reynolds et al., 2009) as well as its overall uptake (Buehler and Pucher, 2012; Cervero et al., 2009). Xing et al. (2010) found that a sense of safety while riding is important, which is related to road design and the presence of dedicated cycling infrastructure.

Cycling may be more popular if it is well integrated with other modes of transport, for example public transportation (Keijer and Rietveld, 2000). Singleton et al. (2014) conducted an empirical evaluation of bicycle uptake in the US and its relationship to local transit conditions. The results indicated that, while cycling acts as a substitute for transit for individual trips, over longer time horizons the two modes are complements.

This study will focus on the role of unalterable geographic conditions rather than endogenous infrastructure and policy variables. The focus on endowed geography aims to answer whether cities are destined to experience a particular level of bike commuting based on their climate and topography, or if the expansion of cycling is viable in areas with seemingly unsuitable geographic endowments.

This paper contributes a more statistically rigorous and data intensive approach than has been available in the prior literature. By combining a national US data set of commute flows with detailed elevation and climate data, I provide new and novel estimates on the role of geography in determining cycling rates across the US. Section 2 will provide details on data sources. Section 3 estimates the role of geography in explaining cycling heterogeneity across US metros. Section 4 investigates heterogeneity within metropolitan areas and Section 5 concludes.

2 Data

I rely on data from five unique sources (1) climate data from the National Oceanic and Atmospheric Administration (NOAA), (2) detailed elevation data from the US Geological Survey (USGS), (3) Core Based Statistical Area (CBSA) level commuting mode share and demographic estimates from the 2016 5-year American Community Survey (ACS) (4) the Census Transportation Planning Products (CTPP) 2012-2016 commuting and workplace data set and (5) a database of travel route

characteristics gathered from an online spatial navigation service.

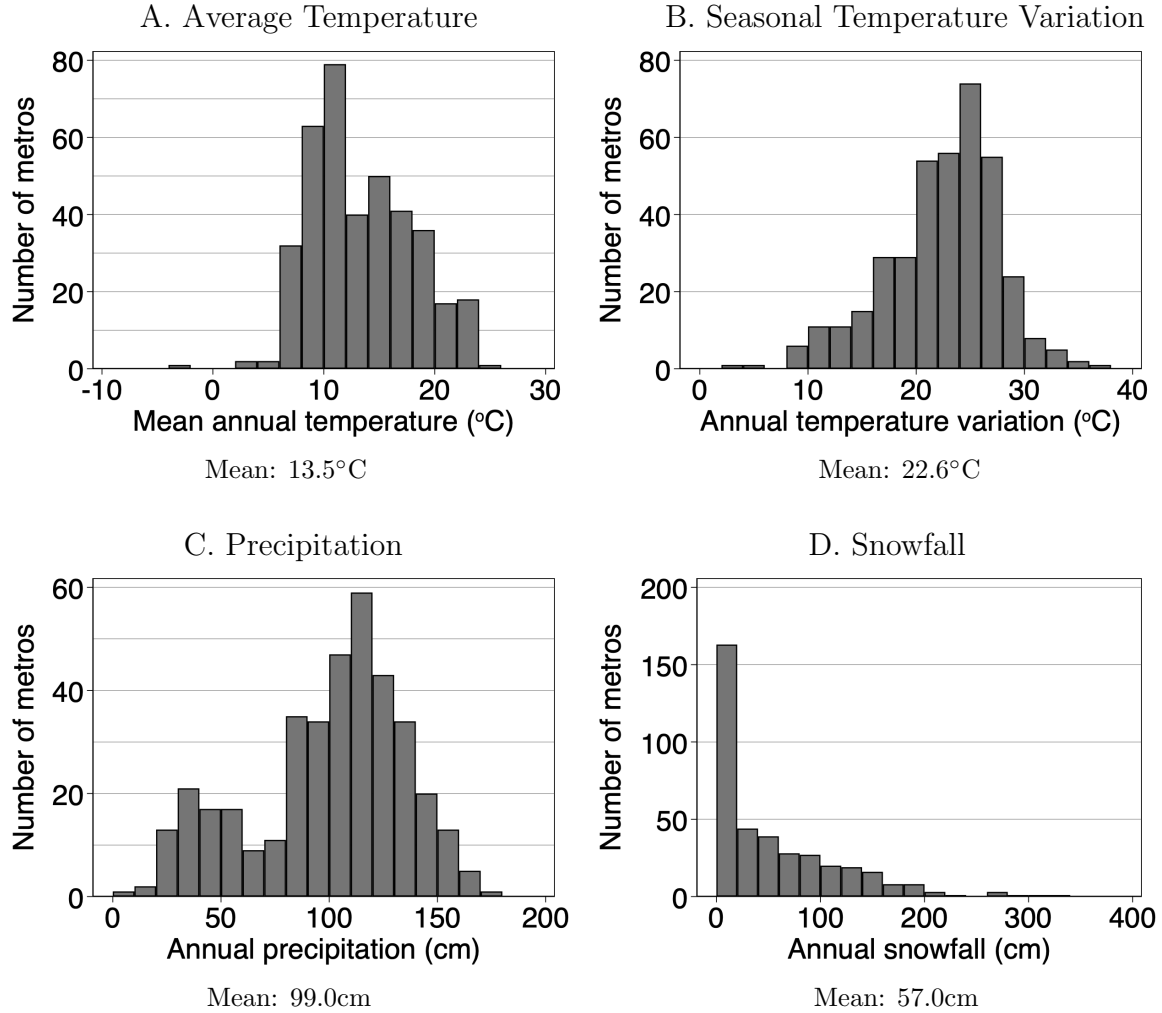
I use averaged climate conditions estimated by NOAA's 1981-2010 Climate Normals data products. The data set contains typical temperature and precipitation conditions across approximately 10,000 weather stations in the US. To calculate metro conditions, I take the average condition across all weather stations that are within the boundaries of a given metro. The majority of metros have several weather stations within their boundaries. Out of the 382 metropolitan CBSAs in the US, there are seven CBSAs that lack snowfall data, one CBSA that lacks rainfall data and one CBSA that lacks temperature data. In such cases I impute the value by using that of the nearest neighboring CBSA. Figure 3 provides information on the distribution of climate variables across metros. The US spans a diverse set of climates, allowing significant variation with which to estimate the role of climate on cycling rates.

In order to conduct analysis on the role of topography in cycling heterogeneity across metros I construct a detailed data set of elevation points for the US. I make use of the USGS Elevation Point Query Service. The web service returns the elevation of any latitude and longitude coordinates for the US. I propose the following metric to measure "hilliness" at the metropolitan level. I plot a 10 km by 10 km square on each metro, centred at the geographic centroid of the metro's principal city. I then divide the resulting box into a grid of 100 1 km by 1 km squares and use the USGS web query service to calculate the elevation at the center of each square. I then take the standard deviation of these 100 elevation points as a measure of regional hilliness. I exclude points that are in an ocean or a Great Lake. While CBSAs can differ dramatically in size, the use of a consistent measure can represent differences in topographical endowment. Figure 4 provides examples of how the measure is calculated.

