ECONOMICS BULLETIN

Testing for Shape Invariance of Semiparametric Equivalence Scales

Thanasis Stengos
University of Guelph

Dianqin Wang
University og Guelph

Abstract

Within a semiparametric framework we propose a test of shape invariance of Engel curves, which is a necessary condition for base independence. Using Canadian family expenditure data for 1996 we reject shape invariance for the fuel and clothing share equations.

The first author would like to acknowledge financial support from the SSHRC of Canada.
Citation: Stengos, Thanasis and Dianqin Wang, (2007) "Testing for Shape Invariance of Semiparametric Equivalence Scales." *Economics Bulletin,* Vol. 4, No. 21 pp. 1-11
Submitted: March 12, 2007. **Accepted:** June 12, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume4/EB-07D10005A.pdf

1 Introduction

Equivalence scales constitute a useful conceptual tool for designing economic policy such as an equitable tax and benefit system that would put on equal footing (bring into equivalence) the living standards of families of different size and composition. Preferences that obey the property that equivalence scales will not vary across household income levels are said to follow base independence. This is an important restriction on preferences that allows for a meaningful interpretation of equivalence scales for policy purposes, (see Blackorby and Donaldson (1993) and Blundell and Lewbel (1991)).

In parametric demand systems base independence is many times rejected. However, it is not clear whether this rejection reflects a genuine dependence of the equivalence scale function on total expenditures, or is the result of restrictive parametric assumptions about the functional form of the share equations. Recently, to deal with this issue Gozalo (1997) and Pendakur (1999), Wilke (2003) and Stengos, Sun and Wang (2006) have introduced nonparametric and semiparametric models of equivalence scales. Gozalo (1997) treats both the share equations and the equivalence scale function as purely nonparametric in the context of a system of Engel curves, whereas Pendakur (1999), Wilke (2004) and Stengos, Sun and Wang (2006) allow for general unspecified share equations that can be estimated nonparametrically, whereas the equivalence scale is estimated as a parameter in each equation and as a result it is directly comparable with parametric estimates. In this paper we will propose a simple test for shape invariance which is a necessary condition for base-independence. This condition simply states that under base independence the (nonparametric) share equation curves for different household types will have the same curvature. As in Gozalo (1997) we use bootstrapping to construct confidence intervals. We find that shape invariance holds for food shares, but not for fuel and clothing shares.

The paper is organized as follows. In the next section we present our test of shape invariance between nonparametric Engel curves. We then proceed to present the results of the estimation of a system of Engel Curves using the 1996 Canadian Family Expenditure Survey.

2 Testing for Shape-Invariant equivalence scales

Base independence implies shape invariance for a system of the equations that belong to a demand system. The same holds true if one looks at the system without temporal variation as a system of Engel curves. The latter are linked through shape invariance for each household type and they have the same curvature. This restriction constitutes the testable part of the base-independence hypothesis. Shape invariance is only a necessary but not sufficient condition for base independence. Hence, rejection of shape invariance is useful since it would directly invalidate base independence, whereas acceptance of shape invariance would not necessarily imply the validity of base independence.

In the analysis below we restrict our attention to a single good in a system of Engel expenditure share equations, where price remains constant. As in Pendakur (1999) for exposition purposes we consider two household types, type a and type b. Let x be the log of total expenditure and let $y = f(x)$ be the share of the single good purchased by the household. Under base-independence, the expenditure shares of type a and type b are related by

$$f_b(x_b) - \eta = f_a(x_b - \delta) \tag{1}$$

where $f_t(x_t)$ is the share of a single good by household type $t = a, b$, δ is the log of base-independent equivalence scales and η is the elasticity of the equivalence scale function with respect to price. If δ and η are independent of x, then they are said to be base independent. We consider a system of Engle curves where prices are given and hence under base independence both δ and η are constants.

Suppose we have n_t observations for household type t, $t = a, b$ on x and y, $x_{t,j}$ and $y_{t,j}$ for $j = 1, 2, \cdots, n_t$. A nonparametric estimate of $f_t(x)$ based on $(x_{t,j}, y_{t,j})$ for $j = 1, \cdots, n_t$ is defined by

$$\hat{f}_t(x) = \frac{r_t(x)}{d_t(x)} \tag{2}$$

where $r(x)$ and $d(x)$ are given by

$$r_t(x) = \frac{1}{n_t h_t} \sum_{i=1}^{n_t} y_{t,i} K\left(\frac{x_t - x_{t,i}}{h_t}\right) \tag{3}$$

$$d_t(x) = \frac{1}{n_t h_t} \sum_{i=1}^{n_t} K\left(\frac{x_t - x_{t,i}}{h_t}\right) \tag{4}$$

$K(.)$ is the kernel function and h_t is the bandwidth. We use the Gaussian kernel, $K\left(\frac{x_t - x_{t,i}}{h_t}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_t - x_{t,i}}{h_t}\right)^2}$. We will determine the optimal bandwidth by cross-validation method over the range $(0, 2sn_t^{-\frac{1}{5}})$, where s is the standard deviation of x.

