Review of Economics & Finance
Submitted on 19/July/2012
Article ID: 1923-7529-2013-01-25-10 Morteza Haghiri

Applied Nonparametric Regression Analysis: The Choice of Generalized Additive Models

Morteza Haghiri
Memorial University – Corner Brook
1 University Drive, Corner Brook, Newfoundland and Labrador, A2H 6P9, Canada
Tel: +1-709-637-2190 E-mail: mhaghiri@mun.ca

***A*bstract:** Literature has documented tremendous changes in classical regression analysis techniques since 1980s. The drawbacks of simple and multiple parametric regression analyses on model specifications and the non-robust assumption of error terms followed by the introduction of a series of diagnostic tests to fix these inevitable pitfalls have made econometricians to develop new methodologies in nonparametric and semi-parametric regressions that either do not have or mitigate the major shortcomings of what their traditional counterparts inherently demonstrate.

The development of the generalized linear models followed by the introduction of generalized additive models and generalized additive mixed models has attracted practitioners to use these methodologies in applied studies. The main objective of this paper is to conduct a comprehensive survey on studies that used generalized additive models as econometric models and show how the parameters of these models are estimated. In particular, it briefly reviews the theory of generalized additive models, and then introduces various techniques to estimate the parameters of the models. Finally, it presents a comprehensive review of studies in which generalized additive models are specified as the econometric model.

Keywords: Generalized additive models, Generalized additive mixed models, Kernel functions, Locally-weighted scatterplot smoothing, Spline smoothing

JEL Classifications: C01, C14; D00

1. Introduction

The drawbacks of simple and multiple regressions on model specifications, i.e., the linearity of regression functions in explanatory variables, the identical and independent distribution assumption of the error terms, the invariant-variance assumption of the dependent variable across sample data, and finally, the non-robust assumption of the random error terms, i.e., normality or Gaussian followed by introducing a series of diagnostic tests to fix these inevitable problems have substantially made econometricians to develop new methodologies in both modeling and estimating nonparametric and semi-parametric regressions, which are relatively less rigid on various assumptions that traditional counterparts rely on by nature. For instance, evidence shows that practitioners must follow the falsification protocols derived from the Popperian principle, which validates studies that focus on deriving and testing refutable hypotheses, such as homogeneity, monotonicity, curvature, and asymmetry (Blaug 1992). As Fox and Kivanda (1994) expressed that not all refereed-published articles have followed this protocol. Haghiri and Simchi (2003a) argued that part of the ongoing discussions between proponents and opponents of the Popper or production ideology stems from a lack of sufficient knowledge about the true relationship (e.g., linear or curvilinear) between response and predictors used in the models that cannot be detected by simply

using the classical regression analysis. To circumvent this problem, econometricians have susceptibly been using non-ad hoc methods of estimating regression analysis in various fields since 1980s (Andersen 2009). Some examples of these methods are generalized linear models (GLMs), additive models, generalized additive models (GAMs), and generalized additive mixed models (GAMMs). Although the last two methods are relatively new and in their primitive stages of their development, there are vast numbers of articles in the literature that use either GAMs or GAMMs as the econometric models (Li and Racine 2006).

The purpose of this paper is to present a compendious review on studies that used GAMs as econometric models. Section 2 presents a theoretical aspect of GAMs and introduces various techniques, which are used to estimate its parameters. Section 3 reviews recent studies in different fields that used GAMs as econometric models. Section 4 provides some remarkable points.

2. Generalized Additive Models

2.1 General Concept

The historical discussion of generalized linear and generalized additive models dates back to forty-five years ago when Cox (1968) laid out several points on regression analyses. It was a general consensus that a fitted regression analysis would possibly provide useful information on the relationship between response and predictors provided that the following two conditions are met: stability and reproducibility (Nelder 1968). Under usual circumstances, this choice theoretically is found either by plotting and examining series of data or investigating through the primitive method of trial and error stems from obtaining unsuccessful analyses (Cox 1968, p.268). For instance, consider a standardized multiple regression function that is specified as equation (1):

$$Y_i = \alpha + X_1\delta_1 + X_2\delta_2 + ... + X_p\delta_p + \varepsilon_i, \qquad i = 1,2,...,n \qquad (1)$$

where $E(\varepsilon) = 0, and\ Var(\varepsilon) = \sigma^2$. In equation (1), we may have n observations on a vector of random response $\boldsymbol{Y}$, denoted by $\boldsymbol{y} = (y_1,..., y_n)'$ and measured either at n separate vectors of predictors $\boldsymbol{x^i}$ = (x_{i1},…, x_{ip}), or designed into a matrix of $\boldsymbol{X} = \left(x_1^i,..., x_p^i\right)$.

