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

Christophe Bontemps, *Toulouse School of Economics, GREMAQ-INRA, France*
Michel Simioni, *Toulouse School of Economics, GREMAQ-INRA-IDEI, France.*
Yves Surry, *Swedish University of Agricultural Sciences, Uppsala, Sweden*

June, 2007[1]

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 has a significant nonlinear impact on house prices.

[1]Correspondance to: Michel Simioni, University of Toulouse (GREMAQ-INRA-IDEI), UMR 1291, 21 Allée de Brienne, F-31000 Toulouse, e-mail: michel.simioni@toulouse.inra.fr.
We thank Anna Alberini, Gabor Kezdi, Pascal Lavergne, Céline Nauges, John Rust, Julien Salanié, two referees, and participants of the EAERE 13^{th} annual conference, Budapest, of the XI^{th} EAAE congress, Copenhagen, and of the AAEA 2005 annual meeting, Providence, for their remarks. All remaining errors are ours.

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 implicit prices of these environmental attributes and, indirectly, an estimate of the consumers' willingness to pay to avoid 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

[2]See Bontemps *et al.* (2006) for an extensive list of references.

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 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 odors and the release/emission of nitrates that pollute the soil, affect water quality and seep 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,[3] and the proportion of permanent grassland converted into cultivated

[3]In studies analyzing this problem, impacts of agricultural pollution (e.g. intensive livestock operations) on house prices are measured using a proximity index reflecting the distance between residential houses and the sites of agricultural pollution (Palmquist al., 1997; Herriges et al., 2005; Ready and Abdalla, 2005). As explained furthering Section 4,

grassland.[4]

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.[5] We choose a partially linear specification, in which all the explanatory variables, apart from the two environmental indicators, are incorporated linearly into the hedonic

this approach cannot be adopted due to the lack of relevant information on the localization of house sales and the sources of agricultural pollution.

[4]The essential and traditional feature of the Brittany countryside is "*bocage*" which is a rural landscape made up of (often small) parcels of land bordered by hedgerows. Permanent grassland is associated with extensive systems that are more respectful of hedgerows, soil and water quality, hence favoring the maintenance of the "*bocage*" landscape in Brittany (Le Goffe, 2000). On the other hand, cultivated grassland (consisting of temporary pastures and artificial meadows) is associated with more intensive agricultural activities which require larger parcels of land and hence the elimination of hedgerows. In addition, it has negative effects on soil and water quality, resulting in a degradation of the environment. Armed with these observations, the use of the proportion of permanent grassland converted into cultivated grassland as an indicator of the degradation of rural landscape is well justified.

[5]The four empirical house price models reported in this article are consistent with the two-stage hedonic model framework developed by Rosen (1974). However, unlike Rosen model, they do not allow in a second stage for a recovery of the consumers' willingness to pay (WTP) functions for environmental nuisances. The reason for this stems from the fact that we do not have at our disposal socio-economic data on each house buyer, hence preventing us from identifying structural parameters (buyers' tastes and habits) that are needed to estimate the WTP functions in the second stage. Given these considerations, the four estimated house price models should be viewed as first-stage hedonic price models from which hedonic price gradients or implicit prices for environmental nuisances are derived.

price function. The two pollution indicators are included in the hedonic price function in a nonparametric or semiparametric way using the three nonlinear 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 residuals of the first-stage estimation procedure. Finally, we perform recent specification tests in order to compare the nonparametric additive, single-index and parametric specifications with the more general specification, which is fully nonparametric.

The specification tests only select the nonparametric additive specification. Implicit prices for pollution reduction are then computed for this selected model specification using a procedure for estimating derivatives for additive separable models. Moreover, we find that pollution resulting from intensive livestock farming in rural districts is a more crucial environmental issue than "pollution" due to the degradation of the rural landscape, 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. All the technical details involved in the computation of the various estimators and test statistics used in this paper, are fully described in a companion working paper (Bontemps *et al.*, 2006).

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. Following Rosen (1974), 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 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. We also assume that ξ enters as an additive term in the hedonic price function as in Bajari and Benkard (2005) and Bajari and Kahn (2005), i.e.

