

Testing for input separability in the Canadian CATV industry: An application of generalized additive mixed models

ABSTRACT

The nature of input relationships has important consequences for regulatory policies since the impact of rate-of-return regulation depends on the extent of substitution that is offered by production technologies which are adopted by regulated firms. The use of classical regression analysis had been criticized for various reasons, such as the linearity of regression functions in predictors, the fixed-variance premise of the response across sample observations, the i.i.d. postulation of the random error terms in addition to their normal distribution, and non-complying data in real cases. Instead, earlier research suggested the use of Generalized Additive Models (GAMs) that don't have the aforementioned problems. Nevertheless, in the presence of complex covariate effects among data series, the use of Generalized Additive Mixed Models (GAMMs) is recommended. A GAMM is a non-ad hoc generalization of a Generalized Linear Mixed Model (GLMM) except that an additive unknown combination of nonparametric functions of covariates and random effects replace the predetermined linear predictor in the GLMM. We develop a nonparametric cost function using GAMMs. First, we estimate the model parameters using general spline smoothing techniques. Second, we apply residual deviance analysis to test the separability and complementarity between inputs. We use an unbalanced panel of operating data from the regulated Canadian cable television (CATV) firms from 1990 to 1996. This period is of particular interest to policymakers because the Canadian CATV was a rate- and entry-regulated industry prior to 1997.

Keywords: Generalized additive mixed models, generalized additive models, spline smoothing, input separability, Canadian cable television industry.

JEL Codes: C13, C14, L51.

1. INTRODUCTION

One of the most challenges that researchers face in applied microeconomics is to identify the nature of substitutability or complementarity amongst inputs used in the production process. For instance, in a noncompetitive market structure the knowledge of input substitutability provides sufficient incentives for producers to integrate with their customers [1]. However, the outcome of this integration will entirely depend on the type of market structure. For example, in a monopoly condition, it is unknown whether such integration increases the industry output or reduces the total production. An important concept in production theory is output elasticity which is defined as the percentage change of output divided by the percentage of an input when input prices are held constant [2]. It is important for researchers to know how sensitive total production is with respect to a change in the input prices. When inputs are perfect complements, all variable inputs are output elastic [3]. This means that for a given level of output, marginal cost will raise in a rate more than average cost for an increase in input prices. Moreover, previous studies showed that complementarity characteristics of factors of production could help

downstream firms purchase intermediate goods from various wholesale markets to make their final products [4]. One of the industries that gain from using complement inputs is the information and communication technology sector where knowledge-based inputs are used in combination with each other. The supermarket industry is another example that offer bundle products to reduce its customers' acquisition costs. As a result, testing relationships among inputs can have particular policy relevance for many industries including the CATV industry because the rate of return regulation can take different forms [5]. This challenge becomes much bigger when classical regression analysis is used to estimate parameters of the production/cost function in econometric models. It has been a general consensus among practitioners for a long time that classical regression analysis has been estimating the parameters of regression functions in applied studies. In spite of being widely used, this method of analysis is not without major drawbacks. Classical regression analysis had been criticized for various reasons, such as the linearity of regression functions in predictors, the fixed-variance premise of the response across sample observations, the i.i.d. postulation of the random error terms in addition to their normal distribution, and uncomplying data in real cases, which might lead to lose the validity of these types of models [6]. Alternatively, precedent studies suggested the use of GAMs, proposed by [7] that don't have the aforementioned problems. The use of GAMs is not recommended in the presence of complex covariate effects among data series. Instead, under these circumstances literature suggests a broader methodology, known as GAMMs, proposed by [8], which is a non-ad hoc extension of GLMMs except that an additive unknown combination of nonparametric functions of covariates and random effects replace the predetermined linear predictor in GLMM. A GAMM encompasses GLMMs and GAMs as special cases.

To our knowledge no studies are available in the literature to examine input relationships in the Canadian CATV industry where the correlation among covariates is predicted. In this paper, we use the theory of GAMMs and develop a nonparametric cost function to test input separability amongst factors of production used in the model. First, we estimate the GAMM model parameters using general spline smoothing techniques. Second, we apply the residual deviance analysis [9] to test either substitutability or complementarity amongst the factors of production. To do this, we utilize an unbalanced panel of operating data collected from regulated CATV firms between 1990 and 1996. This period is of particular interest to policymakers because the Canadian CATV was a rate- and entry-regulated industry prior to 1997.

