Vrije Universiteit Amsterdam

Household preferences and hedonic pricing*

Hans R.A. Koster† Jan Rouwendal‡

1 Introduction

When economists consider choices of people, they usually assume that the behaviour of people can be modelled as the outcome of the maximisation of a *utility function*. When applied to the housing market this concept seems to make sense: when a person considers to choose a location within, say, Amsterdam, she considers a wide range of characteristics of the available houses and their location such as floor space, the number of rooms, the access to urban amenities, the crime level, access to open space, etc. Eventually she will choose the property that makes her most happy and so maximizes her utility, *given* her available budget. Hence, location choices by people and the amount of money they pay for a property provide useful information on the preferences of people for locations.

In this chapter we outline one of the most popular ways in the field of urban economics to measure those preferences and quantify the so-called *marginal willingness to pay* (MWTP) of people for location and housing characteristics. This method is referred to as *hedonic*

*A shorter version of this chapter has been published as Koster et al. (2020). We thank Luca Bertolini, Jos van Ommeren, Erik Verhoef, and Nanke Verloo for useful comments on previous versions of this chapter.

†Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081HV Amsterdam, email: h.koster@vu.nl. Hans is also research fellow at the Higher School of Economics in St. Petersburg, Tinbergen Institute (Gustav Mahlerplein 117 1082MS Amsterdam) and affiliated with the Centre for Economic Policy Research.

‡Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081HV Amsterdam, email: j.rouwendal@vu.nl. Jan is also research fellow at Tinbergen Institute (Gustav Mahlerplein 117 1082MS Amsterdam).

pricing analysis and requires information on many location choices, the prices of properties, as well as location and housing characteristics. Given the surge in the availability of *big data*, these data are often readily available. The biggest advantage of estimating the MWTP of households for location and housing characteristics is that benefits can be calculated for public goods provided in cities, for which it is otherwise nearly impossible to calculate the benefits. For example, the costs of investing in urban renewal, open space, infrastructure, and historic amenities, are usually pretty straightforward to calculate. The benefits, however, are much harder to estimate. The hedonic price approach provides a useful method to do this by estimating the MWTP of households for these public goods.

So what is a hedonic price function precisely?[1]

A hedonic price function is a description of the equilibrium prices of varieties of a heterogeneous good, which is influenced by supply and demand.

An immediate implication of the above definition is that if preferences, quantities or qualities of the heterogeneous goods on offer change, the hedonic price is also likely to change.

Pioneering work on hedonic price functions dates back to Court (1939) who was the first to model the prices of automobiles as a function of their characteristics. However, hedonic pricing was popularised by Griliches (1961) who showed how hedonic price functions could be used to develop a quality-adjusted price index for cars. In 1974, Rosen formalised the method by clarifying the relationship with conventional supply and demand analysis, thereby providing the link with standard micro-economic theory. In particular, he showed tat the marginal price of a characteristic implied by the hedonic price function - that is, the first derivative of the hedonic price function with respect to that characteristic - could be interpreted as the marginal willingness to pay. The hedonic price method has been widely used since then to measure, among other things, the costs and benefits of good air quality (Chay and Greenstone, 2005; Bajari et al., 2012), hazardous waste sites (Greenstone and Gallagher, 2008), historic amenities (Ahlfeldt and Maennig, 2010; Koster and Rouwendal, 2017), open space (Irwin, 2002; Anderson and West, 2006), school quality (Black, 1999; Bayer et al., 2007; Gibbons et al., 2013), proximity to wind turbines (Gibbons, 2015; Dröes and Koster, 2016), power plants (Davis, 2011), and the effectiveness of place-based policies (Ahlfeldt et al., 2017; Koster and Van Ommeren, 2019). In all these studies the heterogeneous commodity studied is housing.

As these examples suggest, the hedonic price method is often used for policy relevant research. The underlying reason is that it provides a way to investigate the welfare consequences of external effects and public goods for which no markets exist. The powerful characteristic of hedonic price techniques is that house prices provide information on the *marginal willingness to pay* - the equivalent of the price of market goods - of people for public goods and location characteristics, for which it is otherwise hard to measure the benefits.

[1]While the current chapter focusses on the housing market, the hedonic price approach can be also applied to other markets, such as the labour market and market for automobiles.

Historic amenities and house prices (1). In the Netherlands governments devote considerable amounts of money to invest in the preservation of historic buildings and neighbourhoods. The idea is that historic amenities may attract shops, restaurants and other modern urban amenities, and thereby contribute to a higher quality of life in cities (Brueckner et al., 1999; Glaeser et al., 2001). Historic amenities will most likely imply positive *external effects*; benefits are not only enjoyed by, say, the inhabitants of a historic building, but also by visiting tourists and residents living close to a historic building (Koster and Rouwendal, 2017). By looking at housing prices, we may come up with an estimate of the *external* benefits of historic amenities and compare those to the costs of investing in cultural heritage. This implies that we can undertake a social cost benefit analysis using changes in house prices.

As an application of the hedonic price methods in this chapter we discuss the study by Koster and Rouwendal (2017) who investigate the impacts of investments in cultural heritage in the Netherlands. We use data on about 650 thousand housing transactions between 1985 and 2011. Most of these transactions (96%) took place in urban areas (*i.e.* in areas with a population density exceeding 1000 people per km^2). Moreover, we obtain data on almost 12 thousand investment projects in cultural heritage since the 1980s. The total investments are about €3 billion of which about €1 billion is a subsidy. We focus on smaller scale projects that are in the vicinity of residential properties. In total we analyse the effects of a total of €1.63 billion investments in historic buildings.

Hence, we aim to identify the MWTP of homeowners for historic amenities, and thus employ a hedonic price approach. The idea is then to compare price changes of otherwise identical properties, located on sites that differ in the amount of investments they received. From the spatio-temporal differences in prices between the properties one may then infer the *marginal willingness to pay* for historic amenities.

When estimating a hedonic price function one relies on standard regression techniques such as ordinary least squares by regressing the total price of a property on its characteristics.[2] However, although this may seem straightforward, there are at least two issues when estimating a hedonic price function:

1. When the hedonic price regressions is linear in the characteristics, the marginal prices are constants.[3] Hence one implicitly assumed that homebuyers all have the same marginal willingness to pay, while people are likely heterogeneous in their preferences and hence have different MWTPs. When assuming a linear hedonic price function (or alike), the hedonic price function will therefore probably be *misspecified*.

2. There is also the issue of omitted variable bias. That is, the estimated marginal willingness to pay for a certain characteristic may depend on characteristics that are omitted

[2]For a review of standard regressions techniques we refer to standard textbooks such as Cameron and Trivedi (2005), Angrist and Pischke (2008), Stock and Watson (2011) or Wooldridge (2015).

[3]Recall that these are the first derivatives with respect to the characteristics.

in the regression.

In this chapter we will discuss how existing studies have dealt with both issues.

Historic amenities and house prices (2). When aiming to measure the MWTP for historic amenities, we may suffer from misspecification and omitted variable bias, which may imply that the estimated impact of historic amenities on house prices will be wrong.

Let's consider the simple example where you compare two properties with differences in the presence and quality of nearby historic amenities. The question is, however, if those two properties are really identical. It is for example likely that historic amenities are disproportionately clustered in city centres, where prices are higher for other reasons. If the properties that are compared are not really the same except for access to historic amenities, our estimates suffer from an omitted variable bias and, hence, our estimate of the MWTP will be incorrect.

The second issue implies that people in the two properties considered may have a different preferences for historic amenities. For example, houses in the city centre are more often occupied by richer people that may like historic amenities more than families with lower incomes that live more often in the suburbs (Koster et al., 2016). Hence, even if you compare two identical properties, a hedonic price function may not identify the marginal willingness to pay of either of these households, but some kind of average.

This chapter is structured as follows. In Section 2 we will provide the economic foundations for hedonic pricing theory, by relying on standard micro-economic theory. In Section 3 we discuss the main empirical issues when bringing hedonic price techniques to the data, followed by a summary in Section 4.

2 MICRO-ECONOMIC FOUNDATIONS

2.1 THE MARGINAL WILLINGNESS TO PAY

A standard assumption in textbook microeconomics is that commodities are available at a given quality. In reality we see a lot of heterogeneity in almost every commodity purchased by households. For example, a large home means that the house has a higher quality. Let us consider the case where people derive utility from a composite good q (think of food, healthcare, etc.) and housing, denoted by k. Hence, k denotes the quality of housing. In what follows we usually treat k as a scalar, but it could also be a vector (so that houses have multiple characteristics). We now specify the utility function as:

$$u = u(q, k). \tag{2.1}$$

The price of the heterogeneous commodity is a function of the quality chosen. We therefore write the budget constraint as:

$$q + p(k) = y. \tag{2.2}$$

This formulation assumes that the unit price of the composite good equals 1, while $p(k)$ is the house price or housing expenditures. This model was used by Rosen (1974) in his analysis of heterogeneous goods markets.

