Nonparametric vs parametric binary choice models:
An empirical investigation

Christophe Bontemps[1], Jeffrey S. Racine[2], and Michel Simioni[3]

(1)   Toulouse School of Economics, INRA-GREMAQ
(2)   McMaster University, Canada
(3)   Toulouse School of Economics, INRA-GREMAQ and IDEI
      21, Allées de Brienne, 31000 Toulouse, France

Nonparametric vs parametric binary choice models: An empirical investigation

1 Introduction

Recent advances in nonparametric kernel estimation have challenged conventional (parametric) approaches towards modelling empirical relationships. Historically, attention has mainly focused on cases where the response and explanatory variables are continuous; see Pagan (1999) The literature dealing with categorical response variables is relatively scarce. In this application domain, efforts have been mostly confined to semiparametric approaches. These semiparametric methods relax rigid parametric specifications that are typically found in applied settings, an example being the single index model which allows for an unknown link function while maintaining the assumption that the explanatory variables enter the probability through a linear parametric index; see Horowitz (1998). In other models, explanatory variables enter through a general nonparametric additive structure, the link function being specified as the usual logistic or normal cumulative density function. As is the case with their fully parametric counterparts, semiparametric models will be inconsistent unless one has correctly specified the underlying data generating process, which is a drawback. Fully nonparametric methods, however, are immune to this drawback. However, few attempts to introduce purely nonparametric estimation methods are to be found in the literature on discrete choice; see Briesch et al. (2002). While these methods produce results that are consistent with the economic foundations of choice theory, they are difficult to implement in practice.

The estimation of conditional PDFs in a nonparametric kernel framework has recently been studied by Hall et al. (2004). The modelling of conditional PDFs may prove extremely useful for a range of tasks including modelling and predicting consumer choice which is the focus of this paper. The aim of this paper is threefold. First, we employ a fully nonparametric model of a conditional PDF comprised of a binary response (choice) variable and continuous and discrete explanatory variables. Second, we address the issue of the performance of this nonparametric estimator relative to the parametric Probit specification which is dominant in applied settings and evaluate these estimators in a variety of ways. Third, we provide a detailed discussion of the results focusing on environmental insights provided by the two estimators, emphasizing how particular patterns detected using the nonparametric estimator are masked by the parametric specification.

The empirical application reported here concerns an environmental question that was addressed in Bontemps and Nauges (2009). In France, despite access to safe public drinking water, and in spite of its excessively high price compared to tap water, 42% of the population still regularly drink bottled water. Using scanner data on French consumption combined with raw water quality and other environmental data, and using a Probit model, Bontemps and Nauges (2009) show that poor raw water quality seems to be the most important factor driving the decision not to drink tap water. The estimated effect is found to be stronger for

low-income households. Significant direct impacts of socioeconomic and demographic households' characteristics, as well as the role of cultural/regional factors are revealed. Overall, this study shows that pollution of raw water implies indirect costs for households who instead of drinking water from the tap spend up to 100 times more for bottled water. We revisit this question using the same dataset and using a nonparametric kernel estimator of the conditional PDF in order to recover estimates of the probability to drink tap water. We address the three issues mentioned above, and show that the nonparametric estimator outperforms the parametric one based upon a variety of measures.

The nonparametric specification we consider uses the identical information set as the Probit specification, while we use recent developments on generalized kernel estimation to deal with the presence of both continuous and discrete variables; see Li and Racine (2007). Moreover, the nonparametric framework allows us to assess the relevance of the explanatory variables through the data-driven bandwidth selection method. We find comparable in-sample predictive performance as measured by the overall correct classification ratio (70.04% for the nonparametric specification, 69.17% for the Probit). We then address the issue of selecting a preferred model specification for the data at hand. We take the view that fitted statistical models are approximations, a perspective that differs from that of consistent models selection which posits a finite dimensional 'true model'. Recently, Racine and Parmeter (2009) propose a test based on this idea that we implement here. The test results indicate that the nonparametric specification possesses an expected 'true error' that is statistically significantly lower than that for the parametric specification and is therefore to be preferred. Finally, we contrast the parametric specification with the nonparametric specification via an examination of some "interesting" cases.