The rest of the paper is organized as follows. Section 2 briefly presents a theoretical aspect of GAMMs and demonstrates the residual deviance statistic test used to examine relationships amongst input used in the model. Section 3 provides the empirical results and the final section concludes the study and outlines directions for further research.

2. GENERALIZED ADDITIVE MIXED MODELS

2.1 Model Specification

Generalized Additive Mixed Models (GAMMs) are nonparametric extensions of GLMMs of [10]. A GLMM is an extension of a Generalized Linear Model [11] in which covariate effects are modeled by a parametric mean function through inclusion of random effects to the linear predictor to estimate the parameters of the econometric model with overdispersed and correlated

outcomes. Generalized Linear Mixed Models have been used widely in different fields, including clinical trials and disease mapping, longitudinal studies, survey sampling, fisheries and aquaculture, etc. [8,10,12,13]. The use of GLMM, however, is not without drawbacks. The shortcoming of GLMMs relates to the prespecified functional form of the mean parametric function that is used to model covariate effects. The knowledge of the true functional form is always unknown to researches, and, as a result, the outcome variable would be varied had different functional forms been selected. To circumvent the aforementioned problem, GAMMS was proposed by [8], which are basically nonparametric regression models with an unspecified nonparametric mean function that can be incorporated into GLMMs to provide more flexibility and mitigate the dependency of outcome variable on the covariates.

A brief explanation of GAMMs is as follows [8]. Consider a standardized multiple regression function with a series of predetermined random or non-random m explanatory variables (covariates) $x_i = [1, x_{i1}, \dots, x_{im}]'$ associated with fixed effects and a $w \times 1$ vector of predictors q_i independent variables associated with random effects that projects collectively variations in the response (dependent variable) y_i with i^{th} number of observations from n unit of firms. Assuming y_i to be conditionally and independently distributed with the expected values of $E(y_i|c) = \mu_i^c$ where c consists of $w \times 1$ vector of random effects and variances $var(y_i|c) = \delta g_i^{-1} \sigma(\mu_i^c)$ that follow a GAM of which $\sigma(\cdot)$ shows a predetermined variance function, g_i is a specified weight, such as a binomial denominator, and δ represents a scale parameter. Given the above expressions, equation (1) specifies a GAMM:

$$s(\mu_i^c) = \gamma_0 + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_m(x_{im}) + q_i'c \quad (1)$$

in which $s(\cdot)$ shows a twice differentiable monotonic link function, $f_j(\cdot)$ represents a smooth function that is twice-differentiated, c , i.e., the random effects are assumed to be normally distributed with zero mean and variances $V(\emptyset)$ where $\emptyset$ is a $d \times 1$ vector of variance components [8]. The key element of equation (1) is that there is no *a priori* on the type of relationship between response and the covariates. For instance, if it is linear, i.e., $f_j(\cdot)$ are linear functions, then equation (1) changes to GLMMs. Under ordinary circumstances, the nonparametric functions are used in equation (1) to express relationships amongst covariate effects, and correlation among the sample data are modeled by the random effects component of equation (1). According to [8] statistical inference in equation (1) includes (i) the estimation of the nonparametric functions $f_j(\cdot)$, which, in turn, depends upon the choice of span degree, also known as the smoothing parameter [6], and (ii) the estimation of the variance component, $\emptyset$, by using the marginal quasi-likelihood method. Interested readers can find more about different derivation of equation (1) in [14,15].