If we maximise utility (2.1) using a Lagrangian subject to the budget constraint (2.2) by choosing q and k, the first-order conditions are:

$$\frac{\partial u}{\partial q} = \lambda, \tag{2.3}$$

$$\frac{\partial u}{\partial k} = \lambda \frac{\partial p(k)}{\partial k}, \tag{2.4}$$

as well as the budget constraint (2.2). Note that λ is the Lagrange multiplier and represents that marginal utility of income.

We can combine equations (2.3) and (2.4) to get:

$$\frac{\partial u / \partial k}{\partial u / \partial q} = \frac{\partial p(k)}{\partial k}, \tag{2.5}$$

This equation tells us that the marginal willingness to pay for housing quality k is equal to the partial derivative of the price specified as a function of quality. Or in other words, the derivative of the hedonic price function with respect to housing quality is equal to the marginal rate of substitution of housing quality with respect to a composite good.[4] This relationship plays a key role in hedonic price studies.

Historic amenities and house prices (3). To illustrate this with a simple example, let's consider the following hedonic price function:

$$p_i = \alpha_0 + \alpha_1 z_i + \xi_i, \tag{2.6}$$

where p_{ij} is the house price of property i, and z_i denotes the presence and quality of historic amenities. In Koster and Rouwendal (2017) we proxy this by past cumulative investments in cultural heritage in the neighbourhood. Further, α_0 and α_1 are the parameters to be estimated and ξ_i is a residual. Given this linear hedonic price function, the
marginal willingness to pay for a one unit increase in historic amenities is given by $\frac{\partial p_i}{\partial z_i} = \hat{\alpha}_1$ (see equation (2.5)). Because homeowners are expected to value historic amenities positively, we expect $\alpha_1 > 0$. Since this specification implies that the marginal willingness to pay is identical for all households, it is not very useful for empirical work, as noted

[4]The marginal rate of substitution is the rate at which someone gives up some amount of one good in exchange for another good while maintaining the same level of utility.

earlier.

2.2 THE VALUE FUNCTION

To analyse the equilibrium in a market for a heterogeneous good, Rosen (1974) introduced the concept of a value function. It gives the willingness to pay of the consumer for a variety of the good with a particular quality, that is the maximum amount of money the consumer is able to pay for that variety, while still being able to reach a particular utility level. The value function is related to the bid rent function that plays an important role in other urban economic models, such as the monocentric city model (see *e.g.* Alonso, 1964; Mills, 1967; Muth, 1969). It gives combinations of price and quality that provide the consumer a given level of utility and can therefore be considered as a kind of indifference curve.

To define the value function, let P denote the willingness to pay for the heterogeneous commodity. We substitute housing expenditures in (2.2), and use the resulting equation to rewrite the utility function (2.1) as:

$$u = u(y - P, k). \tag{2.7}$$

We now ask the question: what is the price the consumer is willing to pay for quality k if her utility must be equal to a predetermined value, say u^*? To answer this question, we write this condition as:

$$u(y - P, k) = u^*, \tag{2.8}$$

and invert the utility function with respect to $y - P$. After rewriting the resulting equation, we find:

$$P = y - u^{-1}(u^*, k). \tag{2.9}$$

The above equation gives us a relationship between the price the consumer is willing to pay and the quality of the heterogeneous commodity for given income y and utility level u^*. In other words, it gives the value a consumer attaches to quality k of the heterogeneous commodity at the given utility and income levels.

It is useful to consider the relationship between P and the variables on the right-hand side of the equation. Starting from (2.9), we can verify the following relationships:[5]

$$\frac{\mathrm{d}P}{\mathrm{d}y} = 1. \tag{2.10}$$

$$\frac{\mathrm{d}P}{\mathrm{d}k} = \frac{\partial u/\partial k}{\partial u/\partial q}. \tag{2.11}$$

[5] To do so you could apply the concept of the total differential to (2.8). For how to derive the equations below, we refer to Appendix A.1.

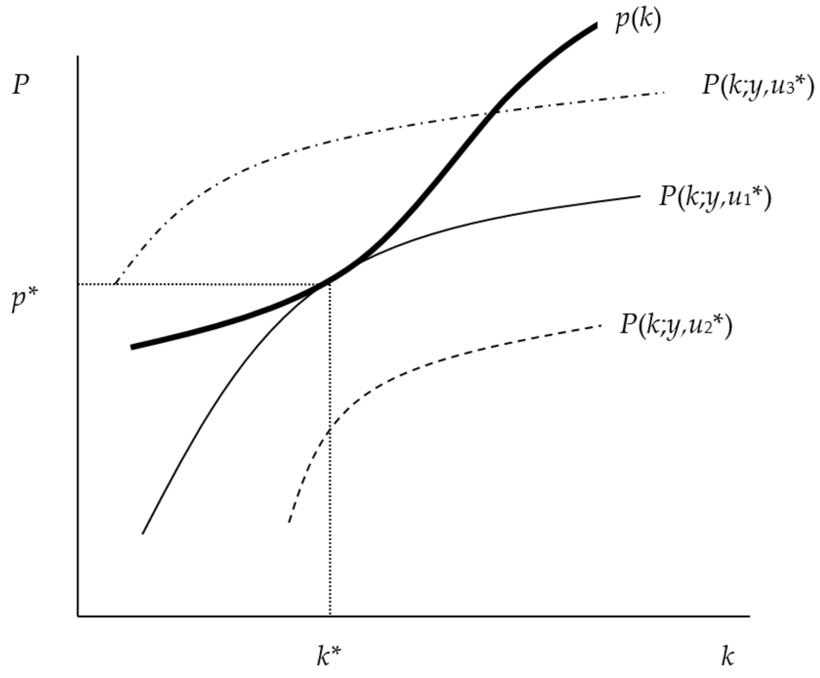


Figure 2.1 – UTILITY MAXIMIZATION WITH VALUE FUNCTIONS

$$\frac{dP}{du^*} = -\frac{1}{\partial u / \partial q} < 0 \tag{2.12}$$

Equation (2.10) states that changes in income are completely translated into changes in the willingness to pay for housing quality. The reason is that consumer's utility utility is constant on each value function. Equation (2.11) says that the value attached to the heterogeneous good is increasing in its quality. Note further that the denominator of (2.11) is the marginal utility of income; a lower value at a higher income means a higher MWTP.

The right-hand side of this equation is indeed again the marginal willingness to pay for housing quality (which is identical to (2.5)). Equation (2.12) shows that the willingness to pay for the heterogeneous good decreases if the consumer has to reach a higher level of utility at a fixed income level.

The solid line in Figure 2.1 shows the willingness to pay P as a function of quality k for a given utility level u_1. The function is increasing and we know from (2.11) that its slope equals the marginal willingness to pay for quality. The marginal willingness to pay for quality is equal to minus the slope of the indifference curve for quality and the composite good that corresponds to the same utility level. The indifference curve is convex, which implies that the marginal willingness to pay for quality is decreasing in quality. For this reason, the value function P is concave in quality.

The two dashed lines in Figure 2.1 show the value functions of the same consumer for a higher utility level u_2 and a lower one u_3. That the value function corresponding to a *higher* utility level lies below the one corresponding to a *lower* level can be understood as follows: if a consumer chooses quality k of the heterogeneous good and the price P to be paid for it goes down, then (2.8) shows that (all else equal) more will be spend on other consumption goods and utility goes up. Equation (2.9) shows that when income increases, the value function shifts upwards parallel to itself. The shift in the value function is equal to the shift in income. The reason is that with given utility, consumption of the composite good should remain unchanged, which is only possible when total change in income is spent on the composite good.

The value function must be distinguished from the hedonic price function. The latter gives the market price for the heterogeneous good as a function of its quality. The value function gives the willingness to pay for the heterogeneous good of a *single* consumer as a function of her income and a given utility level. However, it will not come as a surprise that there is a relationship between the hedonic price function and the value function.

The bold line in Figure 2.1 shows the hedonic price function and three value functions. We can analyse consumer behaviour with respect to the heterogeneous product in a similar way as is done in the standard textbook treatments of utility maximizing behaviour if we consider the value function as a kind of indifference curve. Each value function gives the combinations of quality and price that allow the consumer to reach a particular level of utility. The hedonic price function shows the combinations of quality and price that are available to the consumer. Just as the consumer in the textbook case attempts to reach the highest possible indifference curve, the consumer in Figure 2.1 attempts to reach the *lowest* possible value function. In the situation shown in the Figure, this is the value function corresponding to utility level u_1. The optimal combination of price and quality is found at the point where this value function is tangent to the hedonic price function. Hence the utility maximizing consumer pays a price p^* for her optimal quality k^*.

