4-20-04

SUMMARIZING CURVATURE CONDITIONS FOR FLEXIBLE FUNCTIONAL FORMS

Adam Zhuo Chen

Research Assistant, Department of Economics,
Iowa State University
zchen@iastate.edu

Wallace E. Huffman

C.F. Curtiss Distinguished Professor of Agriculture
and Professor of Economics, Iowa State University
whuffman@iastate.edu

Preliminary draft: Please do not quote without permission of the authors

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, U.S.A., August 1-3, 2004

Authors thank an anonymous referee for his/her helpful comments. The usual disclaimer applies.

Summarizing Curvature Conditions for Flexible Functional Forms

Introduction

Curvature conditions of nonlinear function form usually vary across individual observations and may not be readily obtainable. In the practice of production function estimation, the marginal products and/or elasticities are calculated based on the parameter estimates of the flexible function form. In order to compare the curvature conditions, we need to summarize the information into readable dimensions. Even in a very restrictive case, where we compare estimates of different function forms using the same dataset, it is nearly impossible to compare the curvature conditions at all sample points if there are significant number of observations. To compare curvature conditions for function forms estimated with different datasets is more complicated. Ben-Akiva and Lerman (1985) discussed how the curvature conditions at different sample points should be summarized and presented in the context of discrete choice analysis. However, in the production economics literature, this topic has received inadequate attention.

Literature provided two approaches to summarize marginal effects for production flexible function forms (Greene, 2003). The first is to calculate the marginal effects for individual and then present the summary statistics. The intuition underlying this approach is that these statistics will provide a picture of how the aggregated dependent variable will response to marginal aggregated changes of the explanatory variables. However, we argue that this may not be true. The second approach is to evaluate the marginal effects/ marginal productions/ elasticities) at a sample point, e.g. mean/median/ geometric mean of explanatory variables. This approach has been widely used. Diewert and Wales (1987) compared three flexible functional forms by evaluating curvature conditions at the first and last sample points. Anderson & Newell (2003) proposed a method to simplify the calculation of marginal effects at a certain data point for discrete choice models. Meanwhile, it is noted that this approach hinges on strong distribution assumptions of explanatory variables.

This article proposes two methods to address the issue of summarizing curvature conditions for flexible functional forms in the practice of production function estimation. The first approach is to improve the averaging approach by incorporating a weighting scheme according to the contribution of an individual observation. The second is to strengthen the representativeness of central points. We can either use central points that is more robust to outliers and non-normal distribution in providing a typical individual firm/household/person or to group the data points and summarize the curvature conditions for each group. The two new methods are more intuitive and robust to outliers and abnormal explanatory variable distribution.

The rest of this article is organized as following. Section 2 critiques the usual practice of summarizing the curvature conditions in the context of production study. We propose our methods in Section 3. Section 4 concludes.

Critiques on the Common Practices

Greene (2000) stated that: "For computing marginal effects, one can evaluate the expressions at the sample means of the data or evaluate the marginal effects at every observation and use the sample average of the individual marginal effects. The functions are continuous, so Slutsky theorem applies; in large sample they will give the same answer. But that is not so in small or moderate sized samples. Current practice favors averaging the individual marginal effects when it is possible to do so." This statement is likely to be true when evaluating the marginal effects for discrete choice problems, where the exogenous variables are less correlated to each other and can be approximated as normal in large samples. We argue, however, that when we are evaluating the curvature conditions of flexible functional forms in the context of production, this result may not hold. There are two reasons underlie our argument: irregularity of the input quantities distribution; and possible correlation pattern between input usage. We describe them in the following subsections.

Averaging Approach

In the production economics context, it is not necessary that input quantities in large sample be normally distributed. U.S. Congress, Office of Technology Assessment (1984) claimed American farm size is distributed as a bi-model. In United States, there is an increasing trend that while average farm size is enlarged, the number of small farms (of which the purpose is for entertainment rather than income-generating) is increasing at the same time. In developing countries, such trend exists as well due to the limitation of resource and restriction on farm size, i.e., the Household Responsibility System in China and Land to the Tillers Program in south Asia. The land ownership are consisted of large number of existing small farms and an increasing trend of land consolidation due to size economies and the labor migration from agriculture sector to manufacture and service sectors. Therefore large sample theory may not apply in the agricultural production context.

