Forecasting the impact of infrastructure on Swedish commuters' cycling behaviour

Gunilla Björklund – VTI
Gunnar Isacsson – VTI

CTS Working Paper 2013:36

Abstract

In this paper we investigate the impact of four cycling environments on the propensity to cycle to work. The types of infrastructure investigated were mixed traffic, bicycle lane in the road way, bicycle path next to the road, and bicycle path not in connection with the road. In the mode choice model we combined three different data sets, two with stated preference data and one with revealed preference data, restricted to only include journeys of 12 km or less. At baseline, 24% of the cycling time was spent in mixed traffic, 2% in bicycle lanes, 42% on a bicycle path near the road way, and 31% on a bicycle path not in connection to a road way. Values of travel time savings for bicycling independent on infrastructure (based on revealed preference data) was 176 SEK/h, for cycling in mixed traffic it was 241 SEK/h, for cycling on a bicycle lane in the road way it was 249 SEK/h, for cycling on a bicycle path next to the road it was 178 SEK/h, and for cycling on a bicycle path far from the road it was 167 SEK/h (all differentiated values are based on rescaled stated preference data). Using an incremental form of the logit model we found that the biggest shift to cycle that may be possible is if all cycling after the change takes place on the bike path far from the road. The proportion of cyclists in this sample would then increase from 51.0% to 61.3%, i.e. an increase of 20%.

Forecasting the impact of infrastructure on Swedish commuters' cycling behaviour

Gunilla Björklund[1*], Gunnar Isacsson[2]

[1]Swedish National Road and Transport Research Institute (VTI)

& Centre for Transport Studies

[2] Swedish National Road and Transport Research Institute (VTI)

Abstract

In this paper we investigate the impact of four cycling environments on the propensity to cycle to work. The types of infrastructure investigated were mixed traffic, bicycle lane in the road way, bicycle path next to the road, and bicycle path not in connection with the road. In the mode choice model we combined three different data sets, two with stated preference data and one with revealed preference data, restricted to only include journeys of 12 km or less. At baseline, 24% of the cycling time was spent in mixed traffic, 2% in bicycle lanes, 42% on a bicycle path near the road way, and 31% on a bicycle path not in connection to a road way. Values of travel time savings for bicycling independent on infrastructure (based on revealed preference data) was 176 SEK/h, for cycling in mixed traffic it was 241 SEK/h, for cycling on a bicycle lane in the road way it was 249 SEK/h, for cycling on a bicycle path next to the road it was 178 SEK/h, and for cycling on a bicycle path far from the road it was 167 SEK/h (all differentiated values are based on rescaled stated preference data). Using an incremental form of the logit model we found that the biggest shift to cycle that may be possible is if all cycling after the change takes place on the bike path far from the road. The proportion of cyclists in this sample would then increase from 51.0% to 61.3%, i.e. an increase of 20%.

Keywords: Cyclists; commuting; forecasting; infrastructure; stated preference; revealed preference

---

*Corresponding author. E-mail: gunilla.bjorklund@vti.se

1. Introduction

The last decade has seen an increasing focus on environmentally friendly and healthy modes of transport such as cycling. In Sweden, the government has appointed a special cycling inquiry, which recently has been presented. The mission of the inquiry was to review the regulations and conditions in the areas of infrastructure, cycle parking, cycling and public transport, as well as traffic regulations affecting cycling, and to propose legislative amendments and other measures that may be needed to increase cycling and make it safer (SOU, 2012:70).

The aim of this study is to develop models to forecast how Swedish commuters' cycling behaviour depends on different kinds of bicycle environment. The specific question we want to answer is how improvements in the bicycle environment, e.g., building bicycle paths, influence the share of bicycle commuters. The study is to a large extent based on a study by Wardman et al. (2007), who investigated forecasting impacts of different measures regarding cycling, where infrastructure was one of them, among British commuters. Like our precursor, we chose to study the commuting market because "it represents a significant proportion of trips and ones where congestion is worst, environmental problems most concentrated, data availability greatest, and the salient issues can be addressed by analysis of mode choice without the need to consider the more uncertain and complex issues surrounding the generation of new trips" (Wardman et al., 2007, p. 340). Also, there is probably a limited number of ways to reach the work place and we assume that the commuters, at least in the beginning of their commuting period, made a deliberate choice of travel mode and travel route.

Different data sources have their strengths and weaknesses and therefore using a combination of individuals' actual choices, i.e., revealed preference (RP) data, and choices between hypothetical alternatives, i.e., stated preference (SP) data, is recommended in forecasting models (for an overview, see Cherchi & Ortúzar, 2006). Whereas RP data have high reliability and face validity it is also quite inflexible and often inappropriate due to the constraints of existing alternatives, attribute levels, and correlations between attributes (Louviere et al., 2000). SP data, on the other hand, can capture a wider array of preference-driven behaviours than RP data, and models based on SP data tend to be more robust than RP models, but have the disadvantage of being hypothetical and do not take certain types of real market constraints into account (Louviere et al., ibid.). Another disadvantage with SP data is

that the scale of the utilities in the model could be wrong, resulting in too small estimated coefficients (Wardman et al., 1997).

