DTU Library

Our children cycle less - A Danish pseudo-panel analysis

Rich, Jeppe; Myhrmann, Marcus Skyum; Mabit, Stefan Eriksen

Published in:
Journal of Transport Geography

Link to article, DOI:
10.1016/j.jtrangeo.2022.103519

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Journal of Transport Geography 106 (2023) 103519

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Our children cycle less - A Danish pseudo-panel analysis

Jeppe Rich [*,1], Marcus Skyum Myhrmann, Stefan Eriksen Mabit

Transport Division, DTU Management, Technical University of Denmark, Bygningstorvet Building 116B, Kgs. Lyngby, 2800, Denmark

ARTICLE INFO

Keywords:
Cycling
Pseudo-panel analysis
Hurdle and selection models
Dynamic transport preferences
Transport geography
Econometrics

ABSTRACT

In this paper, we investigate the dynamics of cycling demand for the population of Denmark. Using pseudo-panels based on large-scale cross-sectional data, we analyse the temporal stability of cycling demand preferences for different age cohorts in combination with residential city sizes. Cycling demand is decomposed into two effects. Firstly, a population 'selection' effect that explains the probability of being a cyclist, i.e. engaged in cycling activities. Secondly, the conditional demand for cycle mileage provided that the respondent is a cyclist. The joint probability model is estimated as a Gamma Hurdle model. The study reveals several empirical findings, of which three stand out. Firstly, overall cycling demand in Denmark over the period is in decline. Secondly, it is shown that this is mainly a selection effect. Hence, the main driver of the observed decline is essentially a shrinking cycling population rather than a decrease in trip distances for those who travel by bicycle. Thirdly, the decline is strongest for younger generations, particularly those residing outside the larger cities. With Denmark being an international forerunner for bicycling and with a cycling culture developed over many decades, we believe these findings can be relevant to mitigate similar long-term changes in other countries.

1. Introduction

Internationally, Denmark is a successfully bicycle nation (the second largest cycling population in the EU relative to its population ECF, 2017; Harms et al., 2014) with a trip share between 14–18% seen over the period from 2006–2019, see Table 1. However, despite continued efforts to increase the cycling trip share, the demand for cycling in Denmark is declining (Nielsen et al., 2016). With Denmark often mentioned as one of the historical forerunners in promoting bicycling, it is of international interest to understand why this decline is happening, for whom and to what extent? It is interesting because it may allow us to foresee trends in countries with a less developed cycling culture and potentially lead to preventive measures for mitigating such development.

The aim of the present paper is the detection of cycle demand patterns across age cohorts and a discussion of possible underlying reasons for the observed trends. As part of the discussion, we also discuss future research agendas that relate to possible policies.

In the paper, we analyse cycling preferences by applying a pseudo-panel approach based on a large-scale national travel diary. In doing so, it is possible to throw light on whether individuals from a specific age group observed, e.g. in 2006, behave differently from a corresponding age group observed later. Methodologically, the paper decomposes bicycle demand into two main components. Firstly, a selection component measuring the proportion of the population that cycles, and secondly, a mileage component measuring the mileage for each cyclist. These effects are estimated and measured in a Hurdle-type model across pseudo-panels defined according to age groups and residential city size. This approach allows us to investigate the relative importance of the two effects and to examine the impact of various explanatory variables for the selection and the mileage model.

As a research topic, bicycle demand has recently received increasing attention in the scholarly literature. In part, because there is an increasing awareness of the positive external effects related to bicycling (Rich et al., 2021; Ekelund et al., 2015), especially its impact on well-being (Mytton et al., 2016; Martin et al., 2014; Humphreys et al., 2013; Synek and Koenigstorfer, 2019) and health (Mueller et al., 2015; Jarrett et al., 2012; Kyu et al., 2016; Raza et al., 2020; Rezende et al., 2018; Sommar et al., 2021). In part, because it is an obvious solution to mitigate congestion in the increasingly denser and larger urban areas worldwide. While micro-econometric models exist for cycling demand (Hallberg et al., 2021; Rayaprolu et al., 2018) to measure mode and destination substitution effects, there is a lack of attention to the development of cycling demand over time. An exception, however, is Nielsen et al. (2016) who study the drivers of cycling mode share for

* Corresponding author.
E-mail address: rich@dtu.dk (J. Rich).
[1] Current address: Technical University of Denmark, Bygningstorvet Building 116B, Kgs. Lyngby, 2800, Denmark.

https://doi.org/10.1016/j.jtrangeo.2022.103519
Received 18 June 2022; Received in revised form 16 September 2022; Accepted 19 December 2022
Available online 28 December 2022

Table 1
Time trends for number of trips per person for selected transport mode (population between 10–84 years) with estimated level of confidence.

Year	Walk	Bicycle	Car	Public
2006	16.5% ± 0.8%	*18.0% ± 0.9%*	57.2% ± 1.0%	8.2% ± 0.6%
2007	14.5% ± 0.6%	*17.4% ± 0.7%*	59.7% ± 0.8%	8.3% ± 0.4%
2008	14.8% ± 0.7%	*18.2% ± 0.6%*	58.7% ± 0.9%	8.3% ± 0.5%
2009	16.6% ± 0.6%	*17.6% ± 0.6%*	58.0% ± 0.7%	7.7% ± 0.4%
2010	16.6% ± 0.6%	*16.6% ± 0.5%*	58.5% ± 0.6%	8.2% ± 0.3%
2011	16.5% ± 0.6%	*17.2% ± 0.6%*	57.7% ± 0.8%	8.5% ± 0.4%
2012	16.6% ± 0.8%	*17.0% ± 0.7%*	58.5% ± 0.9%	7.9% ± 0.5%
2013	15.4% ± 1.1%	*17.7% ± 0.8%*	58.6% ± 1.1%	8.3% ± 0.6%
2014	16.9% ± 1.0%	*18.1% ± 0.8%*	57.3% ± 1.1%	7.6% ± 0.5%
2015	18.1% ± 0.9%	*16.2% ± 0.8%*	57.9% ± 1.1%	7.7% ± 0.5%
2016	19.9% ± 0.9%	*15.8% ± 0.9%*	56.8% ± 1.0%	7.5% ± 0.5%
2017	20.8% ± 0.9%	*15.8% ± 0.8%*	55.9% ± 1.0%	7.4% ± 0.5%
2018	19.5% ± 1.0%	*14.8% ± 0.8%*	58.7% ± 1.1%	6.9% ± 0.4%
2019	18.7% ± 0.9%	*14.4% ± 0.7%*	59.2% ± 1.0%	7.7% ± 0.5%
2020	26.7% ± 0.8%	*14.4% ± 0.7%*	54.1% ± 0.9%	4.7% ± 0.3%
2021	31.3% ± 1.0%	*12.0% ± 0.5%*	52.2% ± 1.0%	4.5% ± 0.4%

Danes during the period 1996–2013 and point to increasing population density and change of urban development as relevant drivers.

The lack of attention to the dynamic development of bicycle demand is in stark contrast to other areas within the transport research domain. One example is the dynamics of car ownership propensity, which over the last decade has received increasing interest. This include traditional panel studies, e.g. Nolan (2010), age cohort studies, e.g. Newbold et al. (2005); Hjorthol et al. (2010) and Haustein and Siren (2012), and pseudo-panel studies based on cross-sectional data, e.g. Dargay and Vythoulkas (1999); Krueger et al. (2020), and Papu Carrone and Rich (2020). A slightly different approach is presented in Papu Carrone et al. (2021), which applies panel models to cross-section data (also TU data as applied in this paper) by taking advantage of respondents being re-sampled by chance. In this case, the panel is irregularly spaced, which gives rise to interesting dynamics regarding how life-course events may influence, e.g. car ownership. Unfortunately, the panel part of the data set is limited in size and does not allow a cohort analysis as presented here.