Since the literature usually apply the averaging approach for marginal effects but not for elasticity, we focus our discussion on marginal effects henceforth in this subsection. One may argue that the purpose of averaging approach is to provide the sample mean, as well as standard deviation, of individual marginal effects. However, the mean values of marginal effects are not necessarily representative. Neither does the sample mean converge to the true value in large samples since the overall population may be non-normally distributed, e.g., bi-model. The average of individual marginal effects is not equal to the change of the aggregate dependent variable with respect to marginal change of an explanatory variable either. In fact, it is only a specific realization of the change of the aggregate dependent variable (e.g., the output in agricultural production function estimation) when the marginal changes of inputs of all observations in the sample are equally weighted. In the case of production function estimation, it measures the change of aggregate output when all individual observations have equal extent of change in inputs usage. However, in finite sample, small firms and large firms are likely to have different levels of change of their inputs. The averaging approach fails to summarize the marginal effects of aggregate dependent variable as illustrated in the following.

Denote output as *y*, input vector as *x*, while x_{ji} indicates the jth input of observation i. Thus we may define the averaging approach as: $m_1 = \frac{1}{n}\sum_{i=1}^{n}(\frac{\partial y}{\partial x_j})_i$.

The change of aggregated output with respect to the marginal change of the aggregated input is:

$$\frac{\partial \sum_{i=1}^{n} y_i}{\partial \sum_{i=1}^{n} x_{ji}} = \lim_{\Delta(\sum_{i=1}^{n} x_{ji}) \to 0} \frac{\Delta(\sum_{i=1}^{n} y_i)}{\Delta(\sum_{i=1}^{n} x_{ji})}$$

$$= \lim_{\sum_{i=1}^{n} \Delta x_{ji} \to 0} \frac{\sum_{i=1}^{n} \Delta y_i}{\sum_{i=1}^{n} \Delta x_{ji}}$$

However, since there are n independent control variables in the denominator, we *cannot* claim that:

$$\lim_{\sum_{i=1}^{n} \Delta x_{ji} \to 0} \frac{\sum_{i=1}^{n} \Delta y_i}{\sum_{i=1}^{n} \Delta x_{ji}} = \frac{1}{n}\sum_{i=1}^{n} \lim_{\Delta x_{ji} \to 0} \frac{\Delta y_i}{\Delta x_{ji}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_j}\right)_i .$$

Meanwhile, we can assign a weight w_i to the input change of a single observation as its contribution to the aggregate changes such that $w_i = \frac{\Delta x_{ji}}{\sum_{i=1}^{n} \Delta x_{ji}}$, then the marginal effect of the aggregate amount is:

$$\frac{\partial \sum_{i=1}^{n} y_i}{\partial \sum_{i=1}^{n} x_{ji}} = \lim_{\Delta(\sum_{i=1}^{n} x_{ji}) \to 0} \frac{\Delta(\sum_{i=1}^{n} y_i)}{\Delta(\sum_{i=1}^{n} x_{ji})}$$

$$= \lim_{\Delta t \to 0} \frac{\Delta(\sum_{i=1}^{n} y_i)}{\sum_{i=1}^{n} w_i \Delta t}$$

Denote $f(\mathbf{x}) = \sum_{i=1}^{n} y_i$, we know that when there is a change of Δt , then the amount of change of x_i is $w_i \Delta t$, take first order Taylor series approximation of $f(\mathbf{x})$, the approximate change of $\Delta(\sum_{i=1}^{n} y_i)$ is $\sum_{i=1}^{n} w_i \Delta y_i$, where Δy_i is the change corresponding to a change of Δt for x_i, therefore we have:

$$\frac{\partial \sum_{i=1}^{n} y_i}{\partial \sum_{i=1}^{n} x_{ji}} = \lim_{\Delta t \to 0} \frac{\Delta(\sum_{i=1}^{n} y_i)}{\sum_{i=1}^{n} w_i \Delta t} = \lim_{\Delta t \to 0} \frac{\sum_{i=1}^{n} w_i \Delta y_i}{\Delta t \sum_{i=1}^{n} w_i} = \lim_{\Delta t \to 0} \frac{\sum_{i=1}^{n} w_i \Delta y_i}{\Delta t}$$

$$= \sum_{i=1}^{n} \lim_{\Delta t \to 0} \frac{w_i \Delta y_i}{\Delta t} = \sum_{i=1}^{n} w_i \lim_{\Delta t \to 0} \frac{\Delta y_i}{\Delta t}$$

$$= \sum_{i=1}^{n} w_i \left(\frac{\partial y}{\partial x_j} \right)_i$$

We can summarize this as Theorem 1.