We may also assume that the predictors are either predetermined, or random variables, and/or a combination of both. The main goal of equation (1) is to determine the type of relation between response and the vector of predictors. There are several reasons justifying why practitioners are interested in analyzing such relationship of which three of them (i.e., description, inference, and prediction) are very important. First, researchers would like to describe the dependence of the response on the predictors so that they can learn more about the process that produces the response. Second, they usually want to assess the relative contributions of each of the predictors in explaining the variations in response. Finally, practitioners always want to predict the response's value for some set of the values of predictors. Equation (1) indirectly undertakes a strong assumption about the relation between the expected value of the dependent variable E (Y) and each elements of the matrix of predictors $\boldsymbol{X}$. The relation is expressed as follows: there is an additive and linear dependence between the response and each elements of the matrix of predictors. Although the former assumption is quite natural and observed in any regression analyses, the latter premise is very idealistic and may not be true. Nevertheless, if the assumption in question holds, then the linear regression models are extremely useful because (*i*) it provides a simple description of the data, (*ii*) it shows how each of the predictors contributes with a single coefficient, and (*iii*) it establishes a simple method for predicting new observations (Haghiri, 2003).

Generally, we can extend the linear regression models in various methods. Among these, the surface smoothers, which are the nonparametric estimates of the regression models, were used to be so popular prior to 1970s. However, the use of surface smoothers is not with major drawbacks. The first problem is related to the arbitrary choice of the type of the smoother functions, so-called kernel functions, which can be thought as of neighborhood that defines local in *p* dimensions. The second problem is related to the limited number of predictors in the model that causes the curse of dimensionality problem (Hastie and Tibshirani 1990) if they exceed more than two variables when kernel functions are used as the surface smoothers. Bellman (1961) stressed that the identification of neighborhoods in the *p*-dimensional space is imperative; however, neighborhoods with a fixed number of points become less local as the dimension of the model increases. Due to the localness problem of kernel functions for estimation of the mean response function in nonparametric regression analysis, several estimation techniques, such as the recursive-partitioning regression, or the projection-pursuit regression have been introduced. Both models have good prediction power and under proper conditions they are consistent for the true regression surface subject to the adequacy of data. The only problem these models have is that the examination of the effect of any particular predictor on the mean response function is a cumbersome task especially when a complicated surface is fitted (Schimek 2000). Such problems in estimating the mean response function led researchers to focus on new models that do not have relatively the aforementioned limitations. Consider any multiple variable regression models such as equation (1) in which the conditional mean relationship between the mean response, i.e., *E(Y)* and each of the predictors X_i is assumed to be linear and additive. Equation (1) can also be rewritten as

$$Y_i = m(\mathrm{x}_i) + \varepsilon_i, \qquad i = 1,2,...,n \qquad (2)$$

which $m(\mathbf{x}_\mathrm{i}) = \alpha + \mathbf{x}_i'\boldsymbol{\delta}, \boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_p)'$. Other assumptions, including the zero mean and constant variance for the error terms still hold. The structural relationship between the response *Y* and the vector of *p* predictors $\boldsymbol{X} = \left(x_1^i, ..., x_p^i\right)'$ is explained through

$$m(\mathbf{x}) = E(Y \mid X = \boldsymbol{x}) \qquad (2')$$

where $\boldsymbol{x} = (x_1, \ldots, x_p)'$ and $m(\mathbf{x}) = m(x_1, \ldots, x_p)$. Based on the classical linear regression analysis, we may realize that *m(***x***)* is linear and additive with respect to all the predictors in the model. If we relax the linearity assumption and retain the additivity premise, we wind up a group of models, known as additive models. Equation (3) shows an additive model in which each of the predictors is connected to the response through an individual functional form