Hence, the main reason the dynamic aspects of cycling have been largely unexplored in the research literature is not due to the lack of proper methods. Rather, it is likely related to a lack of appropriate data. It is rarely the case that longitudinal data for bicycle demand exist coherently, particularly not as panel data where the same individuals are tracked over time. The problem with monitoring bicycle use, as opposed to car use, is that bicycle demand has more daily variation and significant fluctuations across the year. Hence, a bicycle demand analysis would require a long tracking period of many days, complicating the data collection process. Also, the baseline probability for engaging in cycling activities is low for some segments, suggesting that a relatively large panel would be required to investigate the dynamics of cycling demand in a traditional panel model. However, without panel data, there is the possibility of applying repeated cross-sectional data to understand preference changes for agents belonging to pseudo-panels (Collet, 2012). Although attention should be directed to how cohorts are formed to minimise aggregation bias (Lewbel, 1994; Granger, 1988) and enable identification (Devereux, 2007), it does provide a way of tracking if preferences for different age cohorts change over time. The paper contributes to the literature in the following ways;

- The use of pseudo-panel methods to explore the demand for cycling across age and city size cohorts,
- Decomposition of behavioural changes into a selection effect and conditional mileage effect,
- Evidence that the decline in cycling demand is mostly a selection effect, which is strongest for rural areas and younger generations.

The paper is organised as follows. In Section 2, we describe the data foundation of the study. In Section 3, we consider methodology, after which, in Section 4, we present the empirical modelling and the results. Finally, we offer a combined discussion and conclusion in Section 5.

2. Pseudo panel data

In Denmark, The Danish National Travel Survey (TU) (Christiansen, 2020) has been conducted since 2006 in its current format. The survey collects information concerning the daily travel habits of approximately 11,000–14,000 Danish respondents per year. It constitutes a fairly representative cross-sectional sample of the Danish population between the ages of 10 and 84, where travel habits are revealed from a travel diary for the day before the interview. The survey participants are selected through stratified random sampling from the Danish Civil Registration System to form a representative sample of the population. Only a single respondent from each household participates.

Annual samples of the survey from 1992 to 2001 are also available. These earlier versions of the data are suitable for aggregated analysis but are not compatible with later data sets. As a result, the years predating 2006 are excluded from this study. The years 2020–2021 are also excluded from the analysis to avoid the effects of the Covid-19 lockdown. As can easily be verified from Tables 1 and 2 the years after 2019 are indeed quite different with respect to the overall mode share and mileage per person, which suggest that these years should be excluded from the analysis. The final data set for the study period (2006–2019) includes a total of 175, 683 individuals. In the present study, we consider only a limited number of variables, namely the number of cars available to the household of the respondent, age of the respondent, gender of the respondent, total cycled kilometres on the day of the survey, and the urbanisation level of the residential location of respondents. Sample statistics for the final sample are shown in Table 3.

2.1. Definition and illustration of age cohorts

We consider age-specific cohorts for the categories: 10–20 years, 21–30 years, 31–65 years, and 66 + years and older. The first cohort primarily includes respondents who are economically dependent on their parents and have a high likelihood of being in either school or an early stage of higher education. The second cohort reflects students and early labour market participants, who are likely to be less economically dependent on their parents. The third cohort primarily represents the active working population in Denmark, while the last cohort mainly consists of retired people or people close to retirement.

To track cycle demand across levels of residential urbanisation, the cohorts are further divided into city-size categories linked to the resi-

Table 2
Time trends for mileage per person for selected transport mode (population between 10–84 years) with estimated level of confidence.

Year	Walk	Bicycle	Car	Public
2006	1.8% ± 0.1%	*3.9% ± 0.3%*	82.4% ± 1.4%	11.2% ± 1.7%
2007	1.7% ± 0.1%	*3.7% ± 0.2%*	83.9% ± 1.0%	10.0% ± 1.2%
2008	1.7% ± 0.1%	*4.0% ± 0.2%*	81.7% ± 1.0%	11.9% ± 1.2%
2009	2.1% ± 0.1%	*4.3% ± 0.3%*	83.3% ± 1.0%	9.8% ± 1.2%
2010	2.1% ± 0.1%	*3.9% ± 0.2%*	81.7% ± 1.1%	11.3% ± 1.2%
2011	2.0% ± 0.1%	*3.8% ± 0.2%*	80.6% ± 1.3%	12.5% ± 1.3%
2012	2.1% ± 0.1%	*3.8% ± 0.3%*	82.7% ± 1.4%	10.5% ± 1.6%
2013	2.0% ± 0.1%	*4.2% ± 0.3%*	82.3% ± 1.3%	10.9% ± 1.7%
2014	2.1% ± 0.1%	*4.5% ± 0.3%*	82.8% ± 1.3%	10.0% ± 1.4%
2015	2.2% ± 0.1%	*4.0% ± 0.3%*	82.6% ± 1.4%	10.2% ± 1.6%
2016	2.3% ± 0.1%	*3.9% ± 0.3%*	82.7% ± 1.3%	10.5% ± 1.6%
2017	2.4% ± 0.2%	*4.0% ± 0.3%*	82.9% ± 1.5%	10.1% ± 1.8%
2018	2.1% ± 0.1%	*3.5% ± 0.3%*	84.2% ± 1.2%	9.4% ± 1.6%
2019	2.1% ± 0.1%	*3.3% ± 0.3%*	84.1% ± 1.4%	9.6% ± 1.6%
2020	3.3% ± 0.2%	*4.1% ± 0.3%*	85.5% ± 1.0%	6.5% ± 1.1%
2021	4.2% ± 0.2%	*3.5% ± 0.3%*	85.5% ± 1.2%	6.4% ± 1.3%

Table 3
Descriptive statistics for the TU data for variables conditional on age groups.

Variables	Age classes			
	Age: 10–20	Age: 21–30	Age: 31–65	Age: 66+
Number of respondents	31507	18808	95598	30734
Female Gender share	48.8%	49.9%	51.5%	53.7%
Avg. No. household cars	1.3	0.8	1.2	0.9
CitySize: <200	17.3%	9.8%	16.4%	12.8%
CitySize: 200–25,000	15.2%	15.2%	14.7%	16.1%
CitySize: 25–100,000	43.4%	25.4%	41.7%	46.5%
CitySize: 100–500,000	7.8%	18.0%	7.7%	7.6%
CitySize: Copenhagen	16.4%	31.7%	19.5%	17.0%
Avg. cycling Mileage (km)	1.73	2.10	1.46	0.79
Percentage cycling (%)	31.2	24.7	16.3	10.7
Avg. conditional Cycling mileage (km)	5.55	8.51	8.95	7.38

dential location of the household. That is, cities of size $1{:}< 200$; $2{:}200-25,000$; $3{:}25,000-100,000$; $4{:}100,000-500,000$ and $5{:}> 500,000$ citizens. Category 4 represents Denmark's 2nd, 3rd and 4th largest cities, while Category 5 represents the capital region, including Copenhagen, with more than 1.5 million inhabitants. While, in principle, very detailed cohorts could be formed, it would reduce the number of observations within these cohorts and could cause identification problems (Devereux, 2007). Hence, in pseudo-panel studies, there is a balance regarding the size of the cohorts, aggregation bias, and the number of explanatory variables that can be identified as part of a model applied to the age cohorts. Concerning explanatory variables, we consider only a limited set: the gender, the exact age, and the number of cars in the household. The number of cars per household is included because it has been evidenced in previous Danish studies that car ownership propensity differs across age cohorts and is known to be correlated with the choice of mode (Papu Carrone and Rich, 2020). Moreover, it is a variable strongly correlated with income and, in many ways, represents a better and more useful proxy for the effect of income on travel behaviour. To this end, it is worth noting that the income measurement is often problematic in a questionnaire context, as illustrated by income missing for close to 30% of the households in the TU data set.

