TI 2010-094/3

Tinbergen Institute Discussion Paper

Estimating the Benefits of Improved Rail Access; Geographical Range and Anticipation Effects

Hans R.A. Koster

Jos N. van Ommeren

Piet Rietveld

VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Estimating the benefits of improved rail access; geographical range and anticipation effects

Hans R.A. KOSTER, VU University Amsterdam*

Jos N. VAN OMMEREN, VU University Amsterdam

Piet RIETVELD, VU University Amsterdam

This version: September 13, 2010

Abstract

In this paper we investigate the effects of new railway stations on house prices using an extensive repeated sales dataset over a period of 13 years. We employ semiparametric panel data techniques allowing for anticipation effects of station openings. We show that a kilometre reduction in distance to the nearest railway station increases property values by about 1.5–2 percent. The geographical range of the effect is about 3.5 kilometres. Ignoring anticipation effects in the estimation procedure leads to a large downward bias for short time period datasets. We do not find any significant house price adjustment effects after station openings.

Keywords: House prices; Anticipation effects; Repeated sales; Railway development; Semiparametric estimation;

Acknowledgements
The authors gratefully acknowledge NVM for providing data. We would like to thank NICIS-KEI for financial support.

---

*Corresponding author. Department of Spatial Economics, VU University, De Boelelaan 1105 1081 HV Amsterdam, e-mail : hkoster@feweb.vu.nl.

1. Introduction

Congestion problems on the road and environmental constraints are causing a renewed interest for public transport (Cervero, 2004). In the US as well as in Europe existing public transport systems are upgraded and new lines are built. There exists a large empirical literature measuring the effects of such transport innovations. The economic benefits of these innovations are usually measured by reductions in travel time, direct user costs and accident costs, using stated-choice experiments. However, mainly because of wider economic benefits, these evaluation methods may lead to severe underestimation of the benefits of improved accessibility (Gibbons and Machin, 2005).[1] Therefore, when the economic benefits are mainly captured by households that live close to improved transport infrastructure, hedonic price methods are preferred, because they aim to monetarise preferences of households for neighbourhood attributes. See also Bayer et al. (2007), who argue that for continuous attributes such as distance to railway stations, hedonic price approaches, rather than discrete choice approaches, are a good way to estimate mean willingness to pay (WTP) for attributes.

A general conclusion in the literature is that households are willing to pay a premium to live close to public transport stations as proximity to stations implies shorter travel times (see Gatzlaff and Smith, 2003; Baum-Snow and Kahn, 2000; and the meta-analysis of Debrezion et al., 2007). However, most studies are cross-sectional, which probably leads to biased estimates, as the effects of stations are correlated with unobserved spatial factors. Recent contributions try to avoid the bias inherent to cross-sectional studies by using repeated sales prices (McMillen and McDonald, 2004; Grimes and Young, 2010) or employing a difference-in-difference methodology based on openings of stations (Gibbons and Machin, 2005).

It is well known that people adjust their current behaviour to events that will take place in the future (see Flavin, 1981). When a government announces to improve local infrastructure, this may increase residential property values as future commute times are reduced and more leisure time is available. This will be

[1] For example, improved transport access may reduce frictions in the labour market, increase the intensity of knowledge spillovers, and lower input costs. But also reductions in (local) environmental externalities could be a benefit of transport innovations.

particularly so for railway stations openings, which imply a clear and discrete change in infrastructure supply. Anticipation effects occur when people have information on government plans about stations openings. When effects of stations' openings are identified based on differences over time, as is common in repeated sales prices, the presence of anticipation effects of stations' opening will lead to an underestimate. One difficulty that arises in estimation procedures based on differences over time is that these plans are usually announced several years in advance. So, taking anticipation effects into account seems fundamental, but is not yet included in repeated sales models. Cross-sectional studies by Damm et al. (1980), McDonald and Osuji (1995), Knaap et al. (2001), Bae et al. (2003) and Tsutusmi and Seya (2008) indeed suggest that plans for (light) rail development have a positive effect on land values in proposed station areas.

In this paper we investigate the impacts of railway proximity on house prices in Dutch cities between 1995 and 2007 based on a repeated sales sample. We focus on openings of small stations on existing tracks that are announced only two or three years before the actual opening of the station, which makes it easier to identify anticipation effects. In particular with large stations, and new tracks, anticipation effects may be difficult to deal with as stations' openings are announced decades in advance.[2]

Our paper improves on the existing literature in three important aspects. First, we use an extensive repeated sales dataset (more than 50,000 observations), which allows us to examine house price differences at the level of the individual property. We therefore avoid any bias in the estimates that may arise when using cross-sectional or area-aggregated panel data. Second, we shed some light on the magnitude of anticipation effects and demonstrate that it is important to have data over a long time period. Third, we investigate the effects of distance to station in a semiparametric way, using the two-stage procedure proposed by Robinson (1988). Ekeland et al. (2004), among others, argue that the relationship between house prices and various attributes is complex and nonlinear. It is therefore preferable to use nonparametric or semiparametric specifications to avoid arbitrary functional form assumptions. Furthermore, using these more flexible specifications we are able to determine the geographical range of the effect of stations on

[2] Additionally, the construction of large new stations may involve many years and negative externalities related to the construction (e.g. noise pollution), which makes it hard to identify the real anticipation effects of transport innovations.

property values. Some previous studies report that the effect diminishes after 1 to 2 kilometres (Weinberger, 2001; Gibbons and Machin, 2005; Ossokina, 2010).

This paper proceeds as follows. In Section 2, we discuss the study's context and present the data. Section 3 considers the econometric methodology. In section 4 we present and discuss the result. Section 5 concludes.

2. Rail innovations and data

2.1 Rail innovations in the Netherlands

The rail network of the Netherlands is one of the densest in the world (Claessens et al., 1998). Until 1930 the network has been expanded very quickly. The Dutch railway network was exploited by a large number of different private companies. From 1930 onwards, competition by bus and car led to the closure of a large number of lines and stations (Figure 1). Since the Second World War the network has been mainly upgraded in terms of efficiency and capacity. Today, the network consists of about 2800 kilometres of track.[3] It is mainly used for passenger rail services and almost 1,000,000 people are transported each day.

We focus on the effects of new stations built on existing railroad tracks between 1995 and 2007.[4] These new stations are constructed in cities that have 90,000 to 300,000 inhabitants. Furthermore, the new stations are located in the suburbs of these cities, have comparable railway service levels and provide stops for local commuter trains only. So, these stops may particularly be important for commuters who live in the suburb and work in city centres. Figure 2 presents a map of the cities and the new stations.

[3] Since the 1990s, no major network expansions have taken place, except for a high speed rail connection between Amsterdam and the Belgium border and a freight line between the port of Rotterdam and the Ruhr area in Germany.
[4] This is exactly half of the number of stations opened during that period. The other half are on new tracks or in areas with low population densities where identification of the effects of stations openings is more problematic as the number of observations close to stations is limited.

