Nonparametric Demand Analysis

Matthew Aaron Looney

08/22/2017

Abstract

Nonparametric econometric analysis has been a growing field of study over the past several decades. Many new techniques have been developed within a theoretical framework. However, despite the rapid growth of theoretical results, nonparametric applied research has lagged considerably. This paper employs a nonparametric regression analysis within the context of demand theory. Data on prices and quantites of three commodities (meats, dairy products and beans) from the 2006 Ecuadorian consumer expenditure survey will be evaluated to derive Marshallian price and income elasticities. The nonparametric results will be compared with standard parametric demand analysis tools such as the Log-Log demand model and the Almost Ideal Demand system to guage the effectivness of using nonparametric techniques to estimate demand elasticities.

1 Introduction

Emperical demand analysis has been dominated by the use of parametric functional form models since the appearance of the Working-Lesser Model in the 1940's (Working 1943) (Leser 1963). A 2015 paper by Clements and Gao (Clements and Gao 2015) provide a citation count of journal articles related to four popular parametric demand models (Linear Expenditure, Rotterdam, Translog and Almost Ideal Demand). The authors considered multiple time periods over the date range of 1974-2013. Their results clearly show an upward trajectory for all four parametric demand models with the Linear Expenditure and Almost Ideal Demand (AIDS) models having the highest citation counts, both in the multiple thousands of citations. This result suggests that parametric demand modeling is still very much a hotly researched area.

While parametric form models have clearly dominated the research landscape there are other empirical techniques available to evalaute consumer expenditure datasets. The goal of empirical demand analysis is to answer basic questions about consumer behavior and in many cases, use these answers to design and implement policy (government) or imporove some aspect of business design to increase market exposure/profits (private industry). Beyond these practical objectives Hal Varian (Varian 1982) suggests that, given a consumers expenditure dataset the demand analyst should ask four basic questions concerning the consumers behavior,

(1) Consistency? Is the observerd data consistent with the utility maximizing model?

(2) Structure? Is the observed data consistent with a utility function with some special structure?

(3) Recoverability? How can the underlying utility function be recovered?

(4) Extrapolation? How can we forecast behavior in other circumstances?

With these goals in mind economists are constantly refining their tools to improve their estimates of consumer behavior.

Nonparametric economic analysis has been a growing field of study over the past several decades. Many new techniques have been developed within a theoretical framework. However, despite the rapid growth of theoretical results, nonparametric applied research has lagged considerably. This may be due in part to the advanced mathematical and statistical exposition presented in many nonparametric research papers. It is often difficult for economist, while well trained in advanced mathematical techinques, to fully grasp the significance and translate from purley theoretical results into applied research. Theoretical researchers working within the field of mathematics and statistics often fail to consider applied economic problems and applied economist are not exploring these newly developed theoretical techinques and refining the theory once they have touched real world data. In addition, until recently, nonparametric techniques were substantially more computationally intensive compared to their paramteric counterparts. While the previous statement would seem to imply that nonparametric tecniques have now become computationally less intesive, in reality it is our computer hardware and programing techniques which have improved to the point where nonparametric analysis has become a viable alternative, especially where parametric techniques fall short. These improvemnts in computer hardware have effectivly opened an area of research which, until recently, had remained closed.

Regression analysis is the workhorse technique employed in econometrics. However, in linear regression it is assumed the regressors enter the conditional mean in a linear fashion and each regressor is independent of the other which is often a violation of the data under study. Even when we use nonlinear regression techniques we often still assume we know the functional form for the data generating process (DGP).

The single largest potiential issue with using a parametric form model to evaluate consumer demand is the prior imposition of functional form on the demand model. If a demand system is well represented by a functional form and the econometric model is correctly specified with theoretical assumptions met then a parametric estimator is both consistent and effiecient. In fact if the previous criteria are met parametric regression is superior in just about every way. However, these requirements are excessivly difficult to satisfy when used in the wild to gain insight about real world data and thus the parametric approach can be seroiusly flawed and worst, can lead the researcher to faulty conclusions which can have serious repercussions if used to implement policy.

Nonparametric modelling affords many advantages over their parametric counterparts. Primary amoung them is the nonparametric models ability to help us uncover a more accurate representation of the unknown function, conditioned on the actual data in hand. In section 2 I will explore the methods and models used to estimate a nonparametric regression. In nonparametric analysis a critical step in acheving a reliable kernel estimator is choosing the correct bandwidth estimator. Some time will be spend exploring several different bandwidth estimators which are prominant in the current literature. I will also explore the two most popular kernel regression methods in current use, Local-Constant Least Squares (LCLS) and Local-Linear Least Squares (LLLS). I will summarize the data and quickly review the parametric models being used for comparison purposes. Section 3 explores the results obtained from the nonparametric kernel regression and compares these results with the parametric form models used on the same dataset. Section 4 concludes and details direction for additional studies.

2 Methods, Models and Data

It is well known that a correctly specified parametric model is prefered over nonparametric methods. However, we also know that correct model specification prior to study is a near impossible task. It is important to test the parametric form we would like to employ to assess if the functional form is consistent with the DGP. The test we untilize in this study was developed by Hasiao, Li and Racine and is a test of Consistent Model Specification (CMS-test). (Hsiao et al. 2007)

The test developed by Hsiao et al. is a kernel based test which tests for correct specification of the parametric form model. A short review of the details of the test are given below.

The null hypothesis is given by,

$H_0 : P[E(y_i|x_i) = m(x_i, \beta)] = 1$ for some $\beta \in \mathscr{B}$

where,

$m(\cdot, \cdot)$ is a known function with β being a $p \times 1$ vecotr of unknown paramaters,

$\mathscr{B}$ is a compact subset in $\mathbb{R}^p$.