3. Methodology

As a first analysis, we estimate the probability of cycling and the corresponding average mileage conditional on cycling as weighted averages for the various age cohorts. A rolling three-year time window is used to ensure enough observations per cohort. Hence, we evaluate the cycling behaviour for three years combined with $t = \{2006, 2007, 2008\}$, $\{2007, 2008, 2009\}, \ldots, \{2017, 2018, 2019\}$. This implies a smoothing of the tendencies in the data similar to other studies, e.g. Rich and Vandet (2019) that analyses value-of-time dynamics.

To estimate the variation for the average estimates, we use a 10% leave out re-sampling for the data in the respective cohorts. This approach corresponds to a 10-fold cross-validation (Jiang and Wang, 2017) and was chosen to reduce computation time in this context where many observations are included in the analysis. However, it will slightly overestimate the variance for the mean and thereby represents an upper-bound estimate. For each sample, the process is as follows. Firstly, we compute the mean cycling share and average mileage many times in an iterative process. After this, inference concerning the variance for the mean is calculated across these samples. This process is repeated for all time windows.

From a statistical modelling perspective, the choice of 'selection' and 'mileage' can be modelled in a combined setting. Estimating these effects across age cohorts makes it possible to examine parameter stability over time. It also makes it possible to integrate explanatory variables into the model to identify the 'drivers' of these two processes. In the following, the probability of cycling on the day before the interview is averaged across all trip purposes and all days over the year.

The problem can be tackled using Tobit-type models. The simplest variant, the Tobit type 1, implies limited flexibility in that explanatory variables affect both the probability of bicycle activities and the amount of cycling in the same way. Specifically, it implies that the effect of a variable x_1 on the selection probability $P[y > 0|x_1, x_2, \ldots, x_J]$ and the conditional expectation $E[y|x_1, x_2, \ldots, x_J, y > 0]$ have the same sign. However, this is a strong assumption, which may not hold.

A slightly more flexible model is the Heckman selection model (Heckman, 1979). However, while the Heckman model circumvents the above shortcoming, it is suited to a different type of problem, namely sample selection problems where part of the data is missing. In the case of bicycle demand, data are not missing. Instead, demand is zero for certain parts of the population.

A more suited model that applies to all distributions, is the Hurdle model (Cragg, 1971). In this model, the selection part of the problem is treated as a corner point of the joint distribution, e.g. demand is zero rather than missing. The model typically gives rise to marginal effects that differ from the Heckman model for the conditional regression model. The standard application of the model of Cragg (1971) often targets count data (Zeileis et al., 2008). However, as stated in Wooldridge (2002), the model can accommodate any distribution, continuous or categorical. While normal and truncated normal distributions are commonly applied in Hurdle models, these distributions appear not to be natural choices in this context. This is because of the strict positive nature of trip distances, suggesting that a log-normal distribution (Hsu and Liu, 2011) or Gamma model is a more natural choice.

The model, which is also often referred to as a two-part model, can be described by a two-equation system as shown in Eqs. (1) and (2). The first equation describes the selection process (or the hurdle) and the second equation the measurement part. In our context, the selection process is the probability of cycling, and the measurement is the cycled mileage. As mileage is only measured if a person is engaged in cycling activities, it is necessary to model cycling demand conditional on being active. The selection process is estimated using a Probit regression as shown in Eq. (1). In Eq. (1)$y = 0$ indicates that a person did not cycle, whereas $y > 0$ indicates that the person engaged in cycling activities on the day. In this paper, the set of explanatory variables for the selection process and the conditional mileage model are the same and referred to as x. The parameters for the explanatory variable in the selection response are given by γ, and Φ is the cumulative distribution function of the standard normal distribution. The conditional measurement equation is described by the relation shown in Eq. (2), where $g(.)$ represents a link function, and y is the cycling mileage in kilometres. On the right-hand side, x represents the explanatory variables with a corresponding parameter vector β, while u is a probability distribution with parameters θ_u.

$$P(y = 0|x) = 1 - \Phi(x\gamma) \tag{1}$$

$$g(y) = x\beta + u, \quad u \sim f(\theta_u) \tag{2}$$

The log-normal Hurdle model appears when $g(y) = log(y)$ and $u \sim N(0, \sigma_u)$, which give rise to the joint density distributions for y given x in Eq. (3) below;

$$f(y|x;\theta) = [1 - \Phi(x\gamma)]^{1|y=0}\left\{\Phi(x\gamma)\left[\frac{\phi((log(y) - x\beta)/\sigma)}{\sigma y}\right]\right\}^{1|y=0} \tag{3}$$

The second model tested is the Gamma Hurdle model with a log link. It appears as a special case when $g(y) = \ln(y)$ and $u \sim Gamma(\lambda, \sigma)$. This model gives rise to the joint density distributions shown in Eq. (4) below;

$$f(y|x;\theta) = [1 - \Phi(x\gamma)]^{1|y=0}\left\{\Phi(x\gamma)\frac{1}{\sigma^{\lambda}\Gamma(\lambda)}y^{\lambda-1}e^{y/\sigma}\right\}^{1|y=0} \quad \sigma, \lambda, y > 0 \tag{4}$$

We can rewrite the Gamma density to be a function of the mean and dispersion $Gamma(\mu_i, \nu)$, as opposed to the current shape and scale parameters. This we do by defining $\lambda = \nu^{-1}$ and $\sigma = \mu\nu$, where $\mu_i = E(y_i)$ describes the expectation function and ν the dispersion parameter. Thus the joint density distribution is given by Eq. (5)

$$f(y|x;\theta) = [1-\Phi(x\gamma)]^{1|y=0}\left\{\Phi(x\gamma)\frac{1}{\Gamma(\nu^{-1})}\left(\frac{1}{\mu\nu}\right)^{\nu^{-1}}y^{\nu^{-1}-1}e^{y/\mu\nu}\right\}^{1|y=0} \quad \sigma,\lambda,y > 0 \tag{5}$$

where the expectation function μ is given as $\ln(\mu) = x\beta$.

The best-performing model was the Gamma Hurdle model. Hence, this is the model reported in Section 4. It is worth noting that we seek to detect possibly small nuances regarding average cycling behaviour across cohorts. Therefore, we should expect high standard deviations and relatively low R^2 values for the models. In other words, models are not prediction models but rather attribution models using the classification from Efron (2020). Therefore, the main thing to look for is if Hurdle model parameters are significantly different across cohorts and if there is a systematic pattern in these parameters over time.

4. Results

This section presents the results of the descriptive cohort analysis and the Gamma Hurdle analysis of parameter stability across age cohorts.

4.1. Cycling probability and mileage for age and city size cohorts

The descriptive cohort analysis allows us to investigate if respondents from similar age groups change cycling behaviour across years. In Fig. 1, it is seen that for rural areas and small-sized cities below 25,000 inhabitants, there is an overall decline in the active cycling population. The strongest tendency is observed for the youngest age cohort. For this age cohort, the probability of cycling has declined over the past 13 years regardless of the level of urbanisation. For denser areas (25,000+), the results also indicate a decline in cycling over the entire period for the youngest group. However, for these segments, the decline happened before 2012.

For the four largest cities in Denmark, here denoted as Ar, Od, Aa and Copenhagen, the results appear relatively stable for the other age cohorts, with minor increases for the oldest segment (66+) in Ar, Od, and Aa and the working-age segment in Copenhagen.