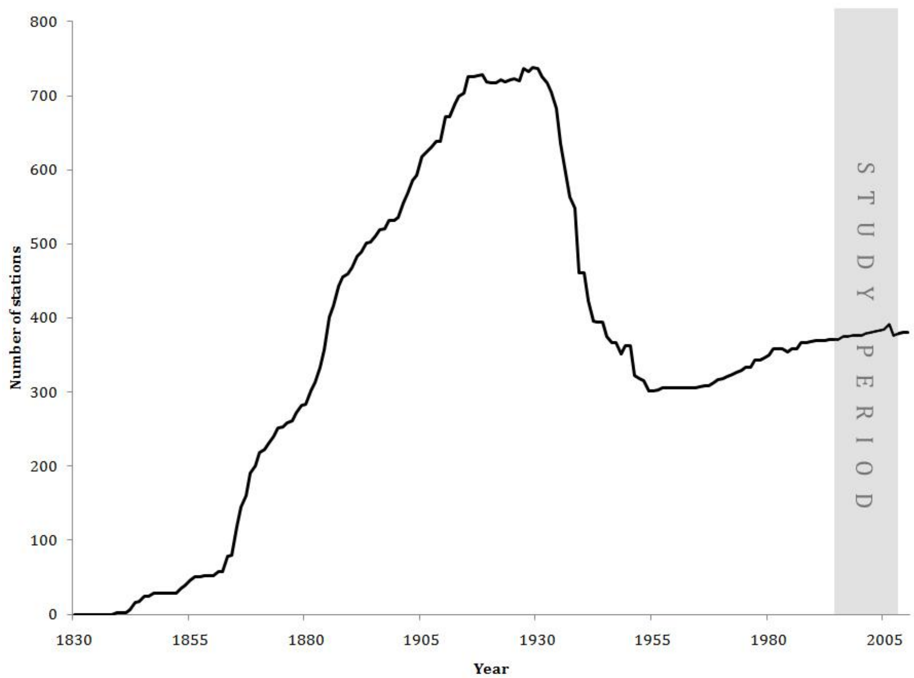

Figure 1: Number of Stations between 1830-2010

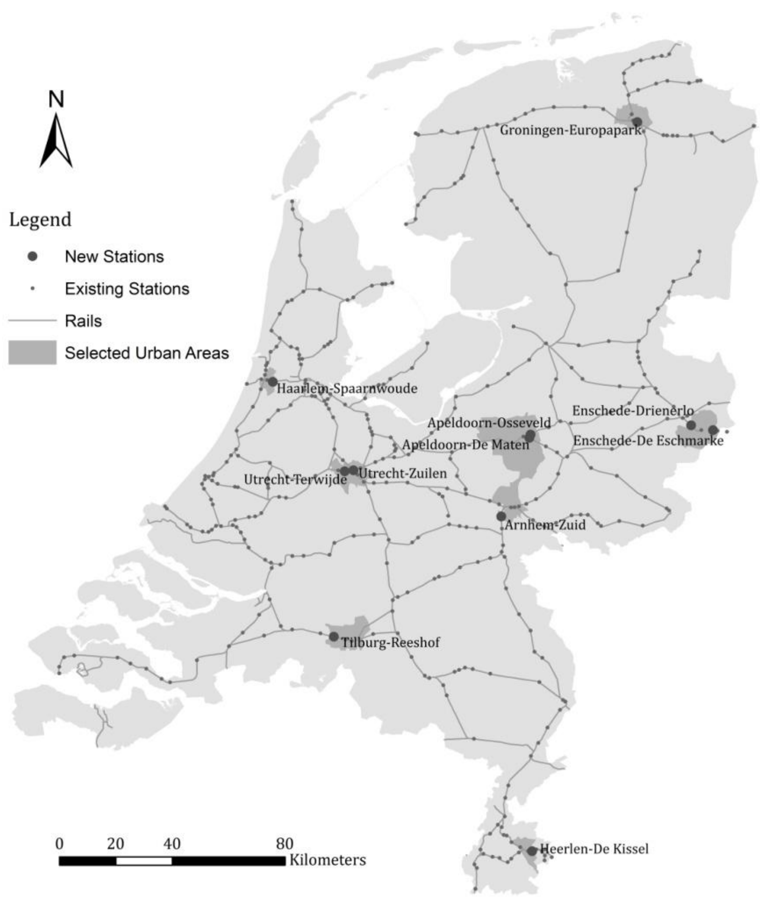

Figure 2: Map of The Netherlands with selected cities and new stations

2.2 Data

Our hedonic price analysis is based upon a house transactions dataset from the NVM (Dutch Association of Real Estate Agents) for those cities where new stations were opened (see Figure 2). It contains information on about 80% of all transactions in these cities between 1995 en 2007. For 157,833 transactions, we know the transaction price, the exact address, and a wide range of house attributes such as size (in square meters), type of house, number of rooms and construction year.[5] For each property we calculate the distance to the nearest railway station using GIS-software. We also assemble neighbourhood level data to correct for aspects of neighbourhoods that change over time.[6] Variables that are included are population density, the share of ethnic minorities, people younger than 25 years and older than 65 years. For 1996, 1998, 2000 and 2002 we do not have information on these attributes, so for these years we impute the average values of the preceding and following year. Most studies that investigate the effects of transport innovations do not correct for changes in neighbourhood attributes, but it may be that it is important to correct for local time-varying variables, because the opening of a station may be correlated with these attributes.

About 35% of the transactions are repeated sales, which we will use in our analysis.[7] The repeated sales dataset consists of 55,823 transactions of 25,270 residential properties.[8] Most properties in this dataset (82%) are transacted twice. Table 1 presents summary statistics for price (in €) and distance to station (in km) before and after a transaction in the same city, as well as their changes over time. The average yearly increase in prices is 9 percent. Only about 10 percent of properties have experienced a reduction in station

[5] We exclude transactions with prices that are above € 1.5 million or below € 25,000 or a square meter price below € 250 or above € 5000. Furthermore, we exclude transactions that refer to properties smaller than 25m^2 or larger than 300m^2. Old houses are sometimes demolished and replaced by new ones, so we delete observations which refer to properties that have changed type of house, or for which the size has changed more than 20m^2.

[6] Neighbourhoods are fairly small: the average distance to the centroid of a neighbourhood is only 286 meter.

[7] Appendix A presents a table with means and standard deviations of the repeated sales sample. We also have compared the descriptives of the repeated sales sample to the full sample and it appears that the means and standard deviations are very similar.

[8] The average distance of a property to the centroid of a city is small and is only 3 kilometres. It would also be an option to select housing transactions within a given distance of a new station (rather than all properties within cities where new stations are opened). However, the samples then largely overlap and the total number of observations will hardly change. For example, when we consider a maximum distance of 4 kilometres, our sample would consist of 51,428 observations.

distance.[9] Therefore, the average reduction in distance to station is low and only 110 meter. For a complete overview of all variables we refer to Table A2 in Appendix A.

Table 1: Descriptive statistics of repeated sales sample before and after transactions

Price (in €)			Distance to Station (in km)		
Before	After	Annual Change (%)	Before	After	Change
127,421	174,652	0.089	1.702	1.592	-0.110
(66,280)	(80,499)	(0.070)	(1.003)	(0.921)	(0.489)

NOTE: Standard deviations are between parentheses. The change in price is the compound annual growth rate.