Using evenly spaced points is intended to generate a representative sample of local elevation. The 10 km by 10 km square is intended to capture an area where the geography is representative for a large sample of the CBSA's commuters. Because the size of the area sampled is somewhat arbitrary, I perform robustness checks using 6 km by 6 km and 14 km by 14 km squares. Results prove to be insensitive to the size of the sampling area. The correlation between the hilliness metric measured with a 6 km or 10 km square is .85 and the correlation between the 10 km and 14 km measure is .98.

The metric's value for the average metro is 24.2 m. There is significant variation in the metric across metros. The "flattest" metros according to the metric are all coastal metros near the Gulf of Mexico. The "hilliest" metros are all located adjacent

Figure 3: Distribution of Climate Variables Across US Metros



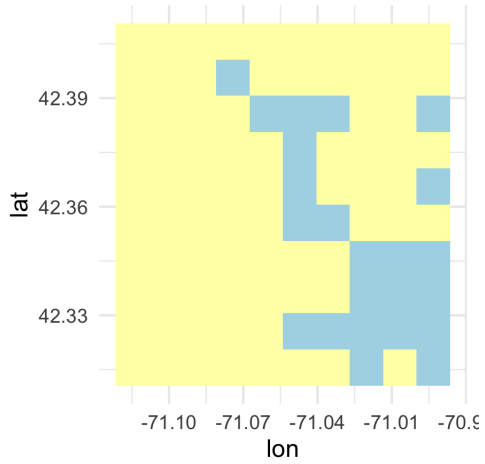
N = 382. Annual temperature variation is measured as the difference in average temperature between the metro’s warmest and coldest month.

to the Rocky Mountains.⁴ The measure appears to do a reasonable job in differentiating between cities built on hills and those built on flat land. While the measure does not capture the endogenous location decisions of workers it does capture endowed topography, which is the variation of interest for this study.

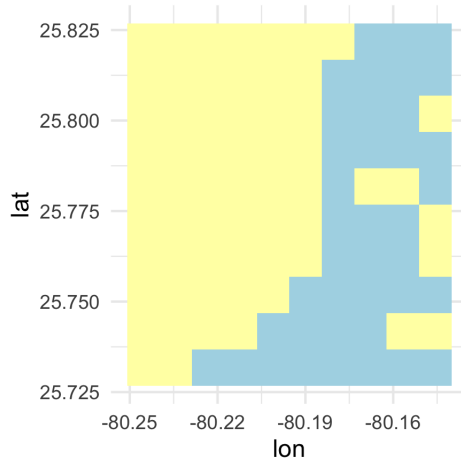
For the metropolitan level analysis I make use of the 2016 5-year ACS. The data source is a representative, 5% sample of the US population with responses collected

⁴The three metros with the highest value for the hilliness metric are Ogden, UT (255.9 m), Boulder, CO (208.3 m) and Provo, UT (205.2 m). The three metros with the lowest value are Brownsville, TX (1.0 m), Cape Coral, FL (0.9 m) and Houma, LA (0.7 m).

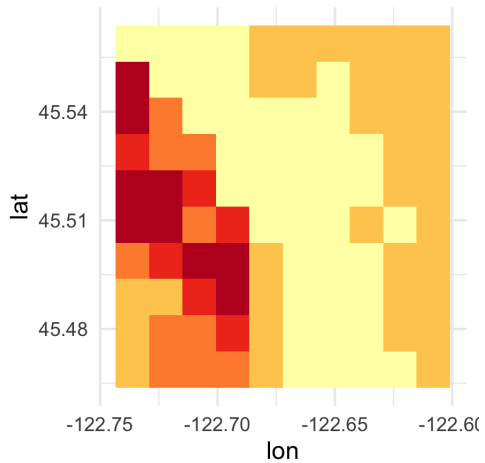
Figure 4: Calculating A Measure of Metro Level Hilliness



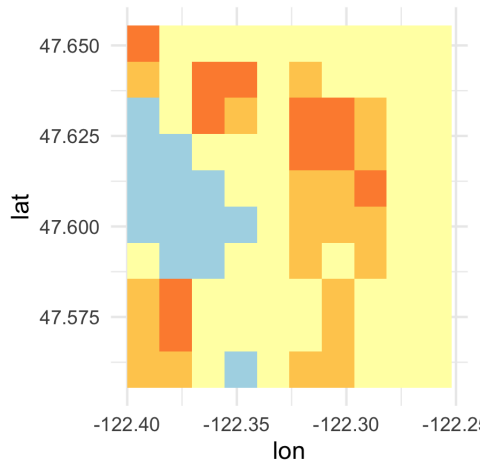
Boston, MA
Hilliness metric = 8.3 m



Miami, FL
Hilliness metric = 4.0 m



Portland, OR
Hilliness metric = 59.9 m



Seattle, WA
Hilliness metric = 36.8 m

Elevation



This figure provides four examples of generating the hilliness metric. I delineate a 10 km by 10 km grid, centered on each metro's principal city. I take the elevation measure from the center of each 1 km by 1 km grid and the metric is calculated as the standard deviation of the 100 points, excluding oceans. I perform the same calculation for each of the 382 metropolitan areas in the sample. The images above show discrete elevation categories, but continuous values are used in the calculation.

from 2012-2016. The survey asks workers their mode of transportation to work. For workers who use multiple modes, they are asked to report the mode that they used for the majority of the commute distance during their most recent week of work.⁵ The ACS also includes various demographic information that will be used in analysis.

In addition to metropolitan level analysis I analyze the behaviour of individual commuters. The CTPP commuting and workplace data provide detailed information on commuting habits across the US. The data includes linked data on the home and work locations of workers in the US, aggregated to census tract pairs. Importantly, the data breaks out commutes by mode, identifying the number of commuters who cycle along each route. The CTPP is nationally representative and relies on a 5% sample of the US workforce collected from 2012-2016. I drop all commuters that live and work in the same tract, as it is impossible to collect geographic trip characteristics for such observations. 13.8% of workers live and work in the same census tract. Cycling is no more common among within-tract commuters (0.59%) than among the overall workforce (0.62%), whereas walking to work is much more common among within-tract commuters (10.0% versus 1.7% for the overall workforce). I also drop all commutes where the home is more than 250 km from the workplace by road, as these outlier observations are unlikely to correspond to actual daily commutes. The final data set includes 3,126,282 unique home tract, work tract pairs.

For every commute route that appears in the CTPP, I calculate an array of route characteristics. I make use of the geocoding service HERE to recover route properties. HERE is a routing service that can return suggested travel routes for a variety of modes, including cycling. The Application Programming Interface (API) provided by HERE is intended to enable software applications that would provide navigation information for urban travelers, for example through in-vehicle GPS navigation systems and smartphone navigation apps. I repurpose this data feed to scrape route characteristics for all 3,126,282 routes in my data set. I use the geographic centroid of origin and destination census tracts to query routing instructions for all routes. Using census tract centroids should provide relatively accurate approximations of starting and ending points, though the assumption will mask the within tract variation. The use of census tract centroids for the individual level analysis, rather than using actual home and work addresses, therefore introduces a source of measurement error. HERE

⁵The precise question in the ACS is: “How did this person usually get to work LAST WEEK? If this person usually used more than one method of transportation during the trip, mark the box of the one used for most of the distance.”

returns step by step navigation instructions for cycling, including a profile of elevation points. I recover from the API the actual road distance required to complete the commute. For a measure of hilliness I take 10 evenly spaced points spanning the trip route and calculate the elevation at each point. HERE provides elevation information spanning the entire route, allowing for detailed estimation of elevation changes. I then sum the absolute value of elevation change between consecutive points to serve as a measure of route level hilliness. The measure approximates the amount of elevation that must be overcome by a commuter who completes a round trip commute along this route.