Consumers differ in tastes and incomes and therefore in the curvature of their value functions.[6] The situation as shown in Figure 2.1 would then refer to one combination of preferences and incomes, and show the relevant price/quality combination for this type of consumers. Other consumers, with different value functions, would have optimal price/quality combinations that are located on different points of the hedonic price function.

A final important aspect of Figure 2.1 is that the value function and the hedonic price function are tangent to each other: they have only one point in common. At that point, the marginal willingness to pay for quality of the consumer equals the slope of the hedonic price function. Other points of the hedonic price function coincide with the marginal

[6] If all consumers had the same preferences and incomes, they would all have the same value functions. This would imply that, with the given hedonic price function, only one variety of the heterogeneous good would be traded. Such a situation is not completely impossible. For instance, if the heterogeneous good is produced in a competitive industry, the market price of all varieties would be determined completely by production costs (the hedonic price function simply reflects the production costs) and it could be the case that only one possible price/quality combination is available on the market. However, in such a case we would not observe a hedonic price function. Instead we would see a homogeneous commodity. It may also be noted that even with identical tastes and incomes a continuum of price/quality combinations can be available on the market if the hedonic price function overlaps with (a part of) the value function. Clearly, this would be a very special situation.

willingness to pay of other consumers. The implication is that it is in general not appropriate to compare different points of hedonic price function and to interpret the difference in price as the willingness to pay for the difference in quality for the same consumer. This is a clear limitation of hedonic price analysis. Formally, its restricts its application to small changes in characteristics. In practice the results of hedonic price analysis are often also applied to evaluate larger changes in air pollution, traffic noise, etc. Although this is not justified by the theory, the hope of pragmatic policy analysts is that the results are still a reasonable approximation to the actual willingness to pay of consumers for the actual changes.

2.3 DEMAND FUNCTIONS FOR QUALITY CHARACTERISTICS

We have seen that the hedonic price function reveals the marginal willingness to pay for quality characteristics of heterogeneous goods (see equation (2.5)). This marginal willingness to pay is important for the purpose of economic welfare analysis. To see its role, return to equation (2.1) and derive the marginal rate of substitution between quality and the composite good. Start by writing down the total differential:

$$\mathrm{d}u = \frac{\partial u}{\partial q}\mathrm{d}q + \frac{\partial u}{\partial k}\mathrm{d}k. \tag{2.13}$$

Then let's find the change in the amount of the composite good that compensates exactly for a change in the quality. We impose $\mathrm{d}u = 0$ and rearrange:

$$-\frac{\mathrm{d}q}{\mathrm{d}k} = \frac{\partial u/\partial k}{\partial u/\partial q} = \frac{\partial p(k)}{\partial k}. \tag{2.14}$$

The second equality sign in this expression uses the first order condition (2.5). In fact, this just repeats our earlier conclusion that the first derivative of the hedonic price function can be interpreted as the marginal willingness to pay of a utility maximizing consumer. This result is of great importance for applied welfare economic analysis. Its significance becomes even clearer if we realise, from the budget constraint, that $q = y - p(k)$. Using this, we can rearrange the first and third parts of (2.14) as:

$$\mathrm{d}y = \frac{\partial p(k)}{\partial k}\mathrm{d}k. \tag{2.15}$$

The left-hand side of this equation gives the change in income that is necessary to compensate the consumer for a small change in quality. However, it should be realised that (2.15) holds only for small changes in quality because the value function and the hedonic price function coincide only in one point.[7]

In fact an analysis of the compensating changes in income that are necessary with finite, but not necessarily small, changes in quality was the main purpose of Rosen (1974). He

[7] Equation (2.15) is in fact always an approximation if $\mathrm{d}k$ is finite, but the implied error is negligible for small changes. The discussion at the end of the previous section is also relevant here.

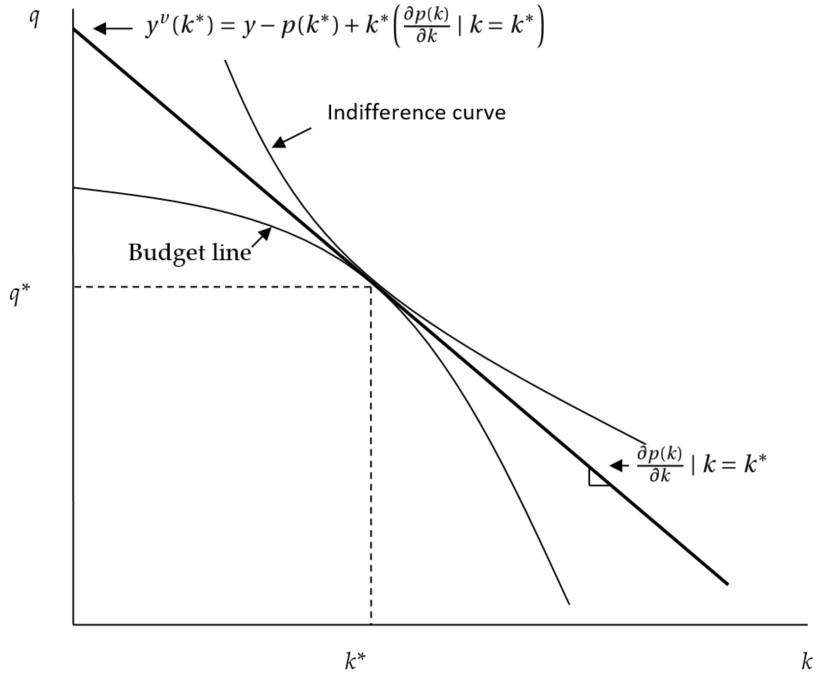


Figure 2.2 – LINEARISING THE BUDGET CONSTRAINT

attempted to interpret the characteristic k as a commodity whose demand function could be found from observations of consumer behaviour. However, this is not a trivial task because demand functions are defined for commodities traded at a given price per unit, whereas this is not the case for the characteristics of heterogeneous good. To see how this idea could – nevertheless – work, consider Figure 2.2. This figure shows an indifference curve, derived from utility function (2.1), and the budget restriction (2.2), which has been rewritten as:

$$q = y - p(k). \tag{2.16}$$

In drawing the budget line, it has been assumed that the hedonic price function is convex, which implies that the marginal price $\partial p/\partial k$ is increasing in k: additional quality becomes more costly when the quality level is already high. It follows than that the amount of money left for consumption of the composite good is a concave function of quality: for each additional unit of quality more consumption of the composite good has to be given up. As Figure 2.2 indicates, there is no constant unit price of quality – the budget line does not have a constant slope – and therefore the concept of the demand function cannot immediately be used.

With some elaboration, we can nevertheless describe the consumer's demand for quality using the concept of a demand function. The trick is that we have to linearise the budget constraint. Figure 2.2 shows that the optimal amount of quality consumed is k^*. The marginal price of quality at this point is $\pi(k^*) = \partial p(k)/\partial k \mid k = k^*$. We can draw the straight line with slope $\pi(k^*)$ that passes through the consumer's optimal combination of quality and the composite good. This is the bold straight line in Figure 2.2. The consumer's choice behaviour with the actual, non-linear budget constraint is clearly identical to what it would have been with this counterfactual linear budget constraint. This observation allows us to describe the behaviour of the consumer using an ordinary demand curve.

We have already determined the slope $\pi(k^*)$ of the linearised budget constraint, which is equal to that of the actual non-linear budget constraint in the consumer's optimum. The intercept of the linearised budget constraint differs from that of the actual budget constraint (for which it is equal to $y - p(0)$). The intercept of the linearised budget constrained is often referred to as virtual income, denoted as y^v, and can be computed as $y^v(k^*) = y - p(k^*) + k^*\left(\frac{\partial p(k)}{\partial k} \mid k = k^*\right)$.

Using a linearisation of the budget line, we can describe quality choice behaviour of the consumer on the basis of a demand function:

$$k = k(\pi, y^v). \tag{2.17}$$

Using this function, we could compute the exact welfare measures that we need for the analysis of non-marginal changes in quality characteristics.

2.4 ISSUES WHEN ESTIMATING DEMAND FUNCTIONS

To empirically estimate *structural* demand functions instead of the *marginal* willingness to pay is not straightforward. Rosen (1974) suggested a two-step analysis. In the first stage the hedonic price function $p(k)$ is estimated, while in the second step the researcher estimates the demand function (2.17). The basic idea is straightforward: if income and quality choice of consumers and the hedonic price function are known, it is possible to compute the marginal price π and the virtual income y^v that are relevant for each consumer. Using this, we can fit a regression line in the usual way and consider the result as the demand function.