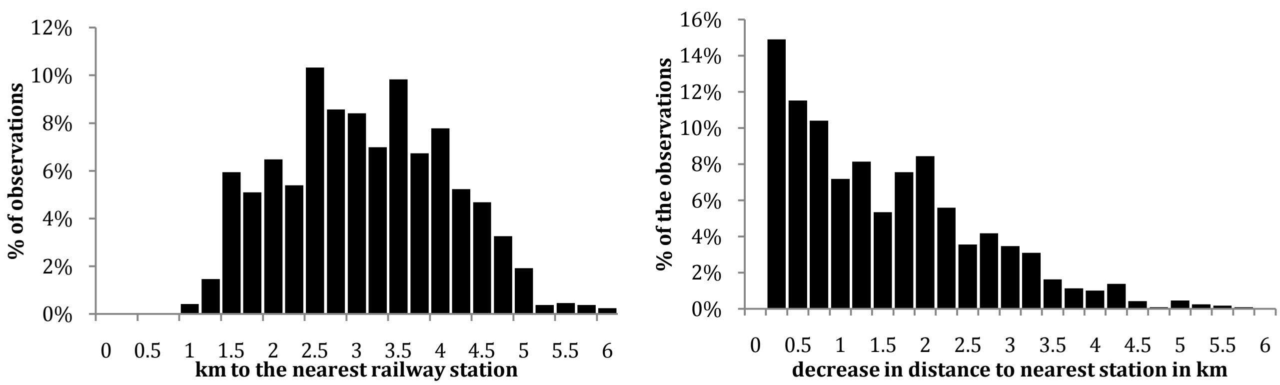

Figure 3: Distance before station openings and distance reductions of properties to railway stations

Figure 3 presents histograms *only* for properties that experienced reductions in distance to stations. Most of these properties are located within 5 kilometres of a railway station (before the opening). Note that properties already located within one kilometre of a station, never experience a reduction in distance. This range of distances is economically meaningful because previous studies have shown that about 80 percent of trips to get to stations originate from locations within 5 kilometres of the station with an average of about 2 kilometres (Keijer and Rietveld, 2000). For these properties, the average distance decrease is 1.40 kilometre with a median of 1.18 kilometre. Other descriptive statistics can be found in Appendix A.

[9] Given the presence of other stations, note that a station opening in a city does not reduce the distance to a station for most properties.

3. Repeated sales models

3.1 Hedonic price function and anticipation effects

A large literature discusses identification problems that may arise in hedonic price models (e.g. Brown and Rosen, 1982; Brown, 1983; Bartik, 1987; Epple, 1987; Bajari and Benkard, 2005). Ekeland et al. (2004) recently showed that identification problems mainly arise because of linearisations which simplify the estimation procedure. Nonparametric estimation methods outperform parametric procedures in describing nonlinear relationships between the price of a house and its attributes. However, a property is generally a function of a many observed attributes. This raises a fundamental issue because unconstrained nonparametric estimation becomes infeasible when the number of independent variables increases, which is known as the 'curse of multidimensionality' (Bontemps et al., 2008). So, constraints on the hedonic price function are needed. Robinson (1988) proposes to estimate a partial linear model to overcome this problem and introduces a two-stage estimation procedure.[10] In the first stage, the linear part of the hedonic price function is estimated, followed by an estimation of the nonparametric part in the second stage.

We assume that price p_{ht} of house h is a function of time-varying x_{ht}, z_{ht} and time-invariant attributes y_h, where t denotes the time period. Let x_{ht} and y_h be attributes which are linearly related to the logarithm of the house price and z_{ht} denote the distance to a station which is nonlinearly related to the house price. Then, the price function can be represented as follows:

$$\log(p_{ht}) = x'_{ht}\beta + y'_h\gamma + m(z_{ht}) + \epsilon_h + \xi_{ht}, \qquad (1)$$

where $m(\cdot)$ is an unknown function of z, β's and γ's are parameters to be estimated, ϵ_h denotes property-specific time-invariant unobserved components and ξ_{ht} is an unobserved component that varies over time. We will distinguish between a fully parametric and semiparametric estimation procedure.

In this study, we explicitly model anticipation effects of station openings on house prices. For now, we assume that there is a linear effect of distance on house prices and that this effect depends on the time between the property transaction and the station opening denoted by τ_{ht}. So, we multiply the distance to

[10] Robinson (1988) demonstrates that the coefficients can be estimated at parametric rates of convergence, despite the presence of a nonparametric part, so they are (surprisingly) efficient.

station variable with $(\tau^*_{ht}+1)$, where τ^*_{ht} denotes the *latent* difference between the opening year of the new station and the property transaction year. When the transaction occurs after the opening year, τ^*_{ht} equals zero.

In the Netherlands, an opening of a station is sometimes announced more than ten years before the actual opening. However, for stations built on existing tracks, the exact location as well as the exact opening date is uncertain up to two or three years before the opening. We therefore expect that the market does not adjust up to two or three years before opening of stations. So, the maximum value of τ_{ht}, denoted by $\tau^{\max}_{ht}$, has to be bounded. If $\tau_{ht} > \tau^{\max}_{ht}$, then $\tau_{ht} = \tau^{\max}_{ht}$, otherwise $\tau_{ht} = \tau^*_{ht}$. In the absence of anticipation effects, $\tau_{ht} = \tau^{\max}_{ht} = 0$.

3.2 Fully parametric fixed effects estimation procedure

The fully parametric estimation starts from the assumption that $m(z_{ht})$ is equal to $z_{ht}(\tau_{ht}+1)\delta$, where δ is a parameter to be estimated. One way to deal with time-invariant unobserved attributes is to use postcode area fixed effects to estimate the effects of stations on property values.[11] This essentially deals with all variation between areas (e.g. quality of housing stock) but assumes that the distribution of unobserved differences of sampled properties within postcodes does not vary over time, or at least is uncorrelated with changes in house prices.[12] When this does not hold, repeated sales on the level of the individual property are preferred. We therefore use fixed effects at the level of the individual property. The fixed effects specification yields:[13]

$$\log\widetilde{(p_{ht})} = \widetilde{x_{ht}}'\beta + \left(z_{ht}\widetilde{(\tau_{ht}} + 1)\right)\delta + \widetilde{\xi_{ht}}. \qquad (2)$$

where the symbol ~ denotes mean-differenced variables of a house.

[11] A postcode contains on average 25 households in The Netherlands. The average distance of a property to the centroid is only 40 meters.

[12] This assumption may not hold for two reasons. The first reason is that the economic value of certain unobserved house attributes may change (e.g. sufficient parking on street may be important for those without private parking space, but less so when the house is close to a station). The second reason is that sales turnover of houses with specific unobserved attributes (e.g. sufficient parking on street) may change over time.

[13] We note that this is not a time-differenced form as in Gibbons and Machin (2005). As almost all properties are transacted twice, this difference in specification is not essential.