$$m(X) \cong \omega(X) = \alpha + \sum_{j=1}^{p} \omega_j(X_j) \qquad (3)$$

in which α is a constant and the $\omega_j s$ are arbitrary univariate smooth functions (one for every predictor). As Schimek and Turlach (2000) suggested, it is required that $E[\omega_j(X_j)] = 0$ to avoid having free constants in each of the functions ω_j . This requirement, which is set in the range of $1 \leq j \leq p$, implies that $E(Y) = \alpha$ and it is necessary for the purpose of identification. If the additivity assumption, defined in equation (3), is correct then we can derive equation (4)

$$E\left[Y - \alpha - \sum_{j \neq q} \omega_j(X_j) \middle| X_q\right] = \omega_q(X_q) \qquad (4)$$

for $q = 1, \ldots, p$. To estimate each of the univariate functions $\omega_1, \omega_2, \ldots, \omega_q$ relating to the predictors in the model we can use an iterative process, known as the backfitting algorithm that was proposed by Friedman and Stuetzle (1981). Since each variable is represented separately in equation (3), the

model interprets the variation of the fitted response surface holding all but one predictor fixed and therefore it does not depend on the values of other predictors. In practice, it means that once an additive model is fitted to data, the *p* coordinate functions can be plotted separately to examine the roles of the predictors in modeling the mean response. Nevertheless, we should not forget that additive models are merely approximations to the true regression surface.

2.2 Model Specification

Generalized additive models (GAMs) are semi-parametric extensions of generalized linear models (GLMs) developed by Nelder and Wedderburn (1972). A GLM is an extension of a classical linear model $E(Y) = m(X) + \varepsilon$ in which the unknown regression function $m(X)$ is specified linearly through a fixed link function *G* in a parametric way. Generalized linear models are very flexible with unnatural scales data and allow for non-linearity and non-uniformity variance structures in the sample observations (Guisan et al. 2002). In addition, GLMs are conformed to several types of probability distributions, such as Gaussian, binomial, Poisson, and/or gamma.

A GAM takes into account an additive relation between the response and predictors of a model and rules out any linearity assumption (Hastie and Tibshirani 1990). Generalized additive models are also known as data-driven models because these models enable practitioners to detect the true relationship between the response and the set of predictors used in the model (Haghiri and Simchi 2003a). Moreover, from the structural point of view, GAMs and GLMs are similar to each other on the grounds of the relation between the response and a series of smoothed function of the predictors in the model. As Guisan et al. (2002, p.90) stressed out this relation makes GAMs to be prudent models that "deal with highly non-linear and non-monotonic relationships between the response and the set of predictors." In this paper, we used prominent works on specifying the theory of GAMs (Hastie and Tibshirani, 1990; and Schimek and Turlach, 2000) followed by reviewing a series of articles that GAMs were specified as the econometric model (e.g., Haghiri 2011; Haghiri et al. 2013; Haghiri, Law, and Nolan 2012; Haghiri, Nolan, and Tran 2004; Haghiri and Simchi 2005; 2006; 2010).

A brief explanation of GAMs follows. Given equations (1) – (2') and their assumptions including zero mean and constant variance for the error terms, a GAM can be specified as follows:

$$E\left[Y_i \middle| \mathbf{x}_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)\right] = \theta_I, \text{ and}$$
$$G(\theta_i) = \vartheta_i = \left[\alpha + \sum_{j=1}^{p} \omega_j\left(x_{ij}\right)\right] = [\omega(\mathbf{x}_i)] \qquad (5)$$

in which *G*(**.**) is a fixed link function, the distribution of *Y* follows an exponential family similar to GLMs, and the rest of the assumptions are the same as equation (3). The estimation of a GAM takes place in a couple of stages. In the first stage, the additive predictors are estimated using a system of normal equations. In the second stage, the estimated additive predictors are linked to the function G(**.**) through an iterative process so-called the local-scoring algorithm (Hastie and Tibshirani 1990, p.141). In this algorithm, a local averaging process is used to generalize the Fisher scoring procedure that is applied to estimate the parameters of a GLM (Schimek and Turlach, 2000, pp.280-297). In practice, the local-scoring algorithm is similar to the Fisher scoring procedure except that in the latter models the least squares step, which is used to update the estimate $\hat{\delta}$ for the linear predictor $\mathbf{x}'\delta$, is replaced by the solution obtained from solving the normal equations part of the model, which is applied by the backfitting algorithm to update the estimates for α and ω_js. To estimate the parameters of equation (5), several smoothing techniques can be used, which are explained in the following section.