Theorem 1: Assume a weight w_i denoting as the contribution of input change of observation *i* to the aggregate changes such that $w_i = \frac{\Delta x_{ji}}{\sum_{i=1}^{n} \Delta x_{ji}}$, then the marginal effect of the aggregate amount is: $\frac{\partial \sum_{i=1}^{n} y_i}{\partial \sum_{i=1}^{n} x_{ji}} = \sum_{i=1}^{n} w_i \left(\frac{\partial y}{\partial x_j} \right)_i$

Lemma: Assuming the contributions to input change are the same across individuals, we have that $\frac{\partial \sum_{i=1}^{n} y_i}{\partial \sum_{i=1}^{n} x_{ji}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_j} \right)_i$.

Greene (2002) argued that by applying large sample theory the averaged marginal effect converges to the marginal effects at a representative central points. However, this hinges on the assumption of normality in large samples. With the irregularity of input distribution in the context of production function estimation, the result does not hold generally. Furthermore, as we have seen from above, the averaging algorithm is a special case of our method with equal weight for each individual observation.

Representative Individual Approach

The representative individual approach relies heavily on the multivariate normal distribution of inputs. It is facing two problems. First, if input distributions deviate from normal distribution, neither mean, geometric mean nor median is a good representative reference point. Median and geometric mean are more robust to the outliers but cannot handle the bi-model case. Second, many built-in functions in econometric computation packages ignore the potential correlation between the exogenous variables and evaluate the mean/median of the inputs respectively. The quantities of different inputs are possibly correlated to each other. The centroid calculated by averaging different input quantities (as calculated by many software packages) may not represent the whole sample well since it ignored the covariance structure. It is common that inputs are constrained or exhibit certain pattern of correlation between each other, especially when there are strong substitution effects. In the practicing of agricultural production function estimation, taking mean/median of inputs, i.e., land, labor, fertilizer, and capital, does not guarante to result a good representative farm. We argue that the usual representative individual may not be a realistic approach. We provide a simple example to illustrate the failure of representative point approach.

Suppose we have two inputs: labor *L* and capital *K*, without random disturbance, the production function is characterized as:

$$y = f(L,K) = \exp(a_1 \ln L + a_2 \ln K + a_3 (\ln L)^2 + a_4 (\ln K)^2 + a_5 (\ln L)(\ln K))$$

therefore we have the output-labor elasticity as:

$$\frac{\partial \ln y}{\partial \ln L} = a_1 + 2a_3 \ln L + a_5 \ln K \ ,$$

assume a correlation pattern between L and K is that:

$$2a_3 \ln L + a_5 \ln K = C \text{ , } C \text{ is a constant.}$$

Then we have the elasticity is a constant for all observations, but obviously when we evaluate at the respective means of the input; we will get a totally different result. In this case, geometric mean can be used and obtain the correct value. However, generally, since we do not have sufficient information during estimation, we cannot decide which central point to use. Furthermore, the non-linearity itself can be a source of the severe bias of the marginal effect estimates. Ben Akiva and Lerman (1985) discussed such bias in details in the context of discrete choice models.

In summary, the representative individual approach may not produce appropriate results.

New Ways to Summarize Curvature Conditions

In this section we propose two new methods to summarize curvature conditions of flexible functional forms in the practice of production function estimation.

Method 1:

We propose a simple solution to improve over the averaging approach. We can use a predetermined weight to adjust the contribution of curvature conditions. Either the ratio of individual input usage to the aggregate sample input usage or ratio of individual output to the aggregate output is potential good candidate. Yet no theory suggests a "best" weighting scheme exists. These firms with small input usage may have a marginal effect large in magnitude, however, giving the market imperfection in real world, their contribution to input change may be relatively small. Which force finally dominates the curvature condition change depends on which one is greater in magnitude. Assuming the individual contributes to the aggregation with a weight equal to the ratio of its own output to the aggregate output, then multiply the weight to the individual marginal effects and obtain the marginal effects of weighted aggregated mean. This is more intuitive and realistic than the approach of assigning equal weight for all individuals. This approach can be extended to the case of elasticities easily. Since the elasticites are unit-free, output percentage as a weight might be a good choice of the weighting scheme.