However, some complications arise. In the conventional situation, where the unit price is a constant, the equivalents of marginal price (the constant unit price) and virtual income (the consumption budget) can be regarded as given for the consumer. But here the situation is different: marginal price and virtual income both depend on the chosen quality, as is apparent from the notation we used above.

The problem is that a consumer may choose a higher than average level of quality k^h for unknown idiosyncratic reasons, for instance because she has a particularly strong taste for quality. With a convex hedonic price function (as we used in Figure 2.2) this would imply that she pays a particularly high marginal price $\pi(k^h)$. Another consumer, with a particularly low taste for quality would have a much smaller k^l and pay a lower marginal price $\pi(k^l)$. This is illustrated in Figure 2.3. If we would use these observations in an attempt to estimate the demand function (2.17), we could easily come to the conclusion that the demand function is upward sloping: a low marginal price is associated with a low demand for quality, and a high marginal price with a high demand, which is unrealistic.

In other words, a challenge when aiming to estimate demand functions is that that unob-

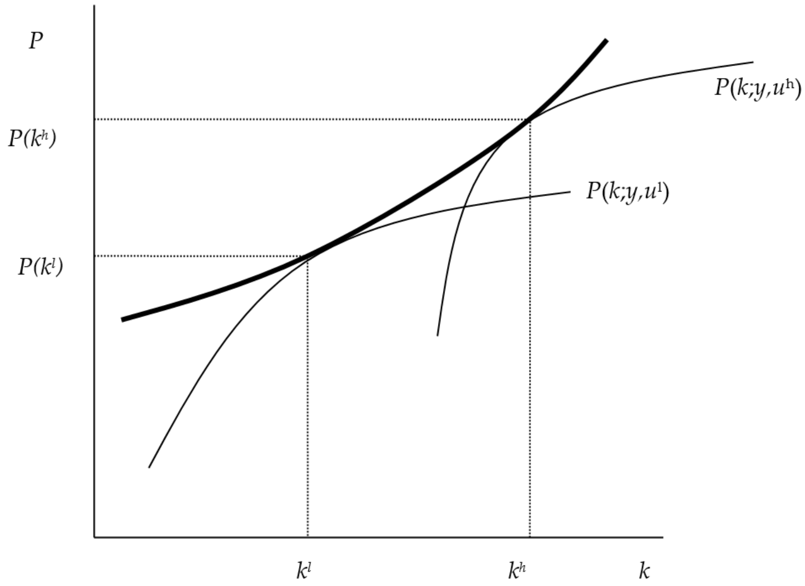


Figure 2.3 – CHOICES OF CONSUMERS WITH DIFFERENT TASTES FOR QUALITY

served heterogeneity among consumers is important, which causes differences in choice behaviour that are difficult to disentangle from price and income effects.[8]

Rosen (1974) realised this problem to some extent and suggested the use of instrumental variables. However, the difficulty of finding good instruments only became fully apparent in later years with analyses by Epple (1987), Bartik (1987) and others. The basic problem is that consumers with identical tastes and incomes will choose the same point on the hedonic price function. It follows that different combinations of marginal prices and quantities correspond in principle all to different demand curves. Of each demand curve only a single point is observed and this severely limits the possibilities to estimate price or income elasticities.

More recent studies by Ekeland et al. (2002), Ekeland et al. (2004), Bajari and Benkard (2005), Bajari and Kahn (2005), and Heckman et al. (2010) show that non-parametric methods can be used to identify individual (inverse) demand functions, even if there is only one observation per individual. One may also use repeated observations of individuals (as in Bishop and Timmins, 2018), but then one has to make the strong assumption that preferences of people do not change between the different decisions. The latter seems unrealistic in the housing market where people mostly move when preferences have changed (*e.g.* because of changes in family situation, job, etc.). Because methods to estimate non-parametric regressions are not easy to apply, in this chapter we will not further pursue the issues associated with estimating demand functions for characteristics. The next section will discuss econometric

[8]Of course, there are also differences in virtual income induced by choice behaviour, and this complicates the story a bit further. It is, for instance, not difficult to check that the virtual income of the consumer with a weak taste for quality is lower than that of the consumer with the strong taste for quality when the actual incomes of both consumers are identical. This means that income effects are also hard to estimate.

issues associated with the hedonic price function.

3 ECONOMETRIC ESTIMATION OF HEDONIC PRICE FUNCTIONS

In Section 2 we showed that we can identify the marginal willingness to pay for a characteristic using the hedonic price function (see equation (2.5)). We aim to discuss two issues when estimating hedonic price functions using real-life data: (*i*) misspecification of the hedonic price function and (*ii*) endogeneity.

3.1 ISSUE 1: MISSPECIFICATION OF HEDONIC PRICE FUNCTIONS

We have already seen that the simplest hedonic price function, which is linear in characteristics, is unlikely to be valid in practice because it does not take into account the possibility that consumers differ in their marginal willingness to pay, or implicit price, for a characteristic. It is therefore important to choose a specification of the hedonic price function that is non-linear in the characteristics. We will now consider how this can be done without making the function non-linear in the parameters to be estimated, which would complicate the econometrics.

Often, the first step to allow for heterogeneity is to estimate a simple log-linear hedonic price function of the following form:

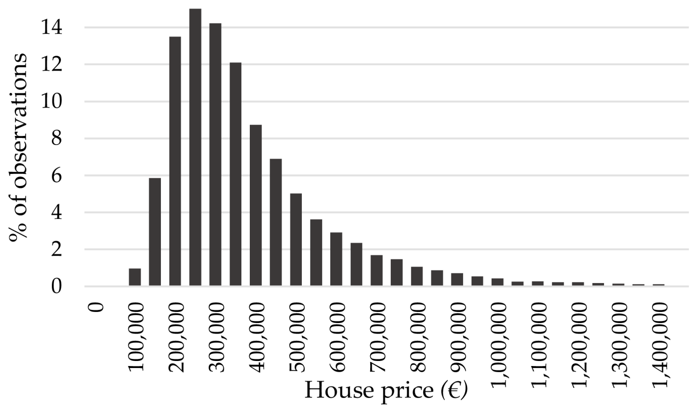
$$\log p_i = \alpha_0 + \alpha k_i + \xi_i, \tag{3.1}$$

Note that one may also take the logarithm of k_i if it is a continuous variable. There are two reasons why in practice one almost always estimates log-linear hedonic price functions:

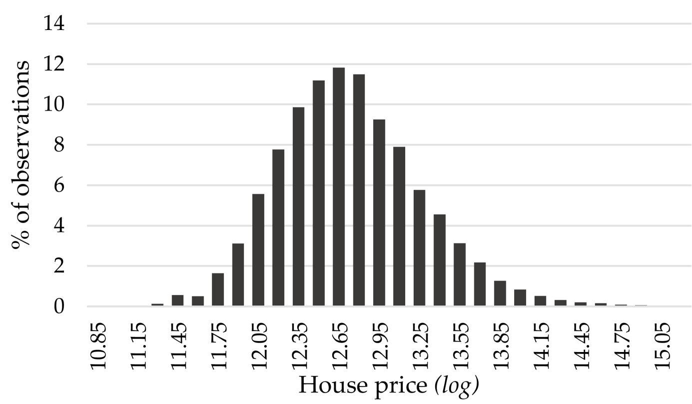
1. House price data are often strongly skewed with a few observations having very high prices. This means that these outliers may disproportionately impact the estimated coefficients.

2. The coefficients are intuitive and easy to interpret as (semi-)elasticities.

Historic amenities and house prices (4). Let us now investigate the impact of investments in historic amenities on house prices in the Netherlands in order to identify the marginal willingness to pay of households for nearby historic amenities. Our analysis is based upon a house transactions dataset from the `NVM` (*i.e.* the Dutch Association of Real Estate Agents). It contains about 75% of all transactions between 1985 and 2011. For 657,574 transactions, we know the sales price, the exact location, and a wide range of housing attributes such as size (in m^2), house type, and construction year. We further gather data on investments in cultural heritage from the Department of Cultural Heritage. Because we are interested in the investments on surrounding properties, we exclude investments to the interior of buildings (about 7% of the projects).



(a) HOUSE PRICES



(b) LOG OF HOUSE PRICES

Figure 3.1 – HISTOGRAMS BASED ON NVM DATA FROM AMSTERDAM AND SURROUNDINGS

We illustrate the issue of skewness of house prices in Figure 3.1 where we plot the distribution of house prices. What we see in Figure 3.1a is that prices are right-skewed. However, when we take logs in Figure 3.1b we find that log house prices are (more or less) normally distributed, so that the issue of outliers is mitigated.