Collecting data through HERE represented a substantial data collection effort. An alternative approach would be to use the “straight-line,” or geodesic distance between home and work, ignoring the actual road network. I show in Appendix A that geodesic routes significantly underestimate the length of actual routes, though the two distance measurements are highly statistically correlated. Hilliness is poorly approximated by using the elevation profile along geodesic lines, likely because roads are oriented to avoid steep hills. The data collection method provides very rich information on route characteristics with which to estimate the partial effect of distance and hills.

For use in control variables I also make use of data on cyclist fatalities from the federal Fatality Analysis Reporting System (FARS) database. For data on the location of bikeshare systems I use the National Association of City Transportation Officials list of bikeshare systems. For the location of rail systems I use the American Public Transportation Association 2014 Public Transportation Fact Book.

3 Geography as a Predictor of Metropolitan Cycling Rates

Cycling can be unpleasant in rain, snow or extreme temperatures. Steep terrain may also act as a deterrent to cycling, as suggested by prior research. Both climate and hilliness differ substantially across US metros, suggesting some metros are more suitable for cycling and allowing for an opportunity to identify these effects empirically. In this section I test for a relationship between geographic endowments across US metropolitan areas and the prevalence of commuting by bicycle.

Figure 5A displays the correlation between bicycle commuting and mean annual temperature. The figure demonstrates there is very little correlative relationship between these two variables. While mean annual temperature in US metros ranges

widely from -2°C (Fairbanks, AK) to 24°C (Miami, FL), average temperature explains very little of the cross metro variation in cycling. In fact, Fairbanks has a higher rate of bicycle commuting (1.1%) than Miami (0.6%). Table 1, column 1 shows the results of a bivariate regression of the share of commuters who bike on the metro's average temperature. All coefficient estimates in Table 1 are inflated by a factor of 100 to improve the readability of the table. Places with higher average temperatures in the US have lower rates of cycling, on average. The result is likely due to omitted variable bias. Metros in the Southern United States have both lower rates of cycling and higher temperatures, on average. I also test for a nonlinear effect of temperature by regressing cycling rate on temperature and temperature squared (column 2). Together, the temperature variables explain only 3% of the variation in cycling.

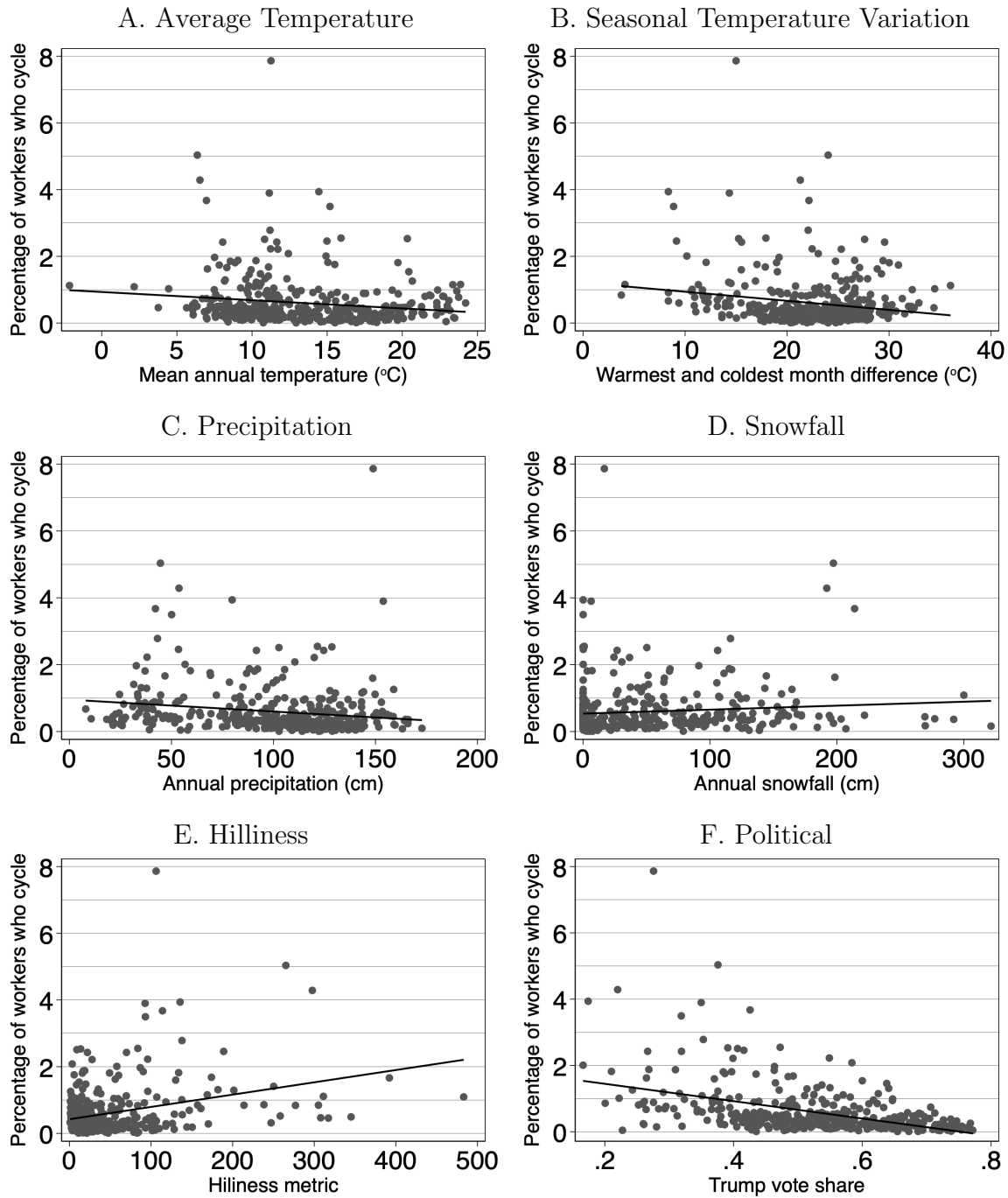
Table 1: Predicting the Share of Metro Workers Commuting by Bicycle

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|--|-------------------|-------------------|-------------------|-------------------|------------------|------------------|-------------------|
| Average temperature ($^{\circ}\text{C}$) | -.025** (.008) | -.100** (.042) | | | | | -.194** (.048) |
| Average temperature squared ($^{\circ}\text{C}$) | | .003* (.001) | | | | | .003* (.001) |
| Temperature variation ($^{\circ}\text{C}$) | | | -.027** (.009) | | | | -.087** (.018) |
| Precipitation (cm) | | | | -.004** (.001) | | | -.002 (.001) |
| Snowfall (cm) | | | | | .001 (.001) | | -.002 (.001) |
| Hilliness metric (m) | | | | | | .006** (.002) | .002 (.002) |
| Constant | .932** (.125) | 1.405** (.299) | 1.211** (.229) | .954** (.138) | .537** (.054) | .440** (.048) | 4.778** (.801) |
| State fixed effects? | N | N | N | N | N | N | N |
| N | 382 | 382 | 382 | 382 | 382 | 382 | 382 |
| R ² | .022 | .030 | .034 | .027 | .009 | .072 | .247 |

Significance levels: * : 5% ** : 1%. Robust standard errors in parenthesis. The dependent variable is the share of the metro's workers who cycle to work. The coefficient estimates have been multiplied by 100 so they correspond to percentage point changes. For example, a coefficient estimate of 0.25 implies a 0.25 percentage point increase in the probability of cycling.

