On the theoretical derivation of a functional form for the hedonic price function

Lisi, Gaetano

CreaM Economic Centre (University of Cassino)

26 February 2012

Online at https://mpra.ub.uni-muenchen.de/37066/
MPRA Paper No. 37066, posted 03 Mar 2012 19:03 UTC
On the Theoretical Derivation of a Functional Form for the Hedonic Price Function

Gaetano Lisi

Abstract

The key issue in the hedonic price theory is that although the literature emphasises the intrinsic nonlinearity in the relationship between house prices and housing characteristics, very little theoretical guidance is provided regarding the more appropriate mathematical specification for the hedonic price function. Thus, most empirical studies make use of flexible functional forms or simple linear models which possess a direct economic meaningfulness. This theoretical paper fills this gap by using the Mortensen-Pissarides matching model to show both the nonlinearity of the hedonic price function and the more appropriate functional relationship between prices and attributes.

Keywords: hedonic price theory, hedonic price function, search and matching frictions

JEL Classification: R21, R31, J63
1. Introduction

Although the economic theory of hedonic prices (Lancaster, 1966; Rosen, 1974) is well known and not in question,\(^1\) it provides very little theoretical guidance on the appropriate functional relationship between prices and attributes in the hedonic price function (Malpezzi, 2003; Taylor, 2003), and thus in empirical studies researchers have used flexible functional forms, such as Box-Cox functions, or simple parametric models (Anglin and Gençay, 1996).\(^2\)

The hedonic price literature almost unanimously underlines the intrinsic nonlinearity in the relationship between house prices and housing characteristics, though nothing is known a priori about a specific functional form (Parmeter, Henderson and Kumbhakar, 2007; Haupt, Schnurbus and Tschernig, 2010). Nevertheless, while the literature suggests that the equilibrium price function is nonlinear, most empirical studies make use of linear models, thus relying on an influential simulation study by Cropper, Deck and McConnell (1988).\(^3\) This is due to the absence of theoretical groundwork regarding the more appropriate functional form to use in the hedonic price models. According to Rosen (1974), there is no reason for the hedonic price function to be linear; in fact, the linearity of the hedonic price function is unlikely as long as the marginal cost of attributes increases for sellers and it is not possible to untie packages. Indeed, Ekeland et al. (2004) demonstrate that nonlinearity is a generic property of the hedonic price function. Hence, a linear model would be a special case for the hedonic price function (Kuminoff, Parmeter and Pope, 2008, 2009). However, the nonlinearity is basically a general concept and may imply the use of several kinds of empirical models.

As a rule, the use of a particular empirical model rather than another should be indicated by economic theory (Stock and Watson, 2003). Indeed, theoretical models are critical in determining an accurate and consistent econometric model: empirical analysis

---

\(^1\) For an exhaustive overview see Sheppard (1999) and Malpezzi (2003).

\(^2\) Often linear, semi-logarithmic or log-log models are chosen. These are characterised as being easily interpretable, and the estimated parameters possess a direct economic meaningfulness (Maurer, Pitzer, and Sebastian, 2004). In particular, in the linear model, the parameters give absolute prices for the unit of the attributes.

\(^3\) Cropper, Deck and McConnell (1988) found that when all attributes are observed, linear and quadratic Box-Cox forms produce the most accurate estimates of marginal attribute prices; whereas, when some attributes are unobserved or are replaced by proxies, linear and linear Box-Cox functions perform best.
alone cannot replace conceptual reasoning when estimating the relationships of most economic phenomena (Can, 1992; Brown and Ethridge, 1995).

This theoretical paper fills this gap by using the Mortensen-Pissarides matching model (see e.g. the textbook by Pissarides, 2000) to show both the nonlinearity of the hedonic price function and the more appropriate functional relationship between house prices and housing characteristics. In particular, under the realistic assumption of decentralised housing markets with important search and matching frictions (Leung and Zhang, 2011), in this model the equilibrium price function is nonlinear with a closed-form solution.