The rest of the paper proceeds as follows. Section 2 outlines the nonparametric estimator and the test procedure. Section 3 presents the empirical application of these methods to the environmental question raised in Bontemps and Nauges (2009). Section 4 presents some concluding remarks.

2 Nonparametric estimation and test procedures

2.1 Nonparametric PDF estimator

Let $f(.)$ and $m(.)$ denote the joint and marginal density of (X, Y) and X, respectively, where Y is a binary discrete variable and where we allow X to include both continuous, unordered, and ordered variables. For what follows, we shall refer to Y as a dependent variable (i.e., Y is explained), and to X as covariates (i.e., X is the vector of the explanatory variables). The density of Y conditional on X is then defined as

$$g(y|x) = \frac{f(x,y)}{m(x)} \tag{1}$$

Consider the kernel estimators of the previous joint and marginal densities we denote by $\widehat{f}$ and $\widehat{m}$. We then estimate the conditional density by replacing the unknown densities in (1) by their estimators, i.e.

$$\widehat{g}(y|x) = \frac{\widehat{f}(x,y)}{\widehat{m}(x)} \qquad (2)$$

As we are facing a mix of discrete (unordered and ordered) and continuous variables when estimating the two unconditional densities, we use the "generalized product kernel" estimator proposed by Li and Racine (2003). A least-squares cross validation approach proposed by Hall et al. (2004) is used to select the bandwidths involved in the computation of the kernel estimators.

2.2   Preferred model selection

We consider two non-nested model specifications, a parametric Probit specification and a nonparametric kernel conditional probability specification. Both models use identical information sets and deliver estimates of the probability that $Y = 0/1$ conditional on the covariates X.

We approach the issue of selecting a parametric versus nonparametric specification from the perspective that fitted statistical models are approximations. Clearly our perspective is distinct from that of consistent model selection which posits a finite-dimensional 'true model'. We consider selection of a parametric versus nonparametric specification not from the perspective of a test that posits that one model is the 'true' one. Rather, both are at best approximations, therefore we select that model that has lowest expected 'true error'.

Our approach is therefore firmly embedded in the statistics literature dealing with 'apparent' versus 'true' error estimation; for a detailed overview of expected apparent and excess error, we direct the reader to (Efron, 1982, Chapter 7). In effect, in-sample measures of fit such as the standard error of the regression or R^2 and so forth measure 'apparent error' which will be smaller than 'true error' which is the expected error when the model is used to predict new draws from the underlying data generating process. For example, for a continuous regression model $Y_i = g(X_i) + \varepsilon_i$, one might compute the Average Square Prediction Error or ASPE given by $ASPE = n_1^{-1} \sum_{i=1}^{n_1} (Y_i - \hat{g}(X_i))^2$ which is a measure of apparent error. But all such in-sample measures are fallible which is why they cannot be recommended as guides for model selection. Our procedure can be thought of as a means of *estimating* a model's *true error* and *testing* whether the true error is statistically smaller for one model than another. We adopt the 'revealed performance', or RP, test proposed by Racine and Parmeter (2009).

Before introducing the RP test, let us recall how predictive performance is measured in binary choice models. Different indices can be used to measure this predictive performance. Efron (1978) gives a detailed discussion of such indices. The most common index used is the Correctly Classified Ratio (CCR) or *accuracy*. This index measures the exact proportion of correct predictions for the data at hand. That is, for each observation i, we compute the

value of the loss function

$$Q(Y_i, \eta_i, \alpha) = \begin{cases} 0 & \text{if } Y_i = 1 \text{ and } \eta_i > \alpha \text{ or if } Y_i = 0 \text{ and } \eta_i \leq \alpha \\ 1 & \text{otherwise} \end{cases}$$