In addition to average annual temperature, annual temperature variation may affect cycling rates. Locations with very moderate climates may provide reasonable cycling conditions year round while places with extreme swings in temperature may prevent cycling for some part of the year. Figure 5B displays the correlation between

Figure 5: Correlation Between Metro Characteristics and Bicycle Commuting Rates



Each dot corresponds to a metropolitan area. Lines are unweighted linear best fit lines.

cycling and the difference in temperature between the warmest and coldest month of the year. The figure indicates that areas with more extreme annual swings in temperature have lower rates of cycling. The estimate in Table 1, column 3 indicates that every 10°C of annual temperature variation correlates with a reduction in the bicycle commuting rate of 0.27 percentage points. Annual temperature differences explain 3% of the variation in cycling across metros.

Cycling is likely more difficult in rain and snow conditions. Figure 5C and D chart the correlation of bicycle commuting with annual precipitation and snowfall respectively. Metros with high annual precipitation show lower levels of cycling. Table 1 shows that this relationship is statistically significant. However, variation in annual rainfall only explains 3% of the variation in cycling across metros. Surprisingly, I find no significant correlative relationship between annual snowfall and cycling.

Metros with flatter topography may be better suited to cycling. I contrast metros with significant variation in elevation to those that are relatively flat, by using the hilliness metric described in the previous section. The correlation is graphed in Figure 5E. A bivariate regression of cycling rate against hilliness reveals that metros constructed on hilly terrain actually have higher rates of cycling, on average (Table 1, column 6). The effect is unlikely to be causal, and is impacted by metros in the West that have both significant hills and high rates of cycling.

Taken together in a regression, average temperature, average temperature squared, temperature variation, precipitation, snowfall and hilliness explain 25% of the variation in the percentage of a metro's workers commuting by bicycle (Table 1, column 7). Results of this multivariate regression suggest that metros with lower mean temperature, and lower annual temperature variation have higher rates of cycling. Precipitation, snowfall and hilliness do not appear as statistically significant in this regression specification. While the modest explanatory power of the model is informative regarding the upper limit of geography's role in determining cycling rates, the coefficient estimates are not reliable due to omitted variable bias.

To deal with omitted variable bias I first add an array of demographic control variables to the regression model (Table 2, column 1). The log of metro population, median income, share of the population with a college degree, gender share, share of the population under age 30, race and ethnicity shares and median home value are all included in addition to the geographic variables of interest. Additionally, I use voting

Table 2: Predicting the Share of Metro Workers Commuting by Bicycle

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Average temperature (°C) | -0.087** (.029) | -0.044 (.057) | -0.045 (.057) | -0.044 (.058) | -0.048 (.059) | -0.064 (.056) | .002 (.064) |
| Average temperature squared (°C) | .002* (.001) | .002 (.002) | .002 (.002) | .002 (.002) | .002 (.002) | .002 (.002) | .001 (.002) |
| Temperature variation (°C) | -.064** (.019) | -.009 (.016) | -.009 (.016) | -.009 (.016) | -.011 (.016) | -.012 (.014) | .004 (.016) |
| Precipitation (cm) | -.003 (.001) | .002 (.002) | .001 (.002) | .001 (.002) | .002 (.002) | .001 (.002) | .002 (.002) |
| Snow (cm) | -.001 (.001) | -.0003 (.001) | -.0003 (.001) | -.0002 (.001) | -.00003 (.0008) | -.001 (.001) | .0004 (.001) |
| Hilliness metric (m) | -.001 (.001) | -.0001 (.002) | -.001 (.003) | -.0002 (.001) | -.00003 (.002) | -.004** (.001) | -.001 (.002) |
| Hilliness metric (m) x West | | | | | | .006* (.003) | |
| Trump vote share 2016 (%) | -1.415** (.393) | -1.291** (.446) | -1.320** (.450) | -1.304** (.441) | -1.228** (.491) | -1.277** (.426) | -2.039** (.633) |
| Log of population | -.061 (.040) | -.077 (.039) | -.077* (.038) | -.077 (.039) | -.101 (.059) | -.072 (.039) | -.101 (.044) |
| Median income (\$1,000) | -.031** (.007) | -.033** (.007) | -.034** (.008) | -.034** (.008) | -.033** (.008) | -.035** (.008) | -.035** (.008) |
| College share (%) | 4.509** (1.252) | 4.676** (1.383) | 4.695** (1.413) | 4.676** (1.385) | 4.750** (1.420) | 4.869** (1.354) | 3.655** (1.105) |
| Male share (%) | 6.711* (3.246) | 3.896 (3.137) | 3.986 (3.325) | 3.936 (3.108) | 3.593 (3.199) | 4.492 (3.085) | 6.852* (3.365) |
| Under 30 share (%) | 2.552** (.689) | 2.534** (.721) | 2.459** (.829) | 2.526** (.705) | 2.635** (.707) | 2.255** (.754) | 1.579* (.771) |
| White share (%) | .160 (.690) | -.378 (.712) | -.379 (.715) | -.363 (.708) | -.428 (.723) | -.407 (.715) | -.029 (.854) |
| Black share (%) | -1.556 (.835) | -2.010* (.873) | -2.043* (.889) | -2.007* (.859) | -2.032* (.883) | -2.194* (.851) | -1.704 (1.004) |
| Asian share (%) | -2.528** (1.507) | -2.622* (1.267) | -2.631* (1.274) | -2.642* (1.282) | -2.995* (1.351) | -1.773 (1.339) | -1.077 (1.279) |
| Hispanic share (%) | -1.265** (.405) | -1.405** (.447) | -1.421** (.443) | -1.410** (.458) | -1.352** (.456) | -1.301** (.421) | -1.448** (.479) |
| Median home value (\$1,000) | .002** (.001) | .003** (.001) | .003** (.001) | .003** (.001) | .003** (.001) | .003** (.001) | .004** (.001) |
| Bike share system (dummy variable) | | | | | .019 (.101) | | |
| Cyclist fatalities (per bike commuters) | | | | | -.001 (.004) | | |
| Rail system (dummy variable) | | | | | .146 (.079) | | |
| State fixed effects? | N | Y | Y | Y | Y | Y | Y |
| Observation geography | CBSA | CBSA | CBSA | CBSA | CBSA | CBSA | County |
| Size of hilliness measure | 10 km | 10 km | 6 km | 14 km | 10 km | 10 km | 10 km |
| N | 382 | 382 | 382 | 382 | 382 | 382 | 382 |
| R ² | .582 | .694 | .695 | .694 | .696 | .701 | .705 |

Significance levels: * : 5% ** : 1%. Robust standard errors in parenthesis. The dependent variable is the share of the metro’s workers who cycle to work. The coefficient estimates have been multiplied by 100 so they correspond to percentage point changes. For example, a coefficient estimate of 0.25 implies a 0.25 percentage point increase in the probability of cycling.

results from the 2016 presidential election as a rough proxy for local political values.⁶