We will estimate the parameter δ, the log of base-independent equivalence scale, using the loss function introduced by Stengos, Sun and Wang (2006) that is an improved variant of the one suggested by Pendakur (1999). If the two households have similar preferences under base independence, we would expect that the ranges of the two sample sets (x_a, y_a) and $(x_b - \delta, y_b - \eta)$ to overlap and the nonparametric regressions based on (x_a, y_a) and $(x_b - \delta, y_b - \eta)$ should be the same. We will define $\hat{g}_b(x)$ as nonparametric estimates based on points $(x_{b,j} - \delta, y_{b,j})$ for $j = 1, \cdots, n_b$, compared to $\hat{f}_b(x)$ based on points $(x_{b,j}, y_{b,j})$ for $j = 1, \cdots, n_b$. The points $(x_a, \hat{f}_a(x_a))$ and $(x_b - \delta, \hat{g}_b(x_b))$ determine the shape of expenditure share curves of household type a and type b. Since $x = \begin{pmatrix} x_a \\ x_b - \delta \end{pmatrix}$ and $n = n_a + n_{b,}$ the sets of points $(x, \hat{f}_a(x))$ and $(x, \hat{g}_b(x))$ can be used to construct a measure of the closeness of the two curves. We run the following artificial regression

$$\hat{g}_b(x) = \eta + \beta \hat{f}_a(x) + u \tag{5}$$

The hypothesis of shape invariance is given as $\beta = 1$. We let a and b denote the OLS estimates of η and β respectively from equation (5) above. To compute the standard error of b, we use a wild bootstrap procedure similar to the one presented in Gozalo (1997)). Using the notation introduced above, under the null hypothesis of shape invariance of the share equations, $(x, \hat{g}_b(x) - a)$ should be similar to $(x, \hat{f}_a(x))$. Let us define the vector of residuals $u_0 = (\hat{g}_b(x) - a) - b\widehat{f_a}(x)$, and let y_0^* denote $\hat{g}_b(x) - a$ and let $y_0^*(i)$ be the $i - th$ element of y_0^*. Similarly we define $u_0(i)$ as the $i - th$ element of u_0 and we also define $\bar{u} = \sum_{i=1}^{n} u_0(i)$. The jth bootstrap observation is produced by the following:

For each i from 1 to n, we

1. randomly generate the bootstrap residual $u(i)$ from a distribution F_i such that for $Z \sim F_i$,

$$\begin{aligned} E_{F_i} Z &= 0 \\ E_{F_i} Z^2 &= (u_0(i) - \bar{u})^2 \\ E_{F_i} Z^3 &= (u_0(i) - \bar{u})^3 \end{aligned}$$

2. set $y_j^*(i)$ to be $y_0^*(i) + (u(i) + \bar{u})$

We use a distribution F_i defined in Gozalo (1997), which has a density function $p_i \phi_{a_i,\kappa} + (1 - p_i)\phi_{b_i,\kappa}$, where $\phi_{c,\kappa}$ is the density function of $N(c\kappa, c^2(1 - \kappa^2))$, and p_i, a_i, and b_i are defined as following:

$$\begin{aligned} p_i &= \left(5 + \sqrt{5}\right)/10 \\ a_i &= u_0(i)\left(1 - \sqrt{5}\right)/2 \\ b_i &= u_0(i)\left(1 + \sqrt{5}\right)/2 \\ \kappa &= (\sqrt{3} - 1)/2 \end{aligned}$$

The bootstrap procedure produces a series of y_j^*, for $j = 1, \cdots, B$, where B is the number of bootstrap replications. Using the series y_j^*, we obtain the empirical distriution of b, using equation (7) above, which gives the bootstrap standard error of b. We proceed to test the hypothesis $H_0 : \beta = 1$, using the bootstrap standard error in the construction of the $t-$ratio for H_0.

3 Empirical Results

We now apply the test for shape invariance to data taken from the 1996 Canadian Family Expenditure Survey to estimate expenditure share curves for cloth and fuel. The data set we examine covers 8741 households. After controlling for regional characteristics we define the household type HT by the number of seniors (ns), the number of adults(na), the number of youths(ny), and the number of children(nc) as

$$HT = 1000ns + 100na + 10ny + nc.$$

For example $HT = 211$ refers to a household with two adults, one youth and one child, while $HT = 1200$ refers to a household with one senior and two adults. We will focus on those types of household with more than 50 observations in the survey. There are 23 such household types, the most popular being the one with two adults, which will serve as the reference household. For a description of the data, see Stengos, Sun and Wang (2006). The distribution of the household types is summarized in Table1.