2.3 Model Estimation

In general, there are two methods of direct and indirect estimating of the mean response function and the parameters of GAMs. As explained in Section 2.1, the indirect approach uses an iterative process known as the backfitting algorithm. The direct method of estimating the parameters of GAMs is called the marginal-integration method that was proposed independently by Newey (1994) and Tjøstheim and Auestad (1994). In the following sections we briefly explain the indirect methods of estimating the parameters of GAMs and refer interested readers to find more about the marginal-integration method in Haghiri and Simchi (2003b).

2.3.1 Locally Weighted Scatterplot Smoothing (LOWESS)

The basic idea of using the LOWESS methodology is to find initially a point in the space of predictors followed by searching the neighborhood points that are smoothed using surface smoothers to estimate the mean response function. For example, consider any point *x*, so-called local observation in the space of the model predictors. Generally, estimating a local regression model can be specified through different approaches attempting to find a neighborhood that contains the initial point *x* in which the regression surface is well approximated by a function from a specific parametric point of view. Therefore, the specification from the local regression model leads to methods of fitting the mean response function that consists of smoothing the response as a function of the predictors.

We can briefly explain the concept of fitting in a local regression model, i.e., to smooth $s(x_0)$, where $s(.)$ depicts the scatterplot smoother functions using *p* nearest neighborhoods, as follows. At first, the *p* nearest neighbors of x_0, denoted by $\Omega(x_0)$, are identified. Then, the furthest near-neighbor observation from x_0 is determined and the distance between these two points is computed. In other words, we, first, calculate $\Delta(x_0) = \max_{\Omega(x_0)} \left| x_0 - x_i \right|$, and then using the tri-cube weight function (Cleveland 1993, p.314) the weights ξ_i are assigned to each point in $\Omega(x_0)$ as

$$\xi\left(\frac{\left| x_0 - x_i \right|}{\Delta\left(x_0\right)}\right) \qquad (6)$$

where

$$\xi(u) = \begin{cases} \left(1-u^3\right)^3, & \text{for } 0 \le u \le 1; \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

Following the computed weights in the previous step, we ultimately obtain the scatterplot smoother $s(x_0)$ value at the initial point x_0 by conducting the weighted least-squares of the response variable *y* to predictors *x* in the domain of $\Omega(x_0)$. The LOWESS approach is popular among the practitioners on the grounds of, at least, two reasons. First, it produces robust results with respect to the outliers. Second, it helps researchers find the neighborhoods for the target point x_0 (Hastie and Tibshirani 1990, p.30-31).

2.3.2 Spline Smoothing

The spline smoothing approach, which provides flexible methodology for fitting data in a nonparametric manner, has gained popularity among researchers in a wide variety of sciences such as analysis of growth data, medicine, remote sensing experiment, and economics (e.g., Haghiri, Law, and Nolan 2012; Haghiri and Simchi 2010). Given equations (2) and (2'), our attempt is to

estimate *m* from the sample observations through minimizing the sum of squares errors (*SSR*) as follows:

$$SSR(m) = \sum_{i=1}^{n} [y_i - \hat{m}(\mathrm{x}_i)]^2, \quad i = 1,2,...,n. \tag{8}$$

where $m(\mathbf{x}_i) = \alpha + \mathbf{x}_i'\boldsymbol{\delta}$. This approach has not without major drawbacks. As Haghiri et al. (2013) stated, a linear relation between the response and predictors of the model may not exist. To circumvent this problem, practitioners use the Taylor-series expansion

$$m(x) = m(x_0) + m'(x_0)(x - x_0) + o\left(\left| x - x_0 \right|^2\right) \tag{9}$$

