INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE

Unité d'Economie et Sociologie Rurales

Centre de Recherches de
l'INRA de TOULOUSE
Auzeville - B.P. 27
31326 CASTANET-TOLOSAN CEDEX
Tel : 33 (0)05.61.28.53.52
Fax :33.(0)05.61.28.53.72
http://www.toulouse.inra.fr/esr/

2006
Cahier de Recherche
2006-02

Semiparametric hedonic price models : assessing the effects of agricultural nonpoint source pollution

Christophe Bontemps
Michel Simioni
Yves Surry

Modèles de prix hédoniques semi paramétriques : évaluation des effets des pollution diffuses d'origine agricole

Christophe Bontemps
Michel Simioni
Yves Surry

Résumé

L'objectif de cet article est d'évaluer l'impact sur les prix des résidences de la pollution d'origine agricole en utilisant différentes spécifications semi paramétriques d'un modèle de prix hédoniques. Les spécifications proposées se décomposent en deux parties : une partie linéaire dans les caractéristiques des maisons et une partie non ou semi paramétrique capturant les effets non linéaires des indicateurs de pollution. Une application à un échantillon de ventes de résidences dans les communes rurales bretonnes montre que la pollution résultant de pratiques intensives dans l'élevage influe de façon significative et non linéaire sur les prix des maisons.

Mots Clés

Pollution agricole ; Agriculture intensive ; Prix ; Test statistique ; Econométrie ; Maison individuelle

Semiparametric Hedonic Price Models : Assessing the Effects of Agricultural Nonpoint Source Pollution[1]

Christophe Bontemps,
University of Toulouse, INRA-GREMAQ, France
christophe.bontemps@toulouse.inra.fr

Michel Simioni,
University of Toulouse, INRA-GREMAQ and IDEI, France
michel.simioni@toulouse.inra.fr

and

Yves Surry,
Swedish University of Agricultural Sciences, Uppsala, Sweden
Yves.Surry@ekon.slu.se

February 2007

[1]We thank Anna Alberini, Gabor Kezdi, Pascal Lavergne, Céline Nauges, and participants of the EAERE thirteenth annual conference, Budapest, of the XIth EAAE congress, Copenhagen, and of the AAEA 2005 annual meeting, Providence, for their remarks. All remaining errors are ours.

Semiparametric Hedonic Price Models :
Assessing the Effects of Agricultural Nonpoint Source Pollution

Abstract:

In the area of environmental analysis using hedonic price models, we investigate the performance of various nonparametric and semiparametric specifications. The proposed model specifications are made up of two parts: a linear component for house characteristics and a non(semi)parametric component representing the nonlinear influence of environmental indicators on house prices. We adopt a general-to-specific search procedure, based on recent specification tests comparing the proposed specifications with a fully nonparametric benchmark model, to select the best model specification. An application of these semiparametric models to rural districts indicates that pollution resulting from intensive livestock farming have a significant nonlinear impact on house prices.

1 Introduction

Intensive agricultural activities generate negative externalities that are becoming increasingly significant over time and space. This phenomenon is reflected in growing concerns about the impacts of intensive livestock farming in rural areas, with populations becoming denser and more urbanized. Agricultural economists have attempted to quantify such effects, using hedonic models of house prices. In such a framework, negative external effects from agriculture (pollution) are measured by relevant indicators that are assumed to be inversely related to house prices. Then, by estimating the first-order derivatives of the hedonic price function with respect to the pollution indicators, we obtain estimates of the prices of these environmental attributes and, indirectly, an estimate of the consumers' willingness to pay for these nuisances (or disamenities). Although well grounded theoretically, the hedonic price model, when implemented empirically, has raised several problems associated with the identification of the parameters in the underlying structural model. In reality, the nature of the relationship between house prices and the various associated attributes is complex and nonlinear, so it would be better represented by nonparametric models rather than the classical parametric specifications (Ekeland *et al.*, 2004).

The aim of this paper is to investigate the performance of various non- and semi-parametric specifications in a conventional hedonic price model. While most of the literature has concentrated on the parametric specifications of the hedonic price model, some recent studies have assessed the advantages of some non- and semi-parametric methods.[2] Comparative studies suggest that these latter methods fit the data better than parametric specifications. In the present study, we propose a twofold approach to research in this field. First, we consider four different specifications (fully nonparametric, nonparametric additive, single-index and parametric). Second, we compare the performances of the three restricted specifications (nonparametric additive, single-index and parametric) with a more general fully nonparametric specification. Thus, our work differs from previous studies by considering the fully nonparametric model as the benchmark, and then performing tests to compare the different specifications against this benchmark.

The empirical application reported here concerns a set of transaction

[2] See for instance Stock (1991), Pace (1993 and 1998), Gençay and Yang (1996), Anglin and Gençay (1996), Iwata *et al.* (2000), McMillen and Thornes (2000), Bin (2004) and Martins-Filho and Bin (2005).

prices for residential houses sold during 1996 and 1997 in Brittany, which is the leading French region for a number of livestock and plant products. Agriculture in this region has two main impacts on the environment. First, the activities of intensive livestock units lead to harmful effects on the environment in various forms, such as the production of unpleasant odours and the release/emission of nitrate that pollutes the soil, affects water quality and seeps into the groundwater. The second effect of agriculture on the environment concerns the degradation of the rural landscape resulting from intensive agricultural practices and activities. In our study, these two effects are assessed by two aggregate environmental indicators: livestock nitrogen emissions per hectare of arable land in rural districts where the residential houses are located, and the proportion of permanent pasture land converted into tilled land.[3]

As in other hedonic price models, we specify the prices of residential houses not only in terms of their physical characteristics and the environmental indicators, but also variables representing the economic structure of the rural districts where the residential houses are located. We choose a partially linear specification, in which all the explanatory variables, apart from the two environmental indicators, are incorporated linearly into the hedonic price function. The two pollution indicators are included in the hedonic price function in a nonparametric or semiparametric way using the three specifications mentioned above. This choice is driven by practical reasons, since many housing characteristics are discrete variables and our empirical objective is to measure the impact of environmental factors on residential housing prices by focusing on possible nonlinearities. Moreover, as shown further below, the specification tests used here involve specifications with only two explanatory variables.

The empirical strategy used here to define the housing price model consists of a general-to-specific specification search involving three stages. In the first stage, the parameters involved in the linear part of the hedonic price models are estimated using Robinson's (1988) partially linear model approach. In the second stage, all four specifications of the nonlinear component of the hedonic price function are determined using the estimated resid-

[3]In studies analysing this problem, impacts of agricultural pollution on house prices are measured using a proximity index reflecting the distance between residential houses and the sites of agricultural pollution (Palmquist et al., 1997; Ready and Abdalla, 2005). As explained further in this paper, this approach cannot be adopted due to the lack of relevant information on the localization of house sales and the sources of agricultural pollution.

uals of the first-stage estimation procedure. The econometric procedures used involve local polynomial regression (Fan and Gijbels, 1996), average derivative estimation (Horowitz and Härdle, 1996) and marginal integration (Linton and Nielsen, 1995). Finally, we perform specification tests similar to those of Fan and Li (1996), Horowitz and Spokoiny (2001) and Gozalo and Linton (2001).[4] The nonparametric additive, single-index and parametric specifications are thus compared with the more general specification, which is fully nonparametric.

The specification tests only select the nonparametric additive specification. Willingness to pay for pollution reduction is then computed for this selected model specification using the procedure for estimating derivatives for additive separable models, as proposed by Severance-Lossin and Sperlich (1999). Moreover, we find that pollution resulting from livestock farming in rural districts is a more crucial environmental issue than pollution due to intensive crop practices, although both have a significant but nonlinear effect on residential housing prices.

The paper is organized as follows: semiparametric house price models are defined in Section 2, the general-to-specific specification search procedure is described in Section 3 and the data are presented in Section 4, while the results of the proposed models are examined and discussed in Section 5. The conclusion highlights the main findings of this empirical exercise.

2 Semiparametric house price models

In this section, we discuss the different specifications for a hedonic residential house price model. This model can be defined as follows. Assume that each residential house can be regarded by economic agents as a "bundle" of different amounts of a vector expressing the various characteristics. All these characteristics are observed by economic agents when making their choices. In the following, we assume that econometricians only observe some of these characteristics, denoted here by a vector X, when considering J characteristics of the house and its surroundings (e.g. number of rooms, state of repair, age of house, population of the rural district, stock of existing houses, etc.). We use a vector Z when considering L environmental characteristics defining the impacts of agricultural pollution. The hedonic price function specifies

[4] To our knowledge, these tests have rarely been used in empirical studies. See the notable exceptions of Horowitz and Lee (2001) and Lavergne and Thomas (2005).

how the price of a house, denoted by Y, varies according to the different characteristics, i.e.

$$Y = m(X, Z, \xi) \tag{1}$$

where ξ is the vector of house characteristics not observed by the econometrician. For simplicity, we assume that this vector is one-dimensional.