where η_i is the probability assigned by the binary choice model to the ith observation, and α is the cut-off-value used to map the classifier, namely the probabilities η_i, to classes of predicted 0 or 1. For a given cut-off-value (usually, $\alpha = 0.5$), the CCR is then computed as

$$CCR(\alpha) = 1 - \frac{\sum_{i=1}^{n} Q(Y_i, \eta_i, \alpha)}{n}. \tag{3}$$

This index can be also linked to the so-called confusion matrix we define in Table 1 where ON and OP are the total numbers of observed 0 and 1 respectively, TN stands for 'true negative', occurring when both the observed value and the prediction outcome ($\eta_i \leq \alpha$) are 0, and FN for 'false negative', when the observed value is 0 and the prediction outcome is 1 ($\eta_i > \alpha$), while FP and TP are the 'false' and 'true positive' respectively.

Table 1: Notation

		Predicted					
		0	1	Total	*Accuracy* (CCR)	=	$\frac{TN+TP}{n}$
Obs.	0	TN	FP	ON	*Sensitivity* (TPR)	=	$\frac{TP}{TP+FN}$
	1	FN	TP	OP			
Total				**n**	*Specificity* (SPC)	=	$\frac{TN}{TN+FP}$

Confusion matrix   **Index**

Though the CCR index depends upon the choice of the cut-off probability, a related and well-known classification performance metric that does not depend on a cut-off value is the "Receiver Operating Characteristic" curve (ROC), described in Egan (1975). This curve is a graphical plot of the *sensitivity* (percentage of predicted true positive) versus $1-$ *specificity* (percentage of predicted false positive) (see Table 1 for more precise definitions), letting the classification cut-off-value vary between its extremes. The AUROC, i.e., the "Area Under the Receiver Operating Characteristic" curve, can then be computed as a summary measure of the classification performance. AUROC lies between 0.5 (worthless classification) and 1 (perfect classification), thereby providing a more comprehensive evaluation ratio than the CCR index alone since it is independent of any particular cut-off-value.

Given the two previous measures (CCR and AUROC), we can define two different ways of measuring the 'apparent error' when estimating a binary choice model. For instance, the second term in the definition of the CCR measure (3) can be viewed as the empirical realization of $E_{n_1, \widehat{F}}[Q(Y^{n_1}, \eta_{n_1}^{n_1}), 0.5)]$ where $E_{n_1, F}$ denotes the expectation over the n_1 observed points $Z^{n_1} = \{Y_i, X_i\}_{i=1}^{n_1}$ which are independently and identically distributed with empirical cumu-

lative distribution function $\widehat{F}$ (we refer Z^{n_1} as the training sample, terminology borrowed from the literature on statistical discriminant analysis), $Y^{n_1} = \{Y_i\}_{i=1}^{n_1}$, and $\eta_{n_1}^{n_1} = \{\eta_i\}_{i=1}^{n_1}$, i.e. the vector of the assigned probabilities to the observations calibrated using the observed sample. In order to implement the RP test, we are interested in estimating a quantity known as 'expected true error'. Following Efron (1982), we can define the 'true error' to be

$$E_{n_2,F}[Q(Y^{n_2}, \eta_{n_1}^{n_2}, 0.5)]$$

where $E_{n_2,F}$ denotes the expectation over the n_2 new points $Z^{n_2} = \{Y_i, X_i\}_{i=n1+1}^{n}$ which are independently and identically distributed with cumulative distribution function F, and are independent of the training sample Z^{n_1} (we refer Z^{n_2} as the evaluation sample), $Y^{n_2} = \{Y_i\}_{i=n_1+1}^{n}$, and $\eta_{n_1}^{n_2} = \{\eta_i\}_{i=n1+1}^{n}$, i.e. the vector of the probabilities assigned to the new points calibrated using the training sample. Next, we define the 'expected true error' as

$$E\left(E_{n_2,F}[Q(.)]\right)$$

where the expectation is taken over all potential classifiers $\eta_{n_1}^{n_2}$, for the selected loss function $Q(.)$. When comparing two approximate models, the model possessing the lower 'expected true error' will be preferred in applied settings.