which assumes that *m* is an unknown function that can be differentiated at least twice and there is a point *x* close to some fixed point x_0. Equation (9) states that for *x* close to x_0, m follows a linear model with an intercept of $m(x_0) - m'(x_0)x_0$ and the slope of $m'(x_0)$. In lieu of this discussion, the following scenarios are worth mentioning. First, if we assume that the unknown function *m* is linear, then it indirectly implies that the regression function has an invariant slope $m'(x_0)$ and small error term $o\left| x - x_0 \right|^2$ that may not be the truly case. This view uses too little information from the sample data because it assumes a constant-slope status; however it presents a comprehensive descriptive analysis from the sample observations. Second, we can postulate the minimization of $SSR(m)$ over functions *m* with variant slopes instead of simply just minimizing the *SSR*. This implies that different slopes exist at each point *x* that connect every two responses by lines associated with their own individual slopes. This scenario uses too much information as it assumes a variant-slope condition; instead, it fails to show a satisfactory description of basic trends in the sample observations. As Wahba (1990) stated, such failure might be caused by the nature of the regression function, defined as equation (2), rather than the normal function of the random-noise component of the model. To fix such failure, we can specify a function *m* and define m'' as the rate of change in the slope of the function. Since its slope varies from one point to another, by taking integral from the entire changes throughout the slope of the function, we wind up

$$\Psi(m) = \int_{x_1}^{x_n} m''(x)^2 \, dx \tag{10}$$

that clearly shows that the slope will change if one of the predictors changes. As Wahba (1990) suggested, we can then define

$$SSR(m) + \eta\,\psi(m) \qquad \eta \geq 0, \tag{11}$$

which can be minimized over all functions provided that they are continuous and can be twice differentiated. In equation (11), η is called the smoothing parameter or the span degree. If η approaches infinity it yields the conventional linear regression with fixed slope, whereas if it approaches zero it generates a regression with flexible slope. If the number of predictors in the model equals two or more, Eubank (2000) showed that a specific parameter λ_τ would minimize equation (11). This parameter is known as the cubic spline smoother and is linear in the sense that there can be found some fixed values like $s_i(x),\ i = 1,...,n$ in equation (12) for each estimation point *x* such that

$$\lambda_\tau(x) = \sum_{i=1}^{n} s_i(x)\, y_i \tag{12}$$

Cubic spline smoothers are not without drawbacks. For instance, they are sensitive to the choice of span degree for the sample data. Precedent studies listed the sources of such sensitivity in twofold: (*i*) the unknown true regression curve, and (*ii*) the inherent variability of the estimator (e.g., Eubank 2000 and Wahba 1990). Few suggestions were made to mitigate, but not to eliminate thoroughly, the aforementioned problem. Some examples of these methodologies are the cross-validation approach (Stone 1974) and the plug-in methods (Gasser et al. 1991).

3. Recent Studies Using GAMs

In this section we review recent studies that used GAMs as econometric models. For example, Haghiri et al. (2013) use GAMs to develop a nonparametric cost function to examine input separability in the Canadian cable television industry (CATV). The researchers used an unbalanced panel data from 1990 to 1996 to address substitutability/complementarity of inputs in the industry and argued that the proposition of a statistical test that enables practitioners to carry out the substitutability/complementarity of inputs in the CATV industry is always of interest to the regulators and policy makers because the rate of return regulation (i.e., one form of industry regulation) will have different effects on input choices. To estimate the parameters of the model, Haghiri et al. (2013) used the general spline smoothing techniques followed by testing the degree of separability among the inputs used in the model. Results showed that the input separability assumption was valid using either parametric or non-parametric cost estimates.

Seniors are often exposed to multimobidity that negatively affects the quality of their lives. Hunger et al. (2011) conducted a research to examine the impact of multimorbidity on a group of people consisting of 4,565 seniors aged 65 years or older in Germany using GAMs. The researchers' main objective was to assess the adverse effects of six prevalent diseases and their combinations on the health-related quality of life measured by the EQ-5D for the participants in the research. Results showed that the interaction effects between three major diseases (i.e., coronary problems, diabetes mellitus, and the history of a stroke) caused negative impact on seniors' health conditions in Germany.