Research combining different kinds of data sets when investigating the demand for different cycling environments is still rare. The study by Wardman et al. (2007), which was based on several data sets, is an exception. They found that the value of travel time savings was about the same for minor roads with no bicycle facilities and for major roads with no bicycle facilities. Travel time savings for cycling on non-segregated on-road bicycle lanes were valued at 0.37 times the value of cycling on a minor road with no bicycle facilities, meaning that the road with more cycle facilities was preferred. Cycling on segregated on-road bicycle lanes was valued at 0.17 times the value of cycling on a minor road with no bicycle facilities, and cycling on completely segregated bicycle ways was valued at 0.14 times the value of cycling on a minor road with no bicycle facilities.

There are other studies of the demand for different cycling environments that only use stated preference data, most of them focusing on valuations of travel time savings (VTTS). Hopkinson and Wardman (1996) found that a route with a cycleway was worth 71 pence[1] compared to a route without cycleway. Stangeby (1997) found that separated bicycle lanes were as important as more than a one hundred per cent reduction in cycle time on short trips. Börjesson and Eliasson (2012) found that for short trips the VTTS for cycling on a bicycle path was valued at 0.70 times the VTTS for cycling on the street, indicating that cyclists prefer cycling on safer paths. Wardman et al. (1997) estimated the VTTS for cycling on an unsegregated bicycle lane at 0.79 times the value of cycling in mixed traffic and the VTTS for cycling on a fully segregated bicycle lane at 0.30 times the value of cycling in mixed traffic. Tilahun et al. (2007) found that for a given individual, keeping utility at the same level and with 20 minutes as base travel time, the off-road facility could be exchanged for 5.13 minutes of travel time, a bicycle lane for 16.41 minutes of travel time, and a no parking facility for 9.27 minutes of travel time.

In the present study we investigate the impact of four cycling environments on the propensity to cycle to work by combining different data sets. The different types of infrastructure are mixed traffic, bicycle lane in the roadway, bicycle path next to the road, and bicycle path not in connection with the road.

---

[1] The exchange rate from British pence to Euro (EUR) is pence 84.0/EUR (March 30, 2013).

Since there is only one previous study (Wardman, et al. 2007) of commuters' cycling behaviour that combines different kinds of data, it seems relevant to conduct another and similar study in this field of research. In addition, the present study is based on data for Sweden whereas Wardman et al. (ibid.) used data for the UK. Since Sweden and the UK are likely to be different in terms of traffic and climate that may impact on unobserved aspects of cycling, conducting a study similar to that of Wardman et al. (ibid.) on Swedish data seems motivated.

The rest of the paper is organised as follows. In section 2, a short description is given of the data and the imputation approach we use to deal with missing values in the data sets. Section 3 reviews the model specification including the estimated utility functions. In section 4 the results from the estimated mode choice models are presented and section 5 presents the results from the forecast applications. Section 6 concludes with a discussion and conclusions.

2. Data

The estimated models are based on three data sets from two surveys made during the summer period year 2011. The first data set (SP1) contains stated preference data where questionnaires were handed out to 3,000 cyclists in four Swedish cities (Karlstad, Luleå, Norrköping, and Västerås[2]) when they actually were cycling. The cyclists were approached at intersections, where traffic signals or stop signs requested them to stop, or at other places where natural stops were supposed to occur. Another 838 persons declined on spot to participate in the study (some of them might have received a questionnaire already), but the others received a prepaid response envelop and the questionnaire to fill in at home. Only persons of age 18 years or older, and understanding Swedish, were recruited. Of the handed out questionnaires, 1,518 (51%) were returned. From this study we use information about cycle time spent in each of the four bicycle environments (mixed traffic, bicycle lane in the roadway, bicycle path next to the road, and bicycle path not in connection with the road; see Figure 1), travel time for car/public transport, travel cost for car/public transport, income, gender, and age. The second data set (SP2) is also from a stated preference study, with the same stated preference questions and socio-economic questions as in the SP1, but where we sent the questionnaires home to 6,000 persons between 18 and 64 years old in the same four cities in order to receive

---

[2] The municipality of Karlstad had about 87,000 inhabitants by the end of 2012, The corresponding figures for Luleå, Norrköping and Västerås were: 75,000; 132,000 and 140,000, respectively.