The alternative hypothesis is given by,

$H_1 : P[E(y_i|x_i) = m(x_i, \beta)] < 1$ for all $\beta \in \mathscr{B}$

The test statistic was originally developed by Fan and Li (1996) and Zheng (1996).

The test statistic is given by,

$I_n = \frac{1}{n^2} \sum_i \sum_{j \neq i} \hat{u}_i \hat{u}_j K_{\gamma j_{ij}}$

where,

$K_{\gamma j_{ij}} = W_{h,ij} L_{\lambda,ij}(\gamma = (h, \lambda))$,

$\hat{u}_i = y_i - m(x_i, \hat{\beta})$ is the parametric null model's residual,

β is a $\sqrt(n)$-consistent estimator of β under H_0.

Hsiao et al. (2007) advocate for the use of cross-validation methods for selecting the kernel smoothing paramater vectors, which is the approach we took in this analysis. Under assumptions (A1)-(A3) of thier paper we can obtain the CV-based test with the new test statistic $\hat{J}_n$:

$n(\hat{h}_1, ..., \hat{h}_q)^{1/2} \hat{I}_n \rightarrow N(0, \Omega)$ in distribution under H_0,

where,

$\Omega = 2E\left[\sigma^4(x_i) f(x_i)\right]\left[\int W^2(v) dv\right]$

A consistent estimator of Ω is given by,

$\hat{\Omega} = n^{-2} 2(\hat{h}_1, ..., \hat{h}_q) \sum_i \sum_{j \neq i} \hat{u}_i^2 \hat{u}_j^2 W^2_{\hat{h},ij} L^2_{\hat{\lambda},ij}$

Which gives the CV-based test statistic as,

$\hat{J}_n = n(\hat{h}_1, ..., \hat{h}_q)^{1/2} \hat{I}_n \hat{\Omega}^{-1/2} \rightarrow N(0, 1)$ in distribution under H_0.

According to the authors, it can be easily shown that the $\hat{J}_n$ test statistic diverges to $+\infty$ if H_0 is false.

In this paper we seek to explore the validity of using nonparametric techniques to estimate demand elasticities. To this end we use the Log-Log demand system developed by Working (1943) and Lesser (1963). The benefit of using this type of parametric form is the simplicity of the model allows us to easily compare the effectivness of our econometric technique without getting bogged down in complicated estimation considerations or theory constraints. The only theory constraints available to the

Log-Log demand model is homogeneity of degree zero in prices and income

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
obs	2,066	1,033.500	596.547	1	2,066
p1	2,066	2.610	0.820	0.365	11.351
p2	2,066	1.231	1.510	0.011	23.736
p3	2,066	1.218	0.444	0.010	4.386
q1	2,066	4.498	3.011	0.001	38.574
q2	2,066	6.244	5.635	0.001	63.762
q3	2,066	0.943	0.929	0.0004	14.074
p1q1	2,066	11.395	7.883	0.003	98.101
p2q2	2,066	4.619	4.215	0.001	51.451
p3q3	2,066	1.073	1.088	0.0005	15.250
X	2,066	17.087	10.679	0.199	128.602
w1	2,066	0.657	0.168	0.001	0.984
w2	2,066	0.269	0.159	0.0001	0.945
w3	2,066	0.074	0.070	0.00003	0.884

good 1 = meats
good 2 = dairy
good 3 = beans

3 Results and Discussion

Table 2: Parametric - Double Log Demand Model

	meats	dairy	pulses	income_elasticity	R-squared
meats_lnq1	-0.9073***	-0.07323***	-0.003844	1.164***	0.74226
dairy_lnq2	-0.05847	-0.8216***	-0.07056*	1.004***	0.64494
pulses_lnq3	-0.3804***	-0.05684*	-0.4394***	0.5658***	0.14743

***Significant at the 1 percent level,
**Significant at the 5 percent level,
*Significant at the 10 percent level.

4 Conclusions

Table 3: Nonparametric Regression using Gaussian Kernel

	meats	dairy	pulses	income_elasticity	R-squared
meats_lnq1	-0.9674***	-0.06514*	-0.04756	1.124***	0.88725
dairy_lnq2	0.002597	-0.843***	-0.0752***	0.9875***	0.71745
pulses_lnq3	-0.3402***	-0.05632**	-0.7777***	0.4965***	0.27586

***Significant at the 1 percent level,
**Significant at the 5 percent level,
*Significant at the 10 percent level.

Table 4: Full AIDS - Marshallian

	meats	dairy	pulses	income_elasticity	R-squared
meats_lnq1	-0.9813***	-0.06048***	-0.03157**	1.073***	0.067727
dairy_lnq2	-0.06908***	-0.8626***	-0.03394***	0.9657***	0.052564
pulses_lnq3	0.1063	0.006219	-0.5657***	0.4532***	0.13459

***Significant at the 1 percent level,
**Significant at the 5 percent level,
*Significant at the 10 percent level.

References

Clements, K. W., & Gao, G. (2015). The Rotterdam demand model half a century on. *Economic Modelling*, *49 IS* -, 91–103.

Fan, Y., & Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, *64*(4), 865.

Hsiao, C., Li, Q., & Racine, J. S. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics*, *140*(2), 802–826.

John Xu Zheng. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, *75*(2), 263–289.

Leser, C. E. V. (1963). Forms of Engel Functions. *Econometrica*, *31*(4), 694–703.

Varian, H. R. (1982). Non-parametric methods in demand analysis. *Economics Letters*, *9*(1), 23–29.

Working, H. (1943). Statistical Laws of Family Expenditure. *Journal of the American Statistical Association*, *38*(221), 43–56.