Rosen (1974) provided a theoretical equilibrium framework in which the interactions of consumers and suppliers determine the hedonic price function. In this approach, we can derive the marginal willingness to pay for a given characteristic by differentiating the hedonic price function with respect to that characteristic. Rosen also proposed a two-step parametric procedure for estimating the demand for each characteristic in cases where the hedonic price function has the following additive structure:

$$Y = m(X, Z) + \xi \tag{2}$$

where we select a parametric functional form of the function $m(\cdot)$. In a first step, the hedonic price function is estimated by regressing the observed house prices, Y, on all their characteristics, X and Z, using the best fitting parametric functional form. Next, we can compute a set of implicit marginal prices for a given characteristic, or marginal willingness to pay, by taking the partial derivative of the estimated hedonic price function with respect to this characteristic evaluated at each observed house sale. Finally, the demand equations for all the characteristics can be established using these estimated implicit marginal prices as endogenous variables. This is performed simultaneously in a second-stage estimation of the system of parametric functional expressions for demand.

Since the work of Brown and Rosen (1982), there has been much debate in the literature on the lack of identification for the marginal willingness to pay or bid. Indeed, the main shortcoming of Rosen's second-stage demand estimation is that the estimated implicit marginal price may not contain any information beyond that provided by the first stage. The only new information is the restriction on the functional expression used in the demand equation (Horowitz, 1987; Chattopadhyay, 1999). As shown by Brown and Rosen (1982), if there is no new information, the estimated demand equation simply reproduces the results of the hedonic regression from which it was initially generated, i.e. the demand equation cannot be identified from the hedonic price function. Only recently, Ekeland, Heckman and Nesheim (2004) showed that the identification problems highlighted in the literature mainly arise from the linearization strategies commonly used for simplifying the estimation procedure. These authors stressed that the hedonic price model is generically nonlinear. In the same way, Bajari and Benkard (2005) considered the identification of hedonic price models in cases where some product characteristics are not observed by the econometrician. Using the results of Matzkin (2003), they showed that, given data on a single market, the hedonic price function and the distribution of the unobserved product characteristic can be identified if the unobserved characteristic is independent of the observed characteristics. Thus, the hedonic price function may have a general non-additive structure.

Based on these observations, it would seem appropriate to consider nonparametric regression estimators as natural candidates for estimating the hedonic price function (1). But, unfortunately, we would face two problems in such a procedure. Firstly, ξ cannot be observed. As pointed out by Bajari and Benkard (2005), the additive error term in the hedonic price function, $Y = m(X, Z) + \xi$, is interpreted as a product characteristic observed by the consumer but not by the econometrician.[5] Then, a second problem occurs due to the "curse of dimensionality", given that the vectors X and Z may involve a large number of characteristics. Indeed, unconstrained nonparametric estimates of the unknown function $m(\cdot)$ deteriorate rapidly as $J + L$ increases. Hence, it is necessary to impose restrictions on $m(\cdot)$. One possibility is to allow $m(\cdot)$ to be nonparametric only in a subset of regressors and specifying a parametric form for the remaining regressors, as proposed by Robinson (1988). Since many housing characteristics are discrete variables, and since the main advantage of this study hinges on measuring the impact of environmental factors on house prices, we thus assume a partially linear specification given by[6]

$$Y = \beta' X + m(Z) + \xi \tag{3}$$

[5]We could use Matzkin's (2003) nonparametric estimation procedure, which deals directly with a non-additive structure. But this procedure assumes that all the observed characteristics are continuous, whereas many housing characteristics in our dataset are discrete.

[6]Note that we assume a strong separability between housing characteristics and environmental variables. Alternatively, we could impose a weak separability as proposed by Pinkse (2001). In that case, we assume that the different regressors are incorporated into the function $m(\cdot)$ through indices summarizing, for instance, the impacts of size, location or amenities. These different factors can freely interact with one another. However, the estimation procedure proposed by Pinkse (2001) also assumes that all the regressors are continuous, while many of the housing characteristics in our dataset are discrete.

Equation (3) represents the conceptual model estimated empirically in the present study using parametric and nonparametric estimation procedures. For this purpose, we propose the following four empirical model specifications:

Specification of $m(z)$	**Resulting model specification**	
Nonparametric specification $m(z) = m(z_1, \cdots, z_L)$	Partially linear and nonparametric model $Y = \beta' X + m(Z_1, \cdots, Z_L) + \xi$	(**M1**)
Additive specification $m(z) = \sum_{l=1}^{L} g_l(z_l)$	Partially linear and additive model $Y = \beta' X + \sum_{l=1}^{L} g_l(Z_l) + \xi$	(**M2**)
Single index specification $m(z) = G(\gamma' z)$	Partially linear and single index model $Y = \beta' X + G(\gamma' Z) + \xi$	(**M3**)
Parametric specification $m(z) = \gamma' z$	Fully parametric model $Y = \beta' X + \gamma' Z + \xi$	(**M4**)

These four empirical models differ from each other according to the way in which the function $m(z)$ is defined. Although model M4 is similar to a typical linear regression, the other model specifications (M1, M2 and M3) include non- and semi-parametric components.[7]

3 Specification search procedure

To estimate these four empirical models, we need to devise a method able to estimate the β coefficients and the function m. To achieve this, we follow a three-stage estimation procedure. In the first stage, we estimate the linear part of the proposed specifications using Robinson's (1988) procedure, where the function $m(\cdot)$ is left unspecified. In the second stage, we estimate this function using the four empirical specifications, with $Y - \widehat{\beta}' X$ representing the dependent variable and $\widehat{\beta}$ being the first-stage estimated value of β. In the

[7]Many other additive model specifications could be adopted to represent $m(z)$. Thus, a quadratic or higher polynomial function could be used in the case of the parametric model specification M4. The formalization of M2 could be enriched by incorporating interaction terms. To restrict the scope of this paper, we do not investigate more "generalized" model specifications and leave this aspect for future research.

third stage, we perform specification tests aimed at selecting an appropriate specification of the function $m(\cdot)$.

3.1 Estimating the linear part of the hedonic price function

The first stage yielding an estimate of β is justified because, if we subtract the conditional expectation value relative to z on both sides of (3), we obtain:

$$Y - E(Y|Z=z) = \beta'\Big(X - E(X|Z=z)\Big) + \xi \tag{4}$$

Let $(Y_i, X_i, Z_i)_{i=1}^n$ be an independently and identically distributed (i.i.d.) sample. Then the estimation procedure can be described as follows:

1. Regress both Y_i and X_i on Z_i nonparametrically, which generates the following residuals $\widetilde{Y}_i \equiv Y_i - E(Y|Z=Z_i)$ and $\widetilde{X}_i \equiv X_i - E(X|Z=Z_i)$.

2. Then perform OLS on these residuals to obtain $\widehat{\beta}$, which is an estimate of β in (4).

Robinson (1988) showed that, under regularity conditions, this procedure yields a $\sqrt{n}$-consistent and asymptotically normal estimator for β, and that a consistent estimator can be determined of its limiting covariance matrix.

We use local polynomial estimators of $E(Y|Z=z)$ and $E(X|Z=z)$. Indeed, this estimator possesses a number of desirable theoretical and practical properties compared with other nonparametric methods, including the widely applied Nadaraya-Watson kernel estimator (Fan and Gijbels, 1996). This local polynomial estimator, denoted by $\widehat{\theta}_0$, is the solution of the following optimization expression:

$$\min_{(\theta_0,\ldots,\theta_L)} \sum_{k=1}^{n} [Y_k - \theta_0 - \sum_{j=1}^{L} \theta_j \ (Z_{k,j} - Z_{i,j})]^2 \ K_h(Z_k - Z_i)$$

where $Z_k = (Z_{k,1}, \ldots, Z_{k,L})$ for every $k = 1, \ldots, n$ and $K_h(\cdot)$ is a multidimensional kernel depending on a vector of bandwidths denoted by h.[8]

[8] The solution of this weighted least squares regression problem is given by

Automatic bandwidth selection criteria, such as plug-in or cross-validation, can be used to choose the vector of bandwidths h for estimating $E(Y|Z=z)$ and $E(X|Z=z)$. In the following, we apply the cross-validation criterion, which has been the most widely used estimation technique for partially linear house price models (see Stock, 1991, Anglin and Gençay, 1996, and, more recently, Kondo and Lee, 2003). We define a set of bandwidths $h_j = a\sigma_j n^{-1/4+L}$, where σ_j denotes the standard deviation of the j-th variable, and then minimize

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{Y}_i - \widehat{\beta}'\widetilde{X}_i)^2 \tag{5}$$

with respect to the coefficient a (bear in mind that $\widetilde{Y}_i$, $\widetilde{X}_i$, and $\widehat{\beta}$ are functions of h).

3.2 Estimating the nonlinear part of the hedonic price function

Our next task consists of estimating the nonlinear part of the hedonic price function $m(\cdot)$. Since $m(z) = E(Y - \beta'X|Z=z)$, we estimate this nonlinear part by regressing the residual $W \equiv Y - \widehat{\beta}X$ on the vector Z according to the specifications M1, M2 and M3. We now consider how to derive estimates of these models.

$$\hat{\theta} = (\Xi'\Upsilon\Xi)^{-1}\Xi'\Upsilon Y$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_L)'$, $Y = (Y_1, \ldots, Y_n)'$, and

$$\Xi = \begin{bmatrix} 1 & Z_{1,1} - Z_{i,1} & \ldots & Z_{1,L} - Z_{i,L} \\ 1 & Z_{2,1} - Z_{i,1} & \ldots & Z_{2,L} - Z_{i,L} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ 1 & Z_{n,1} - Z_{i,1} & \ldots & Z_{n,L} - Z_{i,L} \end{bmatrix},$$

and $\Upsilon = diag(K_h(Z_k - Z_i))$. The estimate of the function $E(Y|Z=Z_i)$ is thus the first component of the vector $\hat{\theta}$, i.e. $E(\widehat{Y|Z=}Z_i) \equiv \widehat{\theta}_0$. Similar formulas can be written for the estimator of $E(X|Z=Z_i)$.