A realization of the 'true error' based upon the observed $z^{n_2} = \{y_i, x_i\}_{i=n1+1}^{n}$ is given by

$$\frac{1}{n_2} \sum_{i=n_1+1}^{n} Q(y_i, \eta_{n_1,i}^{n_2}, 0.5) \tag{4}$$

In the following, we consider the corresponding CCR for ease of interpretation. If we consider S splits of the data into training and evaluation samples, we can then construct the empirical distribution function of loss, or equivalently, of CCR. The algorithm is described in Racine and Parmeter (2009). One can then use the empirical distributions to form a (paired) test for assessing which specification has lower expected true error. We then repeat this procedure with AUROC replacing CCR.

3 Empirical application

3.1 Data

The data set used is based on two main sources. First, data on French households' purchases are provided by TNS Worldpanel, for the year 2001. This database contains information on French households' purchases of food items as well as households' socioeconomic and demographic information.[1] We define each household as a "tap water drinker" or "tap water

---

[1]Purchases of 10,000 surveyed households are recorded all aver the year 2001.

non-drinker" from observed purchases of non-alcoholic drinks.[2] Second we use various sources of environmental information at the municipality level to compute the index of *poor raw water quality* (`PRWQ` hereafter) as in Bontemps and Nauges (2009). Data on price of tap water (IFEN-SCEES-Agences de l'eau, 2001), data on raw water (Ministry of Health, 2001), and data on manure spreading by the Ministry of Agriculture (2000) at the municipality level were collected. Information on water supply management chosen at the municipality level (public versus private) is also included via the dummy variable `deleg`. Finally, we compute the index of poor raw water quality and merge it to the households' panel through the residential address information of each household.

In addition to the *poor raw water quality* index, we observe the following socioeconomic and demographic characteristics at the household level: (1) Head of household's education level (`diploma`) (we distinguish four education levels: head without diploma (reference in the Probit model), head with diploma less than the *baccalauréat* (diplo.L), head with the *baccalauréat* or a higher diploma (diplo.Q), head for whom information is missing (diplo.C)), (2) household's monthly income (before income taxes) (`Income`), (3) rural or urban location (`rural`) (we build an indicator variable which takes the value of 1 if the household lives in a "commune" of less than 2,000 inhabitants, and 0 otherwise), (4) retirement status (`iret`) (we build an indicator variable which takes the value of 1 if household's head is retired and 0 otherwise[3]), and (5) household's geographic location (`region`). We follow the regional division chosen by TNS where France is divided in 8 main zones: Paris, East, North, West, Middle-West, Middle-East, South-East, and South-West.

The sample gathers 4,623 households from 1,282 distinct municipalities all over France. A complete description of the sample is given in Bontemps and Nauges (2009). We briefly summarize the main features present in this sample. There are 68% of households in the sample that are classified as "tap water drinkers". This percentage is close to what is usually found in polls at the national level. Regional difference are also observed in our sample. The highest percentage of tap water drinkers are in the Middle-East (82%), South-West (80%), and South-East (77%). The lowest percentage of tap water drinkers is observed in the North (56%). The average poor raw water quality index is of 0.93, varying from 0.87 in the North of France up to 0.97 in the Paris Region and in the west of France, these two regions being particularly affected by nitrogen pollution. Household monthly income varies from an average of 1,820 euros in Middle-West to an average of 2,490 euros in Paris and its surroundings. The proportion of households living in rural areas varies from 1% (Paris and surroundings) to 15% in the Western region where farming activity dominates. The share of retired households is quite homogeneous across regions, except in the South-East which attracts a high number of retirees due in part to its warmer climate.