3.3 Semiparametric fixed effects estimation procedure

To estimate semiparametric fixed-effects we employ local linear methods (see Fan and Gijbels, 1996). As argued by Bajari and Kahn (2005), local linear methods possess a number of desirable theoretical and practical properties compared to other nonparametric estimation techniques. For example, local linear methods have a lower asymptotic bias than the Nadaraya-Watson and a lower asymptotic variance than the Gasser-Mueler estimator. For each observation, we estimate locally weighted least squares using a Gaussian kernel. We do not impose that the coefficients of distance to station are the same for all households, but these are allowed to vary with the distance to a station.[14] An important parameter of the kernel is the bandwidth. Because Silverman's rule of thumb, which is given by $1.06n^{(-1/6)}$ where n is the number of observations, tends to undersmooth kernel, the bandwidth is set 2.5 times this value (see also Silverman, 1986; Bishop and Timmins, 2008).[15] Including fixed effects in nonlinear models is not standard. However, one may linearise these models (and then include these effects). One way of doing this, is by taking a first-order Taylor series expansion of $m(\cdot)$ around some vector h^* (see Bishop and Timmins, 2008; Ullay and Roy, 1998):

$$\log(p_{ht}) = x_{ht}'\beta + y_h'\gamma + m(h^*) + (z_{ht} - h^*)'m'(h^*) + \epsilon_h + \xi_{ht}. \tag{3}$$

where h^* denotes the property of interest. We use the mean-differenced characteristics of a house h. Then:

$$\widetilde{\log(p_{ht})} = \widetilde{x_{ht}}'\beta + (\widetilde{z_{ht}})'m'(h^*) + \widetilde{\xi_{ht}}. \tag{4}$$

We employ the procedure of Robinson (1988) to estimate (4).[16]

---

[14] So, kernel weights are a function of the distance to the nearest stations. It implies that when two houses are located at the same distance to a station, the coefficients of these houses are the same.

[15] Later on, we will provide a robustness check for the choice of bandwidth. Conventional methods to apply the bandwidth, such as cross validation and minimizing an improved Akaike information criterion, are too computationally intensive, given our data (see Pagan and Ullah, 1998; Hurvich et al., 1998; Kondo and Lee, 2003).

[16] A consistent estimator of the β's is obtained by first regressing $\widetilde{\log(p_{ht})}$ and $\widetilde{x_{ht}}$ on $\widetilde{z_{ht}}$ nonparametrically. This allows us to calculate the residuals $\varrho_{\widetilde{p_{ht}}} = \widetilde{\log(p_{ht})} - E\left[\widetilde{\log(p)}\,|\tilde{z} = \widetilde{z_{ht}}\right]$ and $\varrho_{\widetilde{x_{ht}}} = \widetilde{x_{ht}} - E[\tilde{x}|\tilde{z} = \widetilde{z_{ht}}]$. We then regress $\varrho_{\widetilde{p_{ht}}}$ on $\varrho_{\widetilde{x_{ht}}}$ to obtain a $\sqrt{N}$-consistent estimator of the β's. The next step in the estimation procedure is to estimate the nonlinear part of the hedonic price function $m(\cdot)$. We regress the residual $W_{ht} = \widetilde{\log(p_{ht})} - \widetilde{x_{ht}}\hat{\beta}$ on $\widetilde{z_{ht}}$ nonparametrically, employing local linear methods.

4. Empirical results

4.1 Parametric results

Table 2 presents the results of the parametric specification (2) based on the repeated sales sample. In all models we include dummy variables that capture interactions of city with transaction year. These dummy variables account for unobserved annual changes in the local environment. In Model (1) we do not account for anticipation effects. In the other two models we do, and we set $\tau^{\max}$ respectively to 2 and 3 years. One may argue that it is not likely that the anticipation effect is linear over time. We therefore present another specification (Model 4), where we interact distance to station with dummy indicators for different values of τ, denoted by τ_{dummy}.

Table 2: Estimation results of fully parametric specification of the hedonic price function (repeated sales)

	MODEL (1)			**MODEL (2)**			**MODEL (3)**			**MODEL (4)**		
	$\tau^{\max} = 0$			$\tau^{\max} = 2$			$\tau^{\max} = 3$			$\tau^{\max} = 3$		
	No anticipation effect			*Linear anticipation effect*			*Linear anticipation effect*			*Nonlinear anticipation effect*		
Size in m² (log)	0.149	(0.016)	***	0.148	(0.016)	***	0.148	(0.016)	***	0.148	(0.016)	***
Rooms	0.012	(0.002)	***	0.011	(0.002)	***	0.011	(0.002)	***	0.011	(0.002)	***
Garage	0.015	(0.009)	*	0.015	(0.009)		0.015	(0.009)		0.015	(0.009)	
Central Heating	0.180	(0.005)	***	0.180	(0.005)	***	0.180	(0.005)	***	0.180	(0.005)	***
Listed Building	0.005	(0.013)		0.006	(0.013)		0.006	(0.013)		0.006	(0.013)	
Population density (log)	0.038	(0.020)	*	0.039	(0.020)	*	0.039	(0.020)	*	0.040	(0.019)	**
Share of Ethnic Minorities	0.025	(0.001)	***	0.025	(0.001)	***	0.025	(0.001)	***	0.025	(0.001)	***
Share <25 years	-0.021	(0.002)	***	-0.021	(0.002)	***	-0.021	(0.002)	***	-0.021	(0.002)	***
Share >65 years	-0.025	(0.002)	***	-0.025	(0.002)	***	-0.025	(0.002)	***	-0.025	(0.002)	***
Distance to station$\cdot(1+\tau)$	-0.016	(0.005)	***	-0.005	(0.001)	***	-0.004	(0.001)	***			
Distance to station$\cdot(1+\tau_{dummy})$	No			No			No			Yes		
Property FE (25270)	Yes			Yes			Yes			Yes		
City·year FE (102)	Yes			Yes			Yes			Yes		
Effects for distance to station in year τ before opening												
Distance to station: $\tau = 0$	-0.016	(0.005)	***	-0.005	(0.001)	***	-0.004	(0.001)	***	-0.004	(0.009)	
Distance to station: $\tau = 1$				-0.009	(0.003)	***	-0.008	(0.002)	***	0.009	(0.009)	
Distance to station: $\tau = 2$				-0.014	(0.004)	***	-0.012	(0.003)	***	-0.001	(0.007)	
Distance to station: $\tau = 3$							-0.017	(0.004)	***	-0.020	(0.005)	***
Distance to station: $\tau^{\max}$	-0.016	(0.005)	***	-0.014	(0.004)	***	-0.017	(0.004)	***	-0.020	(0.005)	***
Number of observations	55,823			55,823			55,823			55,823		
R^2-within	0.7815			0.7816			0.7817			0.7820		

*NOTES: The dependent variable is $\log(p_{ht})$. Robust standard errors are between parentheses and are adjusted for clustering of municipality and transaction year interactions. Coefficients are significant on *0.10, **0.05, and ***0.01 levels.*

In Model 1, we do not account for adjustments of house prices before the station's opening. We find that a one-kilometre increase in distance to the station will lower the house price with 1.6 percent. The result is comparable to previous studies that use other methodologies, which found effects ranging from 1.5 to 6 percent. Because we only investigate the effects of relatively small stations with low service levels, our estimates are expected to be on the lower end of the spectrum. The magnitude of the effect seems quite realistic. Let's assume a property with an average value of € 150,000. The annual ownership costs are about 4 percent, so € 6,000.[17] The annual WTP for a kilometre reduction in distance to station is then about € 100. Let's furthermore assume that one person per household drives per bicycle to the station every workday with a speed of 10 kilometres per hour.[18] There are about 200 workdays a year, so the implied value of time is € 2.50 (about 25 percent of the net hourly wage), which is somewhat low, but corroborates previous empirical research (e.g. Mackie et al., 2003).