responses from commuters, both regular cyclists and potential cyclists. We received 1,848 completed questionnaires in return, which implies a response rate of 31%. In addition, 4.5% of the questionnaires were returned due to no regular journey, wrong address, or other reason. The response rate is low, but we assume that many persons who received the questionnaire did not have any regular trip, which was the criteria for participation. Therefore, we can assume that the actual response rate is higher, even if we cannot say how much higher. In addition to the stated preference questions, we asked questions about an ordinary trip, preferable a trip to school or work, how often the respondents use different travel modes on that trip, and time and cost for different modes. The third data set (RP) consists of these revealed (or self-reported) preferences.
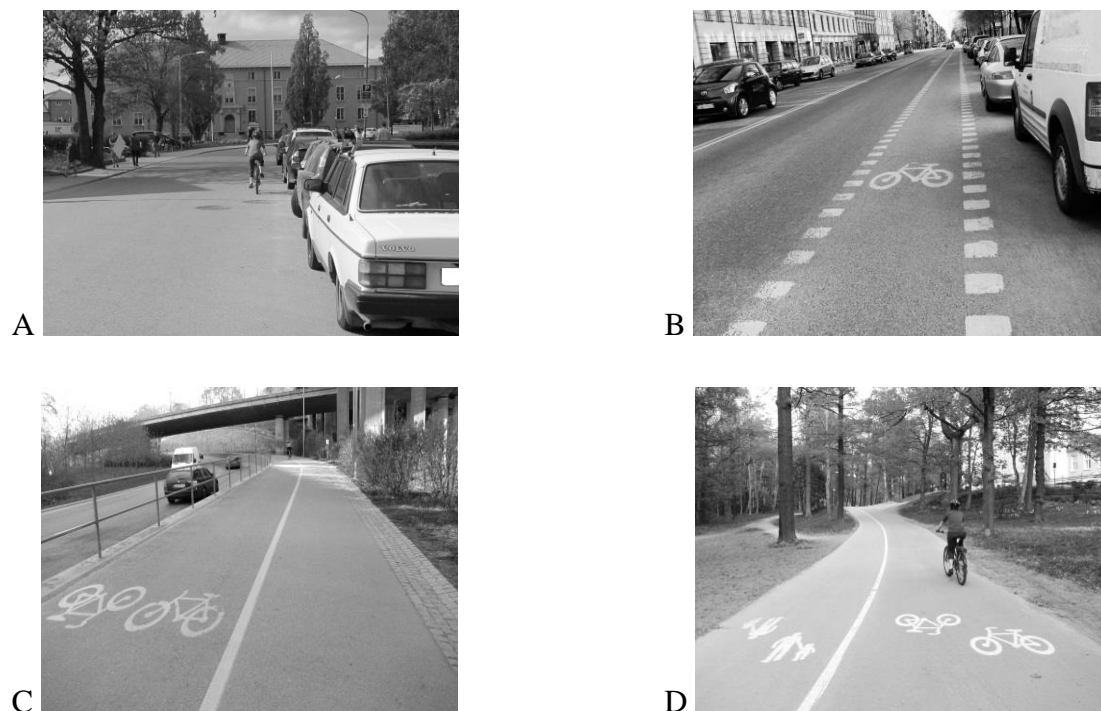

Figure 1. Cycling environments used in the study. (A) Mixed traffic. (B) Bicycle lane in the road way. (C) Bicycle path next to the road. (D) Bicycle path not in connection with the road.

In both SP1 and SP2, each respondent faced twelve stated preference choices between bicycle and an alternative travel mode, where the alternative mode was either car or public transport. Prior to making the choices, the respondents had to indicate whether car or public transport was the best alternative to them. For example, if the respondent had indicated car, each of the

twelve choices was between bicycle or car. To limit the number of choices for each person to twelve, three different versions of the questionnaire were constructed. In each version, three of the four bicycle environments were presented. The bicycle time consisted of the levels 20, 25, and 35 minutes, whereas the time for the alternative travel mode consisted of the levels 10, 13, and 18 minutes, and was thus always the faster mode of travel. The cost of the alternative travel mode varied between 10, 16, and 32 SEK, whereas the cost of the bicycle was assumed to be zero. Figure 2 shows an example of a stated preference choice.

	Bicycle	Alternative travel mode
	The trip takes 20 min	The trip takes 18 min
	The trip takes place in a *bicycle lane in the road way*	The trip costs 10 SEK
I choose:	☐ Bicycle	☐ Alternative travel mode
	☐ I cancel the trip	

Figure 2. An example of a stated preference choice

We make the following restrictions on the sample used to estimate the models. First, in all data sets we only include persons that commute to work, either on the journey they were on when they received the questionnaire (SP1) or who stated this as their ordinary trip (SP2/RP). Furthermore, the journey has to be 12 km or less, and the respondents have to be between 18 and 64 years old to be included in the analysis. In the data sets SP2/RP we impose the restriction that the ordinary trip should be made at least two times a week, and no one of the other travel modes should be used that often. In the RP we also exclude individuals who have an observed trip time and length which implies a cycle speed of less than 5 km/h or more than 30 km/h and we also set a limit of 60 minutes travel time for car and for public transport.

Finally, individuals with cycle trip times above 90 minutes are also removed. The reason for the restrictions was that these values represent outliers or otherwise unrealistic values.

There is quite a lot of missing information in the RP data set so we have imputed values for a number of variables where the relevant information has been missing. First, regarding the mode walking, we have information about the travel length but not the travel time in the RP data set. Since we prefer to use travel time in the empirical models, we assume a walking speed of 5 km/h and calculate the travel time from that. Further, in the RP data set we impute values for time and cost of public transport and car whenever these are missing. The imputed values for cost of public transport are based on the average cost per kilometer in each of the four cities and the length of the persons trip; i.e. we multiply the city's average cost of public transport per kilometer and the length of the persons trip to arrive at the imputed value for cost of public transport for each individual that lacks this information. The imputed values for the cost of car are obtained similarly. In total, we have imputed the cost of public transport for 479 observations (85%) and the cost of car for 236 observations (42%). Including the imputed values in the sample, the average costs per kilometer vary between 1.77 SEK/km and 2.68 SEK/km for car and between 3.44 SEK/km and 4.71 SEK/km for public transport, which we think is not totally unrealistic. The imputed values for travel time with public transport and car are based on the average speed of each mode in each of the four cities and the length of the person's trip. The number of imputed time observations was 239 (42%) for public transport and 59 (10%) for car. Including the imputed variables in the sample, the average speeds vary between 27.7 km/h and 29.7 km/h for car and between 12.2 km/h and 16.6 km/h for public transport.

After imposing the restrictions on the sample outlined above and after imputing values for missing information we end up with a total of 565 individuals/observations in the RP. Of these individuals, 387 are also included in the SP2. The number of individuals in the SP1 and the SP2 are 1,056 (12,479 observations) and 663 (7,459 observations), respectively.[3] These observations are used to estimate the empirical models outlined in the next section.