3.2.1 Fully nonparametric model (M1)

This model is used as a benchmark for assessing the ability of the proposed specifications M2 to M4 to reflect the nonlinear part of the hedonic price function. The function $m(\cdot)$ is estimated using the second stage of Robinson's procedure based on a nonparametric regression of W on Z, leading to an estimate of $m(z) = E(W|Z = z)$.[9] We use a local polynomial estimator (see above) as a nonparametric estimator of $m(\cdot)$.

3.2.2 Additive model (M2)

The additive model is based on the assumption that

$$m(z) = m(z_1, \cdots, , z_L) = c + \sum_{l=1}^{L} g_l(z_l) \tag{6}$$

where c is a constant term, and $g_l(.)$, $l = 1, \ldots, L$, is a set of L unknown functions satisfying the identifiability condition that $E[g_l(z_l)] = 0$, for every l.

Additive models are usually estimated using Hastie and Tibshirani's (1990) backfitting algorithm.[10] While this algorithm converges to a unique solution independent of the starting values, the statistical properties of the resulting estimates are not well understood. An alternative procedure for the estimation of additive models based on marginal integration has been proposed

[9] We could also compute this estimator using our first-step estimates of β and kernel regressions since the conditional expectation could also be written as $m(z) = E(Y|Z = z) - \beta' E(X|Z = z)$. See, for instance, Blundell *et al.* (1998) and Schmalensee and Stoker (1999).

[10] The backfitting algorithm proceeds as follows :

1. *Initialization*: Select initial non parametric estimates of the functions $\left(g_l(.)_{l=1}^{L}\right)$, denoted by $\left(g_l^0(.)_{l=1}^{L}\right)$.
2. *Successive iterations*: Obtain an estimate of each function $g_k^i(.)$ by nonparametric regression of $w - \sum_{l=1, l\neq k}^{L} g_l^{i-1}(Z_l)$ on Z_k.
3. *Convergence*: Continue iterate until there is little change in the individual function estimates.

For some applications of this model, see for instance Pace, 1998, Iwata *et al.* (2000) and Martins-Filho and Bin (2005).

by Linton and Nielsen (1995) in the case of two independent variables, and extended to larger dimensions by Tjøstheim and Auestad (1995) and Chen *et al.* (1996). The idea behind this estimator is quite straightforward. In the case of additivity (6), there exists functions g_l and m_{-l} such that

$$m(Z) = g_l(Z_l) + m_{-l}(Z_{-l}) \tag{7}$$

where Z_{-l} being the vector Z without the component Z_l, the marginal impact of Z_l corresponding exactly to the additive component g_l. The marginal integration estimator is defined noting that

$$\begin{aligned} E_{Z_{-l}}\left[m(z_l, Z_{-l})\right] &\equiv \int m(z_l, z_{-l})\, \varphi_{-l}(z_{-l})\, dz_{-l} \\ &= E_{Z_{-l}}\left[g_l(z_l) + m_{-l}(Z_{-l})\right] \\ &= g_l(z_l) + c \end{aligned} \tag{8}$$

where φ_{-l} denotes the marginal density of Z_{-l}. So marginal integration with respect to this density yields the function g_l up to a constant that can be easily estimated by the average over the observations W_i, we denote by $\widehat{c}$. We estimate the right hand side of equation (8) by replacing the expectation by an average and the unknown multidimensional regression function m by a local polynomial pre-smoother. This method can be applied to estimate all components g_l in equation (6), and finally the regression function m is estimated by summing up the estimator $\widehat{c}$ of c with the estimates $\widehat{g}_l$, $l = 1, \ldots, L$.

3.2.3 Single-index model (M3)

A single-index house price model is based on the assumption that all the information conveyed by the independent variables can be summarized into a single index $\gamma' Z$, where γ is a vector of unknown coefficients linked to the endogenous variable through an unknown link function $G(\cdot)$ given by

$$m(z) = G(\gamma' z) \tag{9}$$

The main idea underlying these models is to avoid the *curse of dimensionality* by reducing the dimensions of the regressor's space to one through the index. However, there is a drawback in terms of identification since equation (9) is equivalent to $m(z) = G^*(\nu + \delta(\gamma' z))$, for any arbitrary value of the location parameter ν and the scale parameter $\delta \neq 0$ and function G^* defined by the relation $G^*(\nu + \delta v) = G(v)$ for all v in the support of $\gamma' z$. Thus some normalizations are required. Location normalisation is achieved by requiring the vector Z to contain no constant component. Scale normalization is achieved by setting the γ coefficient of one component of the vector Z to one.

If $G(\cdot)$ is known, the value of γ can be estimated by solving the following nonlinear least-squares estimation problem:

$$\widehat{\gamma} = arg \min_{\gamma} \sum_{i=1}^{n} \Big(W_i - G(\gamma' Z_i) \Big)^2 \tag{10}$$

where $(W_i)_{i=1,\ldots,n}$ denotes the first-step residuals $(Y_i - \widehat{\beta} X_i)_{i=1,\ldots,n}$. Since $G(\cdot)$ is unknown, the estimation requires an iterative and difficult numerical resolution procedure. Fortunately, as Z only has continuous components, a Density Weighted Average Derivative estimator (DWADE) may be used to estimate γ without solving an optimization problem (see Powell *et al.*, 1989). This estimator is based on the fact that γ is equal to $E\Big[\frac{\partial G(\gamma' Z)}{\partial Z}\, \varphi(Z)\Big]$ up to a multiplicative term, where $\varphi(\cdot)$ denotes the unknown density function of Z.[11] With suitable assumptions on the function $G(\cdot)$ and the density function $\varphi(\cdot)$, integration by parts yields:

$$E\Big[\frac{\partial G(\gamma' Z)}{\partial Z}\, \varphi(Z)\Big] = -2E\Big[Y\, \frac{\partial \varphi(Z)}{\partial Z}\Big] \tag{11}$$

Thus γ can be estimated up to scale by the following estimator :

$$\gamma_{DWADE} = -\frac{2}{n} \sum_{i=1}^{n} Y_i \widehat{\frac{\partial \varphi(Z_i)}{\partial Z}} \tag{12}$$

where we replace in (11) the expectation by the average over the observations and the derivative $\partial \varphi(Z_i)/\partial Z$ by a nonparametric estimate. We use the derivative of the usual leave-one-out Parzen-Rosenblatt estimator as an estimator of this derivative, with higher-order kernels as proposed by Powell *et al.* (1989). Under conditions involving the use of such kernels, it can be shown that γ_{DWADE} is an asymptotically normal distributed consistent estimator of γ.

[11] This density is taken into account in the expectation value to avoid the usual random denominator problem involved in nonparametric kernel estimation.

3.3 Specification tests

As stated in the introduction, one of the contributions of this study is to adopt specification tests for the various empirical specifications M2, M3 and M4 of the function $m(\cdot)$ and compare them with the most general benchmark model M1.

3.3.1 Additive vs. nonparametric

Gozalo and Linton (2001) develop several kernel-based consistent tests of the hypothesis of addivity in nonparametric regression. Their framework allows for a very general additive structure involving discrete covariates and parameters to be estimated. Hereafter, we test the simple null hypothesis H_0 where it is assumed that $m(Z) = c + \sum_{l=1}^{L} g_l(Z_l)$. Among the various tests proposed by Gozalo and Linton (2001), we implement the following:

$$\widehat{\tau}_2 \;=\; \frac{1}{n^2 h^L} \sum_i \sum_{j \neq i} K_{ij}\, \widetilde{u}_i\, \widetilde{u}_j\;\; \pi(Z_i)\pi(Z_j), \tag{13}$$

where h denotes the bandwidth involved in the estimation of the fully nonparametric specification (the unrestricted one), the weights K_{ij} are defined as $K_{ij} = K((Z_i - Z_j)/h)$ where $K(.)$ is a multivariate kernel function, $\widetilde{u}_i$ are the residuals from the estimation of the restricted specification, i.e. the additive model, and $\pi(\cdot)$ is a trimming function ensuring that the density of the vector Z at a given point is bounded from zero. This test has some analogy with the Lagrange Multiplier test of classical statistics as it looks for a correlation between restricted residuals.