When we account for anticipation effects (Models (2)-(4)), the results hardly change, irrespective of how anticipation effects are modelled. The robustness of our result with respect to the inclusion of anticipation effects is not surprising, given our data. About 75 percent of observations with changes in distance are sampled at least two years before the station opening, so any bias of ignoring anticipation effects must be small. However, our conclusion may not apply to studies that use a much shorter time period (or for which the announcement period is likely much longer). To investigate how large the bias would have been in our study when we only would have had observations for a short time period, we have re-estimated Model (2), whereas we exclude observations referring to transactions more than two years before the station opening. For this sample, the station effect is only 0.8 percent, which underestimates the true effect by about 50 percent.

In Model (4), we interact τ and distance to station, which leads to similar results. The results suggest that an announcement of a station leads to a sudden increase in property values, as is suggested by economic

[17] This is lower than in for example the US (see Gyourko and Tracy, 1991), because home-owners are allowed to deduct the full costs of the mortgage from their taxable income and marginal income taxes are higher (about 50%)
[18] In the Netherlands, the bicycle plays a large role as an access mode and accounts for about 35 percent of the access trips (Rietveld, 2000), and likely even higher in our sample.

theory. This may explain the lack of any effect for low values of τ in this specification. It may also be difficult to estimate the small effect of stations on property values for low values of τ, due to a lack of data.

The coefficients of the control variables appear to be robust to specification. We find intuitive signs for the control variables, which are linearly related to the house price. A ten percent increase in size of a house leads to a 1.5 percent increase in the house price. An increase in the number of rooms has a slightly positive impact on property values.[19] The neighbourhood variables have a statistically significant impact on housing values. Population density is positively related with house prices: ten percent increase in population density leads to an increase in house price of 0.38 percent. However, this effect is only marginally significant. People are willing to pay slightly more for neighbourhoods with a higher share of minorities, suggesting more ethnic diversity. As is argued by Jacobs (1961) and Florida (2002), people may value cultural diversity, because it represents viability of urban environments. Households prefer to live in neighbourhoods with high shares of people between 25 and 65: a ten percent point increase in the share of young and old people leads to a decrease in the house price of respectively 0.21 and 0.25 percent.

4.2 Robustness checks for parametric regression

In this subsection we provide some robustness checks by excluding neighbourhood variables, by relaxing assumptions regarding $\tau^{\max}$, by including postcode fixed effects instead of property fixed effects and by relaxing the assumption that after a station opening the price does not adjust anymore. We present the most relevant results in Table 3.

When we exclude neighbourhood variables, we find that the effect of stations is a factor 2.75 larger: one kilometre increase in distance to the station leads to a decrease in house prices of 4.7 percent (Model 5). So, it appears that changes in distance to station are correlated with changes in these attributes. Stations' openings may be accompanied by changes in ethnic and age composition of the population, as well as

[19] Note that a central heating dramatically increases house prices with 18 percent. Likely, installation of a central heating goes hand in hand with other even more fundamental improvements, so this effect should not be interpreted as a causal effect of central heating on price.

changes in density. In the Netherlands there are indeed examples of local governments that decide to build a nursing home for elderly or increase population density together with stations' openings. We see that these neighbourhood variables have an impact on housing values, but are correlated with stations' openings. If these changes in neighbourhood variables are not induced by the new station opening, excluding these variables would then lead to an upward bias. However, if these changes are fully induced by a station opening, the estimates provided by Model (5) may be more accurate. More likely however, the estimates of Model (5) are overestimates, and those of Model (3) are somewhat conservative.

Table 3: Estimation results of fully parametric specification of the hedonic price function (repeated sales)

	MODEL (5)			**MODEL (6)**			**MODEL (7)**			**MODEL (8)**		
	$\tau^{\max} = 3$ *Neighbourhood variables excluded*			$\tau^{\max} = 5$			$\tau^{\max} = 3$ *With postcode fixed effects*			$\tau^{\max} = 3$ *With ex-post adjustment effects*		
Effects for distance to station in year τ before opening												
Distance to station: $\tau \leq$ -2										-0.009	(0.010)	
Distance to station: $\tau =$ -1										-0.009	(0.010)	
Distance to station: $\tau = 0$	-0.012	(0.002)	***	-0.004	(0.001)	***	-0.001	(0.001)	*	0.004	(0.011)	
Distance to station: $\tau = 1$	-0.023	(0.003)	***	-0.007	(0.002)	***	-0.003	(0.002)	*	0.007	(0.009)	
Distance to station: $\tau = 2$	-0.035	(0.005)	***	-0.011	(0.002)	***	-0.004	(0.002)	*	-0.002	(0.007)	
Distance to station: $\tau = 3$	-0.047	(0.006)	***	-0.015	(0.003)	***	-0.006	(0.003)	*	-0.019	(0.004)	***
Distance to station: $\tau = 4$				-0.018	(0.004)	***						
Distance to station: $\tau = 5$				-0.022	(0.005)	***						
Distance to station: $\tau^{\max}$	-0.047	(0.006)	***	-0.022	(0.005)	***	-0.006	(0.003)	*	-0.019	(0.004)	***
Number of observations	55,823			55,823			55,823			55,823		
R^2-within	0.6309			0.7818			0.8017			0.7821		

NOTES: See Table 2. We include the same control variables as in Models (1)-(4), except for Model (5).

It may be that our assumptions concerning $\tau^{\max}$ are not correct. For example, the market may already adjust to rather vague announcements about station openings, although full information on the location and opening data of the new station is still absent. We therefore set $\tau^{\max} = 5$ and impose a linear time structure. Although the effect is somewhat higher, it appears that it is similar to the effect found in Model (3). So, we may conclude that our results are robust to the choice of $\tau^{\max}$.

We also have estimated a specification where we include postcode fixed effects instead of property fixed effects (Model (7)).[20] Again we find that distance to stations is negatively related to house prices, although the effect is much smaller compared to Model (3). Apparently, de effect of stations is correlated with unobserved changes in the local environment.

Housing markets generally are not functioning perfectly. In housing markets, delayed reaction to new information is thought to be standard. After the opening of a station, prices may still adjust. Our (implied) assumption up to now is that there is no effect of stations on property values after the station opening. This assumption may lead to underestimates. We therefore re-estimate Model (4) but we also include interactions of distance to station with dummy indicators for values of $\tau = -1$ and $\tau = -2$, so we allow for the possibility that the market adjusts up to two years after the station opening (Model (8)). In Table 3 we observe that this adjustment effect is small and not statistically significant. Because ex-post adjustment effects do not play a role, we will ignore them in the remainder of the paper.