---

[3] We have also tried to estimate a model which include data from the Swedish national travel survey to extend our survey which was made in four Swedish cities. However, because that data set only includes information about travel time and travel length regarding the particular travel mode used we had to engineer travel times for the modes not chosen and also for travel costs for car and public transport. This resulted in a model where some of the coefficients were unrealistic low or had "wrong" signs. Therefore, we do not use that data set in the analyses of the present paper.

3. Model specification

In this paper we will rely on a standard random utility framework. This implies that each person's utility will consist of a systematic component that depends on the attributes of each alternative and a stochastic, unobserved, component. When estimating a joint model on several data sets, each set contains a vector of attributes, some of which are specific to the data set, but information on at least some of them has to be present in all data sets. Following Louviere et al. (2000), we assume that the latent utility underlying the choice process in two combined data sets is given by equations (1) and (2).

$$U_i^{RP} = \alpha_i^{RP} + \beta^{RP} X_i^{RP} + \omega Z_i + \varepsilon_i^{RP}, \forall i \in C^{RP}, \tag{1}$$

$$U_i^{SP} = \alpha_i^{SP} + \beta^{SP} X_i^{SP} + \delta W_i + \varepsilon_i^{SP}, \forall i \in C^{SP}, \tag{2}$$

where i is an alternative in choice sets C^{RP} or C^{SP}, αs are data set-specific alternative-specific constants (ASCs), β^{RP} and β^{SP} are utility parameters for the common attributes and are equal in the model and ω and δ are utility parameters or parameters for the unique attributes in each data set. The errors terms are independently and identically distributed.

The unobserved part in the utility function has variance $\sigma^2 \times (\pi^2/6)$. Because the scale of utility is irrelevant to behaviour, utility can be divided by σ without changing behaviour. The coefficients in the utility functions are therefore scaled by $1/\sigma^{RP}$ and $1/\sigma^{SP}$, respectively, where σ^{RP} and σ^{SP} scale the coefficients to reflect the variance of the unobserved part of the utility and are therefore called *scale parameters*. Because the coefficients and the scale parameters are not separately identified, only the ratio between each coefficient and its scale parameter can be estimated. (Train, 2009)

Because it is not possible to identify a scale parameter within a particular data set direct comparisons between parameters from different models are impossible (Louviere et al., 2000). However, the scale parameter does not affect the ratio of any two coefficients, e.g., willingness to pay, values of time, and other measures of marginal rates of substitution (Train, 2009). These ratio measures can therefore be used in comparisons between different models.

In joint RP/SP estimation it is common to force the SP utility to have the same scale parameter as the RP data because the RP data are assumed to reflect the "correct" scale

associated with the "real market" (Brownstone et al., 2000). This means that the secondary data set (SP) is multiplied by a factor $\phi = \sigma^{RP}/\sigma^{SP}$ to ensure consistency of scale (Cherchi & Ortúzar, 2006a). Because of the reciprocal relationship between scale and variance, values less than one imply that the SP stochastic variance component is larger than the RP component (Brownstone et al., 2000).

The stated choices collected by SP1 and SP2 results in a panel data structure of the data; i.e., we observe several choices by each individual. This may imply a correlation between choices made by each person. To deal with this we let the alternative specific constants also be individual-specific and add individual-specific random coefficients in the utility functions for observations pertaining to SP1 and SP2. These additional error terms are assumed to be normally distributed with zero means and standard deviations $\sigma_{\text{panel\_sp1}}$ and $\sigma_{\text{panel\_sp2}}$, respectively and vary across individuals but are constant for each individual. These mixed logit models for panel data were estimated by a maximum likelihood approach in BIOGEME (Bierlaire, 2003).

In sum, the following utility functions were jointly estimated on the three data sets outlined previously, where V_i is the systematic and measurable component of utility U_i:

$$V_{SP1\_b} = \beta_{time\_mt}Time_{mt} + \beta_{time\_bl}Time_{bl} + \beta_{time\_bpr}Time_{bpr} + \beta_{time\_bpnr}Time_{bpnr}$$

$$V_{SP1\_c\_pt} = ASC_{SP1_{c\_pt}} + \beta_{c\_pt_{time}}Time_{c\_pt} + \beta_{cost}Cost_{c\_pt}$$

$$V_{SP2\_b} = \beta_{time\_mt}Time_{mt} + \beta_{time\_bl}Time_{bl} + \beta_{time\_bpr}Time_{bpr} + \beta_{time\_bpnr}Time_{bpnr}$$

$$V_{SP2\_c\_pt} = ASC_{SP2_{c\_pt}} + \beta_{c\_pt_{time}}Time_{c\_pt} + \beta_{cost}Cost_{c\_pt}$$

$$V_{RP\_b} = \beta_{time\_b}Time_b$$

$$V_{RP\_c} = ASC_{RP\_c} + \beta_{c\_pt_{time}}Time_c + \beta_{cost}Cost_c$$

$$V_{RP\_pt} = ASC_{RP\_pt} + \beta_{c\_pt_{time}}Time_{pt} + \beta_{cost}Cost_{pt}$$

$$V_{RP\_w} = ASC_{RP\_w} + \beta_{time\_w}Time\_w$$

where b is bicycle, c is car, pt is public transport, mt is cycling in mixed traffic, bl is cycling in bicycle lane, bpr is cycling in a bicycle path next to the roadway, and $bpnr$ is cycling in a

bicycle path not in connection to the roadway. Note that the utility functions pertaining to SP1 and SP2 use a pooled alternative for car and public transport (*c_pt*). Alternative specific constants for bicycle are all normalized to zero; i.e. bicycle is treated as the reference category.