Marra and Radice (2011) addressed the endogenicity problem arisen from the statistically independent assumption of the measured and unmeasured predictors used in a model. As a result, the standard estimation methods are not credible especially when these variables collectively affect the response. To circumvent this problem, the classical linear regression analysis recommends the use of instrumental variables. Marra and Radice (2011) suggested a two-stage procedure for instrument variable estimation within the framework of GAMs by combining any penalized regression spline smoothing techniques and a correction producer for confidence intervals. The researchers also conducted a simulation experiment using sample data, which contained information related to households' health status, utilization of health services, types of insurance coverage, and socio-economic variables that were collected from 1,068 households aged 50 years and older who lived in Italy in 2001. Although Marra and Radice (2011) obtained consistent coefficients their results were susceptible to the choice of instrument variables.

Windle, et al. (2010) examined variations in the ecological relationships within management areas assuming spatially stationary processes. To address this matter, the researchers used the geographically weighted regression (GWR) and compared its estimation results with the ones obtained from the global logistic regression and GAMs. The main objective of the authors' research was to assess and predict the distribution of northern Atlantic cod in Newfoundland and Labrador considering both the environmental (temperature and distance from shore) and biological (snow crab and northern shrimp) factors. Windle et al. (2010) specified a binomial GAM by choosing a logit link function and estimated the parameters of the model using the penalizing

regression spline soothing method with automatic smoothness selection. To assess both linear and non-linear effects the researchers considered smoothers and concluded that the GWR was a better approach compared to the global logistic regression and GAMs.

Dan, Gu, and Xu (2005) analyzed the effects of normal functioning of financial derivatives, e.g., hedging activities, on the return of investment and firm value of the companies that were actively operating in the Canadian oil and gas sector from 2000 to 2002 by specifying a semi-parametric GAM. The researchers stated that real world problems in financial markets lowered the credibility of the Modigliani-Miller theorem (Modigliani and Miller 1958) that considers zero value of hedging for a firm operating in a perfect financial market at the absence of asymmetric information, taxes, or transaction costs. The authors' main objective was to assess the nonlinear pay offs of hedging activities and compared the results with the ones obtained from the traditional linear regressions. Results showed that those companies that hedged actively in the oil and gas market were able to reduce risk against volatility in the product prices and the returns of investment on stocks did not fall similar to the reduction in the prices of oil and gas.

Haghiri and Simchi (2005) examined the residual deviance analysis to test the separability among inputs in the Quebec dairy industry. The authors developed a GAM and used a series of unbalanced panel data contained 4,783 observations collected from 527 dairy farms in Quebec between 1990 and 1998. This period of time was important to policy makers in Canada because the industry has been operating under the supply management policy since mid-1970s. Haghiri and Simchi (2005) estimated the parameters of the model using the spline smoothing method, examined the input separability in the model, and concluded that all inputs were additively separable.

Finally, a research on wildlife habitat conducted by Gunn et al. (2004) in Northwest Territories Canada in March 2002 aimed to identify factors affecting boreal caribou habitats, land use planning and recovery planning in the region. These factors were land use jurisdictions (e.g., communities) and land claim organizations. Gunn et al. (2004) specified a GAM to assess occurrence at the landscape scale and estimated the parameters of the model using the spline smoothing techniques. Results showed linear relationship between the mean response function and some predictors in the model including black spruce, low-shrub, and wetlands, whereas other predictors, such as the herbaceous, fire regenerated, and lichen exhibited nonlinear relationships with the mean response function.