Furthermore, the proposed housing market matching model allows a major drawback of the standard hedonic pricing theory to be overcome: the assumption of competitive markets. Indeed, in the standard hedonic pricing theory, markets are assumed to be sufficiently thick (i.e. markets with a large amount of trading) so that implicit or hedonic prices, i.e. the shadow prices of the characteristics, are revealed to economic agents through trades that differ only in terms of a single attribute. However, this is hardly true: markets become increasingly thin when traded goods are increasingly heterogeneous, and the implicit or hedonic prices as well as the "true" market value of the good are not known (Harding, Rosenthal and Sirmans, 2003; Harding, Knight and Sirmans, 2003; Cotteleeer and Gardebroek, 2006). Indeed, the house price realistically depends not only on the housing characteristics but also on the search and matching frictions and bargaining power of the parties.

The rest of the paper is organised as follows: section 2 presents the housing market matching model; section 3 gives insights on the more appropriate functional form to use in the hedonic price models, while section 4 shows the empirical plausibility of the theoretical result; finally, section 5 concludes the work.

2. A baseline matching model of housing market

We adopt a standard matching framework à la Mortensen-Pissarides (see e.g. Pissarides, 2000) with random search and prices determined by Nash bargaining. We believe that the behaviour of the housing market can be properly formalised by the Mortensen-Pissarides
matching model. Indeed, the random matching assumption is absolutely compatible with a market where the formal distinction between the demand and supply side is very subtle; whereas, bargaining is a natural outcome of thin, local and decentralised markets for heterogeneous goods.

Since we are interested in selling price, the market of reference is the homeownership market rather than the rental market. In this market, if a contract is legally binding (as hypothesised) it is no longer possible to return to the circumstances preceding the bill of sale, unless a new and distinct contractual relationship is set up. In matching model jargon this means that the destruction rate of a specific buyer-seller match does not exist and the value of an occupied home for a seller is simply given by the selling price.

Buyers \( b \) expend costly search effort to find a house, while sellers \( s \) hold \( h \geq 2 \) houses of which \( h-1 \) are on the market, i.e. vacancies \( v \) are simply given by \( v = (h-1) \cdot s > 0 \). It is therefore possible that a buyer can become a seller, and that a seller can become a buyer. Indeed, buyers today are in fact potential sellers tomorrow (Leung, Leong and Wong, 2006).

The expected values of a vacant house \( V \) and of buying a house \( H \) are given by:

\[
\begin{align*}
  rV &= -a + q(\theta) \cdot [P - V] \\
  rH &= -e + g(\theta) \cdot [x - H - P]
\end{align*}
\]

where \( \theta \equiv v/b \) is the housing market tightness from the sellers' standpoint, while \( q(\theta) \) and \( g(\theta) \) are, respectively, the (instantaneous) probability of filling a vacant house and of buying a home. The standard hypothesis of constant returns to scale in the matching function, \( m = m[v, b] \), is adopted (see Pissarides, 2000; Petrongolo and Pissarides, 2001), since it is also used in the (recent) search models of housing market (Diaz and Jerez, 2009; Novy-Marx, 2009; Piazzesi and Schneider, 2009; Genesove and Han, 2010; Leung and

\[a\]Matters thus become simpler without loss of generality. Alternatively, one could assume that the sellers hold \( h \geq 1 \) houses of which \( h \) are on the market, and that the buyers are the homeless. This case would not change the results of the analysis.

\[b\]Time is continuous and individuals are risk neutral, live infinitely and discount the future at the exogenous interest rate \( r > 0 \). As usual in matching-type models, the analysis is restricted to the stationary state.
Hence, the properties of these functions are straightforward: \( q'(\theta) < 0 \) and \( g'(\theta) > 0 \). The terms \( a \) and \( e \) represent, respectively, the costs sustained by sellers for the advertisement of vacancies and the effort (in monetary terms) made by buyers to find and visit the largest possible number of houses. If a contract is stipulated, the risk neutral buyer gets a linear benefit \( x \) from the property (abandoning the home searching value) and pays the sale price \( P \) to the seller (who abandons the value of finding another buyer). The buyer’s benefit depends on the value of the house: in fact, the greater the quality/quantity of housing characteristics (i.e. the value of the house), the higher the buyer’s benefit.

The endogenous variables that are determined simultaneously at equilibrium are market tightness (\( \theta \)) and sale price (\( P \)). The “zero-profit” or “free-entry” condition normally used by matching models (see Pissarides, 2000) yields the first key relationship of the model, in which market tensions are a positive function of price. By using the equilibrium condition \( V = 0 \) in [1], we obtain:

\[
\frac{1}{q(\theta)} = \frac{P}{a}
\]

This positive relationship is very intuitive (recall that \( q'(\theta) < 0 \)): in fact, if the price increases, more vacancies will be on the market.