Also, Fig. 1 reveals that in cities with less than 100,000 inhabitants, the baseline probability of cycling for the youngest age cohort (10–20) is much higher than for the other age groups. While for the larger cities, younger inhabitants (10–30) cycle more often compared to older age groups. However, for Copenhagen, it is noticeable that the working-age cohort almost has similar bike shares as the youngest segment. Additionally, it is seen that for the four largest cities, older generations (66 + years old) cycle much less compared to other age groups.

For the mileage driven by cyclists, we see a different pattern in Fig. 2 compared to the cycling probabilities in Fig. 1. The youngest cyclists (10–20) cycle significantly less in terms of mileage than any other age cohort. The only exceptions are the three large cities (Ar, Od, Aa) in recent years. In this period, the mileage for the youngest group is similar to that of cyclists aged between 20 and 30. This is also the only urbanisation level where the mileage for the youngest group appears to have increased.

Overall, the conditional cycling mileage is increasing for most of the other age cohorts across the country when measured for the entire period. The exceptions are the working-age cohorts in rural areas (<200) and Ar, Od, and Aa, as well as young adults in small-sized cities (200–25,000) and Ar, Od, and Aa. The only two urbanisation levels where there has been a steady increase in cycling mileage across all cohorts are Copenhagen (except the youngest age group) and the cities with 25,000–100,000 inhabitants.

4.2. Gamma Hurdle model for selection and mileage

The Gamma Hurdle models are estimated for the various age cohorts over the entire country for the respective time windows. We avoid dividing the cohorts further by city sizes to have sufficient degrees of freedom to estimate models for the various cohorts. Therefore, the city size variable is split into four dummy variables in the Hurdle models, with Copenhagen representing the reference level.

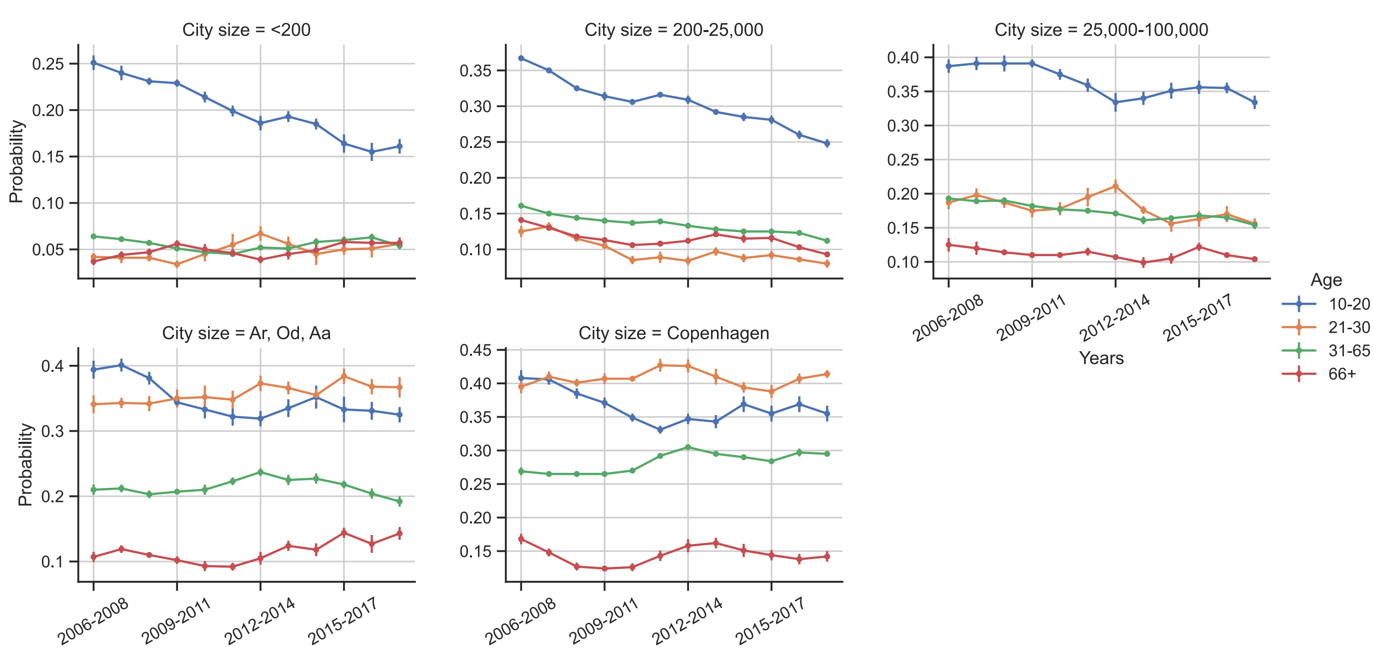

Fig. 1. Empirical probability of cycling, conditional on city-size and age cohort over rolling three-year time windows after 2006.

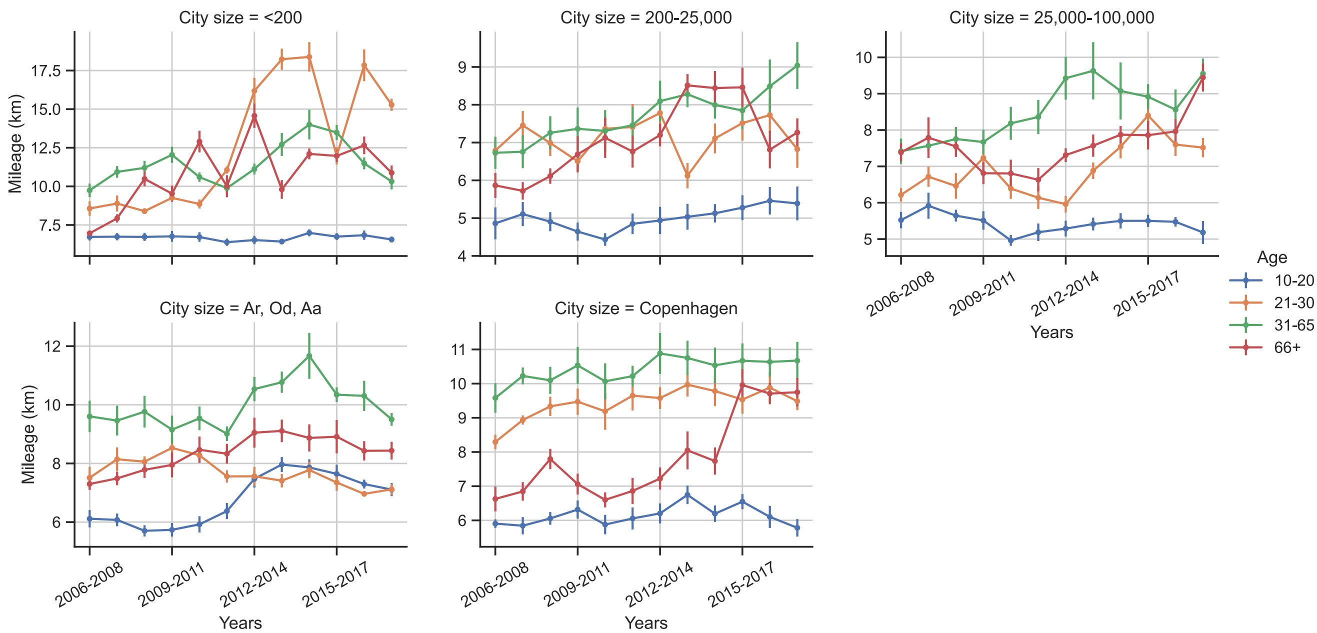

Fig. 2. Mileage cycled among cyclists, conditional on city-size and age cohort over rolling three-year time windows after 2006.

The exogenous variables in both models are: number of cars available to the household, age (starting with 0 for the minimum age in the cohort), gender and city size.