Once demographics are controlled for, precipitation, snowfall and hilliness all appear statistically insignificant, while measures of temperature remain significant. I find that metros with low median income, more college educated residents, more men, more young people, fewer non-white residents and higher median home values tend to have higher cycling mode shares. The share of the metro that voted for Donald Trump has a strong negative effect on cycling, suggesting a cultural determinant. The vote share variable alone explains 19% of the variation in cycling, far more than any of the geographic variables. Voting behaviour may be acting as a proxy for latent cultural variables. Pucher et al. (1999) provided an informative discussion of the cultural determinants of cycling in North America. A local culture that looks favorably on environmentalism may find cycling as relatively desirable due to its low environmental impact. This cultural predisposition may further popularize cycling, creating a virtuous cycle, increasing pressure for infrastructure and regulations that improve cycling. Local voting patterns will reflect the electorate's disposition towards environmental sustainability, providing support for policies and politicians who are seen as supportive of environmentalism. Therefore, an environmentally minded electorate is more likely to dwell in a metro with higher quality cycling infrastructure, have a cultural acceptance of cycling and have a higher cycling rate. An environmentally minded electorate is also less likely to support Donald Trump, whose 2016 presidential campaign was generally skeptical of environmentalist policies.

Despite the use of demographic control variables, there is potentially still residual omitted variable bias affecting the geographic endowment parameter estimates. For example, Western states have high rates of cycling and mountainous terrain, potentially creating a spurious correlation that suggests that hills are conducive to cycling. Table 2, column 2 adds state fixed effects to the regression equation. The inclusion of state fixed effects cause estimates to be based on the difference between metros that share the same state level variables but possess differing topography or climate. The large number of metros relative to states and the significant within-state geographic heterogeneity suggest this specification can focus more closely on the causal effects of climate and topography.

With state fixed effects included, none of the geographic variables predict cycling mode share with statistical significance. Furthermore, the estimates are relatively

⁶Voting data is taken from the MIT Election Data and Science Lab's County Presidential Election Returns 2000-2016. I collapse the county level data to CBSA boundaries.

precise, suggesting the null results are not a matter of statistical power, but that the true impact of these sources of geographic variation are close to zero. Interestingly, the inclusion of state fixed effects has very little impact on the estimated partial effects of demographic variables, many of which remain statistically significant and have large magnitudes. The direction of the partial effects of demographic variables are consistent with expectations and past research. The statistical insignificance of the geographic variables, particularly compared to the statistical importance of the demographic measures, provides strong evidence that geography is not an important determinant of cycling mode share across metros whereas demographics and local political orientation play a strong role.

For the measure of metropolitan hilliness I sample elevation points from a 10 km by 10 km grid. Table 2, columns 3 and 4 repeat the analysis by using 6 km and 14 km grids respectively to test the sensitivity of results to the sample area size. I find no significant differences in results across the various elevation sampling areas.

While potentially endogenous to cycling mode share, urban infrastructure is likely a strong determinant of cycling, as shown in prior literature. Column 5 controls for some proxies of local infrastructure, specifically the presence of a bike sharing system, the rate of cyclist fatalities per cycling commuter, and the presence of a public transit rail system. The presence of rail is meant to be a rough control for the availability of public transit, which may interact with cycling choice (Keijer and Rietveld, 2000; Singleton et al., 2014). The cycling fatality rate controls for some of the variation associated with safety. Without including controls, I find that a high cyclist fatality rate is correlated with lower bike commuting, while the presence of a bike sharing system or rail system is not statistically significantly correlated with the rate of bicycle commuting. Added to the regression model, the effects of geography are not substantially affected by the inclusion of these variables (Table 2, column 5).

Most metros with significant hills are located in the West region of the US. Table 2, column 6 adds an interaction term between the hilliness measure and a dummy variable for the metro being in the West to test for a potentially heterogeneous effect across regions. The results show that, while the effect of hills within the West is zero, the effect for the rest of the US is negative and statistically significant. Outside of the West region, a metro at the 25th percentile of hilliness is predicted to have a cycling rate .078 percentage points higher than a metro at the 75th percentile. The effect is still small when contrasted with the effect of demographic and political variation.

Finally, I perform a robustness check to test the sensitivity of results to the

choice to use CBSAs as the unit of analysis. CBSAs are large and include exurban areas that likely have very low rates of cycling. Potentially, overbounding the study areas in this way will hide meaningful variation that is present in the more urbanized areas. In Table 2, column 7 I rerun the the main regression analysis, but replace all variables that were at the CBSA level with variables derived from only the county of each CBSA that contains the CBSA's primary municipality. This restriction reduces the commuters under analysis by 40%, limiting analysis to more urbanized areas. Overall, results change very little. The partial effects of climate and hilliness remain statistically insignificant and close to zero, while demographic controls retain their significance.

Due to annually reported data I am not able to fully investigate seasonal heterogeneity in cycling rates. It seems very likely that, although climate does not appear to significantly affect cycling rates across metros, daily and seasonal fluctuations in weather could generate daily or seasonal variation in cycling (Miranda-Moreno and Nosal, 2011; Winters et al., 2007). To provide some evidence regarding weather conditions I make use of daily cycling trips taken on New York City's bikeshare system. I regress daily bike use against daily weather conditions. Full results are provided in Appendix B. I find that weather conditions are highly predictive of bikeshare use, with higher temperatures and less precipitation related to higher cycling rates. The effects are significantly less pronounced during rush hour, suggesting the decision to cycle among commuters is less sensitive to weather conditions than among recreational cyclists.

The important role of weather in causing fluctuation in daily cycling rates within a metro may explain the widespread belief that climate differences across metros will be deterministic in setting cycling rates. However, the low explanatory power of climate on cycling uptake across metros suggests that the choice to cycle by workers is influenced by weather only relative to average local conditions.

The result that climate and topographic variation across metros is generally unrelated to overall cycling mode share is potentially important to designing cycling policy. The finding suggests that policies that focus on shifting the underlying societal acceptance of cycling can be successful in shifting commuter mode share towards cycling, even in metros that appear to have geographic endowments that are not conducive to cycling.

4 Geography as a Predictor of Worker Cycling Probability Within Metros

Within metropolitan areas, bicycle commuters are not uniformly distributed but are concentrated within particular census tracts. Potentially, some tracts may be located in areas that provide cycling routes with advantageous geographic conditions, which may partially explain the spatial heterogeneity. This section will test how the geographic route characteristics faced by individual commuters affect the probability of cycling to work. Empirical identification will be based off of workers who live and work in neighbourhoods of similar characteristics but happen to face differing distance and topography between their home and work locations. The effect of route distance and hills will be tested.

The raw data indicates that cycling is unpopular for long commutes. For commutes under 2 km, 3.77% of workers choose to cycle, while for commutes greater than 10 km the figure is only 0.23% and for commutes over 20 km the figure is 0.15%.

Regressing mode choice on route characteristics may encounter omitted variable bias. Workers with particular traits may cluster in particular neighbourhoods, and those neighbourhoods could have certain geographic characteristics. As an example, consider a city where high income residents tend to live in neighbourhoods at higher elevations. If cycling is correlated with income, regressions may suggest strong relationships between cycling and elevation characteristics even if elevation changes are not causally impacting choices. I overcome this type of bias by including regression fixed effects for the worker's home census tract and work census tract. Fixed effects fully remove any heterogeneity in cycling commute likelihood that is correlated with the home neighbourhood or work neighbourhood of a commuter. Regression results therefore capture the role of route distance and hills on cycling mode choice for hypothetical workers that live in similar neighbourhoods and work in similar neighbourhoods.