4. Estimated models

The results from the three separate mode choice models are presented in Table 1.

Table 1. Mode choice models for commuters with 12 km or less to work, separately estimated models

Parameter	SP1		SP2		RP	
	Estimate	t-value	Estimate	t-value	Estimate	t-value
Cost_car_pt	-0.050	-15.75	-0.051	-13.22	-0.111	-4.79
Time_car_pt	-0.075	-8.70	-0.100	-9.42	-0.074	-3.50
Time_bicycle	-0.188	-36.49	-0.164	-25.37	-0.189	0.017
ASC_{car}					-2.37	-8.29
ASC_{pt}					-2.95	-7.21
ASC_{car_pt}	-5.15	-25.36	-1.98	-7.77		
$Sigma_{car\_pt}$	-2.38	-26.95	3.18	21.75		
Number of draws	150		150			
Number of observations	12479		7459		536	
Number of respondents	1056		633		536	
Log-likelihood	-5069.898		-3354.228		-330.277	
Likelihood ratio test	7159.772		3631.914		517.158	
Rho-square	0.414		0.351		0.439	
Adjusted rho-square	0.413		0.350		0.431	
VTTS: car_pt	89 (65 – 113) SEK/h		117 (88 – 145) SEK/h		40 (10 – 70) SEK/h	
VTTS: bicycle	224 (195 – 252) SEK/h		192 (162 – 221) SEK/h		102 (65 – 139) SEK/h	

Note. The standard errors in the calculation of confidence intervals are based on Taylor series expansion.

ASC = alternative-specific constant

We see from Table 1 that all time and cost coefficients have the "right" signs; i.e., they are all negative. This means that when travel time increases for alternative *i* and all other variables are kept constant the probability that alternative *i* is chosen decreases. The same interpretation

applies to the cost parameters; i.e., when the cost of alternative i increases, the individual is less likely to choose that alternative. We can also see that the values of travel time savings in SP1 do not differ significantly from the values in SP2, which means that the participants who were bicycling when they were recruited to the study (SP1) do not differ significantly in their appraisals from the participants (both bicyclists and non-bicyclists) who received the questionnaire by mail (SP2). The fact that the participants in SP1 were bicycling are however reflected by the large constant in that model, indicating a great preference for cycling. It is also shown in the table that the values of travel time savings in the RP model differ, at least for the bicycling time, from the other two data sets. As mentioned before, because the estimated coefficients in SP data might be too small, joint mode choice models with rescaled values according to the RP scale are preferred in forecasts.

The results from the joint mode choice model are presented in Table 2.

Table 2. Joint mode choice models for commuters with 12 km or less to work

Parameter	Estimate	t-value
Cost_car_pt	-0.057	-5.73
Time_car_pt	-0.100	-6.19
Time_walk$_{srp3}$	-0.123	-7.50
Time_bicycle$_{srp3}$	-0.167	-10.86
Time_mixed traffic	-0.228	-5.84
Time_BC lane	-0.236	-5.84
Time_BC path, close to road	-0.168	-5.81
Time_BC path, far from road	-0.158	-5.80
ASC$_{rp\_car}$	-2.08	-7.83
ASC$_{rp\_pt}$	-2.93	-7.40
ASC$_{rp\_w}$	-0.564	-1.32
ASC$_{sp1\_car\_pt}$	-5.16	-5.57
Sigma$_{sp1\_car\_pt}$	2.70	5.77
ASC$_{sp2\_car\_pt}$	-2.85	-5.11
Sigma$_{sp2\_car\_pt}$	3.58	5.69
Scale SP1	1.07	0.36
Scale SP2	1.06	0.32
Number of draws	150	
Number of observations	20503	
Number of respondents	2254	
Log-likelihood	-8091.422	
Likelihood ratio test	13023.606	
Rho-square	0.446	
Adjusted rho-square	0.445	

Note. ASC = alternative-specific constant

The coefficients for bicycle time on the bicycle paths are significantly smaller than the corresponding coefficients for cycling in mixed traffic or in a bicycle lane in the roadway, indicating that the respondents prefer cycling on safer paths.

The coefficient *time_bicycle* was estimated from the RP data. Because not all respondents have reported the infrastructure on their cycle way and we did not want to lose a lot of the observations, we did not separate the cycle time in different environments in the RP data. The coefficient is therefore an overall coefficient, including all types of infrastructure. The size of this coefficient, based on "real" data, is close in magnitude to the bicycle path coefficients based on the two stated preference-data set.

Note also in Table 2 that almost all constants are negative and significant, indicating that in almost the whole sample there is a preference for cycling and that there are additional factors except the ones included in the model that influence mode choice. The only exception is the walking constant, indicating that pedestrians in the RP data do not differ from the cyclists in the same data set. Note that we have a lot of cyclists in all our data sets and the constants must therefore be treated with great caution. For example, Wardman et al. (2007) found a preference for most of the modes over cycling, indicating that cycling was a more unpleasant mode than the others. However, because we use an incremental logit model in the forecasting the constant drops out.

By dividing each of the estimated coefficients pertaining to time with the coefficient pertaining to travel cost, we obtain values of travel time savings for each type of "time coefficient". Table 3 contains these estimates and they have all been expressed in terms of the value of saving an hour travel time.