2.2 Statistical Inference

An inherent assumption in GAMMs is that covariates are additive separable. It was argued that the concept of additively separable of the covariates in a production process might not be always true [16]. When using either GAMs or GAMMs, it is recommended to examine additive separability of the covariates [6,17]. To do this, the use of the residual deviance analysis was suggested, which is simply the logarithm of the likelihood ratio (LR) that follows a chi-squared distribution [9]. Equation (2) shows how the value of the deviance is obtained from taking the

differences between the restricted and unrestricted LR statistic tests. In particular, the computed value of deviance, $\hat{\rho}$

$$Devi(s; \hat{\rho}) = -2\{l(\rho; s) - l(\rho_{max}; s)\} \quad (2)$$

in which $l(\rho; s)$ and $l(\rho_{max}; s)$, respectively, represent the values of the unrestricted and restricted LR. In equation (2), $Devi(s; \hat{\rho})$ has a degree of freedom approximately equal to the differences in the dimensions between the two restricted and unrestricted models, which is the principle of examining input separability in this article [7].

3. EMPIRICAL ANALYSIS

To examine input relationships (i.e., substitutability/complementarity) among factors of production in the econometric model, we utilized the same dataset used in [18], which was collected from the Canadian CATV industry between 1990 and 1996. The data were recorded annually by the Canadian Radio-television Telecommunication Commission (CRTC); an authorized organization for regulating the cable television during the period of the study in Canada. According to [19], the CRTC had divided Canada into several zones, so-called licensed service areas (LSAs), and given exclusive rights to cable television firms throughout the entire country. The data included an unbalanced panel data of 1,041 observations related to the 242 unique undertaking identification (UID) codes for Canadian cable providers during the period of the study. Each UID code was allotted to only one LSA showing the number of individual cable operations in the sample. Thus, each LSA in the data was assumed to be one unit of CATV service production. Assuming that the duality premise holds [20], we specified a nonparametric cost function within a GMM framework and used the statistical program, known as *R* (version 2.12.1.) to estimate the parameters of the model using general smoothing techniques.

Table 1 shows the statistical description of the variables used in the model. The dependent variable was the total cost that was calculated from the summation of operating costs and the user cost of capital for both basic and non-basic services. We subtracted total expenses from those outlays incurred for programming to calculate operating expenses. The covariates of the model consisted of the total number of basic and non-basic cable subscribers as output and the prices of labor, capital, and materials as applied to the industry. The data for each of the covariates was separately calculated for both the basic and non-basic cable subscribers. The first group of subscribers received core channels, arranged by the CATV operator that local television stations would normally broadcast. The non-basic subscribers received more channels (e.g., movie, documentary, sports) than the first group did by paying extra money for each tier. Total payments made to all employees were divided by the number of employees that yielded the price of labor [18]. The summation of depreciation and financial rates was used to obtain the rental price of capital. Finally, the price of materials was obtained by computing the per unit expenditure on intermediate inputs. To do this, [18] spread total costs, excluding those payments made for labor and capital, over a per cable kilometer and per channel basis for all subscribers during the period of the study. To lessen any possible heteroskedasticity in the data, all variables were calculated per unit of output and transformed to logarithms.

Table 1. Statistical description of the variables

Variable	Mean	S. D.	Min.	Max.
Total Costs (\$ 000)	6,975.3	16,533.7	30.6	222,821.3
Price of Labor Shared in Basic Subscribers (\$)	40,388.3	16,275.1	487.6	163,872.6
Price of Labor Shared in Non-Basic Subscribers (\$)	30,790.8	18,458.6	90.3	177,500.0
Price of Capital Shared in Basic Subscribers	0.3211	0.2604	0.1267	7.6618
Price of Capital Shared in Non-Basic Subscribers	0.6045	1.3354	0.1202	40.8860
Price of Material Shared in Basic Subscribers (\$/km 000 per channel)	181.4	128.9	15.1	1,872.6
Price of Material Shared in Non-Basic Subscribers (\$/km 000 per channel)	161.3	143.6	10.2	1,655.9
Number of Basic Subscribers	28,004.0	58,747.0	91.0	586,606.0
Number of Non-Basic Subscribers	19,581.0	51,744.0	14.0	573,420.0

Source: Sample data. Data also used in [18].