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

Only recently, Ekeland, Heckman and Nesheim (2004) showed that the identification problems highlighted in the hedonic price 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 nonparametrically identified if the unobserved characteristic is independent of the observed characteristics. Thus, the hedonic price function may have a general non-additive structure, and it would seem appropriate to consider nonparametric regression estimators as natural candidates for estimating the hedonic price function (2). But, unfortunately, we would face in such a procedure the well-known "curse of dimensionality" problem, 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

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

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	Partially linear and nonparametric model	
$m(z) = m(z_1, \cdots, z_L)$	$Y = \beta' X + m(Z_1, \cdots, Z_L) + \xi$	(**M1**)
Additive specification	Partially linear and additive model	
$m(z) = \sum_{l=1}^{L} g_l(z_l)$	$Y = \beta' X + \sum_{l=1}^{L} g_l(Z_l) + \xi$	(**M2**)
Single index specification	Partially linear and single index model	
$m(z) = G(\gamma' z)$	$Y = \beta' X + G(\gamma' Z) + \xi$	(**M3**)
Parametric specification	Fully parametric model	
$m(z) = \gamma' z$	$Y = \beta' X + \gamma' Z + \xi$	(**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.[6]

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 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:

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

[6] 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.

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 $E(X|Z = z)$ (see Fan and Gijbels, 1996). 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. Automatic bandwidth selection criteria, such as plug-in or cross-validation, can be used to choose the vector of bandwidths for estimating $E(Y|Z = z)$ and $E(X|Z = z)$. In the following, we apply the cross-validation criterion recently proposed by Kondo and Lee (2003). We define a set of bandwidths $h_j = a\ \sigma_j\ n^{\frac{-1}{4+L}}$, $j = 1, \ldots, L$, 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.

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 $m(z) = E(W|Z = z)$. 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 can be estimated using a procedure based on "marginal integration" proposed by Linton and Nielsen (1995) in the case $L = 2$, and extended to higher dimensions by Tjøstheim and Auestad (1995) and Chen *et al.* (1996). The idea behind this estimator is quite straightforward. In the case of additivity, there exists functions g_l and m_{-l} such that

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

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

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

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 sample average over the observations W_i which we denote by $\widehat{c}$. We estimate the left hand side of equation (8) by replacing the expectation by a sample 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 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 dimension of the regressors' space to one through the index. However, there is a drawback in terms of identification since equation (9) is equivalent to $m(z) = G^*(\kappa + \delta(\gamma' z))$, for any arbitrary value of the location parameter κ and the scale parameter $\delta \neq 0$ and function G^* defined by the relation $G^*(\kappa + \delta v) = G(v)$ 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.

As Z only has continuous components, a Density Weighted Average Derivative estimator (DWADE) can be used to directly estimate γ without solving any optimization problem (see Powell *et al.*, 1989). This estimator is based on the fact that γ is proportional to $E\left[\frac{\partial G(\gamma' Z)}{\partial Z}\, \varphi(Z)\right]$ up to a multiplicative term, where $\varphi(\cdot)$ denotes the unknown density function of Z.[7] With suitable assumptions on the function $G(\cdot)$ and the density function $\varphi(\cdot)$, integration

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

by parts yields:

$$E\Big[\frac{\partial G(\gamma' Z)}{\partial Z}\ \varphi(Z)\Big] = -2E\Big[Y\ \frac{\partial \varphi(Z)}{\partial Z}\Big] \tag{10}$$

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

$$\gamma_{DWADE} = -\frac{2}{n}\sum_{i=1}^{n} Y_i \frac{\widehat{\partial\varphi(Z_i)}}{\partial Z} \tag{11}$$

where we replace in (10) the expectation by the sample 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.