Table 3. Values of travel time savings (95% confidence interval in paranthesis)

	SEK/h
Time car_pt	106 (89 – 123)
Time walk	130 (80 – 180)
Cycle time (RP)	176 (127 – 225)
Cycle time mixed traffic	241 (220 – 262)
Cycle time BC lane	249 (227 – 272)
Cycle time BC path, next to road	178 (161 – 194)
Cycle time BC path, far from road	167 (151 – 183)

Note. The standard errors are based on Taylor series expansion.

The estimated VTTS for walking, 130 SEK/h, was shown to be lower than that for cycling suggesting that cycling in this study is considered to be more uncomfortable. Wardman et al. (2007) obtained a similar result with an estimated VTTS for walking at 104 SEK/h[4] while the mean VTTS for cycling was 113 SEK/h (in 1999 prices). In two Norwegian studies higher VTTS for walking than for cycling were estimated: 81 SEK/h for walking and 65 SEK/h for cycling (Stangeby, 1997), and 162 SEK/h for walking and 144 SEK/h for cycling (Ramjerdi et al., 2010)[5]. In a study by Björklund et al. (2013), estimated VTTS for walking as only transport mode varied between 63 SEK/h and 214 SEK/h, depending on type of environment

---

[4] Calculated from pence/min.
[5] The values from both these studies are calculated from Norwegian crownes.

and foot path attributes. The value in the present study is in the same range which may suggest that the assumption of a walking speed of 5 km/h is not too unrealistic.

In previous analyses of data from SP1 and SP2 (Björklund & Carlén, 2012; Björklund & Mortazavi, 2013) we have divided the VTTS based on whether respondents selected car or public transport as an alternative mode of transport to bicycle in the hypothetical choices. The reason for this is that people with car as alternative travel mode tend to have much higher value of travel time, partly because of higher income but also by other, unknown, causes. In the present paper we restricted the VTTS for car and public transport to be the same since we are mainly interested in the coefficients of the model, which we then use in a general prediction model. The model is also based on three different data sets, in contrast to previous analyses where we made separate analyses for each dataset, and the values presented in this paper are therefore a sort of average over all data sets. The bicycle travel times in each environment are based only on data from SP1 and SP2, while we estimated an overall cycle journey from RP data. The value of this overall bicycle travel time saving, 176 SEK/h, is about the same as the VTTS for cycling on bike paths (167 SEK/h and 178 SEK/h). The reason that the general value of cycling is in line with the values of cycling on bike paths could either be that the majority of the cycle time is spent on bike paths (72% of the cycle time in RP was spent on bike paths) or that the estimates based on the hypothetical choices is a bit higher because of strategic bias.

5. Forecasting

Table 4 contains information on the commuting mode shares during the summer period (April to September) from the mailed-out questionnaire. We have excluded observations implying a cycle speed of less than 5 km/h or more than 30 km/h, a walking speed of more than 10 km/h, and trip times for car and public transport of more than 60 minutes. The individuals included are between 18 and 64 years old.

Table 4. Commuting mode shares in the summer period (April to September) in the RP data (made at least two times a week) and in the RVU, trips of 12 km or less.

	Car	PT	Bicycle	Walk
RP data	229 40.5%	19 3.4%	288 51.0%	29 5.1%
RVU 2011	594 47.9%	167 13.5%	311 25.1%	169 13.6%

Compared to the Swedish national survey (RVU, 2011), where the share of cyclists was 25.1%, we obviously have a lot of cyclists in our sample, and a very small share of public transport users. Note, however, that the two samples are not totally comparable. The observations in the mailed-out study are from individuals that stated that they cycle to work at least two times per week (and no other travel mode was used that often), whereas the information on mode choice from the RVU pertains to the chosen mode on the random day of measurement. The mean bicycle time for trips 12 km or shorter was 20 minutes in the RP and 16 minutes in the RVU.

In the forecasting applications, the forecast market share for bicycle $\left(P_b^f\right)$ depends on the base *(b)* market shares and the changes in utility *(ΔU)* for each travel mode. In the subsequent analyses we assume that the utilities are constant for each mode except for cycling, for which the utilities change when we change the time spent in different bicycle environments. Following Wardman et al. (2007), the incremental form of the logit model is:

$$P_b^f = \frac{P_b^b e^{\Delta U_b}}{P_b^b e^{\Delta U_b} + P_c^b e^{\Delta U_c} + P_{pt}^b e^{\Delta U_{pt}} + P_w^b e^{\Delta U_w}}$$

where the subscripts stands for bicycle (b), car (c), public transport (pt), walking (w). The base market shares which we use in this study are taken from the RVU.

For 276 persons in the mailed out-questionnaire we have data on time spent on different bicycle infrastructures. We make an assumption that these market shares are the same for the

whole sample (SP1, SP2, and RP) which is not that unrealistic because the studies are made in the same cities with the same infrastructure.[6]

On average, 24% of the cycling time was spent in mixed traffic, 2% in bicycle lanes, 42% on a bicycle path near the road way, and 31% on a bicycle path not in connection to a road way. In Table 5 the forecasted bicycle shares are based on different combinations of changes in the bicycle infrastructure. The parameters used in the forecast are rescaled SP parameters. Many of the respondents in our sample are probably biased towards cycling, i.e., the respondents in SP1 are all bicyclists and although we did not ask explicitly of cycling to work in the mailed-out study (SP2 and RP), a large part of the questions in the questionnaire concerned cycling which probably attracted more cyclists than non-cyclists. However, almost 32% of the respondents in the RP data have never cycled to work or could not think of themselves cycling to work and because we scale the SP parameters according to the RP data we have taken care of "non-changers" in the model.