4.3 Results from semiparametric regression

In Table 4, the results of the linear part of the semiparametric hedonic price function are presented. The coefficients of the nine control variables are almost identical to those presented in Table 2, which increases our confidence in the estimation procedures. The mean effects for distance to stations are also very similar to those presented in Table 2: a one kilometre increase in distance to a station leads to a decrease in the house price of about 1.7 percent. This result holds irrespective of the value of $\tau^{\max}$. This corroborates the study of Bayer et al. (2007), who find that standard hedonic price regressions reflect mean preferences. Figure 4 visually represents the effects of distance to station for different models.[21]

[20] A number of previous studies based on time differences use aggregate data at the level of the postcode. This is correct given the additional assumption that the distribution of unobserved attributes of sampled properties within the postcode area is constant.

[21] We report the results for observations which are located within a 4 kilometre radius of a station, which is about 97.5 percent of the data. Because the density of data points is relatively low in the upper 2.5 percent, the estimates are somewhat unreliable.

Distance to station is negatively related to house prices for more than 95 percent of the observations. We see that the effect also is nonlinear and is maximally 3 percent in absolute values (Figure 4). The results imply that households that reside close to stations are willing to pay most for a kilometre reduction in distance to the station. After 3.5 kilometres the effect is essentially zero, possibly the likelihood of using the station becomes small. We find a larger geographical range than studies in other countries, which may be related to the popularity of the bicycle. Bicycle use reduces the travel time from homes to stations, compared to walking.

Table 4: Estimation results of semiparametric specification of the hedonic price function (repeated sales)

	MODEL (9) $\tau^{\max} = 0$			**MODEL (10)** $\tau^{\max} = 2$			**MODEL (11)** $\tau^{\max} = 3$		
Size in m² (log)	0.148	(0.011)	***	0.148	(0.011)	***	0.148	(0.011)	***
Rooms	0.012	(0.001)	***	0.011	(0.001)	***	0.011	(0.001)	***
Garage	0.015	(0.008)	*	0.015	(0.008)	*	0.015	(0.008)	*
Central Heating	0.180	(0.003)	***	0.180	(0.003)	***	0.180	(0.003)	***
Listed Building	0.005	(0.012)		0.005	(0.012)		0.005	(0.012)	
Population density (log)	0.037	(0.008)	***	0.039	(0.008)	***	0.039	(0.008)	***
Share of Ethnic Minorities	0.025	(0.000)	***	0.025	(0.000)	***	0.025	(0.000)	***
Share <25 years	-0.021	(0.000)	***	-0.021	(0.000)	***	-0.021	(0.000)	***
Share >65 years	-0.025	(0.001)	***	-0.025	(0.001)	***	-0.025	(0.001)	***
Distance to station$\cdot(1+\tau)$	$m'(\cdot)$			$m'(\cdot)$			$m'(\cdot)$		
Property FE (25270)	Yes			Yes			Yes		
City·year FE (102)	Yes			Yes			Yes		
Average total effects of distance to stations									
Distance to station$\cdot(1+\tau^{\max})$									
Average effect	-0.017			-0.016			-0.018		
S.D. of total effect	0.009			0.006			0.006		
Number of observations	55,823			55,823			55,823		

NOTES: See Table 2.

We verified whether our results are robust to the choice of bandwidth. Our reference bandwidth was equal to 2.5 times the Silverman's rule of thumb. We have re-estimated all models for a relevant range of bandwidths, but it appears that the choice of bandwidth hardly impacts our main conclusions: the effect of stations is the largest for properties located close to stations and converges to zero from 3.5 kilometres onwards. In Appendix B we present two additional figures for different bandwidths.

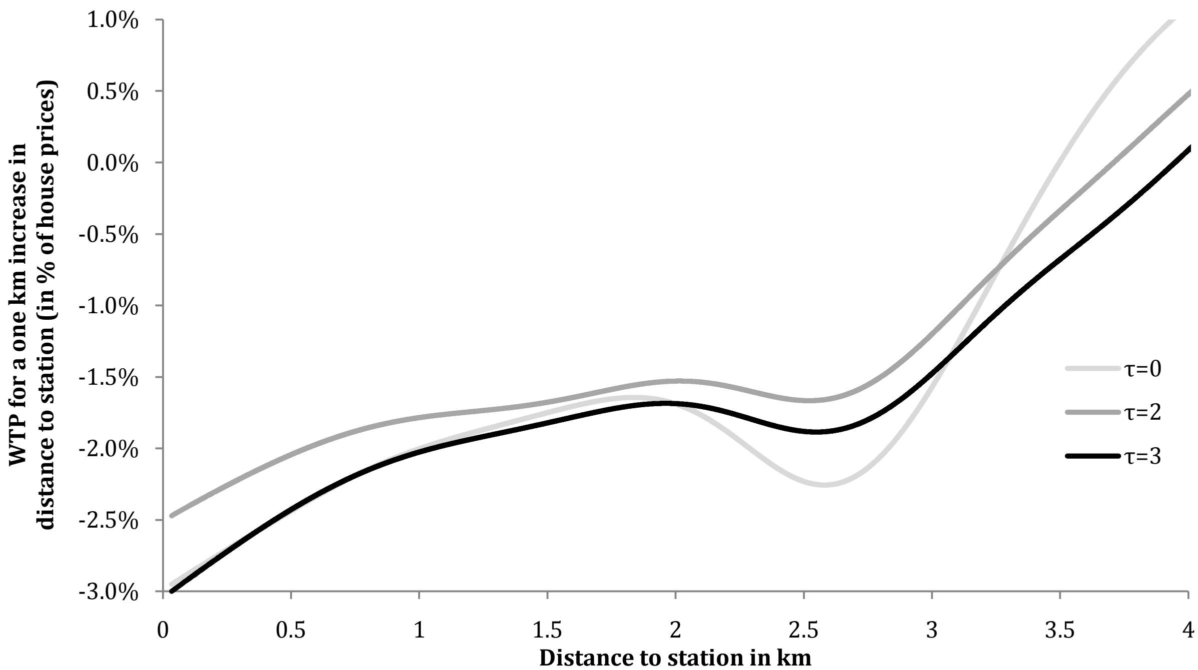

Figure 4: The effect of station proximity on property values

5. Conclusions

In the US, as well as in Europe, existing public transport network are expanded and upgraded. In the current paper, we investigated the WTP for distance to stations by employing hedonic price methods. Using an extensive repeated sales dataset over a period of 13 years, we were able to deal with potential biases, including time-invariant heterogeneity and anticipation effects, inherent to previous studies. The average willingness to pay for a kilometre reduction in distance to station is about 1.5–2 percent of the house price. Our research confirms results of previous studies that found that households place significant value on improvements in transport access, measured by reductions in distance to the nearest station. We show that anticipation effects are important and bias the results up to 50 percent for short time period datasets that exclude observations before the announcement of station openings. In contrast, price adjustment effects after station openings are too small to detect. Using semiparametric techniques, we were able to determine the geographical range of the effect of stations: the WTP for a kilometre reduction in distance to stations on house prices is essentially zero after 3.5 kilometres distance from the station.