4. Summary and Concluding Remarks

Evidence shows that researchers have chosen classical regression analyses and least square method as pivotal mechanisms of estimating parameters of econometric models prior to early 1990s. These approaches, however, have been scrutinized recently on number of controversial grounds such as model specifications, the linearity assumption of regression functions in the predictors, the constant variance assumption of the response across sample observations, the identical and independent distribution assumption of the disturbance terms, and the Gaussian assumption of the random error terms. The introduction of a series of statistical tests, although mitigating the adverse effects of those problems, cannot resolve them thoroughly. To circumvent the above problems, practitioners have started developing nonparametric and semi-parametric regression functions in the framework of generalized linear models, additive models, generalized additive models, and generalized additive mixed models. Several estimation techniques, including the locally weighted scatterplot smoothing and the spline smoothing approach have widely been used to estimate the parameters of the newly developed models. Generalized additive models are flexible models, which can accommodate many predictors, specifying trends, simplifying the interpretation of a series of regression models, such as simple linear, multiple, semi-parametric and

nonparametric regression functions (Andersen 2009) and interaction terms in both the parametric and nonparametric models (Haghiri et al. 2013). Although the development of generalized additive models is at the early stage from both theoretical and applied aspects, numerous articles have already been documented that demonstrated the extensive use of these models (see, e.g., Pagan and Ullah 1999, p.157-159).

References

[1] Andersen, R. (2009), "Nonparametric Methods for Modeling Nonlinearity in Regression Analysis," *Annual Review of Sociology*, 35(1): 67–85.

[2] Bellman, R. E. (1961), *Adaptive Control Processes: A Guided Tour*, 1st Edition, Princeton, New Jersey: Princeton University Press.

[3] Blaug, M. (1992), *The Methodology of Economics*, 2nd Edition, New York, NY: Cambridge University Press.

[4] Cleveland W.S (1993), *Visualizing Data*, 1st Edition, Summit, New Jersey: Hobart Press, p.314.

[5] Cox, D. R. (1968), "Notes on Some Aspects of Regression Analysis (with Discussion)", *Journal of the Royal Statistical Society Series A (General)*, 131(3): 265–279.

[6] Dan, C., Gu, H. and K. Xu (2005), "The Impact of Hedging on Stock Return and Firm Value: New Evidence from Canadian Oil and Gas Companies", Working papers, Department of Economics, Dalhousie University, Halifax, Nova Scotia, Canada.

[7] Eubank, R. L. (2000), "Spline Regression", In: Schimek M.G. (Ed.), *Smoothing and Regression: Approaches, Computation, and Application*, 1st Edition, New York: NY: John-Wiley and Sons, pp.1–16.

[8] Fox, G. and L. Kivanda (1994), "Popper or Production?", *Canadian Journal of Agricultural Economics*, 42(1): 1–13.

[9] Friedman, J. H. and W. Stuetzle (1981), "Projection Pursuit Regression", *Journal of the American Statistical Association*, 76(376): 817–823.

[10] Gasser, T., Kneip, A. and W. Kohler (1991), "A Flexible and Fast Method for Automatic Smoothing", *Journal of the American Statistical Association*, 86(414): 643–652.

[11] Guisan, A., Edwards Jr., T. C. and T. Hastie (2002), "Generalized Linear and Generalized Additive models in Studies of Species distributions: setting the scene", *Ecological Modelling*, 157(2-3): 89–100.

[12] Gunn, A., Antoine, J., Boulanger, J., Bartlett, J., Croft, B. and A. D'Hont (2004), "Boreal Caribou Habitat and Land Use Planning in the Deh Cho Region, Northwest Territories", Working Papers No.153, Department of Resources, Wildlife and Economic Development, Government of the Northwest Territories, Yellowknife, Northwest Territories, Canada.

[13] Haghiri, M. (2003), "Stochastic Nonparametric Frontier Analysis in Measuring Technical Efficiency: A Case Study of the North American Dairy Industry", Unpublished PhD thesis, Department of Bioresource Policy, Business and Economics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[14] Haghiri, M. (2011), "Modeling Stochastic Semiparametric Frontiers using the Generalized Additive Theory", Working paper, Memorial University – Corner Brook, Corner Brook, Newfoundland and Labrador, Canada.

[15] Haghiri, M. and A. Simchi (2003a), "A General Solution to Popper or Production", *Empirical Economics Letters*, 2(2): 51–73.

[16] Haghiri, M. and A. Simchi (2003b), "Estimating Technical Efficiency of Ontario Dairy Producers using the Marginal-Integration Method", *Empirical Economics Letters*, 2(1): 1–18.

[17] Haghiri, M. and A. Simchi (2005), "An Application of the Residual Deviance Analysis in Testing Input Separability Restrictions", *Applied Economics Letters*, 12(12): 755–758.