Given equation (1), we estimated the nonparametric cost function, specified in a GAMM, using the spline smoothing techniques. We labelled this specification as the restricted model and its estimation results were shown in Table 2. The computed values of the log-likelihood and the Akaike Information Criterion were 211.67 and 388.67 (in absolute value), respectively. Among all the coefficients of the model (except the intercept) the price of labor shared in non-basic subscribers (p -value 0.0177), the number of basic subscribers (p -value 0.000), and the price of capital shared in non-basic subscribers (p -value 0.1157) were statistically significant. To check if equation (1) conformed to the inherent premise of additive separability of the covariates in the model we utilized the statistical test that was described in equation (2). In particular, we examined if a change in one of the input prices could have any impact on the total cost through the total numbers of basic and non-basic subscribers. In other words, we wanted to know if changes in the wage rates shared by basic services had any impact on the total cost of providing services for non-basic subscribers. Furthermore, we tested whether the total cost changed as a result of a change in the rental price of capital shared by basic services through non-basic subscribers. To do this, we specified two unrestricted nonparametric cost functions and estimated

the parameters of these models using the same methodology as the restricted model was estimated.

Table 2. Estimation results from GAMMs (restricted model)

Predictors	Estimates	Est. Error	Non-parametric t-value	p-value
Fixed Effects:				
Intercept	14.642	0.0062	2347.374	0.0000
Price of labor shared in basic subscribers	-7.293	10.435	-0.6989	0.4848
Price of labor shared in non-basic subscribers	-0.276	0.1163	-2.3767	0.0177
Price of capital shared in basic subscribers	-0.265	0.3317	-0.7996	0.4242
Price of capital shared in non-basic subscribers	16.822	10.682	1.5748	0.1157
Price of material shared in basic subscribers	-6.046	15.923	-0.3797	0.7042
Price of material shared in non-basic subscribers	0.131	0.1697	0.7769	0.4374
Number of basic subscribers	0.447	0.0908	4.9270	0.0000
Number of non-basic subscriber	-5.666	16.0348	-0.3534	0.7239
Log-likelihood	211.6733	BIC		-300.9235
AIC	-387.3466	Number of observations		899
Degrees of Freedom	18			

Source: Sample data. Data also used in [18].

Table 3 presents the estimation results of the first unrestricted model in which a new variable was obtained from the product of the price of labor shared in basic subscribers and the number of non-basic subscribers. We hypothesized that if the number of non-basic subscribers changes it might affect the amount of labor hired in basic subscribers, which in turn, might change the total cost of operations. The computed value of the log-likelihood is 214.05 and the Akaike Information Criterion (in absolute value) is 388.10. The coefficient of the rental price of capital shared in non-basic subscribers was statistically different from zero (p -value 0.000) with 99 per cent confidence. The new variable (i.e., the interaction between the number of non-basic subscribers and the wage rates shared in basic subscribers) was not statistically significant (p -value 0.276).

Table 3. Estimation results from GAMMs (unrestricted model)^a

Predictors	Estimates	Est. Error	Non-parametric t-value	p-value
Fixed Effects:				
Intercept	14.642	0.0061	2383.0539	0.0000
Price of labor shared in basic subscribers	-4.095	10.9623	-0.3736	0.7088
Price of labor shared in non-basic subscribers	-9.909	8.0792	-1.2266	0.2203
Price of capital shared in basic subscribers	-0.253	0.3282	-0.7716	0.4406
Price of capital shared in non-basic subscribers	13.379	11.2192	1.1925	0.2334
Price of material shared in basic subscribers	-23.175	21.3500	-1.0855	0.2780
Price of material shared in non-basic subscribers	0.151	0.1699	0.8940	0.3716
Number of basic subscribers	0.437	0.0905	4.8362	0.0000
Number of non-basic subscriber	-0.606	16.8540	-0.0360	0.9713
Product of the number of non-basic subscribers & price of labor in basic subscribers	-21.308	19.5545	-1.0897	0.2761
Log-likelihood	214.0513	BIC	-292.0769	
AIC	-388.1026	Number of observations	899	
Degrees of Freedom	20			

Source: Sample data. Data also used in [18].

^aThe interaction term is obtained from the product of the number of non-basic subscribers and the price of labor in basic subscribers.