3.3 Specification tests

As stated in the introduction, one of the contributions of this study is to adopt recent 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 one:

$$\widehat{\tau}_2 = \frac{1}{n^2 h^L} \sum_i \sum_{j \neq i} K_{ij} \, \widetilde{u}_i \, \widetilde{u}_j \; \pi(Z_i)\pi(Z_j), \tag{12}$$

where h denotes the bandwidth involved in the estimation of the fully non-parametric 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 counter-part $\widehat{V}_{2,n}$ 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 $I^c \equiv E[\nu E[\nu|Z]]$. This statistic is used because, by the law of iterative expectations, we observe that $I^c = E[E[\nu|Z]^2] \geqslant 0$, with equality holding 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 this 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:

$$I_n^c = \frac{1}{n}\sum_{i=1}^n \left(\widehat{\nu_i}\ \widehat{f_{h_\gamma}}(\widehat{\gamma}' Z_i)\right)\left(\frac{1}{(n-1)h^L}\sum_{j\neq i}(\widehat{\nu_j}\ \widehat{f_{h_\gamma}}(\widehat{\gamma}' Z_j))\ K_{ij}\right) \quad (13)$$

where $\widehat{\nu_i} = W_i - \widehat{G}(\gamma' Z_i)$ 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

$$T^c = \frac{n h^{L/2}\ I_n^c}{\sqrt{2}\sigma_c} \quad (14)$$

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

3.3.3 Parametric vs. nonparametric

We are now concerned with testing a parametric model of the function $m(.)$ against a nonparametric alternative. Recently, Horowitz and Spokoiny (2001) developed a 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 $M(\cdot, \gamma)$.

$$S_h(\gamma) = \sum_{i=1}^{n} \Big(\widehat{m_h}(Z_i) - \widehat{M_h}(Z_i, \gamma)\Big)^2 \tag{15}$$

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:

$$T^\star = max_{h \in H_n} \frac{S_h(\gamma) - \widehat{N_h}}{\widehat{V_h}} \tag{16}$$

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.[8] This database provides a detailed description of all house sales in France, including sale prices, physical attributes of houses and adjacent lot sizes. 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).

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).[9] The MIN database enables us to determine for each district the proportion of vacant houses (VACANT).

[8] *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.

[9] 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*".

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.[10]

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.[11] Agricultural pollution is thus measured by the following two aggregate indicators.[12] 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 grassland converted into cultivated grassland. A high value associated with this variable would indicate a degradation of the rural landscape.[13]

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

[11]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.

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

[13]The two indicators - NITRO and TMEAD - are interrelated in the sense that they could also capture "common" environmental influences that are difficult to isolate. A good example of such a phenomenon could be the fact that pasture land could support cattle,

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

[Insert Table 1 and Figure 1 here]

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 three steps involved in the

thereby leading to the increases in the emission of nitrogen. Although such "common" environmental influences may exist, they tend not to be harmful to the environment and even are relatively small compared to the pollution impacts of intensive livestock operations that are the major contributors of agricultural pollution in Brittany. As a result, we do not have investigated this matter further and explicitly assume that the impacts of these two environment indicators on the price of houses are independent or separable from each other.

specification search procedure presented in the previous section. At the end of the section, we compute the implicit prices 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. All the estimated parameters belonging to the linear part of the housing price models are statistically significant and have the expected signs and magnitudes.[14] 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,

[14]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.

while opposite effects take place in districts that are either more populated or have households with higher average incomes.

[Insert Table 2 here]

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.

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.[15]

[Insert Figure 2 here]

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

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. A detailed description of these analysis is given in Bontemps *et al.* (2006).

[Insert Table 3 here]

Table 3 summarizes the results of the specification tests. These 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 findings 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 Implicit prices

Based on the specification test diagnosis given above, we report here the implicit prices (IP) for pollution reduction estimated by the nonparametric additive model (M2). These computed values are derived using Severance-Losin and Sperlich's (1999) estimator. 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

$$E_{Z_{-l}}\left[\frac{\partial m(z_l, Z_{-l})}{\partial z_l}\right] \equiv \int \frac{\partial m(z_l, z_{-l})}{\partial z_l}\, \varphi_{-l}(z_{-l})\, dz_{-l} = E_{Z_{-l}}\left[\frac{dg_l(z_l)}{dz_l}\right] = \frac{dg_l(z_l)}{dz_l} \quad (17)$$

An estimate of the derivative $dg_l(z_l)/dz_l$ is thus obtained by estimating the left hand side of equation (17), i.e. by replacing the expectation by a sample 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 mean implicit price function expressed as a percentage of the corresponding housing prices.[16]

[Insert Figures 3.a and 3.b here]

The two Figures reveal that the relationships between implicit prices and the pollution indicators are highly nonlinear for specific ranges of values taken

[16] It is important to note that IP estimates presented in Figures 3 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.