In Table 4, we show an example of estimated parameters for a specific age cohort and time window. The example illustrates the output format of each model segment. For the particular cohort in Table 4, we see how the number of cars available to the household is significant and negatively associated with the selection probability at a 10% significance level. Moreover, age and residential locations in rural areas are negatively associated with the selection at a 5% significance level. For the mileage model for this specific age cohort, we see that males cycle further than females, as do people in small towns with a city size of $200 - 25,000$ inhabitants. Meanwhile, increasing age in this cohort is correlated with higher mileage. The parameter estimates for each cohort and time window, along with the standard errors, are shown in Figs. 3 and 4. These figures thereby illustrate the stability of parameters for cohorts over time.

Starting with the baseline cycling selection probability, described by the *Intercept* variable, there is a decline in the age group 10–20 and a slight increase for the age group 31–65. Other groups fluctuate without a clear tendency. For the variable describing the effect of the number of cars in the household, there seems to be more uncertainty than tendency across age cohorts. Hence, while the impact of car ownership on the selection probability is different across age cohorts, changes over time

Table 4
Example output from a Hurdle model model for the age cohort 10–20 and time window 2006–2008. The table shown in the top section is the parameters for the Probit selection model whereas the conditional Gamma regression model ($y|y > 0$) for distance is shown in the lower section.

Cragg's Hurdle model
6341 observations (4074 censored and 2267 observed)
14 free parameters
Log-Likelihood: −10020
McFaddens's Pseudo R-squared: 0.018

Probit selection equation:						
	Estimate	std err	z	P> |z|	[0.025	0.975]
(Intercept)	0.1171	0.056	2.078	0.038	0.007	0.228
Cars in the household	−0.0430	0.023	−1.871	0.061	−0.088	0.002
Female	0.0179	0.033	0.549	0.583	−0.046	0.082
Age	−0.0585	0.005	−11.251	0.000	−0.069	−0.048
CitySize: Copenhagen	-	-	-	-	-	-
Citysize: 100,000–500,000	−0.0050	0.069	−0.073	0.942	−0.139	0.130
Citysize: 25,000–100,000	−0.0511	0.059	−0.861	0.389	−0.167	0.065
Citysize: 200–25,000	−0.1212	0.048	−2.533	0.011	−0.215	−0.027
Citysize: <200	−0.4416	0.059	−7.482	0.000	−0.557	−0.326
Conditional distance (y—y > 0), gamma regression:						
	Estimate	std err	z	P> |z|	[0.025	0.975]
(Intercept)	1.488	0.070	21.243	0.000	1.351	1.625
Cars in the household	0.0205	0.029	0.698	0.485	−0.037	0.078
Female	−0.094	0.040	−2.359	0.018	−0.172	−0.016
Age	0.055	0.007	7.999	0.000	0.041	0.068
CitySize: Copenhagen	-	-	-	-	-	-
Citysize: 100,000–500,000	−0.011	0.080	−0.135	0.893	−0.168	0.147
Citysize: 25,000–100,000	−0.0671	0.070	−0.963	0.335	−0.204	0.069
Citysize: 200–25,000	−0.162	0.070	−2.866	0.004	−0.272	−0.051
Citysize: <200	0.1473	0.075	1.967	0.049	0.001	0.294

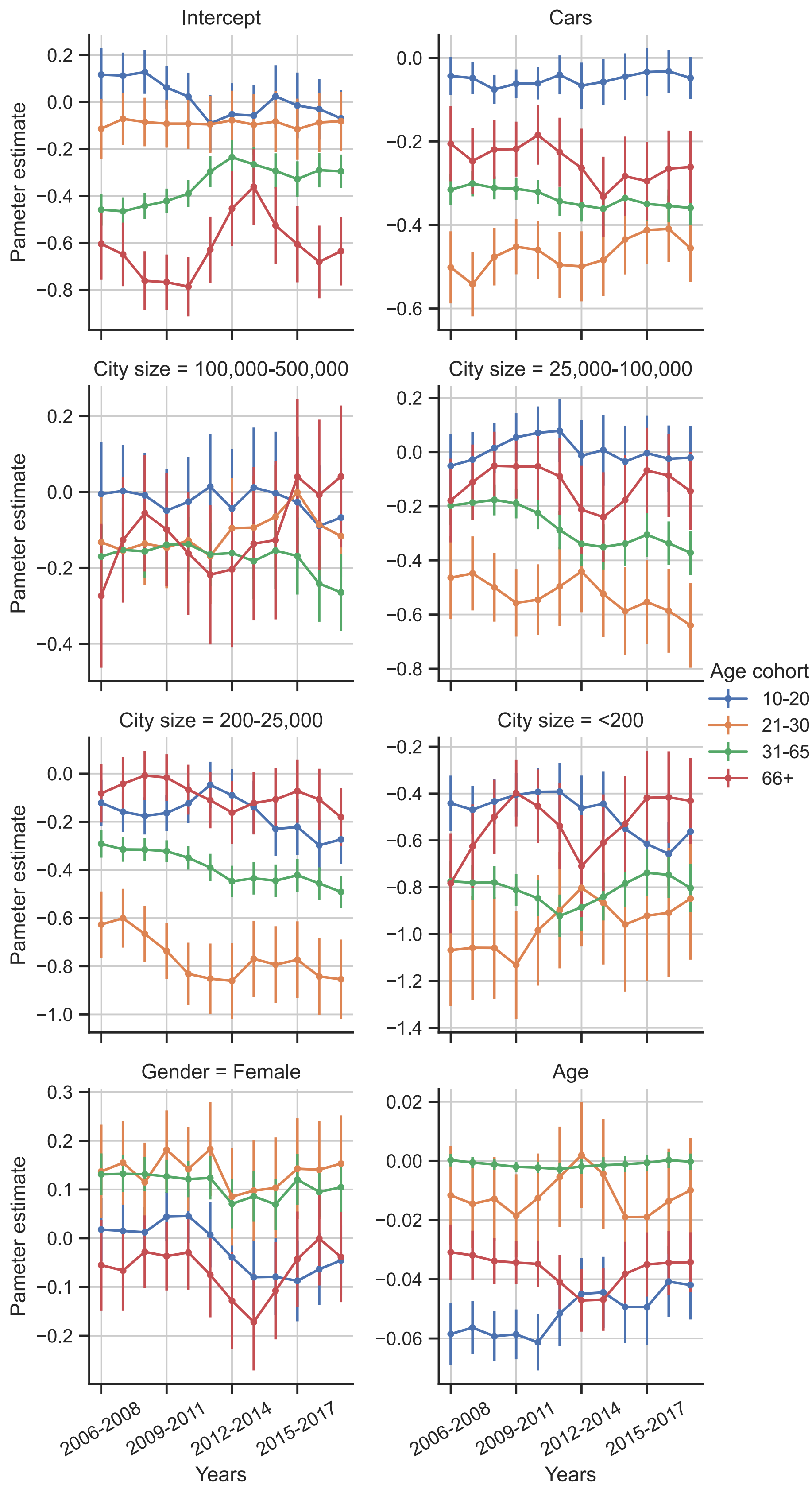

Fig. 3. Evolution of parameter estimates for the selection process (i.e. the hurdle of the Hurdle model) for different age cohorts in Denmark.