The estimation results of the second unrestricted model are shown in Table 4 in which an interaction term was defined as the product of the price of rental capital shared in basic subscribers and the number of non-basic subscribers. We intended to examine if, for instance, the number of non-basic subscribers increases, the total operation costs would also be increased due to the changes in the amount of labor needed to be hired for basic subscribers. The calculated value of the log-likelihood is found to be 212.47 and the Akaike Information Criterion equals 384.49 in absolute value. Among all the coefficients of the model (except the intercept), the price of labor shared in non-basic subscribers (p -value 0.0203) and the number of basic subscribers (p -value 0.000) were statistically different from zero. The coefficient of the new covariate (i.e., the interaction between the number of non-basic subscribers and the rental price of capital shared in basic subscribers) was not statistically significant (p -value 0.285).

Table 4. Estimation results from GAMMs (unrestricted model)^b

Predictors	Estimates	Est. Error	Non-parametric t-value	p-value
Fixed Effects:				
Intercept	14.642	0.0062	2347.5286	0.0000
Price of labor shared in basic subscribers	-3.865	10.9229	-0.3539	0.7235
Price of labor shared in non-basic subscribers	-0.271	0.1164	-2.3256	0.0203
Price of capital shared in basic subscribers	8.674	8.3712	1.0362	0.3004
Price of capital shared in non-basic subscribers	13.301	11.1849	1.1893	0.2347
Price of material shared in basic subscribers	-21.833	21.6913	-1.0065	0.3144
Price of material shared in non-basic subscribers	0.114	0.1720	0.6677	0.5045
Number of basic subscribers	0.444	0.0908	4.8881	0.0000
Number of non-basic subscriber	-0.331	16.8042	-0.0197	0.9843
Product of the number of non-basic subscribers & price of capital in basic subscribers	-20.875	19.5312	-1.0688	0.2854
Log-likelihood	212.2478	BIC		-288.4700
AIC	-384.4957	Number of observations		899
Degrees of Freedom	20			

Source: Sample data. Data also used in [18].

^b The interaction term is obtained from the product of the number of non-basic subscribers and the price of capital in basic subscribers.

It is a general consensus that the incorporation of the interaction term would add more information to the specified model. This implies the initial pair of covariate is not expected to be additive separable. By simply applying the likelihood ratio (LR) statistic test (i.e, equation 2), we examined the null hypothesis of separability amongst additive predictors used in the model. For each of the unrestricted model, we conducted the LR test. Table 5 presents the results of statistical inference obtained from examining the null hypotheses. The computed values of deviance for both models were 4.7559 (model 1) and 1.1490 (model 2), which were less than the critical value of the chi-squared distribution (3.84) with 95 per cent confidence and one degree of freedom. Thus, the interaction term in both models was not statistically different from zero implying that the addition of the new covariate to the primal nonparametric cost function (restricted model) would not be necessary as it added no more information to the model. As a

result, the additive separability assumption between covariates for cost estimation in the CATV industry is valid.

Table 5. Statistical Inference from comparing the restricted and unrestricted GAMMs

New Predictor	DF ^a	Deviance Value	p-value
Model 1:			
Non-basic subscribers & Price of labor shared in basic subscribers	20	4.7559	0.0927
Model 2:			
Non-basic subscribers & Price of capital shared in basic subscribers	20	1.1490	0.5630

Source: Source: Sample data. Data also used in [18].

^a Degrees of freedom

In summary, the result of this study helps policymakers make proper decisions for the industry. As it was argued, the concept of input separability could show how the marginal rate of technical substitution would be changed by any alterations made in other covariates through a third dimension [5]. For instance, if the demand for capital for basic CATV services changes as a result of a change in the number of employees hired for non-basic subscribers, the total operational costs would not be changed. Nevertheless, it is very important for stakeholders in the industry to know which one of the covariates is separable from statistical perspective in the production process. If this piece of information is available then suitable status is provided for firms to make informed decisions.