Gozalo and Linton show that, after performing only a scale adjustment, the test statistic $\widehat{\tau}_2$ becomes asymptotically standard normal. Specifically, under the null hypothesis that the nonparametric additive model is correctly specified, it can be shown that

$$nh^{L/2}\, \widehat{\tau}_2 \;\sim\; N(0, V_{2,n})$$

where $V_{2,n}$ denotes the variance of the test statistic whose empirical counterpart $\widehat{V}_{2,n}$ can be computed as:

$$\widehat{V}_{2,n} \;=\; \frac{2}{n^2 h^L} \sum_i \sum_{j \neq i} K_{ij}^2\, \widetilde{u}_i^{\,2}\, \widetilde{u}_j^{\,2}\;\; \pi^2(Z_i)\pi^2(Z_j)$$

3.3.2 Single index vs. nonparametric

To test the single-index specification of the regression function $m(z)$, we use the procedure proposed by Fan and Li (1996). This is based on the null hypothesis H_0 that $m(z) = G(\gamma' z)$, for some $\gamma \in R^L$ and some unknown real-valued function against the general alternative that H_o is not true. The test is based on the quantity $\nu = W - G(\gamma Z)$. Indeed, under H_0, $E[\nu|Z] = 0$. Then, consider the statistic $E[\nu E[\nu|Z]] = E[E[\nu|Z]^2] \geqslant 0$. The equality holds if and only if H_0 is true. If observations ν_i and $E(\nu_i|Z_i)$ were available, we could use the sample analogue $(1/n)\sum_i \nu_i E(\nu_i|Z_i)$ as an estimator of the statistic. To get a feasible test statistic, we need to estimate ν_i and $E(\nu_i|Z_i)$. To overcome the random denominator problem in the kernel estimation of this conditional expectation, a density-weighted version of the test statistic given by $E[\nu_i f(\gamma' Z_i) E\{\nu_i f \gamma' Z_i)|Z_i\} \varphi(Z_i)]$ where $f(.)$ denotes the density of $\gamma' Z$ and $\varphi(.)$ the density of the vector Z, is estimated. Its expression is given by:

$$I_n^c = \frac{1}{n}\sum_{i=1}^n \left(\widehat{\nu_i}\ \widehat{f_{h_\gamma}}(\widehat{\gamma}' Z_i)\right)\left(\frac{1}{(n-1)h^L}\sum_{j\neq i}(\widehat{\nu_j}\ \widehat{f_{h_\gamma}}(\widehat{\gamma}' Z_j))\ K_{ij}\right) \qquad (14)$$

where $\widehat{\nu_i} = Y_i - \widehat{G}(\gamma' Z_i)$ are the single index residuals, $\widehat{f_{h_\gamma}}(t)$ is a kernel estimator of the density function $f(t)$ of $\gamma' Z$, and $K_{ij} = K((\widehat{\gamma}' Z_i - \widehat{\gamma}' Z_j)/h)$ with $K(\cdot)$ a kernel function. Under the null hypothesis and with suitable conditions on the two bandwidths h_γ and h, it can be shown that the test statistic whose expression is given by

$$T^c = \frac{nh^{L/2}\ I_n^c}{\sqrt{2}\sigma_c} \qquad (15)$$

is asymptotically distributed as $N(0, 1)$.

3.3.3 Parametric vs. nonparametric

Several studies have proposed tests of a parametric specification of a regression model against a nonparametric alternative (see among others, Härdle and Mammen, 1993, or Lavergne and Vuong, 1996). Recently, Horowitz and Spokoiny (2001) developed a new test that is used here. Specifically, we test the null hypothesis, H_0, that $m(\cdot)$ belongs to some parametric family, i.e. there exists some vector of parameters γ such that $m(\cdot) = M(\cdot, \gamma)$, where $M(\cdot, \cdot)$ is known, against the alternative, H_1 in which there is no such γ. The test is based on the distance $S_k(\gamma)$ between the kernel estimation of $m(\cdot)$ and the kernel-smoothed estimation of the parametric regression $M(\cdot, \gamma)$.

$$S_h(\gamma) = \sum_{i=1}^{n} \Big(\widehat{m_h}(Z_i) - \widehat{M_h}(Z_i, \gamma) \Big)^2 \tag{16}$$

where $\widehat{M_h}(Z_i, \gamma) = \sum_{j=1}^{n} W_h(Z_i, Z_j) M(Z_j, \gamma)$ is the kernel-smoothed parametric estimator with kernel weight $W_h(\cdot, \cdot)$. The test statistic, $T^\star$, is computed with a rate-optimal and adaptative bandwidth based on a set of bandwidth values h in some set H_n, and is then centred and studentized:

$$T^\star = max_{h \in H_n} \frac{S_h(\gamma) - \widehat{N_h}}{\widehat{V_h}} \tag{17}$$

where $\widehat{N_h}$ and $\widehat{V_h}$ denote estimates of the mean and variance of $S_h(\gamma)$, respectively. Horowitz and Spokoiny (2001) show that the parametric model is correctly specified under the null hypothesis, and that the statistic $T^\star$ is asymptotically and normally distributed.

4 Data

The variables used in the hedonic regression analysis fall into three broad categories: (*i*) the price and the physical attributes of the residential houses, *(ii)* the characteristics of the surrounding rural district, and *(iii)* the environmental nuisances. Observations on the first category of variables were taken from the real estate database (known as MIN) maintained by the Association of French Notaries.[12] This database provides a detailed description of all house sales in France, including sale prices, physical attributes of houses and adjacent lot size. Taxes and various fees linked to house sales are incorporated into the computation of the prices actually paid by the buyers. This latter variable is denoted PRICE. Four physical characteristics of the house and the adjacent lot are used in the empirical analysis: age of the house (AGE), state of repair (REPAIR), number of rooms (ROOMS), and lot size (LOT). A description of these five variables is given in Table 1.[13]

[12] *MIN* stands for "*Marché Immobilier des Notaires*". It is important to note that this database does not contain at all any information on the socio-economic situation of the buyers. Hence, we do not have at our disposal data on the disposable income of the buyers.

[13] A total data sample of 2092 sales distributed over 465 different districts with less than 5000 inhabitants were collected for 1996 and 1997 from the MIN database.

The second category of explanatory variables comprises the characteristics of the surrounding districts, which are expressed by four indicators. The first variable is the population of the district (POP). The second indicator is the average taxable family income (AVINC).[14] The MIN database enables us to determine for each district the proportion of vacant houses (VACANT). The fourth explanatory variable in this category (denoted by COUNTY) is a dummy variable indicating whether or not the surrounding district is located in the department of Ille et Vilaine.[15]

To assess the consumer's willingness to pay for environmental nuisances generated by agriculture, we would require not only personal and confidential information on consumers' views on such issues, but also detailed information on the location of livestock farmholdings in relation to consumers' residential houses. Collecting such quantitative information is so sensitive that it is impossible to undertake suitable surveys to generate the relevant data.[16] Agricultural pollution is thus measured by the following two aggregate indicators:[17]

- The first indicator (*NITRO*) is the amount of nitrogen emissions from livestock farming per hectare of arable land in the rural district where the residential house is located.
- The second indicator considered here (*TMEAD*) is the proportion of permanent pasture land converted into tilled land. A high value associated with this variable would indicate a degradation of the rural landscape.

By deriving the first-order derivatives of house prices with respect to these variables, we obtain an estimate of the price of these two environmental at-

[14]The observations concerning POP and AVINC were obtained from the French Census, INSEE, which is a governmental body collecting and producing socio-economic statistical information for France. INSEE stands for Institut National de la Statistique et des Études Économiques.

[15]The Brittany region is composed of four departments, Ille et Vilaine being the least rural and the most urbanized.

[16]Indeed, for reasons of confidentiality, the MIN database does not provide detailed information on the exact location of (residential) house sales. As a result, it is impossible to locate neighbouring livestock farms that are the source of agricultural pollution. To collect such information would have required undertaking a lengthy and extensive survey on residential house sales in Brittany for the period under study.

[17]Sample observations for these two variables were provided by the Regional Branch of the French Ministry of Agriculture, Fisheries and Forestry.

tributes and, indirectly, an estimate of the consumers' willingness to pay.

Table 1 gives a description of the two pollution indicators. We also present in Figure 1 the results of the nonparametric estimation of the joint density of these two indicators.[18] This joint density appears to be single-peaked.

[18] A bivariate normal kernel was used and the bandwidth was chosen using Silverman's rule. See Silverman (1986).

Table 1: Variables and Descriptive Statistics

	Variable	Description	Units	Min	Max	Mean	Std. Dev.
Dependent variable	PRICE	Market Price	Euro	15400	162 583	76 494	33 933
Characteristics	AGE	Age	Year	0	298	47.835	42.018
of	REPAIR	State of repair	= 1 if good	0	1	0.687	0.464
the	ROOMS	Number of rooms	#	1	7	4.429	1.353
house	LOT	Lot size	#	102	21 880	1 793	2 551
Characteristics	COUNTY	County location	= 1 if *"Ille et Vilaine"*	0	1	0.478	0.499
of	VACANT	Vacant Housing	Percent	0.000	20.000	6.275	3.157
the	POP	County population	# (x1000)	0.104	4.972	2.047	1.215
surrounding	AVINC	Average income	Euro	571	2 854	1 082	250
Environmental	TMEAD	Temporary meadows	Percent	0.010	70.143	29.420	9.972
variables	NITRO	Nitrogen concentration	kg/ha	0.000	339.48	45.169	51.118