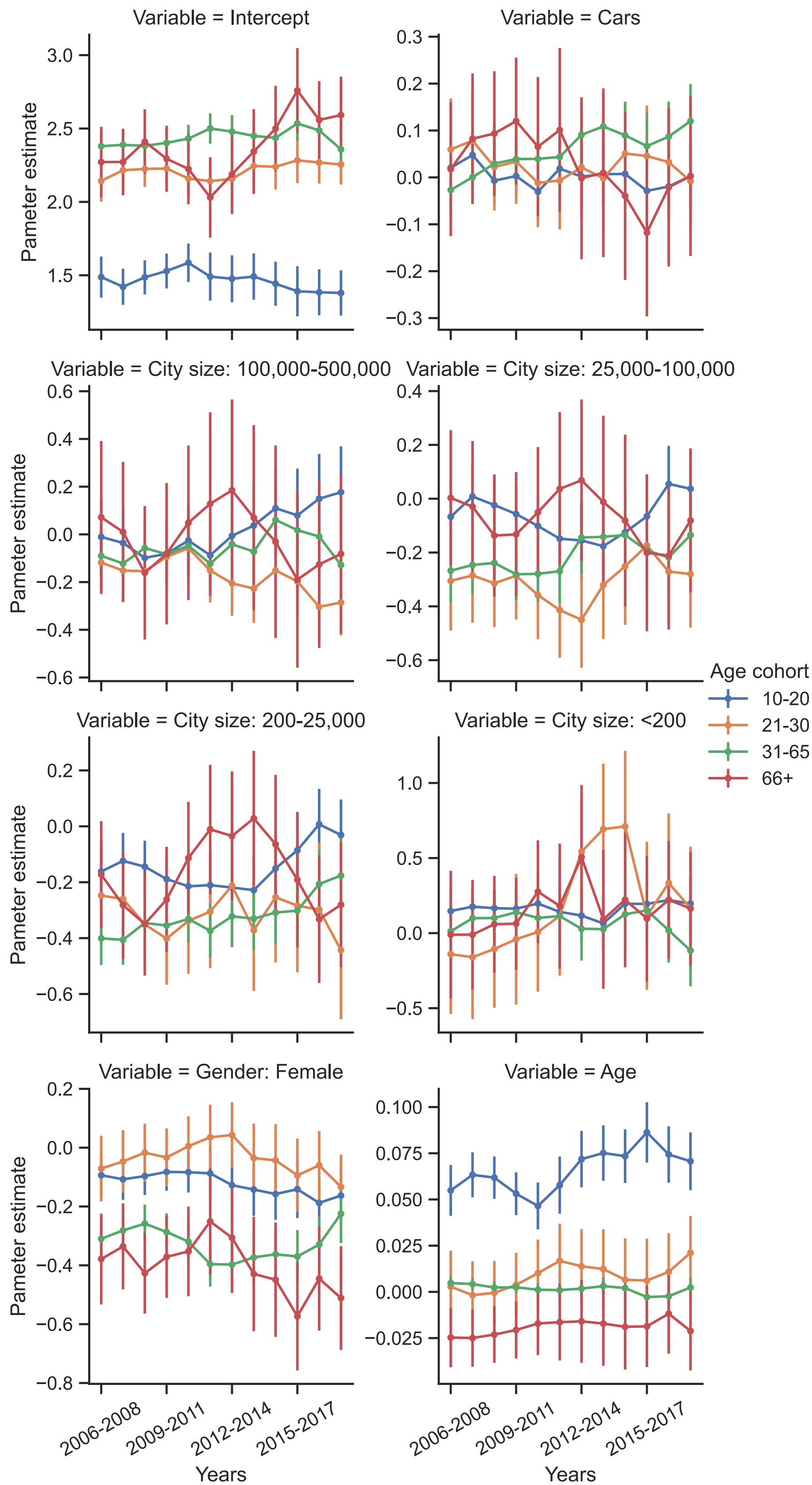

Fig. 4. Evolution of the parameter estimates for the cycling distance/measurement process of the Hurdle models for different age cohorts in Denmark.

are minor and within the uncertainty intervals.

Concerning the urbanisation level, it is noted that for cities with 100, 000 −500, 000 inhabitants, there is almost no statistical change in the cycling probability compared to Copenhagen. The only age cohort being significantly different is the oldest age cohort (66+), where the dummy goes from significantly negative to positive but insignificant. For the remaining city sizes, we see large uncertainties compared to changes. The only exception is a decrease for the working-age cohort in small- and medium-sized cities (200–100,000).

The effect concerning gender is generally inconclusive. For the age effect within the cohorts, we see a tendency that increasing age for the youngest age cohort is associated with a higher degree of cycling activities.

Concerning the parameters for the mileage equation in Fig. 4, the baseline cycling distances described by the *Intercept* variable are hardly changing over the period for each age cohort. For the number of cars, results are uncertain and focused around zero. It suggests that the number of cars in the household may not significantly affect the cycling mileage among cyclists having chosen to cycle. Also, for the city size variables, we observed hardly any change in effect across the cohorts over the period. For the gender effect we see a split between the two youngest cohorts and the older cohorts, where the former two hardly have any gender split in the mileage, while for the older cohorts, we see that women cycle significantly shorter distances than men. For the age effect we see that there is hardly any impact of increasing cyclist age. The only exception is that increasing age within the youngest cohort appears to have an increasing effect (absolute value) on mileage over the period.

5. Discussion and conclusion

This study sets out to explore the dynamics of cycling demand in Denmark. The analysis is based on pseudo-panels formed from a large Danish trip diary, which allows tracking demand through time for age groups and residential city size. By decomposing demand into a 'selection effect' and a conditional mileage effect, we can analyse the cycling uptake and how it develops. In particular, by using a Hurdle-type model across the various age cohorts, we are able to examine parameter stability for the two effects and tie these effects to specific explanatory variables.

The results from the descriptive cohort analysis show that the probability of cycling in Denmark is in an overall decline seen over the period 2006–2019. Two factors primarily drive this decline. Firstly, a declining cycling uptake in smaller cities and the countryside. These segments account for approximately half of the Danish population, making them important drivers of the decline. Secondly, a declining cycling uptake for the youngest generation. It is worth noting that for the larger cities with more than 25,000 inhabitants, the decline for the youngest generations can be attributed to the period before 2012. However, the two effects point to an alarming tendency for younger generations in rural areas whose demand for cycling has declined steadily over the entire period.

The results also show that the cycling probability and conditional mileage model work in opposite directions. The mileage is generally increasing for most generations, while the likelihood of cycling trips is generally decreasing. The Hurdle model shows that the impact of some factors on the cycling probability and mileage vary significantly across age cohorts and over time.

For the selection process, the number of cars in the household affects the likelihood of engaging in cycling activities, and the age cohorts have different parameters. While the selection of cycling activities for the youngest age cohort (10–20) is mainly unaffected by cars in the household, the age cohort from 21–30 years is significantly affected. Hence, the more cars in the household for this group, the fever cycle trips. The age cohort from 31–65 is slightly less affected than the age cohort over 65. If we consider Tables 1 and 2 there are some indications that the demand for car transport remains fairly unchanged in the period. It might suggest that the changes we see in demand for bicycling is the result of other things than increasing demand for car traffic. Again with reference to these figures, it seems as if walking has increased in the period and this could be a potential source of substitution. On the other hand, public transport has dropped by similar proportions as bicycles and the substitution of the missing bicycle trips thereby leaves a question mark which indeed is an interesting research question for the future.

Another interesting finding is that the selection process indicates that the decline in cycling is mainly a 'small-town' issue. This echoes the finding in Nielsen et al. (2016) that geographical differences are increasing. The smaller the city, the less likely people are to engage in cycling activities. For example, in cities with 200–25,000 inhabitants, almost all age cohorts are experiencing a declining selection probability. The analysis also points to an interesting observation in that the age effect for the youngest age cohort is becoming less negative over time. It could suggest that, while fewer teenagers bike, those who do, keep doing it further into their teenage years.

For the mileage effect, the variability over time is generally higher. It points to the fact that the degrees of freedom for this conditional model is significantly lower. The general conclusion regarding the mileage effect is that people, who travel by bicycle, travel largely similar distances as before. Also, the distance effect is not strongly correlated with other explanatory variables except for gender. For the gender variable, however, it seems that males travel longer distances.

At the very overall level, our study may point to the consequences of not prioritising rural areas and small towns when seeking to grow and maintain cycling as a relevant mode of transport.