Note that with log prices, one can derive the marginal willingness to pay $\partial p_i / \partial k_i$ for each observation. We can write (3.1) as $p_i = e^{\alpha_0 + \alpha k_i + \xi_i}$. Now we calculate the derivative using the chain rule of differentiation and use equation (2.5):

$$\frac{\partial u_i / \partial k_i}{\partial u_i / \partial q_i} = \frac{\partial p_i}{\partial k_i} = \mathrm{e}^{\alpha k_i + \xi_i} \cdot \alpha = \alpha p_i. \tag{3.2}$$

The above formula implies the marginal willingness to pay for one unit increase in k is heterogeneous and dependent on p_i. Note that this formula implies that the marginal willingness to pay is higher for households living in more expensive housing, as seems likely. However, note also that we cannot be sure that this specific type of heterogeneity is really present in the data, because it is entirely implied by the (arbitrary) assumption of a log-linear hedonic price function.

Historic amenities and house prices (5). Let's say you estimate the following linear regression using ordinary least squares:

$$\log p_{int} = \alpha_0 + \alpha_1 z_{nt} + \theta_t + \xi_{int}, \tag{3.3}$$

where p_{int} is the house price of property i in neighbourhood n in year t and z_{nt} denotes the cumulative investments in cultural heritage per km^2 in neighbourhood n in year t, and θ_t are year dummies to control for general price trends.

In Table 3.1 we report regression results. The coefficient α_1 in column (1) implies that

Table 3.1 – REGRESSION RESULTS
(Dependent variable: log of house price)

		+ Controls	*+ House f.e.*
	(1)	(2)	(3)
Investments in historic buildings	0.0389**	0.0411***	0.0151***
(in million € per km²)	(0.0160)	(0.0142)	(0.00514)
Housing control variables (17)	No	Yes	Yes
Year fixed effects	No	Yes	Yes
Property fixed effects	No	No	Yes
Observations	657,574	657,574	657,574
R^2	0.400	0.763	0.982

Notes: Standard errors are clustered at the neighbourhood level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

when investments per km^2 increase by €1 million (about 3 standard deviations), prices rise by about 12%. You then may use equation (3.2) as well as observations on prices and the value of investments z_{nt} to calculate the marginal willingness to pay for each property. However, we already anticipate that this effect may be overstated, because we do not include any control variables yet.

It may therefore be preferred to allow for more flexible heterogeneity in $\partial p_i / \partial k_i$ by letting the data tell how housing characteristics are related to prices. It is usually assumed that the observed and unobserved characteristics are additively separable. Hence, the so-called 'non-parametric' hedonic price function is then given by:[9]

$$p_i = \alpha_0 + f(k_i) + \xi_i, \tag{3.4}$$

where the house price of property i, p_i, is some unknown function $f(\cdot)$ of its observed characteristic. The unobserved characteristic is denoted by ξ_i If we want to know the marginal willingness to pay, this is given by: $\partial p_i / \partial k_i = \partial f(k_i)/\partial k_i$. How then to specify $f(k_i)$?

Here we discuss two potential ways:

1. One may use series approximation. This implies that you just add quadratic, cubic, quartic, terms of the variable of interest:[10]

$$p_i = \alpha_0 + \alpha_1 k_i + \alpha_2 k_i^2 + \ldots + \alpha_m k_i^m + \xi_i. \tag{3.5}$$

The main advantage of using series approximation is that one can use linear regression

[9] Methods have been proposed (see *e.g.* Heckman et al., 2010) to estimate 'fully non-parametric' hedonic price functions of the form $p_i = f(k_i, \xi_i)$, but those appears very hard to estimate and interpret

[10] In practice, usually one includes up to five terms, as more terms will imply multicollinearity.

techniques, because the above equation is linear in its parameters. Still, the marginal willingness to pay for a certain characteristic is highly non-linear and dependent on k_i:

$$\frac{\partial p_i}{\partial k_i} = \alpha_1 + 2\alpha_2 k_i + \ldots + m\alpha_m k_i^{m-1}. \tag{3.6}$$

However, there are several disadvantages, which makes series approximation less useful. First, it is hard to estimate equation (3.5) when k is not a scalar, but a vector (*e.g.* size, number of rooms, construction year) and if one also want to take interactions into account between those variables. The number of terms to include in the regressions will grow infeasibly large very quickly. This is known as the *curse of multi-dimensionality*. Second, because regression techniques minimise the sum of the squares of the differences between the house price and those predicted $f(\cdot)$, it will put a large weight on areas where most of the observations are. Hence, series approximation may therefore not be flexible enough and may perform badly in areas where the data is thin.

2. An alternative to estimate $f(k_i)$ is to adopt so-called *locally weighted regression* techniques. The idea is to estimate *for each property* i the following linear regression:

$$p_i = \alpha_0 + \alpha_i k_i + \xi_i, \qquad \forall i. \tag{3.7}$$

Note that α_i is now property-specific. The willingness to pay is now easy to derive as $\partial p_i / \partial k_i = \alpha_i$.

Attentive readers may remark that equation (3.7) is infeasible to estimate because it is impossible to estimate a regression for only one observation, in particular when k consists of multiple housing characteristics.

Instead, let us assume that properties that are similar in k_i also will have a similar α_i. This probably makes sense; people that live *e.g.* in houses with a similar size, with a similar construction year in a similar neighbourhood are probably more alike than people living in very different houses. We then estimate *weighted* regressions for each property i (*i.e.* 'locally' weighted regressions).

We then have to determine the weight for each regression i. This weight function is referred to as the Kernel function. Often, a Gaussian/Normal Kernel function is assumed:

$$\boldsymbol{w}_i = \frac{1}{h\sqrt{2\pi\sigma_k}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\boldsymbol{k}-k_i}{\sigma_k h}\right)^2}, \tag{3.8}$$

where $\boldsymbol{w}_i$ is a i-specific vector of weights, $\boldsymbol{k}$ is a vector of a housing characteristic for all observations, k_i is the observation of that characteristics for i and σ_k is the sample standard deviation of k. Please note that the weight is higher when two observations are more alike.

In equation (3.8) h denotes the so-called bandwidth, which essentially determines how many observations to include in the locally weighted regression of i. Let's consider two polar cases. First, when $h = 0$, the weighted regression (3.7) boils down to estimating separate OLS regressions for each observations. When $h \to \infty$, (3.7) is the same as a linear hedonic price function with identical parameters $\alpha_i = \alpha$. So what is the 'right' bandwidth? This is not an easy question; in the literature on multivariate hedonic price functions, often a value of around 3 is assumed (see Bajari and Kahn, 2005; Bishop and Timmins, 2018). However, although this value appears reasonable in practice, this value is arbitrary. Sometimes, Silverman's (1986) rule of thumb is applied or other methods are used, such as minimising a Cross-Validation or the Akaike Information Criterion (Hurvich et al., 1998). The exact determination of the bandwidth, however, reaches beyond the scope of this chapter. In general it will be small when the dataset is large and larger when the dataset is smaller.

The main advantage of locally weighted regression techniques is that they are very flexible and allow for flexible interactions between all included house and location characteristics, also in areas where data is sparser. This is important, as hedonic price functions may be highly non-linear. The main disadvantage is that the method is computationally intensive. When having many observations (*e.g.* 25 thousand housing transactions) the method is quite slow, because it implies the estimation of a regression for each of those 25 thousand observations. Moreover, as the method is quite data-intensive at the same time, it is tricky to apply the method to a small dataset. Luckily, given the surge in computer power in the last decade, estimating locally weighted regressions for large datasets becomes more feasible.

Historic amenities and house prices (6). We now continue by analysing the relationship between historic amenities and house prices *non-parametrically*. We aim to apply locally weighted regression techniques, so we estimate:

$$\log p_{int} = \alpha_{0i} + \alpha_{1i} z_{nt} + \theta_{it} + \xi_{int}, \qquad \forall i. \tag{3.9}$$

Note that the coefficients to be estimated are now property-specific (as denoted by the subscript i). We set the bandwidth equal to 3.

We show the relationship between the marginal effect of investments on house prices, denoted by α_{1i}, and the level of investments in Figure 3.2. It can be seen that the marginal effect of investments is lower once more investments have taken place. This is suggestive evidence that investments better can be spread over multiple locations, rather than investing in a few locations where the marginal effect of those investments will be lower.