Table 5. Forecast impact in some combinations of improved cycling conditions

Scenario	Cyclist share today	Cyclist share after change in infrastructure
Half of the time which today is spent in mixed traffic is in equally parts transferred to cycle path next to road and cycle path not in connection to road.*	51,0	54,9
All of the time which today is spent in mixed traffic is in equally parts transferred to cycle path next to road and cycle path not in connection to road.	51,0	58,7
All of the time which today is spent in mixed traffic and in cycle lanes is in equally parts transferred to cycle path next to road and cycle path not in connection to road.	51,0	59,4
All of the time which today is spent in mixed traffic, in cycle lanes, and in cycle paths next to road is transferred to cycle path not in connection to road.	51,0	61,3

* I.e., instead of the baseline distribution 24% in mixed traffic, 2% in cycle lane, 42% in cycle path next to road, and 31% in cycle path not in connection to road, the distribution will be in this scenario 12% in mixed traffic, 2% in cycle lane, 48% in cycle path next to road, and 37% in cycle path not in connection to road.

---

[6] In Appendix, we present forecasts based on commuting mode shares from the RVU instead. However, we have no information about infrastructure and other important variables, creating the baseline, in the RVU. This requires strong assumptions, e.g., that the differences between the RVU and our sample depends only on differences in the alternative-specific constants.

Table 5 shows some examples of changes in bicycle infrastructure and how it affects the number of cyclists under the conditions listed above. The biggest shift to cycle that may be possible on this basis is if all cycling after the change takes place on the bike path far from the road and there are no restrictions on non-changers. The proportion of cyclists would then increase from 51.0% to 61.3%, i.e. an increase of 20%.

6. Discussion and conclusions

In this study, we have estimated a mode choice model and applied it to forecast how Swedish commuters' choice of whether to cycle or not changes as a result of changes in the bicycle infrastructure. In so doing we have collected data from two surveys and the estimated mode choice model was based on a combination of self-reported data and SP data. The model includes, *inter alia*, variables that describe the cycling infrastructure. This aspect of the choice model is central to the forecasts on how cycling changes as a result of changes in cycling infrastructure. The main results of the mode choice model suggest that the VTTS are smaller for bicycle time on bicycle paths than the corresponding values for cycling in mixed traffic or in a bicycle lanes in the roadway, indicating that the respondents prefer cycling on safer paths.

When applying this model to forecast changes in the market share for the mode cycle, we found that the largest change in the market share resulted if all of the time which today is spent in mixed traffic, in cycle lanes in the road way, and in cycle paths next to road is transferred to cycle paths not in connection to road. The proportion of cyclists would then increase from 51.0% to 61.3%, i.e. an increase of 20%. In the most optimistic scenario in the study of Wardman et al. (2007), the proportion of cyclists increased from 5.8% to 9.0% at a 60% restriction on non-changers and to 13.8% without restriction. The smallest shift in the present study will take place if half of the time which today is spent in mixed traffic is transferred to cycle path next to road and cycle path not in connection to road. The proportion of cyclists would then increase from 51.0% to 54.9%, i.e. an increase of almost 8%.

It should be noted that this study presents one of the first attempts to forecast cycle share changes among Swedish commuters as a function of the cycle infrastructure. We have also imposed a number of restrictions that may be more or less important to the results. In the

following we therefore discuss these issues and also suggest some directions for future research in this field.

Firstly, to construct our base alternative concerning the bicycle infrastructure we use the average times spent on different kinds of cycle infrastructures. These averages were calculated from the answers provided by the respondents in the mailed-out questionnaire. Even if these averages do not seem completely unrealistic, it is not likely to be representative to the population of cyclists in Sweden.

Secondly, another piece of information that we use to construct the base scenario in the forecasts was the average travel time for trips by cycle of 12 km or less. This was found to be 20 minutes in one of the mailed-out questionnaires but in the National Travel Survey (RVU) this number equals 16 minutes. If, ceteris paribus, the average travel time in this study had been 16 minutes instead of 20 minutes, the increase in cyclist share would have been smaller, e.g., 16% instead of 20%, in the most optimistic scenario.

Thirdly, the proportion of new cyclists depends to a large extent on how many individuals that would consider cycling if conditions were improved. Obviously, some individuals will never choose cycling under any circumstances. Our RP data consisted of 32% respondents who state that they would never choose cycle as the mode of transport. This may be an underestimate of the corresponding share in the population. The reason is that our survey has a specific focus on cycling, so many individuals that would never consider cycling might not have responded at all to the survey.

And, finally, for reasons stated in the introduction we have restricted our study to commuters, who are in a specific age interval. This means that the potential increase of cycling among children and old people when safer cycling infrastructure is offered is not taken into consideration.

It should also be noted that the model only measures how the share of cyclists are affected by new cyclists that previously have used other travel modes: car, public transport, and walking. We have not investigated if existing cyclists will cycle more because of improved bicycle infrastructures. In addition, changing behaviour is not that easy. Even if the bicycle conditions are improved, habits are probably not changed immediately.