N=2092 observations

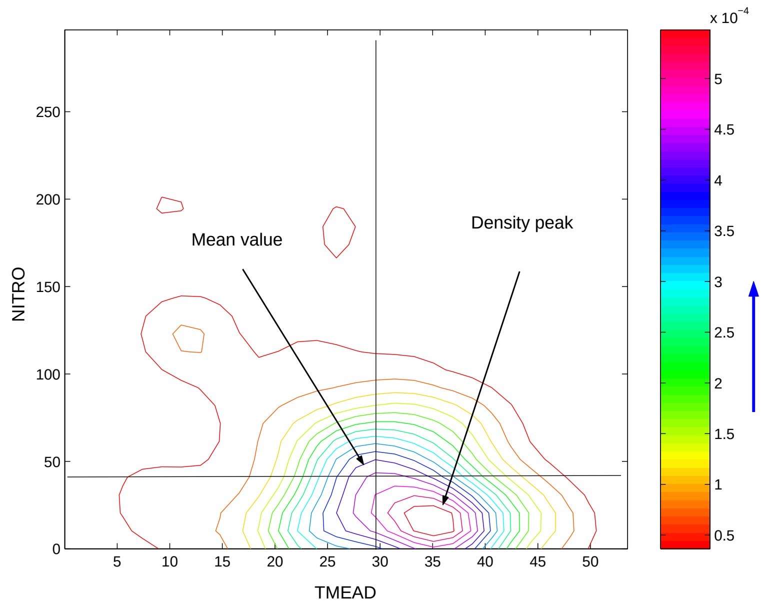


Figure 1: *Joint iso-density curves for the two environmental variables* (z_1, z_2)

5 Empirical results

In hedonic price models, we specify the (logarithm of) prices of residential houses as a function of their physical characteristics, not only the two environmental indicators but also variables representing the economic structure of rural districts where the residential houses are located. As stated above, all the explanatory variables (AGE, REPAIR, ROOMS, LOT, COUNTY, VACANT, POP and AVINC) apart from the two environmental indicators (TMEAD and NITRO) are incorporated into the hedonic price models in a linear fashion. This makes up the linear part of the hedonic price function. The two pollution indicators are involved in the hedonic price function in a nonparametric or semiparametric way, thus forming the nonlinear part of the model.

In the following, we report the results of the specification search procedure presented in the previous section. We first consider the estimated values of the parameters involved in the linear part of the hedonic price function (first stage). We then present a 3D-graphical comparison of the four specifications of $m(z)$ (second stage), followed by the specification test results (third stage). At the end of the results section, we compute the marginal willingness-to-pay for pollution reduction in the case of the selected model specification.

5.1 Linear part

Table 2 reports the estimates of the β parameters involved in the linear part of the first stage estimation. We selected the bandwidth used for estimating the conditional expectation values in equation (4) using Kondo and Lee's (2003) cross-validation criterion.[19] All the estimated parameters belonging to the linear part of the housing price models are statistically significant and have the expected signs and magnitudes.[20] Examining first the influence of the physical characteristics of the houses on prices, we note that a variation of one year in the age of a house yields, all other things being equal, a reduction of 0.2% of its sale price. Undertaking major renovations on a residential house in Brittany leads to a 36% appreciation in its price, everything else being held constant. A larger number of rooms or a bigger lot size are factors contributing to an increase in the value of a house. Explanatory variables characterizing the district where the houses are located have signs that conform to our expectations. Hence, the price of any house located in the districts of the most urbanised county of Brittany (Ille-et-Vilaine) will exhibit an average price increase of 9.1%. By contrast, residential houses located in districts with higher housing vacancy rates will show lower prices, while opposite effects take place in districts that are either more populated or have households with higher average incomes.

The second-stage parameter estimates of model M4 reported in Table 2 show that the environmental indicators have the expected effects on house prices. The estimated coefficients of the indicators are statistically significant and negative. Thus, a single percent point increase in the proportion of permanent pasture converted into tillable land results in a 0.3% decline in the price of houses. A similar interpretation could be made for the effects of (livestock) nitrogen emissions on property values.

[19] Appendix A provides more details about the bandwidth choice involved in the estimators described in section 3. In this appendix, Figure 5 presents Kondo and Lee's cross-validation criterion as a function of the bandwidth exhibiting a clear minimum value.

[20] Since the prices of houses are expressed in a logarithmic form, we could interpret the estimated coefficients as the percentage variation in the house price resulting from one unit change in the explanatory variables.

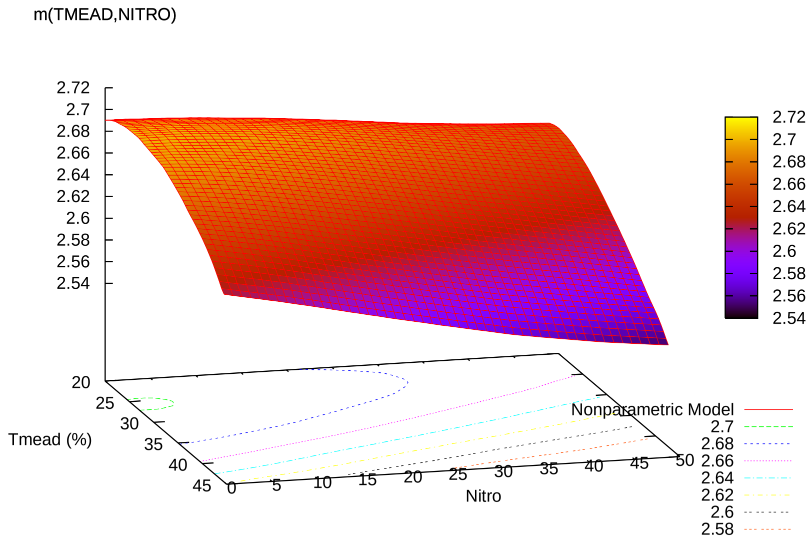
Table 2: First stage parameter estimates and fully parametric model (M4) parameter estimates

		Linear part of models M1 to M4	
	Variable	**Estimated value**	**Standard error**
First stage estimates	Age	-0.002	*0.0002*
	Repair	0.359	*0.0174*
	Rooms	0.140	*0.0057*
	Lot	0.029	*0.0028*
	County	0.091	*0.0169*
	Vacant	-0.017	*0.0032*
	Pop	0.016	*0.0074*
	Avinc	0.050	*0.0061*
		Fully parametric model (M4) parameter estimates	
	Variable	**Estimated value**	**Standard error**
Second stage estimates	Constant	0.889	*0.0253*
	Tmead	-0.003	*0.0007*
	Nitro	-0.0006	*0.0001*

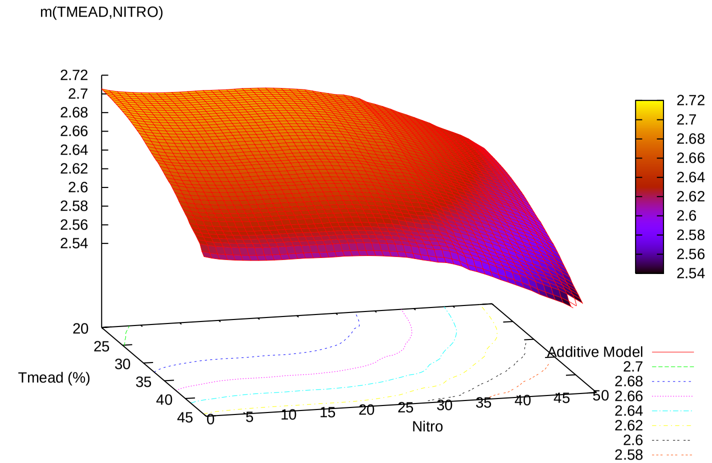
5.2 Nonlinear part of the hedonic price function

As in any empirical study of nonparametric models, we report the role and importance of nonlinearities by means of graphical analysis. Hence, we use Figure 2 to develop the estimated response surfaces linking housing prices to the two pollution indicators for the four specifications of the function $m(z)$. We restrict the representation of these curves to an area with high values of the joint distribution for the environmental variables z_1 and z_2, *i.e.* with values of *TMEAD* and *NITRO* belonging to the $20 - 45\%$ and $0 - 50kg/ha$ intervals respectively.[21] Appendix B describes the choice of bandwidths involved in the estimation of the various specifications of the nonlinear part of the hedonic price function.

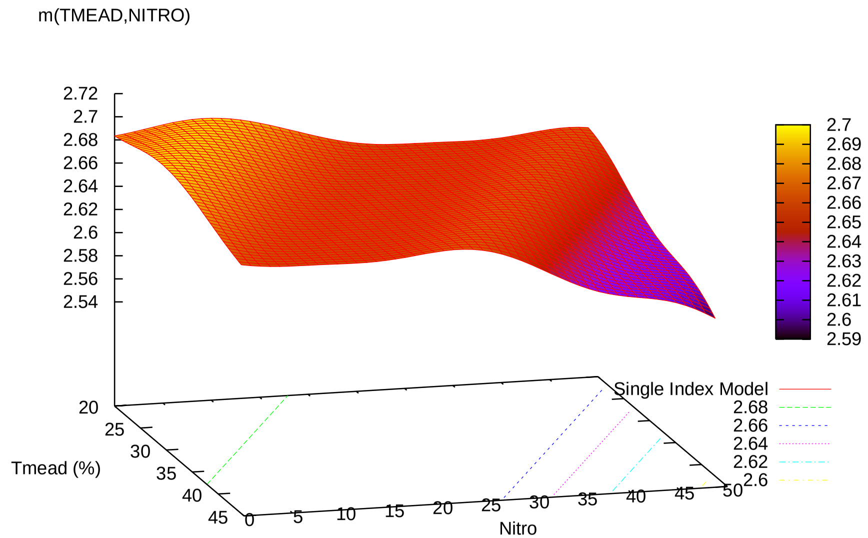
[21] We defined a rectangle such that only very small values of the joint density of TMEAD and NITRO fall outside its area.