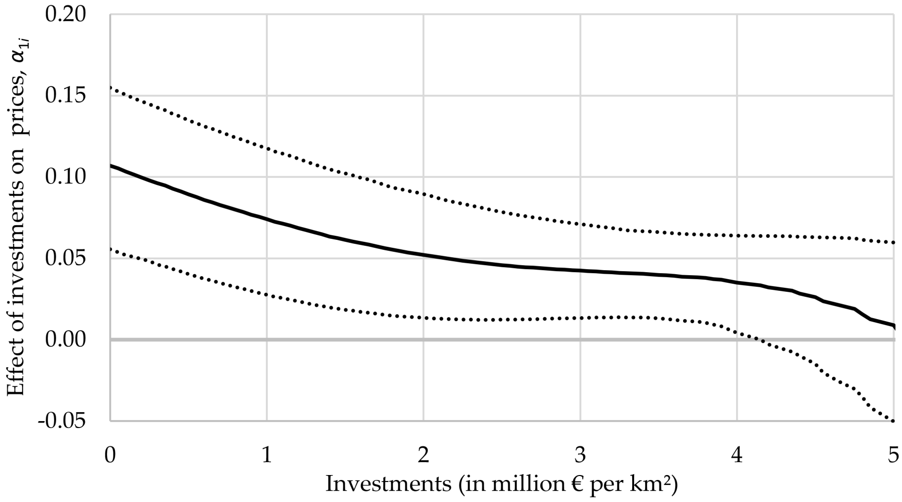


Figure 3.2 – HOUSE PRICES AND INVESTMENTS IN HISTORIC BUILDINGS
Notes: We assume a bandwidth of 3, following Bajari and Kahn (2005). The dotted lines indicate 95% confidence bands.

3.2 ISSUE 2: ENDOGENEITY

So far we have considered houses with only one characteristic k, which is unrealistic. Let's say you aim to estimate a multivariate linear hedonic price function:

$$p_i = \alpha_0 + \sum_{c=1}^{C} \alpha_c k_{ic} + \xi_i, \tag{3.10}$$

where p_i is the price of property i, k_{ic} are housing characteristics, where $c = 1, ..., C$ and ξ_i is a housing characteristic that is unobserved by the econometrician.

An important issue in estimating the preferences for housing and location is the issue of endogeneity: the researcher is unlikely to observe all characteristics that are relevant to buyers, and these omitted variables may lead to biased estimates of the marginal willingness to pay. For example, the economist may observe the size of the house, the number of rooms and the average age of buildings in the neighbourhood. However, she is unlikely to observe the quality of the trees and facades of buildings, and may for example also fail to get data on the crime rate. *If these omitted characteristics are correlated to the observed characteristic of interest*, this implies a bias (Bajari et al., 2012).[11]

There are typically three possible solutions.

1. One may try to find instrumental variables. The instrument should then be sufficiently strongly correlated to the characteristic of interest, but not directly to unobserved housing or location characteristics. In practice this is not so easy, but some examples are given in Koster et al. (2014) and Koster and Rouwendal (2017).

[11] Note that if there is no correlation between unobservable characteristics and the characteristic of interest, the marginal willingness to pay is unbiased.

2. Another solution is to include fixed effects at a low level of spatial aggregation and use temporal variation in the characteristic of interest to identify the effect of interest. Although this controls for all time-invariant characteristics of the house, price trends related to unobservable characteristics of the house may still be correlated to the characteristic of interest.[12]

3. An increasingly popular solution to the problem of omitted variable bias is the use of quasi-experimental methods. Using random variation in the determination of the treatment status to a policy one can identify the causal effect of that policy (Gibbons and Overman, 2012).[13] However, quasi-experimental settings are often not available and also have their disadvantages. For example, one only can identify the average treatment effect based on the assumption that the treatment effect is similar across the population. Furthermore, quasi-experimental settings may not be comparable to situations that occur in everyday life.

Of course, no 'one-size-fits-all' solution is available to endogeneity issues, but every study applying hedonic price techniques should carefully think about and discuss the necessary identifying assumptions under which one can identify causal effects of housing characteristics on prices.

Historic amenities and house prices (7). Let us consider including property fixed effects in the regression where we aim to measure the marginal willingness to pay for investments in historic amenities. Recall that i denotes a property in neighbourhood n sold in year t:

$$\log p_{int} = \alpha_0 + \alpha_1 z_{nt} + \sum_{c=2}^{C+1} \alpha_c k_{itc} + \theta_t + \xi_{int}, \tag{3.11}$$

As discussed before, a problem is that houses may be very different because of other reasons than historic amenity investments, even after including housing control variables k_{itc}, so there is likely omitted variable bias. A solution is to include a fixed effect for each property, implying that we subtract the property-specific mean of each variable.

---

[12]Bajari et al. (2012) recently proposed a new method to control for time-varying unobservable attributes, based on the assumption of home buyer rationality. Then, data on at least two home sales are necessary and preferences should remain relatively stable over time. In the past, the requirement of repeated sales would be probably be an insurmountable hurdle; today excellent panel-data on housing and locational transactions is increasingly available and this approach may be feasible. However, the assumption of home buyer rationality may be considered unrealistic.

[13]For example, a certain place-based policy may target certain deprived neighbourhoods based on a threshold score. By comparing houses in neighbourhoods very close to this threshold score, one approximately identifies the local treatment effect (Imbens and Lemieux, 2008). These so-called regression-discontinuity designs have gained a lot of popularity in the last decade. Examples of quasi-experimental research designs in urban economic applications include Chay and Greenstone (2005) who use the Clean Air Act to provide evidence on capitalisation of air quality into house prices, and Boes and Nüesch (2011) on the effects of aircraft noise on the housing market. Koster and Van Ommeren (2019) apply a regression-discontinuity design to measure the benefits as capitalised in house prices of place-based investments that improve the quality of public-housing.

We then may write:

$$\log p_{int} - \overline{\log p_{in}} = \alpha_1(\log z_{int} - \overline{\log z_{in}}) + \sum_{c=2}^{C+1} \alpha_c (k_{itc} - \overline{k_{ic}}) + (\theta_{it} - \overline{\theta_i}) + (\xi_{int} - \overline{\xi_{in}}),$$
$$\log p_{int} = \alpha_0 + \alpha_1 \log z_{nt} + \sum_{c=2}^{C+1} \alpha_c k_{itc} + \mu_i + \theta_t + \xi_{it}, \tag{3.12}$$

where the bar represent property-specific means and μ_i represents property fixed effects.

Why does the inclusion of fixed effects mitigate omitted variable bias? Well, say that there is an unobserved variable that only varies at the property level (*e.g.* house type, size of the house, access to open space). Subtracting the mean implies that this variable cancels out from equation because $k_{itc} = k_{ic}$. Hence, it is more likely that we isolate the causal effect of historic amenities on house prices. Note that one may also include multiple fixed effects at the same time. Specifically, we include property *and* year fixed effects.

Let's go back to Table 3.1. In column (2) we first include 17 housing control variables, such as house size, number of rooms, maintenance quality, etc. The impact of the listings rate is similar to the previous specification without control variables.

In column (3) we include property fixed effects, which means that we identify the effect of the listings rate and housing prices based on variation over time. Recall that if a variable does not change over time, it will be controlled for by the property fixed effects, see equation (3.12). Hence, property fixed effects control for difficult-to-observe characteristics of houses and locations that are time-invariant. We see that this matters for the results: a one million euro increase in investments in historic amenities increases house prices by 1.5% instead of about 4%.

One may still be worried that this effect is not a causal effect of investments, because, for example, investments mainly take place in areas with many historic buildings. Those areas may have different price trends, *e.g.* due to gentrification. In Koster and Rouwendal (2017) we therefore take a couple of additional steps to further address omitted variable bias.

4 SUMMARY

A hedonic price function is a description of the equilibrium prices of varieties of a heterogeneous good, which is influenced by supply and demand. In this chapter we explained how *hedonic pricing* techniques can be used in order to identify preferences of people for houses and locations. This is important for policy evaluation as the benefits of public investments in say clean air, improved infrastructure, open space, neighbourhoods and historic amenities, among others, can be investigated, which is something that is very hard to do otherwise.

We first discussed the micro-economic foundations of hedonic pricing techniques and showed that the marginal rate of substitution of a housing characteristic with respect to a composite good is equal to the willingness to pay for such a characteristic. Hence, by taking the derivative of the hedonic price function with respect to a characteristic, we can identify the willingness to pay for a housing or location characteristic.

Lastly, we turned our attention to two main issues when estimating hedonic pricing techniques: omitted variable bias and arbitrary functional form assumptions. We argued that one should carefully think of an identification strategy to deal with omitted variable bias bias. We further showed that semi- or non-parametric estimation techniques can be used to allow for flexible data-driven relationships between housing characteristics and the price.