---

[2]We follow the definition of Bontemps and Nauges (2009) by defining the household as "tap water drinker" (resp. "tap water non-drinker") if the average consumption of non-alcoholic drinks by person by day is lower (resp. greater) than 0.5 liters. The set of non-alcoholic drinks includes: bottled water, tea drinks, sodas, tonics, fruits and vegetables juices, etc.

[3]This variable was proven more significant that age of household's head in Bontemps and Nauges (2009).

3.2 In-sample performance

As mentioned in Section 2.2, we estimate two non-nested model specifications, a parametric Probit specification and a nonparametric kernel conditional probability specification. Both models use identical information sets and deliver estimates of the probability of drinking tap water conditional on the covariates $X = \{$PRWQ, Income, diploma, region, deleg, rural, iret$\}$. For the Probit specification, we add the interactions PRWQ $\times$ Income and PRWQ $\times$ iret in order to use the "preferred" parametric specification of Bontemps and Nauges (2009).

Table 2: Probit coefficient estimates and significance vs nonparametric bandwidth estimates and associated scale factors:

	Estimate	$Pr(\mid Z \mid > z)$	Bandwidth	upper bound
(Intercept)	2.3296	0.0000	-	-
PRWQ	-1.8113	0.0021	0.1801905	∞
Income	-0.5492	0.0155	1.294752	∞
diploma	-	-	0.8634835	1
diplo.L	-0.1328	0.0464	-	-
diplo.Q	0.0435	0.4433	-	-
diplo.C	-0.0229	0.5703	-	-
Region	-	-	0.1208747	0.875
Region2	-0.0284	0.7376	-	-
Region3	-0.5879	0.0000	-	-
Region4	-0.0590	0.3836	-	-
Region5	-0.0468	0.5887	-	-
Region6	0.3706	0.0000	-	-
Region7	0.1486	0.0576	-	-
Region8	0.2974	0.0005	-	-
deleg	-0.0178	0.6966	0.5	0.5
rural	0.2397	0.0095	0.0721212	0.5
iret	-1.3491	0.0089	3.253532e-13	0.5
PRWQ×Income	0.5789	0.0166	-	-
PRWQ× iret	0.9461	0.0871	-	-
irob	-	-	9.802058e-15	0.5

The bandwidths are chosen by minimizing a least-square cross-validation criterion.
The upper bound for a bandwidth, is equal to $(c_j - 1)/c_j$ in the case of an unordered discrete variable with c_j categories, and 1 in the case of an ordered one.

Table 2 reports the estimates of the parameters of the Probit specification and the bandwidths selected by least-squares cross-validation for the nonparametric specification.[4] Empirical significance levels allow us to test the significance of the associated variables in the parametric specification while the magnitudes of the selected bandwidths reveal the predictive relevance of the associated variables in the nonparametric specification. We observe that significance tests and relevance assessments provide coherent results. All the continu-

---

[4]All the computations are made using the np package of R software; see Hayfield and Racine (2008).

ous variables have significant effects on the probability to drink tap water in the parametric specification, and are relevant in the nonparametric specification as they are far from being "smoothed-out". The discrete variables, `rural` and `iret` are also relevant as their bandwidths are far from their upper bounds. This is not the case for the variable `deleg`, however. For the ordered discrete variables, the two specifications show that some regional effects exist while the diploma effects seem to be less clear.

Table 3: In-sample confusion matrices ($n = 4623$ and $\alpha = 0.5$)

		Probit specification					Nonparametric specification		
		Predicted					Predicted		
		0	1				0	1	
Obs.	0	143	1324	1467		Obs. 0	134	1333	1467
	1	101	3055	3156		1	52	3104	3156
		244	4379	**4623**			186	4437	**4623**

Table 3 reports the confusion matrices for the two specifications. As usually done when evaluating binary choice models, we use a cut-off value $\alpha = 0.5$ to map the classifiers, namely the estimated probabilities, to classes of predicted 0 ("the household does not drink tap water") or 1 ("the household does"). The comparison of these confusion matrices reveals that both specifications have a tendency to over-predict the fact drinking tap water, while the two specifications produce similar values for the usual in-sample performance measures as shown in Table 4.