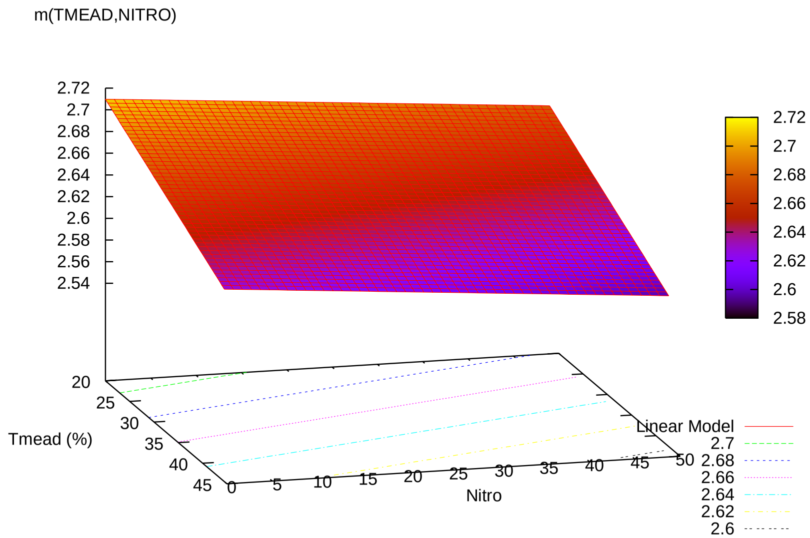
Nonparametric specification (M1)



Nonparametric additive specification (M2)



Single index specification (M3)



Parametric specification (M4)

Figure 2: Estimation of $m(z_1, z_2)$ for the four specifications

A visual inspection of the four estimated surfaces provides a first impression of the responses of house prices to the two environmental indicators in terms of shape and steepness of the curve. As expected, all the surfaces exhibit a decrease in the house price with increasing values of environmental indicators. While the M2 specification closely resembles the benchmark M1, the fully parametric and single-index specifications of the hedonic price function (M3 and M4) seem unable to represent all the features of our data. Therefore, we need specification tests to go beyond this eyeball analysis.

5.3 Specification tests

Each specification test presented in section 3 was performed to compare a restricted specification with the more general nonparametric model M1. All these tests require a choice of bandwidths. In the absence of practical guidelines for the choice of these parameters, we perform, for each specification test, a sensitivity analysis of the test statistics for the bandwidths. We first test the null hypothesis that the nonparametric additive model (M2) is correctly specified. Following Gozalo and Linton (2001), we use the same bandwidth for the restricted and unrestricted estimators (see equation (13)). This bandwidth is defined as $h_n = c \times n^{-1/4}$, with c varying over the range [0.4 1.4] by steps of 0.2, using normalized variables. We report in appendix C (Table 4) the results for the test statistic $\widehat{\tau}_2$. These values clearly do not vary much with the bandwidths, showing that, even in the least favourable case (see the values of the statistic and its empirical significance level reported in Table 3), we cannot reject the null hypothesis.

Let us now consider the single-index specification (M3). Practical implementation of Fan and Li's (1996) test requires a very smooth kernel function, so we use a standard normal density function as the kernel. Two smoothing parameters h_γ and h are required to compute the estimators involved in the test statistic (see equation (15)). For the former parameter, we use the usual bandwidth for nonparametric one-dimensional density estimation, $h_\gamma = c \times n^{-1/5}$, while for the latter parameter, we choose $h = c \times n^{-1/4}$ according to theoretical guidelines given by Fan and Li (1996). We compute the values of the Fan and Li's test statistic for various values of c over the range [0.4 1.4] (see Table 4 in appendix C). Varying c has very little effect on the test statistic, whose smallest value clearly shows that the Fan and Li test rejects the null hypothesis even in the least favourable case (see Table 3).

Consider now the parametric specification (M4). Horowitz and Spokoiny's test is based on the studentized distances between a kernel nonparametric estimate of the conditional mean function and a kernel-smoothed parametric estimate. These estimates and studentized distances are computed for a variety of bandwidths, while the test statistic is taken as the maximum of the studentized distances over the set of bandwidths (see equation (17)). For this set, we select a bandwidth $h = c \times n^{-1/6}$ where c varies in the set $\{0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7, 1.9\}$ and using normalized variables. As recommended by Horowitz and Spokoiny, we use a bootstrap procedure to obtain the empirical level of significance of this test statistic. The test rejects the null hypothesis that the parametric model is correctly specified at conventional level, since the empirical significance level is 0.006 (see Table 3).

Table 3: A summary of the specification tests results

H_0:	Test Statistic	p-value
Nonparametric additive specification (M2)	$\widehat{\tau}_2 = 0.049$	0.480
Single index specification (M3)	$T_c = 3.342$	0.001
Parametric specification (M4)	$T^* = 6.107$	0.006

Note: We report the least favourable case for the statistic $\widehat{\tau}_2$ and T_c using the sensitivity analysis provided in appendix C. The p-values are either the asymptotic or bootstrapped ones

To sum up, the results show that the nonparametric additive specification (M2) is clearly not rejected, in contrast to the two others (M3 and M4). This result is consistent with the informal graphical finding that the parametric and single-index specifications fail to reflect important nonlinear features of the data, while the nonparametric additive specification fits the data satisfactorily.

5.4 Marginal prices

Based on the specification test diagnosis given above, we report here the willingness–to-pay (WTP) for pollution reduction estimated by the nonparametric additive model (M2). These computed values are derived using the estimator developed by Severance-Losin and Sperlich (1999).

This estimator can be motivated in the same way the marginal integration estimators of the functions $g_l(.)$, $j = 1, \ldots, L$, are, by noting that

$$E_{Z_{-l}}\left[\frac{\partial m(z_l, Z_{-l})}{\partial z_l}\right] \equiv \int \frac{\partial m(z_l, z_{-l})}{\partial z_l}\, \varphi_{-l}(z_{-l})\, dz_{-l} \tag{18}$$
$$= E_{Z_{-l}}\left[\frac{dg_l(z_l)}{dz_l}\right]$$
$$= \frac{dg_l(z_l)}{dz_l}$$

An estimate of the derivative $dg_l(z_l)/dz_l$ is thus obtained by estimating he right hand side of equation (18), i.e. by replacing the expectation by an average and the unknown derivative function $\partial m(z_l, z_{-l})/\partial z_l$ by a local polynomial pre-smoother.

Figures 3a and 3b report the estimated values of the willingness-to-pay for each housing transaction, as well as a nonparametric estimation of the mean willingness-to-pay function and its 95% confidence limits.[22] Figures 4a and 4b report the same mean willingness-to-pay function, but expressed as a percentage of the corresponding housing prices.

[22]It is important to note that WTP estimates presented in Figures 3 and 4 are computed assuming the following units of measurement for the environmental indicators: TMEAD: 10% and NITRO: 100 kg/ha. Thus, parameter estimates presented in Table 2 associated with TMEAD and NITRO should be interpreted bearing in mind these new units of measurement.

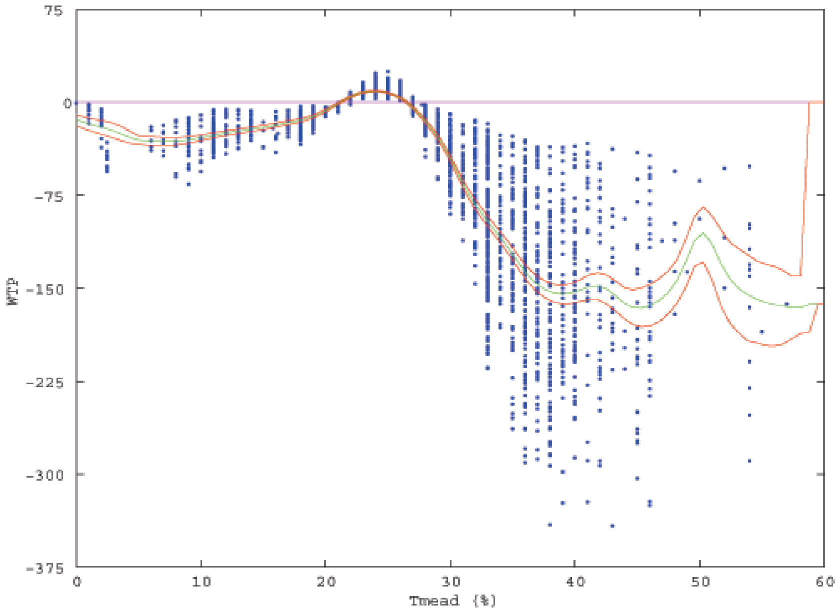


Figure 3.a : WTP for landscape degradation (Tmead)

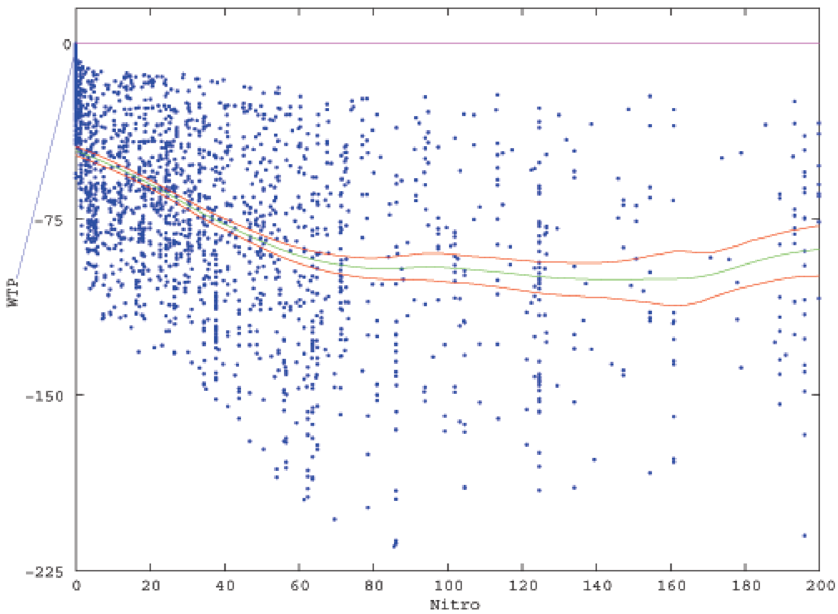


Figure 3.b : WTP for livestock nitrogen emission (Nitro)

Note: In both figures, dots represents WTP estimated using model M2 for each observation of the data sample, while solid lines represent WTP nonparametric estimation and bootstrapped 95% confidence bounds.