References

Bajari, P., Benkard, C.M. (2005). Demand Estimation with Heterogeneous consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy 113(6): 1239-1276.*

Bajari, P., Kahn, M.E., (2005). Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. *Journal of Business & Economic Statistics 23(1): 20-35.*

Bartik, T.J. (1987). The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy 95(1): 81-88.*

Bayer, P., Ferreira, F., McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighbourhoods. *Journal of Political Economy 115(4): 588-638.*

Baum-Snow, N., Kahn, M.E. (2000). The Effects of New Public Projects to Expand Urban Rail Transit. *Journal of Public Economics 77: 241-263.*

Bishop, K., Timmins, C. (2008). Simple, Consistent Estimation of the Marginal Willingness to Pay Function: Recovering Rosen's Second Stage Without Instrumental Variables. *Working Paper.*

Bontemps, C., Simioni, M., Surry, Y. (2008). Semiparametric Hedonic Price Models: Assessing the Effects of Agricultural Nonpoint Source Pollution. *Journal of Applied Econometrics 23: 825-842.*

Brown, J.N., Rosen, H.S. (1982). On the Estimation of Structural Hedonic Price Models. *Econometrica 50(3): 765-768.*

Brown, J.N. (1983). Structural Estimation in Implicit Markets. *In: Triplett (ed.). The Measurement of Labour Cost. Chicago University Press, Chicago.*

Cervero, R. (2004). Transit-Oriented Development in the United States: Experiences, Challenges and Prospects. *TCRP Report 102, Transportation Research Board, Washington DC.*

Claessens, M.T., Van Dijk, N.M., Zwaneveld, P.J. (1998). Cost Optimal Allocation of Rail Passenger Lines. *European Journal of Operational Research 110: 474-489.*

Damm, D., Lerman, S., Lerner-Lamm, E., Young, J. (1980). Response of Urban Real Estate Values in Anticipation of the Washington Metro. *Journal of Transport Economics and Policy 14: 315-336.*

Debrezion, G., Pels, E., Rietveld, P. (2007). The Impact of Railway Stations on Residential and Commercial Property Value: a Meta Analysis. *Journal of Real Estate, Finance and Economics 35: 161-180.*

Ekeland, I., Heckman, J.J., Nesheim, L. (2004). Identification and Estimation of Hedonic Models. *Journal of Political Economy 112(1): S60-S109.*

Epple, D. (1987). Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy 95(1): 59-80*

Fan, J., Gijbels I. (1996). Local Polynomial Modeling and Its Applications. *In: Monographs and Statistics and Applied Probability 66. Chapman and Hall, England.*

Flavin, M. (1981). The Adjustment of Consumption to Changing Expectations about Future Income. *Journal of Political Economy 89(5): 974-1009.*

Florida, R. (2002). The Rise of the Creative Class. *Basic Books, New York.*

Gatzlaff, D., Smith, M. (1993). The Impact of the Miami Metrorail on the Value of Residences Near Station Locations. *Land Economics 69: 54-66.*

Gibbons, S., Machin, S. (2005). Valuing Rail Access Using Transport innovations. *Journal of Urban Economics 57: 148-169.*

Grimes, A., Young, C. (2010). Anticipatory Effects of Rail Upgrades: Auckland's Western Line. *Motu Working Paper, Waikato New Zealand.*

Gyourko, J., Tracy, J. (1991). The Structure of Local Public Finance and the Quality of Life. *Journal of Political Economy 91: 774-806*

Hurvich, C.M., Simonoff, J.S., Tsai, C.L. (1998). Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society B 60(2): 271-293.*

Irwin, E.G. (2002). The Effects of Open Space on Residential Property Values. *Land Economics 78(4): 465-480.*

Jacobs, J. (1961). The Death and Life of Great American Cities. *Vintage, New York.*

Keijer, M.J.N., Rietveld, P. (2000). How do People Get to the Railway Station? The Dutch Experience. *Transportation Planning and Technology 23 : 215-235.*

Knaap, G.J., Ding, C., Hopkins, L. (2001). Do Plans Matter? *Journal of Planning Education and Research 21(1): 32-39*

Kondo, Y., Lee. M.-J. (2003). Hedonic Price Index Estimation Under Mean-Independence of Time Dummies from Quality Characteristics. *Econometrics Journal 6: 28-45.*

Mackie, P., Fowkes, S., Wardman, M., Whelan, G., Nellthorp, J., Bates, A. (2003). Value of Travel Time Savings in the UK - summary report. *Report to Department of Transport, Institute of Transport Studies, University of Leeds.*

McDonald, J.F., Osuji, C.I. (1995). The Effect of Anticipated Transportation Improvement on Residential Land Values. *Regional Science and Urban Economics 25: 261-278.*

McMillen, D.P., McDonald, J.F. (2004). Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983-1999. *Real Estate Economics 32(3): 463-486.*

Ossokina, I.V. (2010). Geographical Range of Amenity Benefits: Hedonic Price Analysis for Railway Stations. *CPB Discussion Paper 146.*

Pagan, A., Ullah, A. (1999). Non-parametric Econometrics. *Cambridge University Press, Cambridge*

Rietveld, P. (2000). The Accessibility of Stations: The Role of the Bicycle in The Netherlands. *Transportation Research D: 71-75.*

Robinson, P.M. (1998). Root-N-Consistent Semi-Parametric Regression. *Econometrica 57: 1403-1430.*

Silverman, B. (1986). Density Estimation for Statistics and Data Analysis. *In: Monographs on Statistics and Applied Probability 26. Chapman and Hall, London.*

Ullah, A. Roy, N. (1998). Nonparametric and Semiparametric Econometrics of Panel Data. *In: Ullah, A., Giles, D.E.A. (eds.). Handbook of Applied Economic Statistics. Marcel Dekker, New York.*

Weinberger, R. (2001). Light rail proximity: Benefit or detriment, the case of Santa Clara County, California. *Transportation Research Board (presented at Transportation Research Board 80th Annual Meeting, January 7–11), Washington, DC.*

Appendix A: Descriptive statistics

Table A1: Descriptive statistics

	MEAN	ST.DEV.
Price	152,542	(78,790)
Size in m^2	97.813	(30.338)
Rooms	3.754	(1.167)
Garage	0.022	(0.146)
Central Heating	0.852	(0.355)
Monument	0.006	(0.074)
Population density (persons per km^2)	7,388	(3,396)
Share of Ethnic Minorities	0.141	(0.116)
Share <25 years	0.304	(0.058)
Share >65 years	0.127	(0.071)
Distance to station (km)	1.649	(0.968)
Number of observations	55823	
Number of properties	25270	

Table A2: Change of variables before and after transactions

	BEFORE		AFTER		CHANGE	
	MEAN	ST.DEV	MEAN	ST.DEV	MEAN	ST.DEV
Price	127,421	(66280)	174,652	(80499)	47,231	(41,330)
Size in m^2	96.766	(30.136)	97.317	(30.067)	0.551	(7.784)
Rooms	3.712	(1.142)	3.741	(1.183)	0.029	(0.642)
Garage	0.024	(0.151)	0.019	(0.138)	0.157	(0.191)
Central Heating	0.808	(0.394)	0.897	(0.303)	0.089	(0.326)
Monument	0.005	(0.070)	0.006	(0.078)	0.001	(0.077)
Population density (persons per km^2)	7,331	(3,368)	7,442	(3,388)	110	(810)
Share of Ethnic Minorities	0.099	(0.100)	0.181	(0.115)	0.082	(0.068)
Share <25 years	0.302	(0.058)	0.305	(0.059)	0.003	(0.021)
Share >65 years	0.130	(0.072)	0.126	(0.069)	-0.004	(0.018)
Distance to station (km)	1.702	(1.003)	1.592	(0.921)	-0.110	(0.489)
Number of observations	30533		30533		30533	
Number of properties	25270		25270		25270	