The role of hills may be asymmetric in regards to whether the home or work is at a higher elevation. This may be because workers prefer not to exercise immediately before work. For example, arriving at work after perspiring might be unappealing. Consistent with perspiration being a concern, Abraham et al. (2002) found workers are more likely to commute by bicycle if there are showers at their workplace. Other causes could be considered. For example, workers may face a commitment problem in which they would choose to commit themselves to exercise in the future, but when

faced with the immediate task they would shirk responsibility (Strotz, 1955). Therefore, workers may be more likely to cycle if their work is located downhill from their home, so that the majority of the exercise can be deferred until the end of the work day.

To test for the effect of distance and hills I execute a linear regression model as represented in Equation 1, where B is a dummy variable that takes a value of one if the commuter cycles, D is the road distance between home and work, E is the route level measure of elevation change described in Section 2, δ is a dummy variable that takes a value of one if the worker’s home location is at a higher elevation than the work location, ϕ is a fixed effect for home tract and ψ is a fixed effect for work tract. i subscripts denote an individual, h denotes the tract where the worker lives and w denotes the tract where the worker is employed. I include a squared term for distance due to the potentially nonlinear relationship between distance and mode choice. Because my primary interest is overall explanatory power and estimating partial effects, rather than developing a predictive choice function, I elect to use OLS estimation, rather than a probit or logit model, in order to minimize estimation bias (Angrist, 2001). The regression is executed on CTPP commuting data that covers 5,438,420 survey respondents.

$$B_i = \beta_0 + \beta_1 D_i + \beta_2 D_i^2 + \beta_3 E_i + \beta_4 \delta_i + \phi_{h(i)} + \psi_{w(i)} + \varepsilon_i \quad (1)$$

The regression approach aims for a causal interpretation of regression coefficients as any spurious correlation between home and work location choice and mode choice are fully controlled for. Identification is based only on variation in route level characteristics. However, I am unable to control for the possibility that home and work location decisions are influenced directly by the hilliness of the connecting route. If those predisposed to cycling are more likely to make location choices that provide relatively flat commute routes, this would put a downward bias on the estimate of β_3 . Because there is no reason that the presence of hills would cause an increase in cycling, and the result I find is close to zero, the possible presence of this bias does not present a serious identification barrier, but rather suggests that hills are even less relevant to cycling choice than the coefficient estimate will suggest.

Results of the micro level analysis are displayed in Table 3. An important conclusion from the analysis is that the large majority of variation in the choice to cycle is unrelated to distance and hills. Table 3 gradually introduces fixed effects with

Table 3: Probability of a Worker Commuting by Bicycle, Worker Level Regressions

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|-------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-----------------|-------------------|-----------------|
| Commute distance (km) | -.049** (.0003) | -.045** (.0003) | -.042** (.0003) | -.031** (.0003) | -.032** (.0003) | | -.093** (.015) | |
| Commute distance squared (km) | .0003** (.000002) | .0003** (.000002) | .0002** (.000002) | .0002** (.000002) | .0002** (.000002) | | .001** (.0002) | |
| Hilliness metric (100 m) | .005** (.002) | -.059** (.002) | -.052** (.003) | -.032** (.003) | -.035** (.004) | -.007 (.004) | -.174 (.161) | -.183 (.178) |
| Downhill to work (dummy) | .049** (.007) | .022** (.007) | .025** (.007) | .122** (.008) | .020** (.012) | .030* (.012) | .216 (.161) | .252 (.165) |
| State fixed effects? | N | Y | N | N | N | N | N | N |
| County fixed effects? | N | N | Y | N | N | N | N | N |
| Home tract fixed effects? | N | N | N | Y | Y | Y | Y | Y |
| Work tract fixed effects? | N | N | N | N | Y | Y | Y | Y |
| Distance fixed effects? | N | N | N | N | N | Y | N | Y |
| Sample | Full | Full | Full | Full | Full | Full | Partial | Partial |
| R ² | .0048 | .0083 | .0148 | .0489 | .0633 | .0657 | .142 | .142 |
| N | 5,438,420 | 5,438,420 | 5,438,420 | 5,438,420 | 5,438,420 | 5,438,420 | 429,858 | 429,858 |

Significance levels: * : 5% ** : 1%. Robust standard errors in parenthesis. The dependent variable is a binary cycling choice variable. The coefficient estimates have been multiplied by 100 so they correspond to percentage point changes. For example, a coefficient estimate of 0.25 implies a 0.25 percentage point increase in the probability of cycling. The partial sample includes only routes that had at least 10 survey respondents.

column 5 showing results that correspond to Equation 1. Column 1 does not include any fixed effects, likely yielding coefficients that are subject to omitted variable bias. The R² value of column 1 indicates that less than one half of one percent of the variation in cycling mode choice across individuals can be explained by the combined impacts of commute distance, commute distance squared, the route level hilliness measure and a dummy variable for the home being uphill from work. In the overall choice to cycle, route level geographic characteristics have almost no ability to explain behaviour. Column 2 introduces fixed effects at the state level, column 3 tightens fixed effects to the county, column 4 further tightens fixed effects to the home tract level and column 5 includes both home tract and work tract fixed effects, corresponding to the specification represented by Equation 1. The differences in coefficient estimates between columns 2-5 are relatively modest, suggesting the state level fixed effects are able to pick up much of the omitted variable bias.

According to Table 3, column 5, distance between home and work appears to be a significant factor. Doubling a commute distance from 2 to 4 km, while holding hilliness, home tract and work tract conditions constant results in the probability of commuting by bicycle to fall by 0.062 percentage points. Note that the analysis omits workers who live and work within the same census tract, which effectively ignores the

role of very short commutes. The extent of hills along the commute route is also statistically significant. I find that every 100 m of elevation change along the route reduces the probability of cycling by 0.035 percentage points. For context, the average commute has an elevation change of 114 m. Consistent with the proposed motivations for the asymmetric role of hills, I also find that cycling is more common when the home is located uphill, rather than downhill from work, though the effect is small. For two commutes that have identical distance and elevation conditions and identical home and work neighbourhood characteristics, the probability of cycling to work is 0.020 percentage points higher if the home is uphill from work, rather than downhill.

Equation 1 models the effect of distance on commute choice as a quadratic function. Equation 2 proposes an alternative functional form where the effect of distance is allowed to affect mode choice nonparametrically. Equation 2 introduces a separate dummy variable for every possible distance, at the granularity of a kilometer. For example, a dummy variable is included for commutes between 0 and 1 km, a separate dummy is included for commutes between 1 and 2 km, continuing as such to cover all commutes in the data set. The elements of Equation 2 are identical to Equation 1 except for the addition of χ_i , which is a vector of distance dummy variables. This method tightly controls for the effect of distance, though the method prevents the estimation of the partial effect of distance.

$$B_i = \beta_0 + \beta_1 E_i + \beta_3 \delta_i + \chi_i + \phi_{h(i)} + \psi_{w(i)} + \varepsilon_i \quad (2)$$

The result of Equation 2 is displayed in Table 3, column 6. When the distance fixed effects are included the hilliness variable is no longer statistically significant. The measure of hilliness is correlated with distance, meaning the original negative effect found from Equation 1 may be because the hilliness metric is carrying some of the distance effect that was not fully controlled for. With fully controlling for distance, I find that hills at the route level appear to be even less relevant. The effect of the home being uphill from work remains statistically significant.