Figures 3 and 4 clearly indicate that the relationships between willingness-to-pay and the pollution indicators are highly nonlinear for specific ranges of values taken by z_1 and z_2. Up to a certain threshold that is significantly different from zero, the derivative of the hedonic price function with respect to the "landscape degradation" indicator "TMEAD" is rather small relative to average observed house prices. Nevertheless, the hedonic price function derivative exhibits a marked and sharp decline when the proportion of permanent pasture converted into tillable land increases from 20 to 40%. Then, the willingness-to-pay tends to flatten out when this indicator rises to values greater than 40%. In addition, Figures 3.a and 4.a clearly show changes in the degree of curvature of the relationship between WTP and "landscape degradation".

On the other hand, a different pattern seems to emerge for WTP estimates associated with livestock nitrogen emissions (NITRO). An examination of Figures 3.b and 4.b reveals that the relationship between the mean WTP function and NITRO is steep and convex for small values of nitrogen emissions until it reaches 80 kg per hectare of arable land. Then, the mean WTP function for nitrogen emissions tends towards an asymptotic value that is equal to 7% of the residential house prices.

It is interesting to compare these WTP estimates from the nonparametric additive model specification with similar estimates from a parametric specification. If we perform this exercise with the model specifications (M2) and (M4) assessed in this study, we note that the WTP estimates obtained with the two model specifications are comparable and very similar for large values of the two environmental indicators. For instance, in the case of the landscape degradation indicator, the WTP estimates obtained with model M4 are constant and equal to 3% of the house price (assuming a 10 percent change in the proportion of pasture land converted into tillable land). On the other hand, the estimate obtained with model M2 is equal to 3.5% when TMEAD is greater than 50%. Similar conclusions can be proposed in the case of WTP for livestock nitrogen emissions.

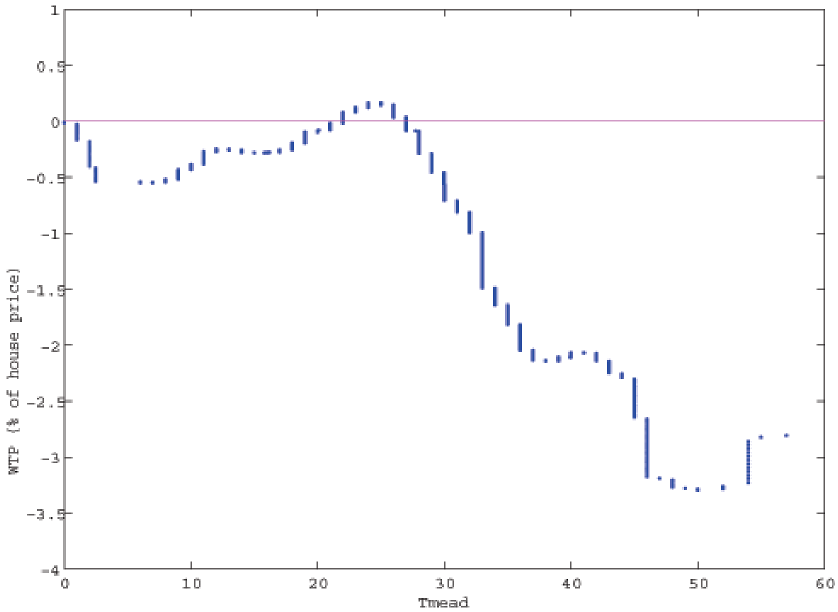


Figure 4.a : WTP for landscape degradation - Additive model M2.
(expressed as a percentage of house prices)

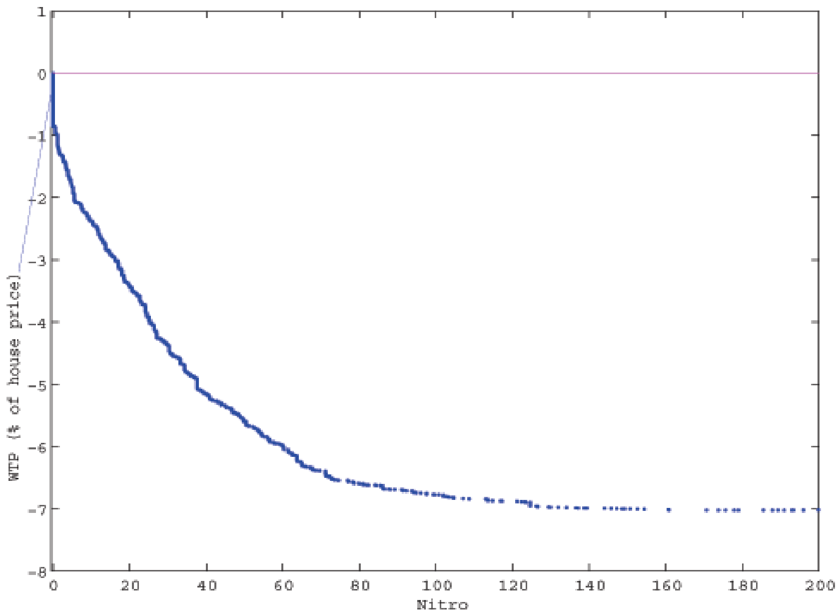


Figure 4.b : WTP for livestock nitrogen emission - Additive model M2
(expressed as a percentage of house prices))

Note: In both figures, dots represent WTP estimated for each observation of the data sample

6 Concluding remarks

The main objective of this paper is to show the relevance of semiparametric models in studying the relationship between agricultural pollution and property values. For this purpose, semiparametric hedonic price models are estimated in an intensive livestock farming region of France to study the influences of landscape degradation and livestock nitrogen emissions on house prices. Using appropriate specification tests, we conclude that a nonparametric additive expression is the most appropriate model to explain the nonlinear relationships between property values and agricultural pollution. Estimates of the willingness to pay for agricultural pollution seem reasonable and compatible with *a priori* expectations, being in conformity with estimates obtained using a parametric (semi-log) model specification.

The application of these various nonparametric models to an agricultural-related hedonic pricing case appears as a promising approach to represent complex nonlinearities. However, it is still too early to give a definitive appraisal of its merits. We require further research and applications to other agriculture-related situations. Along these lines, it would be fruitful to analyse the role of positive and negative agricultural amenities in a common (semiparametric) model framework (Ready and Abdalla, 2005). In such a way, we could compare this approach with a more conventional parametric model. Moreover, this common framework could be further refined by taking into account recent advances in nonparametric econometric estimation. These could then be used in hedonic price models to overcome problems such as the curse of dimensionality, the existence of discrete (dummy) variables, the need for more general non-linear functional expressions and spatial considerations that are crucial in predicting house prices.

References.

Anglin, P.M., and R. Gençay (1996), "Semiparametric Estimation of a Hedonic Price Function", *Journal of Applied Econometrics* **11**: 633-648.

Bajari, P., and C.L. Benkard (2005), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, forthcoming.

Baltagi, B., Hidalgo, J., and Q. Li (1996), " A Nonparametric Test for Poolability using Panel Data," *Journal of Econometrics*, **75**: 345-367.

Bin, O. (2004), "A Prediction Comparison of Housing Sales Prices by Parmetric versus Semi-parametric regressions", *Journal of Housing Economics*, **13**: 68-84.

Blundell, R. , Duncan, A., and K. Pendakur (1998), "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, **13**: 435-461.

Brown J. N. and H.S. Rosen (1982), "On the estimation of structural hedonic price models," *Econometrica*, **50**: 765-68.

Chattopadhyay, S. (1999), "Estimating the demand for air quality: new evidence based on the Chicago housing market," *Land Economics*, **75**: 22-38.

Chen, R., Härdle, W., Linton, O.B., and E. Severance-Lossin (1996), "Nonparametric Estimation of Additive Separable Regression Models", in: *Statistical Theory and Computational Aspects of Smoothing*, Physica-Verlag, 247-265.

Ekeland, I., Heckman, J.J., and L. Nesheim (2004), "Identification and Estimation of Hedonic Models", *Journal of Political Economy*, **112**s: 60-109.