Instead, the selling price is obtained by solving the following optimisation condition, the so-called Nash bargaining solution usually used for decentralised markets:

\[
P = \arg \max \{ P - V \} \cdot (x - H - P)^{1-\gamma}
\]

with \( V = 0 \) \( \Rightarrow P = \frac{V}{(1-\gamma)} \cdot (x - H - P) \)

\( \Rightarrow P = \gamma \cdot (x - H) \)

---

\(^{6}\) The main difference between our model and the quoted studies is that we closely track the standard matching framework \( \text{à la} \) Mortensen-Pissarides, without any deviation from the baseline model.

\(^{7}\) Standard technical assumptions are postulated: \( \lim_{\theta \to -\infty} q(\theta) = \lim_{\theta \to +\infty} g(\theta) = \infty \), and \( \lim_{\theta \to -\infty} g(\theta) = \lim_{\theta \to +\infty} q(\theta) = 0 \). By definition, markets with frictions require positive and finite tightness, i.e. \( 0 < \theta < \infty \), since for \( \theta = 0 \) the vacancies are always filled, whereas for \( \theta = \infty \) the home-seekers immediately find a vacant house.
where \( 0 < \gamma \leq 1 \) is the bargaining power of sellers. By using the previous result \((x-H-P)=(1-\gamma)P\) in equation [2], eventually we get:

\[
P = \frac{\gamma \cdot (rx + e)}{r + g(\theta)(1 - \gamma)}
\]

Hence, if \( \gamma = 1 \), then \( P = x + \frac{e}{r} > x \); while it is unrealistic to assume \( \gamma = 0 \), since the price cannot be negative or null: in fact, entering into a contractual agreement obviously implies that \( x > B, \forall \theta \). As regards the economic meaning of equation [4], if the market tightness increases, the effect of the well-known congestion externalities on the demand side (see Pissarides, 2000) will lower the price (recall that \( g'(\theta) > 0 \)).

This simple model is able to reproduce the observed joint behaviour of prices and time-on-the-market (see e.g. Leung, Leong, and Chan, 2002). In fact, with a probability of filling a vacant house of \( q(\theta) \), the (expected) time-on-the-market is \( q(\theta)^{-1} \). Hence, from equation [3], the house with a higher selling price has a longer time on the market; whereas, from equation [4], the longer the time-on-the-market the lower the sale price (since \( q(\theta)^{-1} \) is increasing in \( \theta \)).

It is straightforward to obtain from [3] that when \( P \) tends to zero (infinity), \( \theta \) tends to zero (infinity), as \( q(\theta) \) tends to infinity (zero). Consequently, given the negative slope of [4] and the fact that price is always positive, only one long-term equilibrium deriving from the intersection of the two curves exists in the model. Finally, normalising the population in the housing market to the unit, \( 1 = b + s \), and using the definition of vacancies, \( v = (h-1) \cdot s \), and the value of equilibrium tightness \( (\theta \equiv \theta^* = v/b) \), the model is closed in a very simple manner.

3. Hedonic price and functional form specification

The two key equations of the model are the free-entry condition, i.e. equation [3], and the Nash bargaining solution, i.e. equation [4]. Indeed, the latter is none other than the hedonic price function of the model. As suggested by the hedonic price theory (Rosen,
the selling price is a (positive) function of housing characteristics. From [4], in fact, \( P \) depends positively on \( x \), which in turn depends positively on the housing characteristics (the value of the house). Hence, the hedonic or implicit price is positive and the equilibrium hedonic price function has a closed-form solution.