Method 2:

To improve the representative individual method, we need to consider how to reduce the dimension of the inputs thus to find an appropriate representative point. One way is to calculate the distance of individual observations to a reference point, i.e., the origin or the centroid, then locate the representative individual(s) using the usual mean, median, or geometric mean. When the sample is severely clustered, the curvature conditions should be evaluated at multiple representative individuals for the existing clusters, respectively.

The distance can be defined in various ways. Two commonly used distance measures are Euclid distance and Markov distance.

Euclid distance is defined as:

$$d(\mathbf{x},\mathbf{x}_0)=\sqrt{(\mathbf{x}-\mathbf{x}_0)'(\mathbf{x}-\mathbf{x}_0)}$$

Markov distance is defined as:

$$d(\mathbf{x},\mathbf{x}_0)=\sqrt{(\mathbf{x}-\mathbf{x}_0)'\Lambda(\mathbf{x}-\mathbf{x}_0)}\text{, where }\Lambda=S^{-1}.$$

Markov distance is more frequently used since it is invariant to the unit of the variables under study.

A recipe of locating representative individual(s) can be described as:

Step 1: Calculate the distance of individual points to a reference point (e.g., centroid);

Step 2: Graph the histogram of the distance for the whole sample and decide whether there are clusters according to the graph (or clustering can be applied directly, then make the corresponding judgment whether the sample appears to be clustered or not);

Step 3.1: If it appears to be a uni-model, then simple statistic procedure can be applied to locate the representative individual;

Step 3.2: If it appears to be a clustered sample, then apply clustering algorithm, e.g., Hierarchical Clustering Methods, to group the observations into *G* groups, and compute the representative points within each group. The overall summary statistics can be a weighted average of these points or they can be presented directly since researchers may be interested in the marginal effects of different clusters.

Note that we propose to use clustering rather than the classification procedure used in Ben-Akiva & Lerman (1986). The difference of classification and clustering is that classification "pertain to a known number of groups, and the operational objective is to assign new observations to one of these groups" while clustering is "a more primitive technique in that no assumptions are made concerning the number of groups or the group structure" (Johnson & Wichern, 2001). In most production studies, we do not have a predetermined *G*, therefore clustering is more applicable in these studies. Though we need to set a cut-off distance for the dendrogram (tree diagram) to decide how many clusters we keep, the number of clusters is ex post rather than ex ante pre-set in classification problem.

Another clustering method may be used is non-hierarchical clustering method, e.g., K-means algorithm. It is computationally convenient but need a predetermined number of clusters, which is usually obtained from preliminary clustering, graphical observation, or simply intuition.

Conclusion

In this article, we critiqued the usual practices of summarizing the curvature conditions of flexible functional forms. We also proposed two new methods to accomplish that goal. Theoretically, the new methods produce more robust and more accurate estimates for the curvature conditions of aggregated variables, as well as the curvature conditions of these variables at the representative points. Both methods provide policy makers a better picture of how the dependent variable may response to the marginal change of explanatory variables.

Meanwhile, when applying the first approach, alternative weighting-schemes are possible with different interpretation. In clustering algorithm, not only we need to select a distance measure, but also need to choose which point the distance may refer to. With the importance of marginal effects in inferring policy implications, these works obviously deserve further exploration.

References:

Anderson, S. and R. G. Newell (2003) “Simplified marginal effects in discrete choice models” *Economic Letters* 81 (2003) pp.321-326

Ben-Akiva M., and Lerman (1985) Discrete Choice Analysis: Theory and Application to Travel Demand, Chapter 6: Aggregate Forecasting Techniques, The MIT Press, Cambridge MA

Diewert W. E. and T. J. Wales (1987) “Flexible functional forms and global curvature conditions” *Econometrica* Vol. 55, No. 1 (1987) pp.43-68

Greene, W.H. Econometric Analysis, Prentice Hall, Upper Saddle River NJ, 4th ed. (2000)

Johnson RA and DW Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, New Jersey, 5th edition, 1995

Stoker T. M. (1996) “Smoothing bias in the measurement of marginal effects” *Journal of Econometrics* 72 (1996) pp.49-84

U.S. Congress, Office of Technology Assessment: “Technology, Public Policy and the Changing Structure of American Agriculture” Washington D.C., U.S. Government Printing Office, OTA-F-285 March 1986