It is worth noting that using the above classification scheme to select data into different cells allows for controls for age affects, household size and the number of children which are among the most important characteristics in consumer demand. In effect our approach is similar in spirit to the partial linear semiparametric model of Blundell, Duncan and Pendakur (1998), Lyssiotou, Pashardes and Stengos (1999) and Yatchew, Sun and Deri (2003), where characteristics enter the linear part of partial linear specification.[1]

[1] Note that since we are only using one-dimensional kernels our estimates do not suffer from the curse of dimensionality problem that typically plagues multidimensional kernel estimates. That is true even in the case of the smallest cell of 50 observations.

Table 1: Distribution of Household Types

HT	Members	Obs	HT	Members	Obs
10	1 youth	52	210	2 adults, 1 youth	442
20	2 youths	91	211	2 adults, 1 youth, 1 child	249
100	1 Adult	1064	212	2 adults, 1 youth, 2 children	100
101	1 adult, 1child	182	220	2 adults, 2 youths	352
102	1 adult, 2 children	128	221	2 adults, 2 youths, 1 child	122
110	1 adult, 1 youth	254	300	3 adults	142
111	1 adult, 1 youth, 1 child	99	1000	1 senior	907
120	1 adult, 2 youths	66	1100	1 senior, 1 adult	350
200	2 adults	1542	1200	1 senior, 2 adults	84
201	2 adults, 1 child	566	2000	2 seniors	678
202	2 adults, 2 children	861	2100	2 seniors, 1 adult	67
203	2 adults, 3 children	343			

We use the loss function introduced by Stengos, Sun and Wang (2006) to estimate δ of the reference household of two adults (HT=200) and the other household types. We then proceed to test for shape invariance for the system of fuel and clothing share equations. Table 2 presents the results of these tests. It is clear that the hypothesis of shape invariance and consequently of base independence is rejected in most cases, except for a few categories for fuel.[2] This suggests that shape invariance and hence base independence does not hold for these share equations and as such equivalence scales will be dependent on expenditures.

[2]To compute standrad error of the coefficient estimate b we use 999 bootstrap replications.

Table 2: The bootstrap testing results contain the mean of b and *p-values* for the null hypothesis of shape invariance

HT	$\bar{b}_{cloth}$	$p-value$ for $\bar{b}_{cloth}$	$\bar{b}_{fuel}$	$p-value$ for $\bar{b}_{fuel}$
10	-0.0326	0.0000	-0.0053	0.0000
20	0.0661	0.0000	0.1978	0.0000
100	0.0088	0.0000	1.0130	0.74345
101	0.1164	0.0000	0.1637	0.0000
102	0.0426	0.0000	0.1956	0.0000
110	0.2413	0.0000	0.5219	0.0000
111	0.0269	0.0000	-0.3654	0.0000
120	0.1043	0.0000	0.1336	0.0000
201	0.2251	0.0000	0.8950	0.0000
202	0.0904	0.0000	1.0885	0.2378
203	-0.2112	0.0000	1.0820	0.2846
210	0.0195	0.0000	0.5607	0.0000
211	0.1294	0.0000	0.7432	0.0000
212	0.1646	0.0000	0.4066	0.0000
220	0.0098	0.0000	0.8348	0.1123
221	0.1387	0.0000	0.5108	0.0000
300	0.0958	0.0000	0.8896	0.2223
1000	-0.0726	0.0000	0.8516	0.1423
1100	-0.0437	0.0000	0.9893	0.7135
1200	0.1352	0.0000	0.5794	0.0000
2000	0.1513	0.0000	0.7365	0.0000
2100	0.0395	0.0000	0.2192	0.0000

References

[1] Blackorby C. and D. Donaldson, 1993, Adult-equivalence scales and the economic implementation of interpersonal comparisons of well-being, *Social Choice and Welfare* 10, 335-361.

[2] Blundell R. and A. Lewbel, 1991, The information content of equivalence scales, *Journal of Econometrics* 50, 49-68.

[3] Blundell R., Duncan, A. and K. Pendakur, 1998, Semiparametric estimation and consumer demand, *Journal of Applied Econometrics* 13, 453-461.

[4] Gozalo, P., 1997, Nonparametric bootstrap analysis with applications to demographic effects in demand functions, *Journal of Econometrics* 81, 357-393.

[5] Lyssiotou, P., Pashardes, P. and T. Stengos, 1999, Preference Heterogeneity and the Rank of Demand Systems, *Journal of Business and Economic Statistics*, 17, 248-252.

[6] Pendakur, K., 1999, Estimates and tests of base-independent equivalent scales, *Journal of Econometrics* 88, 1-40.

[7] Stengos, T., Sun, Y. and D. Wang, 2006, Estimates of Semiparametric Equivalence Scales, *Journal of Applied Econometrics,*21, 629-639.

[8] Wilke, R., 2003, Semiparametric Estimation of Regression Functions under Shape Invariance Restrictions. *ZEW*-Discussion Paper 03-64.

[9] Yatchew, A., Sun, Y. and C., Deri, 2003, Efficient estimation of semi-parametric equivalence scales with evidence from South Africa. *Journal of Business and Economic Statistics,* 21, 247-257.