It is interesting to note the relatively small magnitude of the effect of hills. For example, consider two commuters who live and work in identical neighbourhoods and face commutes of identical distance. The first commuter's route is perfectly flat, while the second commuter's workplace is at 100 m higher elevation relative to their home. According to Equation 2 point estimates, if the first commuter's likelihood of commuting were 0.60%, which is the sample mean of US workers, the probability the

second commuter would choose to cycle would be 0.59%. The presence of substantial hills appears to have a very small effect on the overall decision of whether to cycle.

One limitation of using route level data is the considerable sampling error that may be introduced due to the small number of respondents corresponding to a given route. As a robustness check, I rerun Equations 1 and 2 on the subsample of routes where there were at least ten survey respondents. The restriction lowers the number of survey observations in the analysis from 5,438,420 to 429,858. Results are displayed in Table 3, columns 7 and 8. I find that hills have a statistically insignificant effect on the cycling rate. The effect of the workplace being downhill from the home is positive, consistent with the full sample analysis, but falls short of statistical significance. The significant effect of distance (column 7) increases substantially relative to the full sample estimate, though the mean cycling probability also increases substantially, from 0.68% to 1.64%, meaning that in percent terms the partial effects are comparable.

If route characteristics are uncorrelated with idiosyncratic worker mode preferences, the above estimated coefficients can be interpreted as the causal effect of geographic route characteristics on mode choice. To the extent that workers select their home and work neighbourhoods jointly, in order to facilitate their preferred commute mode, OLS estimates overstate the causal effects, as cyclists' route choice would be positively correlated with short, flat routes. However, the OLS results indicate that cyclists do not locate along routes that are particularly flat. To the extent that the estimates are inflated by endogenous route selection, the true causal effect of hills is even smaller and less relevant than the above results indicate.

When planning for cycling routes and infrastructure, planners may tend to ignore areas of steep topography, assuming hills would hinder cycling uptake. Results from this section suggest that hills are not a significant deterrent to potential cyclists. However, the distance from home to work is important, suggesting that urban design that allows for shorter distance commutes in general are compatible with encouraging cycling.

5 Conclusion

Prior research has argued that climate and topography play roles in the attractiveness of cycling as a commuting option. I leverage large data sets of commuting flows, cycling navigation route data, climate data and elevation data to

explain spatial heterogeneity in cycling uptake with geographic variables. Analysis demonstrates that geographic endowments are not an important determinant of cycling heterogeneity across metros. Once cultural, demographic and cross state variations are controlled for, the role of both climate and topography is close to zero. I do find that local demographics are important. Low median income, high rates of college education, high shares of young people, low black and Hispanic population shares and high property values all significantly relate to higher cycling uptake. I also find presidential voting data to be highly predictive of cycling rates, suggesting a strong role of local culture in determining cycling.

Within metros, cycling is found to be more common for short commutes, with cycling very rare for commutes with one way distances above 10 km. The magnitude of hills along commute routes are found to be essentially irrelevant to cycling mode choice. Cycling is found to be slightly more common when the workplace is located downhill from the home. Together, the distance and hilliness of a commute explain less than half of one percent of the variation in cycling mode choice.

The role of geography in cycling uptake is frequently discussed in relation to the construction of bicycle infrastructure such as bike lanes. Opponents of bicycle infrastructure often point to hills or unsuitable weather as evidence that cycling can not be locally popular. The findings in this study have a potentially important lesson for policy: climatic and topographical endowments are unimportant to the general uptake of cycling. The exogenous cause of spatial heterogeneity in cycling appears to be related to local demographic and cultural idiosyncrasies.

This study is limited by looking only at commuters. The spatial causes of cycling for non-work trips and for recreation may be different. Possibly, local geographic conditions could play a different role in the decision to participate in these types of cycling.

Smiley et al. (2016) provides an interesting case study of cycling uptake in central Memphis, Tennessee. The study catalogs the parallel increases in cycling and cycling infrastructure, demonstrating the apparent malleability of Memphis's cultural and political disposition towards cycling. Heinen et al. (2013) provides some evidence from the Netherlands that firms with positive cycling cultures generate higher rates of employee cycling. In the UK, Dickinson et al. (2003) argues that cycling uptake will remain limited unless the culture around cycling is made to be more inclusive, particularly towards female cyclists. The role of underlying social and cultural variables, rather than geography, appear to be the major determinate in spatial

heterogeneity in cycling uptake and should be studied more closely if policy is to increase the share of workers who cycle.

References

- Abraham, J. E., McMillan, S., Brownlee, A. T., and Hunt, J. D. (2002). Investigation of cycling sensitivities. In *81st Annual Meeting of the Transportation Research Board, Washington, DC*.
- Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *Journal of Business & Economic Statistics*, 19(1):2–28.
- Beenackers, M. A., Foster, S., Kamphuis, C. B., Titze, S., Divitini, M., Knuiman, M., van Lenthe, F. J., and Giles-Corti, B. (2012). Taking up cycling after residential relocation: built environment factors. *American Journal of Preventive Medicine*, 42(6):610–615.
- Buehler, R. and Pucher, J. (2012). Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*, 39(2):409–432.
- Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American Housing Survey. *Transportation Research Part A: Policy and Practice*, 30(5):361–377.
- Cervero, R., Sarmiento, O. L., Jacoby, E., Gomez, L. F., and Neiman, A. (2009). Influences of built environments on walking and cycling: lessons from Bogotá. *International Journal of Sustainable Transportation*, 3(4):203–226.
- De Hartog, J. J., Boogaard, H., Nijland, H., and Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, 118(8):1109–1116.
- Dickinson, J. E., Kingham, S., Copsey, S., and Hougie, D. J. P. (2003). Employer travel plans, cycling and gender: will travel plan measures improve the outlook for cycling to work in the uk? *Transportation research part D: transport and environment*, 8(1):53–67.
- Dill, J. and Carr, T. (2003). Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. *Transportation Research Record*, 1828(1):116–123.
- Heinen, E., Maat, K., and van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation*, 40(1):23–43.
- Heinen, E., Van Wee, B., and Maat, K. (2010). Commuting by bicycle: an overview of the literature. *Transport reviews*, 30(1):59–96.
- Hunt, J. D. and Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, 34(4):453–470.