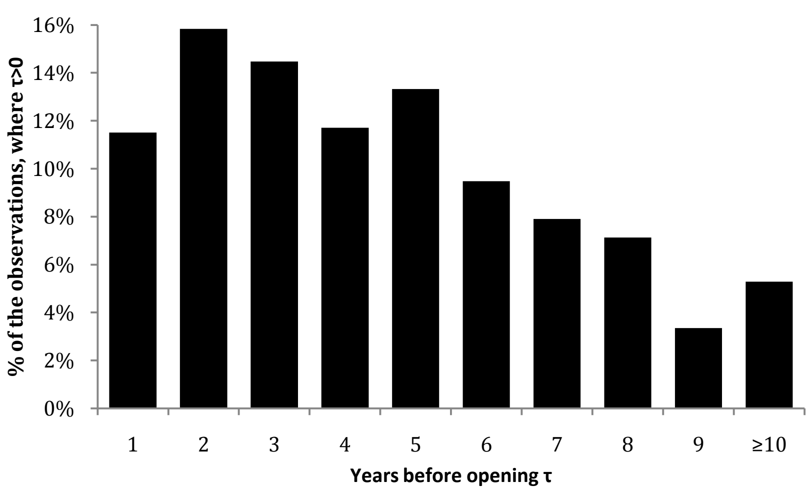

Figure A1: Distribution of τ

Appendix B: Robustness check of bandwidth choice

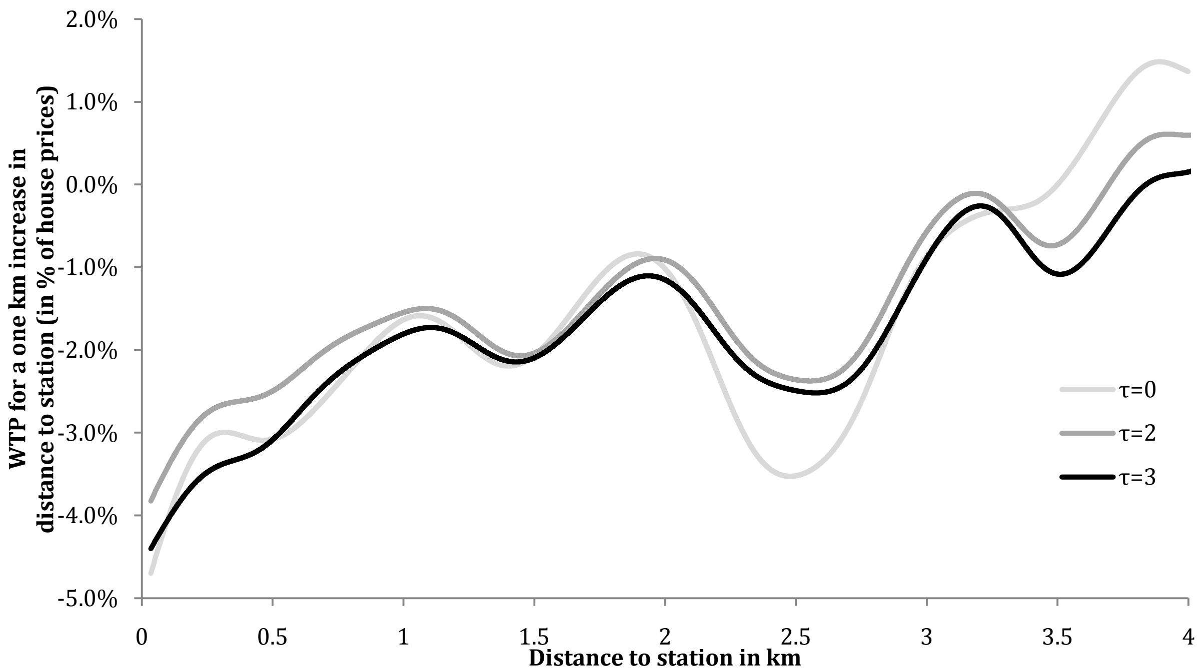

Figure B2: Results for models with a bandwidth which is equal to the Silverman's rule

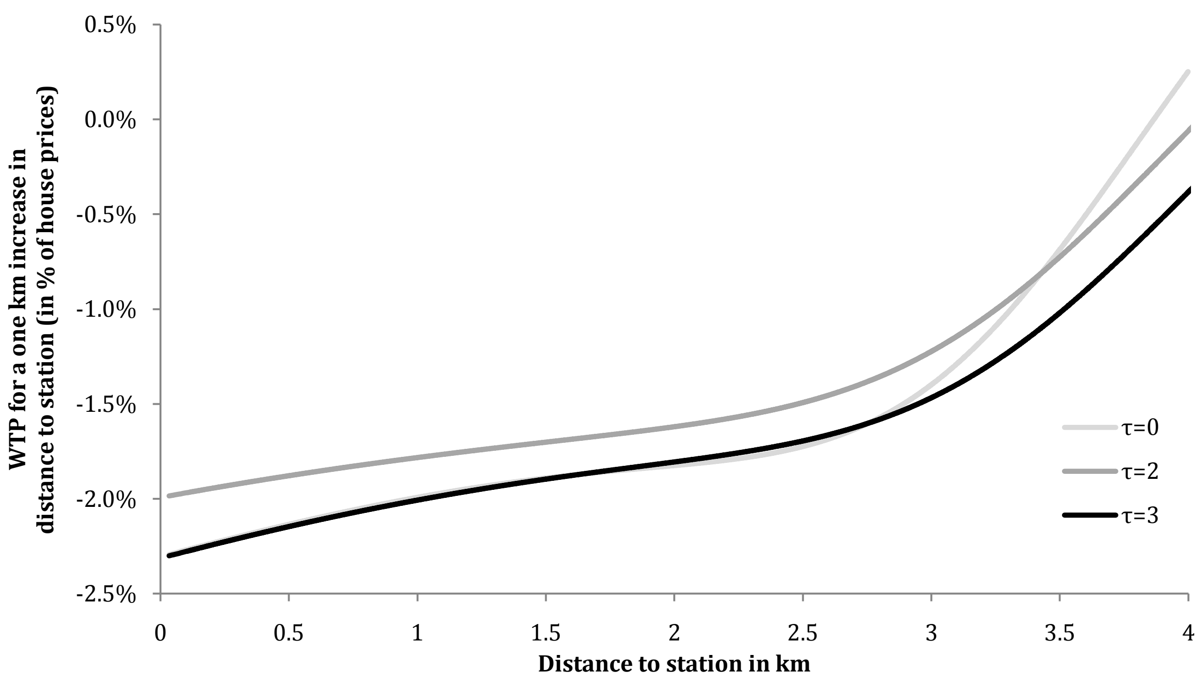

Figure B3: Results for models with a bandwidth which is equal to five times the Silverman's rule