Furthermore, in the incremental model used to forecast bicycle shares among commuters the new cyclists are taken proportionally from each of the other travel modes. However, in reality this is probably not the case. Other studies have indicated that new cyclists probably to a large extent have switched from public transport. For example, Björklund and Mortazavi (2013), who analyze data from two of the data sets used in the present paper (SP1 and SP2), report that individuals with public transport as alternative travel mode to cycling have much lower valuations of travel time savings for cycling than persons with car as the alternative mode, suggesting that potential cyclists are likely to be found among public transport users. Rietveld and Daniel (2004) found that public transport had a low share in the Netherlands whereas the bicycle share was the highest among the European countries which according to the authors indicated that potential cyclists are likely to be found among public transport users. However, they also concluded that cycling and public transport may be complements, not only competitive transport modes. It is also unlikely, although not impossible, that new cyclists have switched from walking to cycling. Because of the small proportion of pedestrians in our study, the share of pedestrians changing to cycling in the incremental model is consequently also small.

One of the advantages with stated preference data is that preferences and valuations of infrastructure, travel modes, travel routes etc. that do not exist today can be investigated. However, because of the hypotetical nature of this kind of data it should be combined with revealed preference data, at least when forecasts are made. In the present study we combined the stated preference data with self-reported data on the respondents' commuting behaviour. This is a step in the right direction, but self-reported data are not really revealed preference data for a number of reasons like problems of recalling past events correctly. Because of the increasing interest in cycling it should perhaps be relevant in the future to collect data of bicycle infrastructure at different places, which in combination with logging the number of cyclists on the different spots and cyclists' route choices could be a very good base for future forecasts.

Acknowledgements

References

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models, Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Björklund, G. & Carlén, B. (2012). Värdering av restidsbesparingar vid cykelresor [Valuation of travel time savings in bicycle trips]. (VTI notat 26-2012). Stockholm, Sweden: Swedish National Road and Transport Research Institute.

Björklund, G., Mellin, A., & Odolinski, K. (2013). Fotgängares värderingar av restidsbesparingar vid olika typer av gångvägar [Pedestrians' valuation of travel time savings in different kinds of walking paths]. Unpublished manuscript.

Björklund, G. & Mortazavi, R. (2013). Influences of infrastructure and attitudes to health on value of travel time savings in bicycle journeys. Unpublished manuscript.

Brownstone, D., Bunch, D. S., & Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Transportation Research Part B, 34, 315-338.

Börjesson, M. & Eliasson, J. (2012). The value of time and external benefits in bicycle appraisal. Transportation Research Part A, 46, 673-683.

Cherchi, E. & Ortúzar, J. de D. (2006a). On fitting mode specific constants in the precense of new options in RP/SP models. Transportation Research Part A, 40, 1-18.

Cherchi, E. & Ortúzar, J. de D. (2006b). Use of mixed revealed-preference and stated-preference models with nonlinear effects in forecasting. Transportation Research Record, 1977, 27-34.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2002). Stated choice methods: Analysis and application. Cambridge: Cambridge University Press.

Ramjerdi, F., Flügel, S., Samstad, H. & Killi, M. (2010). Den norske verdsettingsstudien – Tid [The Norwegian valuation study – Time]. (TØI rapport 1053B/2010). Oslo, Norway: Transportøkonomisk institutt.

SOU (2012:70). Ökad och säkrare cykling – en översyn av regler ur ett cyklingsperspektiv [Increased and safer cycling – a review of regulations from a cycling perspective]. Statens Offentliga Utredningar.

Stangeby, I. (1997). Attitudes towards walking and cycling instead of using a car. (TØI report 370/1997). Oslo, Norway: Transportøkonomisk institutt.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. Transportation Research Part A, 41, 287–301.

Wardman, M., Hatfield, R., & Page, M. (1997). The UK national cycling strategy: can improved facilities meet the targets. Transport Policy 4 (2), 123-133.

Wardman, M., Tight, M. & Page, M. (2007). Factors influencing the propensity to cycle to work. Transportation Research Part A, 41, 339-350.

Appendix

In Table A1, we present forecasts based on commuting mode shares from the RVU for a mean bicycle trip of 20 minutes. However, we have no information about infrastructure and other important variables, creating the baseline, in the RVU. This requires strong assumptions, e.g., that the differences between the RVU and our sample depends only on differences in the alternative-specific constants. Given that the assumptions are fulfilled the most optimistic scenario implies a shift from 25.1% cyclists to 33.7%, i.e., an increase of 34% in comparison to the 20% increase in our sample.

Table A1. Forecast impact in some combinations of improved cycling conditions, base market shares from the RVU

Scenario	Cyclist share today	Cyclist share after change in infrastructure
Half of the time which today is spent in mixed traffic is in equally parts transferred to cycle path next to road and cycle path not in connection to road.*	25.1	28.1
All of the time which today is spent in mixed traffic is in equally parts transferred to cycle path next to road and cycle path not in connection to road.	25.1	31.4
All of the time which today is spent in mixed traffic and in cycle lanes is in equally parts transferred to cycle path next to road and cycle path not in connection to road.	25.1	32.0
All of the time which today is spent in mixed traffic, in cycle lanes, and in cycle paths next to road is transferred to cycle path not in connection to road.	25.1	33.7

* I.e., instead of the baseline distribution 24% in mixed traffic, 2% in cycle lane, 42% in cycle path next to road, and 31% in cycle path not in connection to road, the distribution will be in this scenario 12% in mixed traffic, 2% in cycle lane, 48% in cycle path next to road, and 37% in cycle path not in connection to road.