However, unlike the standard hedonic price theory, the sale price of this model depends not only on the housing characteristics but also on the market tensions, bargaining power of the parties and search costs. In particular, market tensions are an endogenous variable of the model. Hence, in order to express the hedonic price function only in terms of exogenous variables we need to combine the equations [3]-[4]. By using a Cobb-Douglas functional form (also used by Novy-Marx, 2009; Piazzesi and Schneider, 2009; Peterson, 2012), i.e. \( m = v^{1-a} \cdot b^a \), where \( \alpha \) is the elasticity of the matching function with respect to the share of buyers, we get the following implicit function for selling price:

\[
\left[ r + \left( \frac{P}{a} \right)^{\frac{1-a}{a}} \cdot (1-y) \right] \cdot P = y rx + ye
\]

being \( q(\theta) = \frac{v^{1-a} \cdot b^a}{v} = \theta^{-a} \), \( g(\theta) = \frac{v^{1-a} \cdot b^a}{b} = \theta^{1-a} \), and \( \theta = \left( \frac{P}{a} \right)^{\frac{1}{a}} \) from [3]. Total differentiation of equation [5] with respect to \( P \) and \( x \) thus yields:

\[
\frac{dP}{dx} = \frac{yr}{\left[ r + \left( \frac{P}{a} \right)^{\frac{1-a}{a}} \cdot (1-y) + P \cdot \left( \frac{1-a}{a} \right) \cdot \left( \frac{P}{a} \right)^{\frac{1-a}{a} - 1} \cdot (1-y) \right]} \cdot dP = yr \cdot dx
\]

For the sake of simplicity, we use a reasonable and common value of \( \alpha = 0.5 \).\(^8\) Hence, the hedonic or implicit price of this model collapses to:

\[
\frac{dP}{dx} = \frac{yr}{\left[ r + \left( \frac{P}{a} \right)^{\frac{1-a}{a}} \cdot (1-y) + P \cdot \left( \frac{1-a}{a} \right) \cdot \left( \frac{P}{a} \right)^{\frac{1-a}{a} - 1} \cdot (1-y) \right]}
\]

\(^8\) See Petrongolo and Pissarides (2001). Indeed, Piazzesi and Schneider (2009) use a very similar value for the U.S. housing market, namely 0.57. It is straightforward to show that the result holds for any value of \( \alpha \).
As a result, the hedonic price function is non-linear even if the buyer is risk neutral and acquires a linear benefit from the property: in fact, the implicit or hedonic price $p$ depends on $x$, since $p = f(P)$ and $P = f(x)$. This is in line with the hedonic price literature which suggests that the equilibrium price function should be nonlinear. Furthermore, we may also state that $\frac{d^2 p}{dx^2} \equiv p''(x) < 0$, since the selling price is increasing in the house value (namely, the hedonic price is positive). Hence, this theoretical model also gives a precise statement about the form of the hedonic price function: it in fact suggests an increasing relationship at decreasing rates between selling price and housing characteristics.

4. Empirical testing

In order to test the empirical plausibility of an increasing relationship at decreasing rates between selling price and housing characteristics, we rely on the (structure of the) benchmark parametric model proposed by Anglin and Gençay (1996) – and also considered by Parmeter, Henderson and Kumbhakar (2007), and Haupt, Schnurbus and Tschernig (2010).\(^9\)

The Anglin-Gençay benchmark parametric model is characterised by many binary variables and the relationship between the dependent variable (selling price), the continuous regressor (the lot size) and the discrete variables is represented in terms of relative changes (elasticity). Data on housing characteristics, in fact, typically consists of one continuous regressor (the lot size) and many ordered and unordered categorical variables (Parmeter, Henderson and Kumbhakar, 2007; Haupt, Schnurbus and Tschernig, 2010). Indeed, it is possible to transform all the ordered categorical variables into unordered categorical variables (binary regressors), thus saving degrees of freedom. Furthermore, the use of many binary regressors may be very useful to determine the most appropriate functional form, since dummy variables, by definition, cannot be transformed.

\(^9\) Haupt, Schnurbus and Tschernig (2010) show that the null hypothesis of correct specification of the linear parametric model proposed by Anglin and Gençay (1996), against the alternative of parametric misspecification, cannot be rejected at any reasonable significance level. Furthermore, they also show that the parametric model predicts better than the nonparametric specification proposed by Parmeter, Henderson and Kumbhakar (2007).
Given our data availability, the econometric model is thus the following:\(^{10}\)
\[
\ln(p_i) = \beta_0 + \beta_1 \cdot \ln(LOT_i) + \beta_2 \cdot \ln(BTMS_i) + \beta_3 \cdot \ln(WIN_i) + \beta_4 \cdot STAT_i + \beta_5 \cdot LOC_i + \beta_6 \cdot ARS_i + \beta_7 \cdot QVW_i + \beta_8 \cdot ELV_i + \beta_9 \cdot NWC_i + \epsilon_i
\]