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 cultivated grassland increases from 20 to 40%. Then, the implicit prices tend to flatten out when this indicator rises to values greater than 40%. In addition, Figure 3.a clearly show changes in the degree of curvature of the relationship between IP and "landscape degradation".

On the other hand, a different pattern seems to emerge for IP estimates associated with livestock nitrogen emissions (NITRO). An examination of Figure 3.b reveals that the relationship between the mean IP 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 IP 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 IP 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 IP 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 IP estimates obtained with model M4 are constant and equal to 3% of the house price (assuming a 10 percent change in the proportion of cultivated grassland). 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 IP for livestock nitrogen emissions.

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 implicit prices 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.[17].

References.

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

Bajari P., Kahn M.E. 2005. Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities. *Journal of Business and Economic Statistics* 23: 20-33.

Bontemps C., Simioni M., Surry Y. 2006. Semiparametric Hedonic Price Models : Assessing the Effects of Agricultural Nonpoint Source Pollution. Working paper 2006/02, INRA-ESR , Toulouse.

(Available at http://www.toulouse.inra.fr/ centre/esr/wpRePEc/bon200602.pdf.)

Brasington, D.M., Hite, D. 2005. Demand for Environmental Quality: a Spatial Hedonic Analysis. *Regional Science and Urban Economics* **35**: 57-82.

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

Chernih A., Sherris M. (2004). Geoadditive Hedonic Price Models. Working paper 2004-5, Department of Actuarial Studies, Faculty of Business, University of New South Wales, Sidney.

Ekeland I., Heckman J.J., Nesheim L. 2004. Identification and Estimation of Hedonic

[17] On all these matters, see the recent works of Chernih and Sherris, 2004, Brasington and Hite, 2005, and Parmeter *et al.*, 2007

Models. *Journal of Political Economy* **112**: 60-109.

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

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

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

Herriges J.A., Secchi S., Babcock B.A. 2005. Living with Hogs in Iowa: The Impact of Livestock Facilities on Rural Residential Property Values. *Land Economics* **81**: 530-545.

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

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

Le Goffe, P. 2000. Hedonic Pricing and Agricultural and Forestry Externalities. *Environment and Ressource Economics* **15**: 397-401.

Linton, O.B., Nielsen J.P. 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.

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

Parmeter, C.F., Henderson, D.J., Kumbhakar S.C. 2007 . Nonparametric Estimation of a Hedonic Price Function. *Journal of Applied Econometrics* **22**: 695-699.

Powell J.L., Stock J.H. Stoker T.M. 1989. Semiparametric Estimation of Index Coefficients. *Econometrica* **57**: 1403-1430.

Ready R., Abdalla C. W. 2005. The Amenity and Disamenity Impacts of Agriculture: Estimates from a Hedonic Pricing Model. *American Journal of Agricultural Economics*, **87**: 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.