Table 4: In-sample performance ($n = 4,623$ and $\alpha = 0.5$)

Model	Sensitivity	Specificity	CCR
Probit	96.79 %	9.74 %	69.17 %
Nonparametric	98.35 %	9.13 %	70.04 %

In Figure 1 we report two graphs representing the in-sample performance of the two models when the cut-off-value (α) is varying. The nonparametric specification provides better accuracy for a given range of the cut-off-values, roughly $\alpha \in [0.45, 0.70]$, than the parametric specification, the difference between the CCR values being very small outside this interval. In the same way, the ROC curve for the nonparametric specification always dominates the ROC curve for the parametric specification but the difference between the areas under these two curves is quite small (0.023).

3.3 Out-of-sample performance

We consider the RP test using $S = 10,000$, where S is the number of splits of the data into two independent samples of size $n1 = n - n_2$ and n_2. In our example we have $n = 4,623$

Figure 1: In-sample performance with varying cut-off-values α

(a) Correct Classification Ratio   (b) ROC Curves
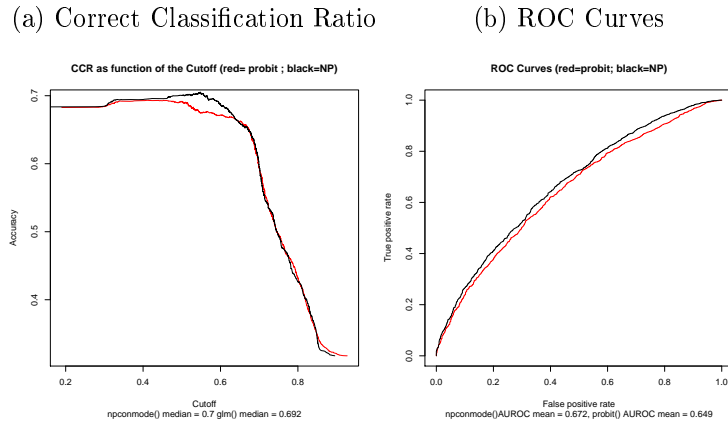
and we report results for $n_2 = 250$.[5] For each split into two independent samples, we fit each model to the n_1 observations, obtain predictions for the values of the covariates in the independent sample of size n_2, then compute the error associated with the response in the independent sample. We repeat this S times, then test whether the expected error on the independent data is equal nor not. The null is that the expected true error is equal for both models, the alternative that is is smaller for the nonparametric model. We use both a paired t-test and also a paired Mann-Whitney-Wilcoxon test. The P-values for each test are $2.816224e-12$ and $1.397127e-11$, respectively, indicating that the nonparametric specification possesses expected true error that is statistically significantly lower than that for the parametric specification, possesses in-sample performance that also dominates the Probit specification, and is therefore preferred.

Figure 2: Boxplots of the RP test statistics for the $S = 10,000$ splits of the data

(a) Boxplots for CCR   (b) Boxplots for AUROC

---
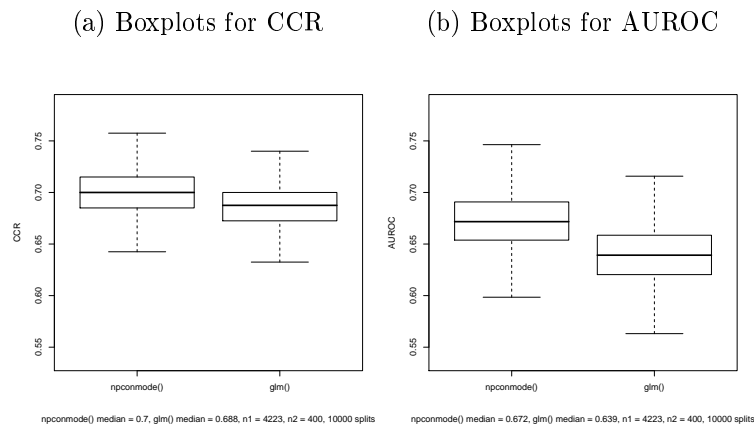
[5]Results are qualitatively unchanged for other choices of n_2 and are available from the authors upon request.