5.1. Causes and limitations of the study

Since the model and the data indicates that children and younger people cycle less frequently on average, particularly in rural areas, it is relevant to ask the question if this effect could be related to the global trend of people (and in particular young people) moving to urban areas. We believe there is some truth in this but in some indirect way. First, the rural areas and the smaller cities are those areas for which the bicycle infrastructure is less developed and this could raise possible traffic safety concerns for parents. With people moving away from these areas the incentive to further invest in these areas is reduced. Hence, the observed effect might point to a lack of infrastructure in these areas and could be a focal point in the future for planners and politicians. Clearly, a relevant question is how infrastructure has been maintained or improved during the period and if such changes could influence the travel pattern. Another relevant observation which is also related to the urbanisation trend is that schools in Denmark, and in particular in rural areas, have been centralised in the past 20 years. With the number of pupils being reduced this is a logical consequence. However, it increases the cycling distance to schools, which in the period has increased significantly. The crow flight distance for school trips has increased from around 4 km in 2006–2008 to around 7–8 km in 2019 and has almost doubled. Although it may not alone be the consequence of school centralisation (it could also be the result of parents being more picky with schools in general) it could induce a lower cycle uptake for this age segment.

In a longer perspective, it is relevant to examine if the reduction in cycle uptake develops into long-term negative cycling habits when the younger generations grow older. However, the confirmation of this hypothesis would require panel-type data over a long period of time.

For older generations over 65, there has been a tendency of increased distance in recent years (2012–2019). This could be due to the increasing uptake of electric bikes and higher awareness of the health-related benefits of cycling for these older generations. The fact that the general health of older generations improves over time, with an increasing life expectancy, could also play a role.

Finally, when considering the selection effect, it is relevant to acknowledge that this effect might not solely result from a declining bicycle population but could result from less frequent bicycle use. In other words, if younger generations are cycling on Monday and Wednesday but skipping the Tuesday trip, this will emerge in the selection process due to how data are collected. Because we do not have panel data, we are not in a position to estimate the degree of this frequency effect.

As a final statement, the most important conclusion of the paper is that, while total cycling demand across the population declines only slightly over time, we see a general decline across the country as a whole for the youngest age group (10–20). Most noticeably, the cycling demand for younger age groups living outside the cities is strongly declining. The second most important conclusion is that the observed decline is mainly a selection effect. Hence, the drop in cycling demand essentially results from a decline in the general cycling uptake, not from people cycling shorter distances.

5.2. Future research

An investigation of the causal relationship between the observed trends over time, explanatory variables and policy initiatives is a research area that is both relevant and largely unexplored. It would require, however, that the various effects be disentangled appropriately, which is often tricky. Ways to accommodate such investigation could be to narrow in on specific areas or periods to assess the impact of changes to the infrastructure (Skov-Petersen et al., 2017) or specific socio-economic changes that happens locally or nationally. Likewise, it could relate to changes in the public transport system or the diffusion of new technology such as electric bikes. An overarching question relating to the causality discussion is the question of where the missing bicycle trips are substituted. It could be to other modes of transport or a more inactive lifestyle?

As stated above, it is also relevant to investigate if age cohort effects, as found in this paper, turn into more long-term attitudinal effects. In other words, will a decline in the cycling uptake for specific age cohorts translate into a long-term generational effect that will affect the life-long cycling uptake for generations to come? Moreover, if this is the case, what are the implied consequences of such development for the physical and mental health of the population.

It is also of great interest to study if the Danish results are unique or can be found in other countries. Although there are evidence that bicycling in rural areas are declining in The Netherlands as well (Harms et al., 2014) it still remains to be seen if this correlates with younger generations. Clearly, similar findings are observed in other countries, it would allow a comparison of how variables such as car ownership influence the development of cycling demand and if mitigating policies in one country can be an inspiration for other countries.

Finally, it could be relevant to analyse how the selection and mileage effects are distributed across more detailed cohorts to pinpoint specific focus areas or socio-economic groups for which attention is required. Previously, model-based machine learning has been used to examine trends across detailed synthetic pseudo-panels (Borysov and Rich, 2020) by modelling the entire joint data distribution. While this requires heavy statistical machinery, it could represent a valuable addition in this context by singling out specific groups with the highest reduction in cycling uptake.

Data availability

The authors do not have permission to share data.

References

Borysov, S.S., Rich, J., 2020. Introducing synthetic pseudo panels: application to transport behaviour dynamics. Transportation 48, 2493–2520. URL:https://link.springer.com/article/10.1007/s11116-020-10137-5.

Christiansen, H., 2020. National Travel Survey, 2020. URL: www.cta.man.dtu.dk/english/national-travel-survey.

Collet, R., 2012. Household Car Use in France: A Demographic and Economic Analysis. Econ. Bull. 32, 475–485.

Cragg, J.G., 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica 39, 829. https://doi.org/10.2307/1909582. URL:https://www.jstor.org/stable/1909582?origin=crossref.

Dargay, J.M., Vythoulkas, P.C., 1999. Estimation of a dynamic car ownership model: A pseudo-panel approach. J. Transp. Econ. Policy 33, 287–302. https://doi.org/10.2307/20053816.

Devereux, P.J., 2007. Small-sample bias in synthetic cohort models of labor supply. J. Appl. Econom. 22, 839–848. https://doi.org/10.1002/jae.938.

ECF, 2017. Cycling facts and figures. URL:https://ecf.com/resources/cycling-facts-and-figures.

Efron, B., 2020. Prediction, estimation, and attribution. J. Am. Stat. Assoc. 115, 636–655 arXiv:https://doi.org/10.1080/01621459.2020.1762613.

Ekelund, U., Ward, H.A., Norat, T., Luan, J., May, A.M., Weiderpass, E., Sharp, S.J., Overvad, K., Østergaard, J.N., Tjønneland, A., Johnsen, N.F., Mesrine, S., Fournier, A., Fagherazzi, G., Trichopoulou, A., Lagiou, P., Trichopoulos, D., Li, K., Kaaks, R., Ferrari, P., Licaj, I., Jenab, M., Bergmann, M., Boeing, H., Palli, D., Sieri, S., Panico, S., Tumino, R., Vineis, P., Peeters, P.H., Monnikhof, E., Bueno-de Mesquita, H.B., Quirós, J.R., Agudo, A., Sánchez, M.J., Huerta, J.M., Ardanaz, E., Arriola, L., Hedblad, B., Wirfält, E., Sund, M., Johansson, M., Key, T.J., Travis, R.C., Khaw, K.T., Brage, S., Wareham, N.J., Riboli, E., 2015. Physical activity and all-cause mortality across levels of overall and abdominal adiposity in European men and women: the European Prospective Investigation into Cancer and Nutrition Study (EPIC). Am. J. Clin. Nutr. 101, 613–621. https://doi.org/10.3945/ajcn.114.100065. URL:https://academic.oup.com/ajcn/article/101/3/613/4569416.

Granger, C.W.J., 1988. Aggregation of time series variables - A survey. Discussion Paper/ Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

Hallberg, M., Rasmussen, T.K., Rich, J., 2021. Modelling the impact of cycle superhighways and electric bicycles. Transp. Res. Part A: Policy Pract. 149, 397–418. https://doi.org/10.1016/J.TRA.2021.04.015.

Harms, L., Bertolini, L., te Brömmelstroet, M., 2014. Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. J. Transp. Health 1, 232–242. https://doi.org/10.1016/j.jth.2014.09.012. URL: https://www.sciencedirect.com/science/article/pii/S2214140514000802. walking and Cycling: The contributions of health and transport geography.