Historic amenities and house prices (8). Are the total benefits of investments in historic amenities – measured by the willingness to pay – on surrounding houses larger than the costs? In Koster and Rouwendal (2017) we calculate that the total external benefits of investments in cultural heritage are €1.85 billion. This is more than the €1.63 billion of investments in cultural heritage. More specifically, the results suggest that the benefits are about 14% higher than the costs. If we take the upper bound estimated price effect, the benefits-to-costs ratio is considerably higher. These calculations provide suggestive evidence that investments in cultural heritage generate substantial positive benefits for homeowners.

REFERENCES

Ahlfeldt, G. and W. Maennig (2010). "Substitutability and Complementarity of Urban Amenities: External effects of Built Heritage in Berlin". In: *Real Estate Economics* 38.2, pp. 285–323 (cit. on p. 2).

Ahlfeldt, G., W. Maennig, and F. Richter (2017). "Urban Renewal after the Berlin Wall: A Place-based Policy Evaluation". In: *Journal of Economic Geography* 17.1, pp. 129–156 (cit. on p. 2).

Alonso, W. (1964). *Location and Land Use. Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press (cit. on pp. 6, 26).

Anderson, S. and S. West (2006). "Open Space, Residential Property Values, and Spatial Context". In: *Regional Science and Urban Economics* 36.6, pp. 773–789 (cit. on p. 2).

Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press (cit. on p. 3).

Bajari, P. and C. Benkard (2005). "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach". In: *Journal of Political Economy* 113.6, pp. 1239–1276 (cit. on p. 12).

Bajari, P. and M. Kahn (2005). "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities". In: *Journal of Business and Economic Statistics* 23.1, pp. 20–33 (cit. on pp. 12, 17, 18).

Bajari, P. et al. (2012). "A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution". In: *American Economic Review* 102.5, pp. 1898–1926 (cit. on pp. 2, 18, 19).

Bartik, T. (1987). "The Estimation of Demand Parameters in Hedonic Price Models". In: *Journal of Political Economy* 95.1, pp. 81–88 (cit. on p. 12).

Bayer, P., F. Ferreira, and R. McMillan (2007). "A Unified Framework for Measuring Preferences for Schools and Neighborhoods". In: *Journal of Political Economy* 115.4, pp. 588–638 (cit. on p. 2).

Bishop, K. and C. Timmins (2018). "Using Panel Data to Easily Estimate Hedonic Demand Functions". In: *Journal of the Association of Environmental and Resource Economists* 5.3, pp. 517–543 (cit. on pp. 12, 17).

Black, S. (1999). "Do Better Schools Matter? Parental Valuation of Elementary Education". In: *Quarterly Journal of Economics* 114.2, pp. 577–599 (cit. on p. 2).

Boes, S. and S. Nüesch (2011). "Quasi-experimental Evidence on the Effect of Aircraft Noise on Apartment Rents". In: *Journal of Urban Economics* 69.2, pp. 196–204 (cit. on p. 19).

Brueckner, J., J. Thisse, and Y. Zenou (1999). "Why is Central Paris Rich and Downtown Detroit Poor? An Amenity-based Theory". In: *European Economic Review* 43.1, pp. 91–107 (cit. on p. 3).

Cameron, A. and P. Trivedi (2005). *Microeconometrics: Methods and Application.* New York: Cambridge University Press (cit. on p. 3).

Chay, K. and M. Greenstone (2005). "Does Air Quality Matter ? Evidence from the Housing Market Michael Greenstone". In: *Journal of Political Economy* 113.2, pp. 376–424 (cit. on pp. 2, 19).

Court, A. (1939). "Hedonic Price Indexes with Automotive Examples". In: *The Dynamics of Automobile Demand*. New York: General Motors, pp. 98–119 (cit. on p. 2).

Davis, L. (2011). "The Effect of Power Plants on Local Housing Values and Rents". In: *Review of Economics and Statistics* 93.4, pp. 1391–1402 (cit. on p. 2).

Dröes, M. and H. R. A. Koster (2016). "Renewable Energy and Negative Externalities: The Effects of Wind Turbines on House Prices". In: *Journal of Urban Economics* 96, pp. 121–141 (cit. on p. 2).

Ekeland, I., J. Heckman, and L. Nesheim (2002). "Identifying Hedonic Models". In: *American Economic Review* 92.2, pp. 304–309 (cit. on p. 12).

— (2004). "Identification and Estimation of Hedonic Models". In: *Journal of Political Economy* 112.S1, S60–S109 (cit. on p. 12).

Epple, D. (1987). "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products". In: *Journal of Political Economy* 95.1, pp. 59–80 (cit. on p. 12).

Gibbons, S. (2015). "Gone with the Wind: Valuing the Local Impacts of Wind Turbines through House Prices". In: *Journal of Environmental Economics and Management* 72, pp. 177–196 (cit. on p. 2).

Gibbons, S., S. Machin, and O. Silva (2013). "Valuing School Quality using Boundary Discontinuities". In: *Journal of Urban Economics* 75.1, pp. 15–28 (cit. on p. 2).

Gibbons, S. and H. Overman (2012). "Mostly Pointless Spatial Econometrics?" In: *Journal of Regional Science* 52.2, pp. 172–191 (cit. on p. 19).

Glaeser, E., J. Kolko, and A. Saiz (2001). "Consumer City". In: *Journal of Economic Geography* 1.1, pp. 27–50 (cit. on p. 3).

Greenstone, M. and J. Gallagher (2008). "Does Hazardous Waste Matter ? Evidence from the Housing Market and the Superfund Program". In: *The Quarterly Journal of Economics* 123.3, pp. 951–1004 (cit. on p. 2).

Griliches, Z. (1961). *Hedonic Price Indexes for Automobiles: An Econometric of Quality Change*. NBER, pp. 173–196 (cit. on p. 2).

Heckman, J., R. Matzkin, and L. Nesheim (2010). "Nonparametric Identification and Estimation of Nonadditive Hedonic Models". In: *Econometrica* 78.5, pp. 1569–1591 (cit. on pp. 12, 15).

Hurvich, C., J. Simonoff, and C. Tsai (1998). "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion". In: *Journal of the Royal Statistical Society B* 60.2, pp. 271–293 (cit. on p. 17).

Imbens, G. and T. Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". In: *Journal of Econometrics* 142.2, pp. 615–635 (cit. on p. 19).

Irwin, E. (2002). "The Effects of Open Space on Residential Property Values". In: *Land Economics* 78.4, pp. 465–480 (cit. on p. 2).

Koster, H. R. A. and J. Rouwendal (2017). "Historic Amenities and Housing Externalities: Evidence from the Netherlands". In: *Economic Journal* 127, F396–F420 (cit. on pp. 2, 3, 5, 18, 20, 21).

Koster, H. R. A. and J. N. Van Ommeren (2019). "Place-based Policies and the Housing Market". In: *Review of Economics and Statistics* 101.3, pp. 1–15 (cit. on pp. 2, 19).

Koster, H. R. A., J. N. Van Ommeren, and P. Rietveld (2014). “Agglomeration Economies and Productivity: A Structural Estimation Approach using Commercial Rents”. In: *Economica* 81.321, pp. 63–85 (cit. on p. 18).

— (2016). “Historic Amenities, Income and Sorting of Households”. In: *Journal of Economic Geography* 16.1, pp. 203–236 (cit. on p. 4).

Koster, H. R. A., N. Volkhausen, and J. N. Van Ommeren (2020). “Short-term Rentals and the Housing Market: Quasi-Experimental Evidence from Airbnb in Los Angeles”. In: *CEPR Discussion Paper #13094* (cit. on p. 1).

Mills, E. (1967). “An Aggregative Model of Resource Allocation in a Metropolitan Area”. In: *American Economic Review* 57.2, pp. 197–210 (cit. on p. 6).

Muth, R. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* Chicago: University of Chicago Press (cit. on p. 6).

Rosen, S. (1974). “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition”. In: *Journal of Political Economy* 82.1, p. 34 (cit. on pp. 2, 5, 6, 9, 11, 12).

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall (cit. on p. 17).

Stock, J. and M. Watson (2011). *Introduction to Econometrics.* 3rd editio. Cambridge, MA: Pearson (cit. on p. 3).

Wooldridge, J. (2015). *Introductory Econometrics: A Modern Approach.* Scarborough ON: Nelson Education (cit. on p. 3).