9

In Figure 2a, results are presented in the form of boxplots of out-of-sample ASPE for each of the two specifications. It can be seen from this figure that a stochastic dominance relationship exists between the nonparametric specification and the Probit one, confirming the previous RP test result and again indicating that the nonparametric model is to be preferred on the basis on independent draws from the data. Does this dominance depend on the choice of a cut-off value $\alpha = 0.5$? We address this question by computing the AUROC value for each of the S replications involved in the RP test. These areas do not depend on any chosen cut-off value sand thus provide a more robust indicator of the performance of the classification than ASPE alone. In Figure 2b we compare the empirical distributions of the AUROC for each specification. The two boxplots overlap even less than the boxplots in Figure 2a, indicating again that the nonparametric specification is to be preferred.

3.4   Environmental issues

We now focus on the respective insights on tap water consumption afforded by the two models. By way of example, a graphical comparison of the estimated probabilities of drinking tap water expressed as functions of the two continuous variables PRWQ and Income is provided in the 3-Dimensional surface plots (Figure 3 and 4) where we fix the variables diploma, Region, deleg, and rural at chosen values (resp. for household with diploma lower than *Baccalauréat*, in the North of France, no delegation for tap water distribution, not in a rural area) and we let the iret variable change from retired to non-retired. This comparison allows us to investigate how the variables iret and PRWQ interact given the manner in which this interaction is admitted in the two specifications. Both probabilities shift downwards for retired people. We observe that changing from retired to non retired induces a flip in the shape of the Probit probability, being increasing with respect to Income and decreasing with respect to PRWQ for non-retired consumers and being increasing with respect to both Income and PRWQ for retired consumers. The nonparametric probability is more flat whatever the value of Income and PRWQ for non retired people, but it exhibits a similar pattern to the Probit one for retired people. This effect for retired people is surprising, and a closer look shows that the probability is varying more in the PRWQ dimension than in the Income dimension, capturing a reverse environmental effect. One still observes that the richer the household, the higher the probability of drinking tap water, whatever the environmental quality, but to a lower extent. The nonparametric probability is always bigger than the parametric one for retired people, while the two surfaces cross for non retired people. Note that the surfaces for the nonparametric specification never cross the 0.5 cut-off value, being always bigger for retired people and smaller for non retired people. That is, for this example, the iret variable fully discriminates between the tap water drinkers and non-drinkers when using the usual 0.5 cutoff value.