where \(p_i\) is the selling price of house \(i\); \(STAT, LOC, ARS, QVW, ELV,\) and \(NWC\) are dummy (binary) variables for excellent state of real estate unit \((1 = yes; \ 0 = no)\), valuable area \((1 = yes; \ 0 = no)\), elegant architectural style \((1 = yes; \ 0 = no)\), excellent quality of view \((1 = yes; \ 0 = no)\), presence of elevator \((1 = yes; \ 0 = no)\), and new construction \((1 = yes; \ 0 = no)\); \(BTMS\) and \(WIN\) are the number of bathrooms and windows, respectively; and \(LOT\) is the lot size (in square feet).

Neglecting the binary variables,\(^{11}\) we focus on \(\beta_j\), with \(j = 1,2,3\). It follows that with \(0 < \beta_j < 1\) the relationship is increasing at decreasing rates, while with \(\beta_j > 1\) the relationship is increasing at increasing rates, finally with \(\beta_j = 1\) the relationship is linear.

Using data on several Italian local property markets, the OLS results show that the coefficients \(\beta_j\) have positive signs and are often statistically significant, i.e. \(\beta_j \neq 0\) (in particular, \(\beta_1\) always has a strong statistical significance). As regards the coefficients \(\beta_2\) and \(\beta_3\), they range between 0.072 and 0.277 and the null hypothesis of \(\beta_j = 1\) is rejected at any reasonable significance level (thus confirming the nonlinearity of the hedonic price function); whereas, the value of coefficient \(\beta_1\) ranges between 0.761 and 1.025 and thus in some cases the null hypothesis of \(\beta_1 = 1\) can not be rejected. Nevertheless, it is very likely that the coefficient of lot size “absorbs” the effect of other quantitative variables which depend on the house size (number of bedrooms, garage places, stories, etc.), thus being overestimated. Indeed, in a regression without “number of windows” and “number of bathrooms”, the coefficient of lot size is always higher and ranges between 0.835 and 1.173. Hence, an increasing relationship at decreasing rates may be the most appropriate functional form for the hedonic price function (as suggested by the theoretical model).

\(^{10}\) It has not been possible to closely track the Anglin-Gençay model because of the different database used.

\(^{11}\) The coefficients for the binary variables give the surcharge which is to be paid relative to a property without those attributes. For more details about the economic interpretation of the effect of dummy variables on the dependent variable in natural logarithmic form see Halvorsen and Palmquist (1980).
Finally, this theoretical framework may be used to study how errors in measuring marginal attribute prices vary with the form of the hedonic price function; in this way, the simulation strategy developed by Cropper, Deck and McConnell (1988), and updated by Kuminoff, Parmeter and Pope (2008, 2009), may take the (equilibria of the) housing market with search and matching frictions into account, thus relaxing the unrealistic assumption of competitive housing markets.12

5. Conclusions

The hedonic price literature emphasises the intrinsic nonlinearity in the relationship between house prices and housing characteristics, though nothing is known a priori about a specific functional form. Indeed, the economic theory of hedonic prices provides very little theoretical guidance on the appropriate functional relationship between prices and attributes in the hedonic price function. This is a very significant shortcoming for empirical studies, since theoretical models are critical in determining accurate and consistent econometric models and the use of a particular empirical model rather than another should be indicated by economic theory. As a consequence, most empirical studies make use of flexible functional forms or simple models which possess a direct economic meaningfulness. This theoretical paper fills this gap by using the Mortensen-Pissarides matching model for insights regarding the nonlinearity of the hedonic price function. In particular, by relaxing the unrealistic assumption of competitive housing markets, the relationship between selling price and housing characteristics is increasing at decreasing rates.

References


12 In the quoted studies, the marginal bid of consumers (namely, the “true” marginal price paid) for each attribute is obtained by simulations of housing market equilibria. Subsequently, equilibrium housing prices, together with housing attributes, provide the data used to estimate various functional forms for the hedonic price function. Finally, errors in estimating marginal prices are calculated by comparing the consumer’s equilibrium marginal bid with the gradient of the hedonic price function.