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

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

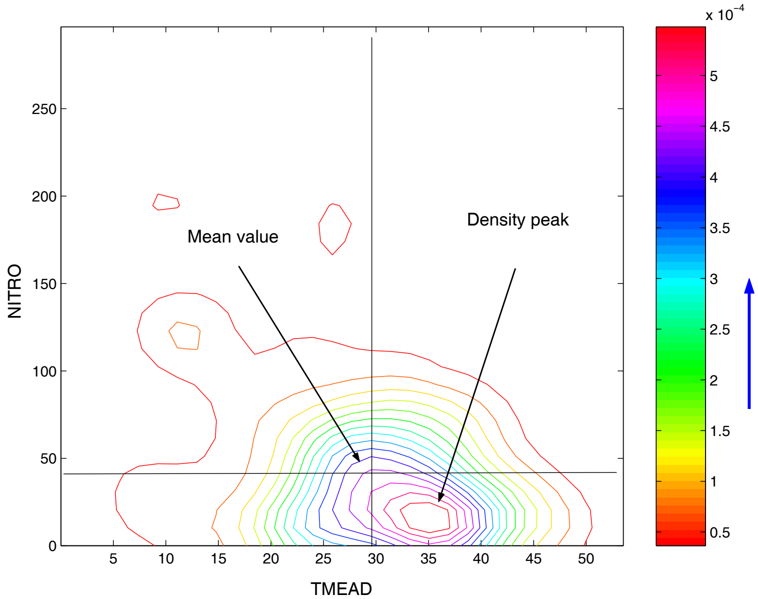
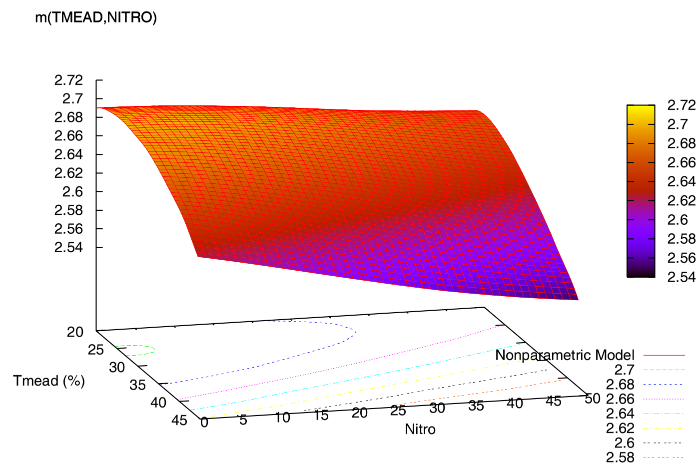
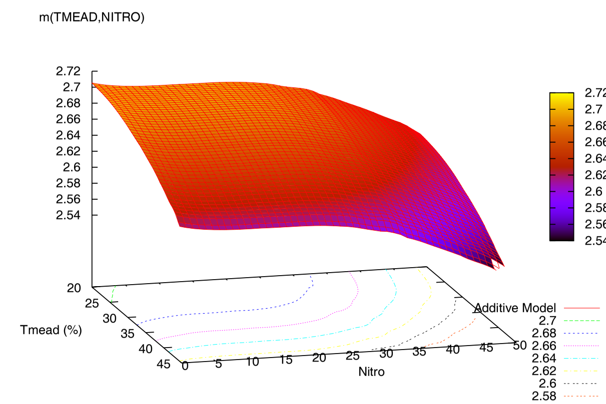


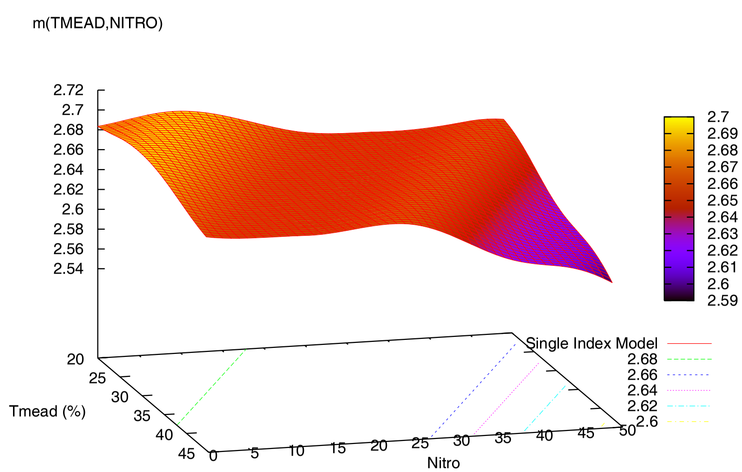
Figure 1 : Joint iso-density curves for the two environmental variables (z_1, z_2)



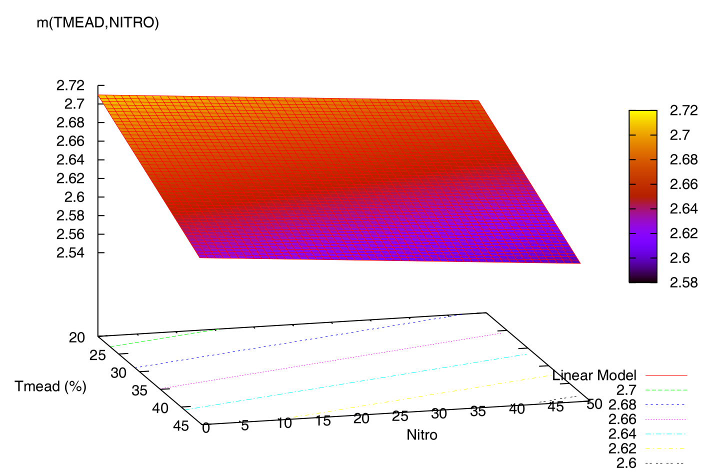
Nonparametric specification (M1)