Fan, J., and I. Gijbels (1996), *Local Polynomial Modelling and Its Application*, London: Chapman and Hall.

Fan, Y., and Q. Li (1996), "Consistent Model Specification Tests : Ommitted Variables and Semiparametric Functional Forms," *Econometrica* **64**: 865-890.

Gençay R., and X. Yang (1996), "A Forecast Comparison of Residential Housing Prices by Parametric vs Semiparametric Conditional Mean Estimators", *Economic Letters* **52**: 129-135.

Gozalo P. L. and O. B. Linton (2001), "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," *Journal of Econometrics*, **104**: 1-48.

Härdle, W. and E. Mammen (1993), "Comparing non parametric versus parametric regression fits ", *Annals of Statistics*, **21**: 1926-1947.

Härdle, W. and A. B. Tsybakov (1993), "How Sensitive Are Average Derivative?", *Journal of Econometrics*, **58**: 31-48.

Hastie, T.J., and R. Tibshirani (1990), *Generalized Additive Models*, London: Chapman and Hall.

Horowitz, J.L. (1987), "Identification and stochastic specification in Rosen's hedonic price model," *Journal of Urban Economics*, **22**: 165-173.

Horowitz, J.L. (1998), *Semiparametric Methods in Econometrics*, Springer-Verlag, Berlin.

Horowitz, J.L., and W. Härdle (1996), "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates", *Journal of the American Statistical Association*, **91**: 1632-1640.

Horowitz, J.L., and S. Lee (2001), "Semiparametric Methods in Applied Econometrics," *Statistical Modelling* **2**: 3-22.

Horowitz, J.L., and V. G. Spokoiny (2001), "An Adaptative, Rate-optimal Test of a Parametric Mean-regression Model against a Nonparametric Alternative." *Econometrica* **69**: 569-631.

Iwata, S., Murao, H., and Q. Wang (2000), "Nonparametric Assessment of the Effects of Neighborhood Land Uses on Residential House Values," in: Fomby T.B. and R. Carter Hill (eds), *Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics*, JAI Press Inc., Stamford, 229-259.

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

Lavergne, P. and A. Thomas (2005) : "Semiparametric estimation and testing in a model of environmental regulation with adverse selection," *Empirical Economics*, **30**:171Ű192.

Lavergne, P. and Q. Vuong (1996) : "Nonparametric selection of regressors," *Econometrica*, 64(1), 207-219.

Linton, O.B., and J.P. Nielsen (1995), "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration", *Biometrika*, **82**: 93-100.

Matzkin, R..L. (2003), "Nonparametric Estimation of Nonadditive Random Function," *Econometrica*, 71: 1339-1375.

Martins-Filho, C., and O. Bin (2005), "Estimation of Hedonic Price Functions via Additive Nonparametric Regression", *Empirical Economics*, **30**: 93-114.

McMillen, D.P., and P. Thorsnes (2000), "The Reaction of Housing Prices to Information on Superfund Sites: A Semi-parametric Analysis of the Tacoma, Washington Market," in: Fomby T.B. and R. Carter Hill (eds), *Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics*, JAI Press Inc., Stamford, 201-228.

Pace, K.R. (1993), "Nonparametric Methods with Applications to Hedonic Models," *Journal of Real Estate Finance and Economics* **7**: 185-204.

Pace, K.R. (1998), "Appraisal using Generalized Additive Models," *Journal of Real Estate Research* **15**: 77-100.

Palmquist, R.B., Roka, F.M., and T. Vukina (1997) "Hog Operations, Environmental Effects, and Residential Property Values," *Land Economics* **73**: 114-124.

Pinkse, J. (2001), "Nonparametric Regression Estimation using Weak Separability," mimeo, Department of Economics, The Pennsylvania State University.

Powell J. L., Stock J. H. and T. M. Stoker (1989) "Semiparametric estimation of index coefficients," *Econometrica* **57**: 1403-1430.

Ready R. and C. W. Abdalla (2005), "The amenity and disamenity impacts of agriculture: estimates from a hedonic pricing model," *American Journal of Agricultural Economics*, **87(2)**: 314-326.

Robinson, P.M. (1988), "Root-N-Consistent Semi-parametric Regression," *Econometrica* **56**: 931-954.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* **82**: 34-55.

Schmalensee, R., and T. M. Stoker (1999) "Household Gasoline Demand in the United States," *Econometrica* **67**: 645-662.

Severance-Lossin, E., and S. Sperlich (1999), "Estimation of Derivatives for Additive Separable Models," *Statistics* **33**: 241-265.

Silverman, B.W. (1986), *Smoothing Methods in Statistics*, London: Chapman and Hall.

Stock, J.H. (1991), "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits," in: Barnett W.A., Powell J. and G. Tauchen (eds), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge, 77-98.

Tjøstheim, D., and B.H Auestad (1995), "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of the American Statistical Association* **89**: 1398-1409.

Appendices

A Bandwidth choices for the first-step parameter estimates

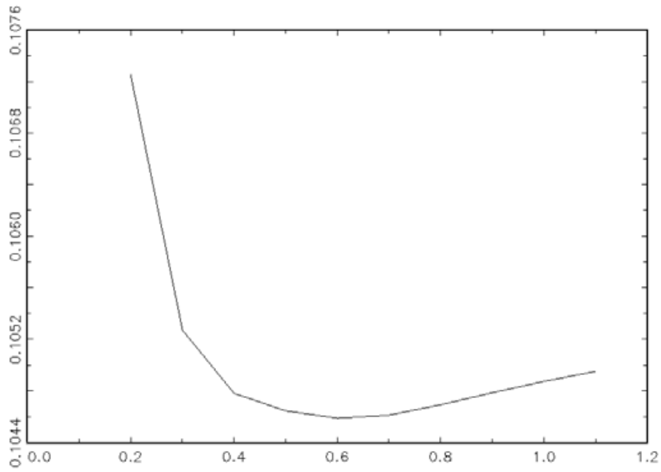


Figure 5 : Kondo and Lee (2003) cross-validation criterion.

B Bandwidth choices for the estimation of the nonlinear part of the hedonic price function

B.1 Fully nonparametric specification

For the nonparametric specification, we built a grid search and use a cross-validation criterion for the bandwidth choice, solving the following equation:

$$\min_h CV(h) = \frac{1}{n} \sum_{i=1}^{n} \left(W_i - \widehat{m}_h^{-i}(z_i) \right)^2 \tag{19}$$

where:

- $(W_i)_{i=1,\dots,n}$ are the first-step residuals $(Y_i - \widehat{\beta} X_i)_{i=1,\dots,n}$.

- $\widehat{m}_h^{-i}(z_i)$ denotes the *leave-one-out* local polynomial estimator used for the cross validation.

- Each bandwidth is computed with respect to the distribution of z_i (using its standard deviation $\sigma(z_i)$) and the theoretical rate of convergence, so that $h_i = h_{0i}\ \sigma(z_i)\ n^{-1/6}$ for $i = 1, 2$

B.2 Nonparametric additive specification

For the additive specification, derivation of the nonparametric estimator for the functions $g_l(.)$ $l = 1, \cdots, L$ involves the choice of a vector of bandwidths. As $L = 2$ in our application, we follow Linton and Nielsen's (1995) rule of thumb, controlled by the theoretical guidelines given by their theorem, which yields the following bandwidth h_l for $\widehat{g}_l(\cdot)$ and $l = 1, 2$:

$$h_l = h_{0l} \left\{ \frac{\widehat{\sigma}^2 \nu(K)(max(z_l) - min(z_l))}{\mu(K)^2(\widehat{\theta}_1 + \widehat{\theta}_2)^2} \right\}^{\frac{1}{5}} n^{-\frac{1}{5}}$$

where:

- h_{0l} is a multiplicative factor used to tune the bandwidth according to Linton and Nielsen's rule of thumb.

- $\widehat{\theta}_l$ are respective coefficients associated with $z_l^2/2$, $l = 1, 2$, obtained by regressing W against a constant, z_1, z_2, $z_1 z_2$, $z_1^2/2$ and $z_2^2/2$, while $\widehat{\sigma}^2$ is the variance of the residuals of this latter regression.

- The terms $\nu(K)$ and $\mu(K)$ are constant parameters that depend only on the kernel being used.

B.3 Single-index specification

For the single-index specification of $m(z)$, we follow the guidelines given by Härdle and Tsybakov (1993), and used by Horowitz and Härdle (1996), to choose the bandwidth in the single-index estimation. These latter authors showed that the asymptotically optimal bandwidth h_{DWADE} for the DWADE estimator has the following form:

$$h_{DWADE} = h_0\ n^{-\frac{2}{2p+L+2}} \tag{20}$$

where L is the dimension of vector Z, and assuming that all the partial derivatives of the density of Z exist up to order $p + 1$. In our empirical application, $p = 1$ and $L = 2$, which implies that $h_{DWADE} = h_0 \, n^{-1/3}$. We then use an empirical rule of selection with a grid search on h_0 to select the bandwidth for our application.

C Sensitivity analysis for the specification tests

Table 4: Tests statistics $\widehat{\tau}_2$ and T^c for various choices of the bandwidth

c	0.4	0.6	0.8	1.0	1.2	1.4
$\widehat{\tau}_2$	0.032	0.040	0.048	**0.049**	0.047	0.046
T^c	3.500	**3.342**	3.393	3.434	3.438	3.409