[18] Haghiri, M. and A. Simchi (2006), "On the Estimation of Stochastic Semiparametric Frontiers in Measuring Technical Efficiency", Memorial University – Grenfell Campus, Newfoundland and Labrador, Canada.

[19] Haghiri, M. and A. Simchi (2010), "A Nonparametric Extension of Generalized Quadratic Box-Cox Models in Measuring Technical Efficiency", *Applied Economics*, 42(1): 1–8.

[20] Haghiri, M., Law, S. M. and J. F. Nolan (2012), "Nonparametric Estimates of Technology using Generalized Additive Models: The Case of the Canadian Cable Television Industry," *International Journal of Economics and Finance*, 4(1):36–47.

[21] Haghiri, M., Law, S. M., Nolan, J. F. and A. Simchi (2013), "Comparing Nonparametric to Parametric Interaction Terms in Generalized Additive Models: Production Technology in the Canadian Cable Television Industry", *International Journal of Data Analysis Techniques and Strategies*, (forthcoming).

[22] Haghiri, M., Nolan, J. F. and K.C. Tran (2004), "Assessing the Impact of Economic Liberalization across Countries: A Comparison of Dairy Industry Efficiency in Canada and the United States", *Applied Economics*, 36(11): 1233–1243.

[23] Hastie, T. J. and R. J. Tibshirani (1990), *Generalized Additive Models*, 1st Edition, London, UK: Chapman and Hall.

[24] Hunger, M., Thorand, B., Schunk, M., Döring, A., Menn, P., Peters, A. and R. Holle (2011), "Multimorbidity and health-related quality of life in the older population: results from the German KORA-Age study", *Health and Quality of Life Outcomes*, 9(53): 1–10.

[25] Li, Q. and J. S. Racine (2006), *Nonparametric Econometrics: Theory and Practice*, 1st Edition, Princeton, New Jersey: Princeton University Press.

[26] Marra, G. and R. Radice (2011), "A Flexible Instrumental Variable Approach", *Statistical Modeling*, 11(6): 581–603.

[27] Modigliani, F. and M. H. Miller (1958), "The Cost of Capital, Corporate Finance and the Theory of Investment", *American Economic Review*, 48(3): 261–297.

[28] Nelder, J. A. (1968), "Regression, model-building and invariance", *Journal of the Royal Statistical Society Series A (General)*, 131(3): 303–315.

[29] Nelder, J. A. and R. W. M. Wedderburn (1972), "Generalized Linear Models", *Journal of the Royal Statistical Society Series A*, 135(3): 370–384.

[30] Newey, W. K. (1994), "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10(2): 233–253.

[31] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, 1st Edition, New York, NY: Cambridge University Press.

[32] Schimek, M. G. (2000), *Smoothing and Regression: Approaches, Computation, and Application*, (Ed.), 1st Edition, New York, NY: John-Wiley & Sons.

[33] Schimek, M. G. and B. A. Turlach (2000), "Additive and Generalized Additive Models", In: M. G. Schimek (ed.), *Smoothing and Regression: Approaches, Computation, and Application*, 1st Edition, New York, NY: John-Wiley & Sons, pp.277–327.

[34] Stone, M. (1974), "Cross-validatory choice and assessment of statistical predictions", *Journal of the Royal Statistical Society Series B*, 36(2): 111–147.

[35] Tjøstheim, D. and B. Auestad (1994), "Nonparametric Identification of Nonlinear Time Series: Projections", *Journal of the American Statistical Association*, 89(428): 1398–1409.

[36] Wahba, G. (1990), *Spline Functions for Observational Data*, 1st Edition, Philadelphia: CBMS-NSF Regional Conference Series SIAM (Society for Industrial and Applied Mathematics).

[37] Windle, M. J. S., Rose, G. A., Devillers, R. and M. J. Fortin (2010), "Exploring Spatial Non-Stationarity of Fisheries Survey Data using Geographically Weighted Regression: An Example from the Northwest Atlantic", International Council for the Exploration of the Sea *(ICES) Journal of Marine Science*, 67(1): 145–154.