GLOSSARY

The adapted definitions in this glossary are obtained from `Wikipedia`.

budget constraint In economics, a budget constraint represents all the combinations of goods and services that a consumer may purchase given current prices within her given income. Consumer theory uses the concepts of a budget constraint and a preference map to analyze consumer choices. 5, 9, 10, 11

composite good In economics, a composite good is an abstraction that represents all but one of the goods in the relevant budget. 4, 5, 7, 8, 9, 10, 21

concave In mathematics, a function is called concave if the line segment between any two points on the graph of the function lies below the graph. More specifically, if the second derivative of a variable is always negative on its entire domain, then the function is conxave. An implication is that a function is increasing or decreasing with an decreasing rate. 7, 10

control variables A control variable in a regression is a variable which is kept constant throughout the course of the investigation. Control variables could strongly influence the regression outcome, were they not held constant. The control variables themselves are not of primary interest to the econometrician . 19

control variable A control variable is a variable which is held constant in order to assess the relationship between multiple variables. Its unchanging state allows the relationship between the other variables being tested to be better understood. 15, 20

convex In mathematics, a function is called convex if the line segment between any two points on the graph of the function lies above or on the graph. More specifically, if the second derivative of a variable is always non-negative on its entire domain, then the function is convex. An implication is that a function is increasing or decreasing with an increasing rate. 7, 10, 11

endogeneity Endogeneity broadly refers to situations in which an explanatory variable is correlated with the residual. 18, 19

equilibrium In economics, an equilibrium is a situation in which economic forces such as supply and demand are balanced and in the absence of external influences the (equilibrium) values of economic variables will not change. For example, in the standard text perfect competition, equilibrium occurs at the point at which quantity demanded and quantity supplied are equal. If the market is not in equilibrium, as a result of the price being above the market price, there are natural forces at work (*i.e.* downward price changes) to bring the market back into the situation of a market equilibrium. 2, 6, 20

fixed effects A fixed effects model refers to a regression model in which the group means are fixed. Generally, data can be grouped according to several observed factors (*e.g.*

locations, households). In a fixed effects model each group mean is a group-specific fixed quantity. 19, 20

heterogeneous A product is homogeneous is uniform in composition or character (*i.e.* color, shape, size, weight, height, distribution, texture, architectural design, etc.); one that is heterogeneous is distinctly nonuniform in one of these qualities. 2, 3, 5, 6, 7, 8, 9, 10, 14, 20, 29

indifference curve An indifference curve shows any combinations of two products that will provide the consumer with equal levels of utility. The consumer has no preference for one combination or bundle of goods over a different combination on the same curve. 6, 7, 8, 10

Kernel A Kernel regression is a non-parametric technique in statistics to estimate the conditional expectation of a random variable. The objective is to find a non-linear relation between random variables X and Y. 16

locally weighted regression Locally weighted regression, also known as moving regression, is a generalisation of moving average regressions. Its most common methods, initially developed for scatterplot smoothing, are LOESS (locally estimated scatterplot smoothing) and LOWESS (locally weighted scatterplot smoothing). They are two strongly related non-parametric regression methods that combine linear regression models in a k-nearest-neighbor-based meta-model. 16, 17

log-linear A log-linear model is a mathematical model that takes the form of a function whose logarithm equals a linear combination of the parameters of the model, which makes it possible to apply (possibly multivariate) linear regression. 13, 14

marginal willingness to pay The marginal willingness to pay (MWTP) is the maximum amount of money a consumer is willing to pay for the last unit of a particular characteristic of a product she consumes. If the product is supplied on a market at a given price per unit a consumer is willing to pay another unit as long as her marginal willingness to pay exceeds the price. This is simular for characteristics of a heterogeneous commodity. For instance, a consumer prefers a larger house as long as het marginal willingness to pay exceeds the additional amount of money – the marginal price – that has to be paid for a house with one additional square meter floor area. 1, 2, 3, 4, 5, 7, 8, 9, 13, 14, 15, 16, 18, 19

misspecification Misspecification occurs when the functional form or the choice of independent variables poorly represent relevant aspects of the true data-generating process. 4, 13

monocentric city model Since its formulation in 1964, Alonso's monocentric city model of a disc-shaped Central Business District (CBD) and the surrounding residential region has served as a starting point for urban economic analysis. 6

multicollinearity Multicollinearity (also collinearity) is a phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. 15

MWTP See *marginal willingness to pay*. 1, 2, 3, 4

non-parametric A non-parametric test (sometimes called a distribution free test) does not assume anything about the underlying distribution (for example, that the data comes from a normal distribution). That is compared to parametric tests, which makes assumptions about the parameters (for example, that an explanatory variable relates linearly to a dependent variable). 12, 15

omitted variable bias Omitted variable bias (OVB) occurs when a statistical model leaves out one or more relevant variables. The bias results in the model attributing the effect of the missing variables to the estimated effects of the included variables. 3, 4, 19, 20, 21

ordinary least squares Ordinary least squares (OLS) is a type of linear least squares method for estimating the unknown parameters in a linear regression model. OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function. 3, 14

regression Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors', 'covariates', or 'features'). The most common form of regression analysis is linear regression, in which a researcher finds the line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion. 3, 11, 14, 15, 16, 17

residual residuals capture the deviation of an observed value of an element of a statistical sample from its 'theoretical value'. More specifically, the residual is the difference between the observed value and the estimated value of the quantity of interest (for example, a sample mean). 5

scalar A scalar is a number and is used to indicate a variable. In a scalar you can save one value, number or string. 4, 16

total differential The total derivative of a function $f(\cdot)$ at a point is the best linear approximation near this point of the function with respect to its arguments. Unlike partial derivatives, the total derivative approximates the function with respect to all of its arguments, not just a single one. In many situations, this is the same as considering all partial derivatives simultaneously. 6, 9

utility In economics, the concept of utility is used to model worth or value. Its usage has evolved significantly over time. The term was introduced initially as a measure of pleasure or satisfaction within the theory of utilitarianism by moral philosophers such as Jeremy Bentham and John Stuart Mill. The term has been adapted within economics, as a utility function that represents a consumer's preference ordering over a choice set. 1, 4, 5, 6, 7, 8, 9, 29

utility function In economics, a utility function is an important concept that measures preferences over a set of goods and services. Utility represents the satisfaction that consumers receive for choosing and consuming a product or service. 1, 4, 6, 10, 29

vector A vector contains multiple numbers or scalars. In the context of hedonic price function, it means that one house containes multiple housing characteristics. We may summarise this in a vector $k = k_0, k_1, ..., k_I$. 4, 16

willingness to pay Willingness to pay (WTP) is the maximum price at or below which a consumer will definitely buy one unit of a product. This corresponds to the standard economic view of a consumer reservation price. See also *marginal willingness to pay*. 6, 7, 8, 9, 16, 21, 29

APPENDIX

A.1 DERIVATION OF THE VALUE FUNCTION

Let P denote the willingness to pay for the heterogeneous commodity. We substitute housing expenditures in (2.2), and use the resulting equation to rewrite the utility function (2.1) as:

$$u = u(y - P, k). \tag{A.1}$$

We now ask the question: what is the price the consumer is willing to pay for quality k if her utility must be equal to a predetermined value, say u^*? We then write:

$$u(y - P, k) = u^*, \tag{A.2}$$

and invert the utility function with respect to $y - P$ as to get $y - P$ on the left-hand side. How this works is best illustrated with a simple example of a Cobb-Douglas utility function:

$$\ln u^* = \alpha \ln(y - P) + (1 - \alpha)\ln k. \tag{A.3}$$

When we put P to the left-hand side we obtain:

$$\begin{aligned} \ln(y-P) &= \frac{1}{\alpha}\big(\ln u^* - (1-\alpha)\ln k\big), \\ \ln(y-P) &= \frac{1}{\alpha}\ln u^* + \frac{\alpha-1}{\alpha}\ln k, \\ y - P &= (u^*)^{\frac{1}{\alpha}} k^{\frac{\alpha-1}{\alpha}}, \\ P &= y - (u^*)^{\frac{1}{\alpha}} k^{\frac{\alpha-1}{\alpha}}, \end{aligned} \tag{A.4}$$

which is equivalent to $P = y - u^{-1}(u^*, k)$. We then totally differentiate this equation, which implies that you investigate the change in the price p as a results of another variable (say k), holding other variables constant. For example, for equation (2.11), we have:

$$\frac{\partial u}{\partial q}\mathrm{d}(y - P) + \frac{\partial u}{\partial k}\mathrm{d}k = 0, \tag{A.5}$$

because $q = y - P$ and u^* are held constant. Because y is also held constant, this means that $\mathrm{d}(y - P) = -\mathrm{d}P$. Using this and rearranging yields:

$$\begin{aligned} \frac{\partial u}{\partial q}\mathrm{d}(-P) + \frac{\partial u}{\partial k}\mathrm{d}k &= 0, \\ \frac{\mathrm{d}P}{\mathrm{d}k} &= \frac{\partial u/\partial k}{\partial u/\partial q}. \end{aligned} \tag{A.6}$$

Equations (2.10) and (2.11) can be derived in a similar fashion.