- Keijer, M.J.N. and Rietveld, P. (2000). How do people get to the railway station? The Dutch experience. *Transportation planning and technology*, 23(3):215–235.
- Mattingly, K. and Morrissey, J. (2014). Housing and transport expenditure: Socio-spatial indicators of affordability in Auckland. *Cities*, 38:69–83.
- Miranda-Moreno, L. F. and Nosal, T. (2011). Weather or not to cycle: Temporal trends and impact of weather on cycling in an urban environment. *Transportation research record*, 2247(1):42–52.
- Parkin, J., Wardman, M., and 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., Komanoff, C., and Schimek, P. (1999). Bicycling renaissance in North America?: Recent trends and alternative policies to promote bicycling. *Transportation Research Part A: Policy and Practice*, 33:625–654.
- Reynolds, C.C.O., Harris, M.A., Teschke, K., Cripton, P.A, and Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental health*, 8(1):1–19.
- Rietveld, P. and Daniel, V. (2004). Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A: Policy and Practice*, 38(7):531–550.
- Rodriguez, D. A. and 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.
- Saelens, B. E., Sallis, J. F., and Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, 25(2):80–91.
- Singleton, P. A., and Clifton, K. J. (2014). Exploring synergy in bicycle and transit use: Empirical evidence at two scales. *Transportation research record*, 2417(1):92–102.
- Smiley, K. T., Rushing, W., and Scott, M. (2016). Behind a bicycling boom: Governance, cultural change and place character in Memphis, Tennessee. *Urban Studies*, 53(1):193–209.
- Stinson, M. and Bhat, C. (2004). Frequency of bicycle commuting: internet-based survey analysis. *Transportation Research Record: Journal of the Transportation Research Board*, (1878):122–130.
- Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180.

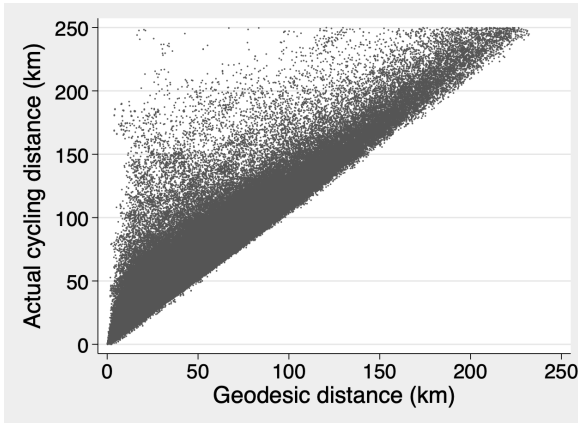
- Vandenbulcke, G., Dujardin, C., Thomas, I., de Geus, B., Degraeuwe, B., Meeusen, R., and Panis, L. I. (2011). Cycle commuting in Belgium: Spatial determinants and ‘re-cycling’ strategies. *Transportation Research Part A: Policy and Practice*, 45(2):118–137.
- Walsh, C., Jakeman, P., Moles, R., and O’Regan, B. (2008). A comparison of carbon dioxide emissions associated with motorised transport modes and cycling in Ireland. *Transportation Research Part D: Transport and Environment*, 13(6):392–399.
- Winters, M., Friesen, M. C., Koehoorn, M., and Teschke, K. (2007). Utilitarian bicycling: a multilevel analysis of climate and personal influences. *American Journal of Preventive Medicine*, 32(1):52–58.
- Xing, Y., Handy, S. L., and Mokhtarian, P. L. (2010). Factors associated with proportions and miles of bicycling for transportation and recreation in six small US cities. *Transportation Research Part D: Policy and Practice*, 15(2):73–81.
- Zhao, P. (2014). The impact of the built environment on bicycle commuting: Evidence from Beijing. *Urban Studies*, 51(5):1019–1037.

Appendix A

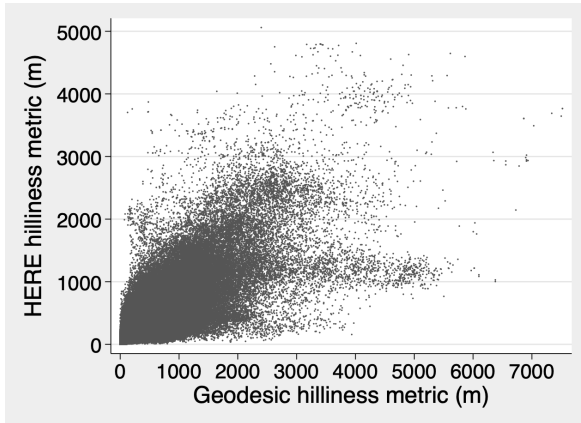
This study makes use of data on cycling routes which follow roads and paths as suggested by the HERE routing API. Collection of HERE data required an extensive data collection effort. A simpler alternative would be to use straight line distances between origin and destination points and assume that these straight lines would provide reasonable approximations of actual routes executed by bicycle commuters. In this appendix I show how the two alternative data collection approaches compare. Figure A1.i shows the correlation between the geodesic distance and actual travel distance for all 3,126,282 commute routes in the data. The overall correlation is 0.97. The result suggests that straight line distance captures a very high share of the meaningful statistical variation of actual cycling routes. The average cycling route distance collected from the HERE API is 26% longer than the average geodesic distance, representing the extent to which the actual road network adds circuitry to cycling routes.

Figure A1. Route Level Data, Comparing HERE and Geodesic Approaches

i. Distance of Commute



ii. Hilliness Metric



The figures display the correlation in route characteristics between a data collection approach that approximates routes using geodesic lines and an approach that uses the actual road routes suggested by the HERE API. $N = 3,126,282$.

Figure A1.ii shows the correlation between the hilliness metric using the geodesic routes versus the actual route data collected from the HERE API. As described in Section 2, the metric is the sum of the absolute value of elevation change between 10 evenly spaced points along the route. The correlation is 0.82. The geodesic approach

captures most of the actual variation in hilliness, though less than for distance. The hilliness metric is estimated to be 16% larger when using the geodesic approach as compared to the HERE API. Road networks avoid steep hills, causing the geodesic approach to overstate the extent of hills experienced by commuters.

Appendix B

I obtained data from the Citi Bike bike sharing system in New York City to test the role of daily weather on daily cycling activity. The data is made publicly available and covers every trip in the system from the start of 2014 until the end of 2019. For weather data, I use daily weather station data conditions as reported from the John F. Kennedy International Airport.

I regress the number of trips in a day against the mean temperature for that day, the mean temperature squared, the amount of rainfall and the amount of snowfall. I limit the sample to weekdays. Table B1 displays results. Column 1 includes all bike share trips and indicates that weather conditions are highly predictive of system usage. The temperature parameters indicate that hotter temperatures relate to higher system usage, with a diminishing marginal effect. Precipitation and snowfall have strong negative effects on the number of trips. Overall the weather variables predict 45% of the daily variation in trips.

Table B1: Impact of Daily Weather on Bike Share Ridership in New York City

| | All Trips (1) | Rush Hour Trips (2) |
|----------------------------------|--------------------------|-------------------------|
| Average temperature (°C) | 1946.044** (88.376) | 335.987** (17.013) |
| Average temperature squared (°C) | -26.279** (3.762) | -5.985** (.725) |
| Precipitation (cm) | -5573.422** (516.201) | -859.388** (97.599) |
| Snowfall (cm) | -635.179** (123.452) | -134.801** (37.452) |
| Constant | 25351.280** (524.475) | 4884.815** (109.125) |
| N | 1,558 | 1,558 |
| R ² | .447 | .323 |

Significance levels: * : 5% ** : 1%. Robust standard errors in parenthesis. The dependent variable is the number of bike share trips completed in the day.

Table B1, Column 2 alters the dependent variable to include only rush-hour trips.⁷ While the partial effects of temperature and precipitation remain statistically significant, the explanatory power of the model falls significantly, from .45 to .32.

⁷I define a rush hour trip to be one that begins between 8 am and 10 am or between 4 pm and 6 pm.

This suggests that bike share trips among commuters are significantly less dependent on daily weather conditions than non-commuting trips.