4. CONCLUSION AND OUTLOOK

Whether an input has a substitute or complementarity relationship with other inputs used in the production process is a valuable knowledge for policymakers who are interested in understanding better such relationship because it affects the rate of return regulation. Prior to the development of nonparametric and semiparametric regression analyses, the most widely used method of estimating the parameters of model was the classical regression analysis. This type of quantitative analysis, however, has been undergoing criticisms for its rigidity on prespecified presumptions, axioms, and postulations. To work with a relatively non-ad hoc method of regression analysis, generalized additive models and generalized additive mixed models are suggested. Generalized additive models are not usually recommended when complex covariate effects among data series exist. Both models can be estimated using a series of iterating and smoothing techniques such as the backfitting algorithm, generalized spline smoothing approaches, and the locally weighted scatterplot smoothing [5]. This paper demonstrates one of the applications of the generalized additive mixed models in testing substitutability or

complementarity in the Canadian cable television industry; a rate- and entry-regulated industry prior to 1997. The result of this study showed that the industry was characterized by separable factors of production during 1990 to 1996. The cable television industry in Canada is using a highly labor savings type of technology. In addition, the findings of the research help policymakers to have an in-depth understanding of such a network-based industry and its current nature of providing services in the country.

REFERENCES

1. Hamilton JL. Input substitutability and vertical integration. *Review of Industrial Organization*. 1992;7(1):29-38.
2. Silberberg E. Theory of the firm in long-run equilibrium. *American Economic Review*. 1974;64(4):734-741.
3. Wohlgenant MK. Input complementarity implies output elasticities larger than one: Implications for cost pass-through. *Theoretical Economic Letters*. 2012;2(1):50-53.
4. Reisinger M, Tarantino E. Vertical integration with complementary inputs. TILEC Discussion Paper No. 2011-004, WHU - Otto Beisheim School of Management, Vallendar, Germany. 2013; <http://dx.doi.org/10.2139/ssrn.1743483> (accessed May 5, 2013).
5. Haghiri M, Law SM, Nolan JF, Simchi A. 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*. 2013;5(3):229-251.
6. Haghiri M. Applied nonparametric regression analysis: The choice of generalized additive models. *Review of Economics and Finance*. 2013;3(1):25-34.
7. Hastie TJ, Tibshirani RJ. **Generalized Additive Models**. 1st edition. London: Chapman and Hall; 1990.
8. Lin X, Zhang D. Inferences in generalized additive mixed models by using smoothing splines. *Journal of Royal Statistical Society Series B*. 1999;61(2):381-400.
9. Bowman AD, Azzalini A. Applied smoothing techniques for data analysis: The Kernel approach with S-Plus illustration. 1st edition. New York: Oxford University Press;1997.
10. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993;88(421):9-25.
11. McCullagh P, Nelder JA. **Generalized Linear Models**. 2nd edition. London: Chapman and Hall;1989.
12. Kurnia A, Notodiputro KA. Generalized additive mixed models for small area estimation. 2nd International Conference on Mathematical Sciences, Universiti Teknologi, Malaysia; 2007.
13. Mamouridis V. Additive mixed models applied to the study of red shrimp landings: Comparison between frequentist and Bayesian perspectives. 3rd Advances in Marine Ecosystem Modelling Research Symposium. June 27-30. Plymouth. United Kingdom;2011.
14. Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*.1994;50(3):689-699.
15. Zhang D, Lin X, Raz J, Sowers M. Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*. 1998;93(442):710-719.

- 491 16. Kneip A, Simar L. A general framework for frontier estimation with panel data. *Journal of*
492 *Productivity Analysis*. 1996;7(2/3):187- 212.
- 493 17. Haghiri M, Simchi A. An application of the residual deviance analysis in testing input
494 separability restrictions. *Applied Economics Letters*. 2005;12(12):755-758.
- 495 18. Haghiri M, Law SM, Nolan JF. Nonparametric estimates of technology using generalized
496 additive models: The case of the Canadian cable television industry. *International Journal*
497 *of Economics and Finance*. 2012;4(1):36-47.
- 498 19. Law SM, Nolan JF. Measuring the impact of regulation: A study of Canadian basic cable
499 television. *Review of Industrial Organization*. 2002;21(3):231-249.
- 500 20. Varian HR. *Intermediate microeconomics: A modern approach*. 7th edition. New Jersey:
501 W.W. Norton & Company; 2005.