Haustein, S., Siren, A., 2012. Cohort analysis of older adults' travel patterns in Denmark. Technical Report. Department of Transport, DTU.. URL: https://orbit.dtu.dk/en/publications/cohort-analysis-of-older-adults-travel-patterns-in-denmark

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47, 153. https://doi.org/10.2307/1912352. URL:https://www.jstor.org/stable/1912352?origin=crossref.

Hjorthol, R.J., Levin, L., Sirén, A., 2010. Mobility in different generations of older persons. The development of daily travel in different cohorts in Denmark, Norway and Sweden. J. Transp. Geogr. 18, 624–633. https://doi.org/10.1016/j.jtrangeo.2010.03.011.

Hsu, A.C., Liu, S.C., 2011. The hurdle models choice between truncated normal and lognormal. https://doi.org/10.1080/00036840600749581 40, 201–207. URL:https://www.tandfonline.com/doi/abs/10.1080/00036840600749581, doi:10.1080/00036840600749581.

Humphreys, D.K., Goodman, A., Ogilvie, D., 2013. Associations between active commuting and physical and mental wellbeing. Prev. Med. 57, 135–139. https://doi.org/10.1016/j.ypmed.2013.04.008. URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3722482/pmc/articles/PMC3722482/?report=abstract.

Jarrett, J., Woodcock, J., Griffiths, U.K., Chalabi, Z., Edwards, P., Roberts, I., Haines, A., 2012. Effect of increasing active travel in urban England and Wales on costs to the National Health Service. doi:10.1016/S0140-6736(12)60766-1.

Jiang, G., Wang, W., 2017. Error estimation based on variance analysis of k-fold cross-validation. Pattern Recognit. 69, 94–106,. https://doi.org/10.1016/j.patcog.2017.03.025. URL:https://www.sciencedirect.com/science/article/pii/S003132031730136X.

Krueger, R., Rashidi, T., Vij, A., 2020. X vs. Y: an analysis of intergenerational differences in transport mode use among young adults. Transportation 47, 2203–2231.

Kyu, H.H., Bachman, V.F., Alexander, L.T., Mumford, J.E., Afshin, A., Estep, K., Veerman, J.L., Delwiche, K., Iannarone, M.L., Moyer, M.L., Cercy, K., Vos, T., Murray, C.J.L., Forouzanfar, M.H., 2016. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ 10, i3857. URL:https://www.bmj.com/lookup/doi/10.1136/bmj.i3857.

Lewbel, A., 1994. American Economic Association Aggregation and Simple Dynamics. Am. Econ. Rev. 84, 905–918.

Martin, A., Goryakin, Y., Suhrcke, M., 2014. Does active commuting improve psychological wellbeing? Longitudinal evidence from eighteen waves of the British Household Panel Survey. Prev. Med. 69, 296–303. https://doi.org/10.1016/j.ypmed.2014.08.023.

Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., de Nazelle, A., Dons, E., Gerike, R., Götschi, T., Int Panis, L., Kahlmeier, S., Nieuwenhuijsen, M., 2015. Health impact assessment of active transportation: A systematic review. Prev. Med. 76, 103–114. https://doi.org/10.1016/J.YPMED.2015.04.010. URL:https://www.sciencedirect.com/science/article/pii/S0091743515001164?via%3Dihub.

Mytton, O.T., Panter, J., Ogilvie, D., 2016. Longitudinal associations of active commuting with wellbeing and sickness absence. Prev. Med. 84, 19–26. https://doi.org/10.1016/j.ypmed.2015.12.010.

Newbold, K.B., Scott, D.M., Spinney, J.E.L., Kanaroglou, P., Páez, A., 2005. Travel behavior within Canada's older population: A cohort analysis. J. Transp. Geogr. 13, 340–351. https://doi.org/10.1016/j.jtrangeo.2004.07.007.

Nielsen, T.A., Mulalic, I., Christiansen, H., 2016. Drivers of cycling mode-share: Analysis of danes travel behavior 1996–2013. Transp. Res. Proc. 14, 2284–2288. https://doi.org/10.1016/j.trpro.2016.05.244. URL:https://www.sciencedirect.com/science/article/pii/S2352146516302502. transport Research Arena TRA2016.

Nolan, A., 2010. A dynamic analysis of household car ownership. Transp. Res. Part A: Policy Pract. 44, 446–455. https://doi.org/10.1016/j.tra.2010.03.018.

Papu Carrone, A., Monteiro, M.M., Rich, J., 2021. Modelling car ownership dynamics based on irregularly spaced panel data. Travel Behav. Soc. 25, 223–232. https://doi.org/10.1016/J.TBS.2021.07.008.

Papu Carrone, A., Rich, J., 2020. A micro-model of car ownership dynamics: are preferences changing? 13, 138–147. URL:https://www.tandfonline.com/doi/abs/10.1080/19427867.2020.1716473, doi:10.1080/19427867.2020.1716473.

Rayaprolu, H.S., Llorca, C., Moeckel, R., 2018. Impact of bicycle highways on commuter mode choice: A scenario analysis. Environ. Plann. B: Urban Anal. City Sci. 1–16. https://doi.org/10.1177/2399808318797334.

Raza, W., Krachler, B., Forsberg, B., Sommar, J.N., 2020. Health benefits of leisure time and commuting physical activity: A meta-analysis of effects on morbidity. J. Transp. Health 18, 100873. https://doi.org/10.1016/J.JTH.2020.100873.

Rezende, L.F.M.d., Sá, T.H.d., Markozannes, G., Rey-López, J.P., Lee, I.M., Tsilidis, K.K., Ioannidis, J.P., Eluf-Neto, J., 2018. Physical activity and cancer: an umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. Br. J. Sports Med. 52, 826–833. https://doi.org/10.1136/BJSPORTS-2017-098391. URL: https://pubmed.ncbi.nlm.nih.gov/29146752/.

Rich, J., Jensen, A.F., Pilegaard, N., Hallberg, M., 2021. Cost-benefit of bicycle infrastructure with e-bikes and cycle superhighways. Case Stud. Transp. Policy 9, 608–615. https://doi.org/10.1016/j.cstp.2021.02.015.

Rich, J., Vandet, C.A., 2019. Is the value of travel time savings increasing? Analysis throughout a financial crisis. Transp. Res. Part A: Policy Pract. 124, 145–168. https://doi.org/10.1016/J.TRA.2019.03.012.

Skov-Petersen, H., Jacobsen, J.B., Vedel, S.E., Thomas Alexander, S.N., Rask, S., 2017. Effects of upgrading to cycle highways - an analysis of demand induction, use patterns and satisfaction before and after. J. Transp. Geogr. 64, 203–210. https://doi.org/10.1016/j.jtrangeo.2017.09.011. URL:https://www.sciencedirect.com/science/article/pii/S0966692316304008.

Sommar, J.N., Johansson, C., Lövenheim, B., Schantz, P., Markstedt, A., Strömgren, M., Stigson, H., Forsberg, B., 2021. Overall health impacts of a potential increase in cycle commuting in Stockholm, Sweden. Scand. J. Public Health, 14034948211010024. https://doi.org/10.1177/14034948211010024. URL:http://www.ncbi.nlm.nih.gov/pubmed/33977822.

Synek, S., Koenigstorfer, J., 2019. Health effects from bicycle commuting to work: Insights from participants of the German company-bicycle leasing program. J. Transp. Health 15, 100619. https://doi.org/10.1016/j.jth.2019.100619. URL:https://linkinghub.elsevier.com/retrieve/pii/S2214140519301963.

Wooldridge, J.M., 2002. Conditional Expectation Related Concepts in Econometrics. Econ. Anal. Cross Sect. Panel Data 30–31.

Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression Models for Count Data in R. J. Stat. Softw. 27, 1–25. https://doi.org/10.18637/JSS.V027.I08. URL:https://www.jstatsoft.org/index.php/jss/article/view/v027i08.