Figure 3: Surface plots for the two models (retired=0)
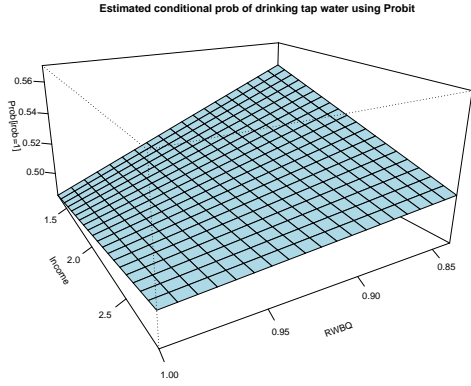
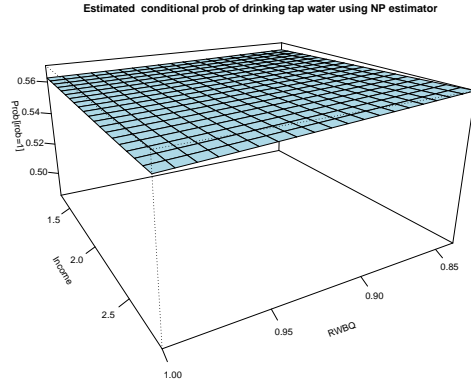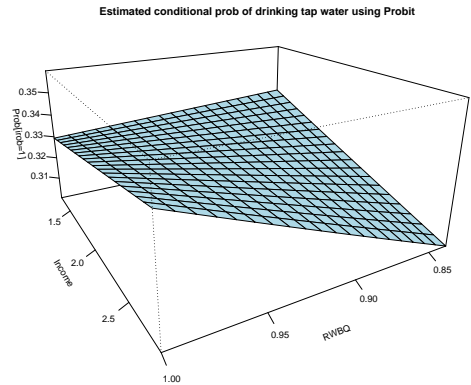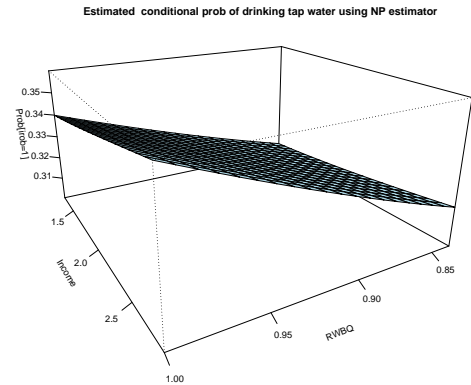(a) Probit model   (b) Nonparametric model

Figure 4: Surface plots for the two models (retired=1)

(a) Probit model   (b) Nonparametric model

4   Concluding remarks

The aim of this paper is threefold. First, we employ a fully nonparametric model of a conditional PDF comprised of a binary response (choice) variable and continuous and discrete explanatory variables. Second, we address the issue of the performance of this nonparametric estimator relative to the parametric Probit specification which is dominant in applied settings and evaluate these estimators in a variety of ways. Third, we provide a detailed discussion of the results focusing on environmental insights provided by the two estimators, emphasizing how particular patterns detected using the nonparametric estimator are masked by the parametric specification.

    To sum up, we show using a real dataset that the nonparametric specification weakly

outperforms the Probit specification when considering in-sample performance measures. The Probit specification with the chosen interactions between explanatory variables was selected to fit the data best using these criteria in Bontemps and Nauges (2009). But, as emphasized by Racine and Parmeter (2009), there is no guarantee that the nonparametric specification will perform any better than the Probit specification, even though the former may indeed exhibit an apparent marked improvement in (in-sample) fit according to the chosen performance measures. Out-of-sample predictive ability seems to be a promising tool for discriminating among specifications.

References

Bontemps, C., Nauges, C., 2009. Carafe ou bouteille ? le rôle de la qualité de lŠenvironnement dans la décision du consommateur. Economie et Prévision 188.

Briesch, R. A., Chintagunta, P. K., Matzkin, R. L., 2002. Semiparametric estimation of brand choice behavior. Journal of the American Statistical Association 97 (460), 973–982.

Efron, B., 1978. Regression and anova with zero-one data: Measures of residual variation. Journal of the American Statistical Association 73, 113–121.

Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Society for Industrial Mathematics.

Egan, J. P., 1975. Signal Detection Theory and ROC Analysis. Vol. 195. New York : Academic Press.

Hall, P., Racine, J., Li, Q., 2004. Cross-validation and the estimation of conditional probability densities. Journal of the American Statistical Association 99 (468), 1015–1026.

Hayfield, T., Racine, J. S., 7 2008. Nonparametric econometrics: The np package. Journal of Statistical Software 27 (5), 1–32.

Horowitz, J., 1998. Semiparametric Methods in Econometrics. Springer-Verlag, New-York.

Li, Q., Racine, J., 2003. Nonparametric estimation of distributions with categorical and continuous data. Journal of Multivariate Analysis 86, 266.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.

Pagan, A., 1999. Nonparametric Econometrics. Cambridge University Press.

Racine, J. S., Parmeter, C. F., 2009. Data-driven model evaluation:a test for revealed performance, mac Master University.