Nonparametric additive specification (M2)



Single index specification (M3)



Parametric specification (M4)

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

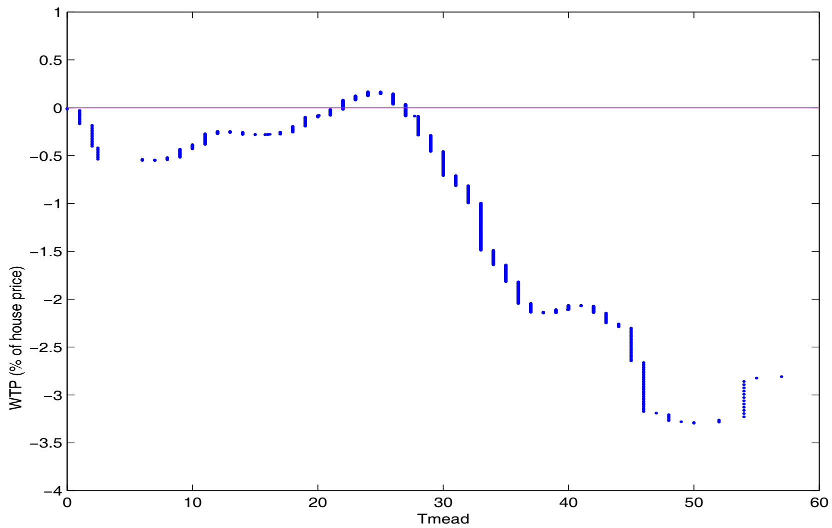


Figure 3.a : IP for landscape degradation - Additive model M2.

(expressed as a percentage of house prices)

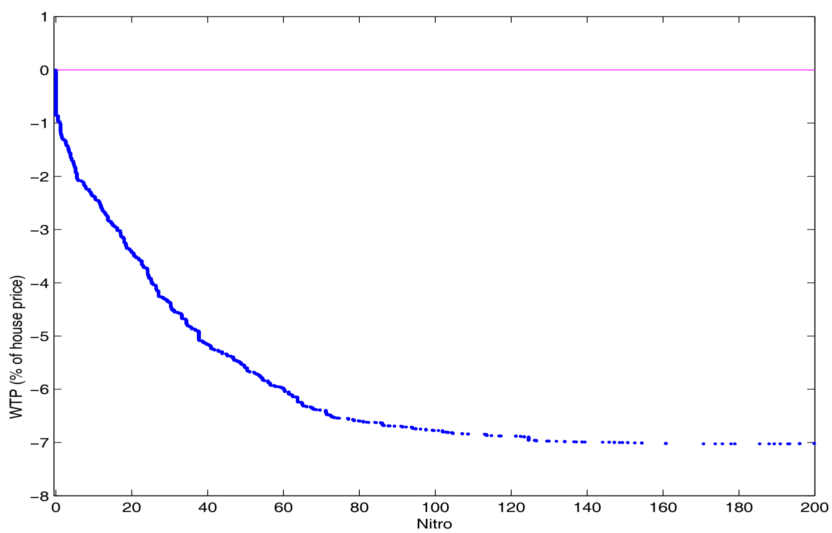


Figure 3.b : IP for livestock nitrogen emission - Additive model M2

(expressed as a percentage of house prices))

Note: In both Figures, dots represent IP estimated for each observation of the data sample

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	m^2	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

Total data sample of 2092 observations distributed over 516 districts with less than 5000 inhabitants

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*

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 favorable case for the statistic $\widehat{\tau}_2$ and T_c using the sensitivity analysis provided in appendix C of Bontemps *et al* (2006). The p-values are either the asymptotic or bootstrapped ones