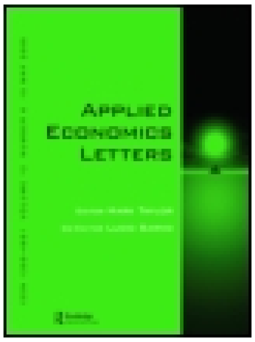

Applied Economics Letters

ISSN: 1350-4851 (Print) 1466-4291 (Online) Journal homepage: http://www.tandfonline.com/loi/rael20

Empirical issues in the computation of Stone–Lewbel price indexes in censored micro-level demand systems

Daniele Moro, Elena Castellari & Paolo Sckokai

To cite this article: Daniele Moro, Elena Castellari & Paolo Sckokai (2017): Empirical issues in the computation of Stone–Lewbel price indexes in censored micro-level demand systems, Applied Economics Letters, DOI: 10.1080/13504851.2017.1346353

To link to this article: http://dx.doi.org/10.1080/13504851.2017.1346353

Published online: 04 Jul 2017.

APPLIED ECONOMICS LETTERS, 2017
https://doi.org/10.1080/13504851.2017.1346353

Routledge
Taylor & Francis Group

ARTICLE

Empirical issues in the computation of Stone–Lewbel price indexes in censored micro-level demand systems

Daniele Moro, Elena Castellari and Paolo Sckokai

Dipartimento di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza, Italy

ABSTRACT
In this article, we propose an empirical method for the computation of the Stone–Lewbel (SL) price index for product aggregates, when censored samples with zero expenditures are available from household budget surveys. The proposed technique is based on a regression imputation method that takes into account the price dynamics, therefore, allowing to disentangle the role of demographics from the role of prices in computing the SL index. Our simulations seem to indicate that our method is a valuable alternative.

KEYWORDS
Stone–Lewbel price indexes; censored data; household budget survey; demand estimation

JEL CLASSIFICATION
C24; D11; D12

I. Introduction

Demand systems estimation allows to obtain information that can be used for policy and welfare analysis. Nowadays, demand analysis is largely based on micro-level household data, which provide a huge amount of information on consumption habits of a large sample of individuals/households. These data sets often lack price information. Sometimes, unit values as proxies for prices can be computed, although some caution should be taken (e.g. Cox and Wohlgenant 1986; Deaton 1988); but in many other cases, when only expenditures are collected, no price information can be obtained.

To circumvent the problem, it is common to resort to external sources of prices, such as consumer price indexes (CPIs). However, while accounting for variability through time, CPIs do not usually allow to introduce household price variability (e.g. Hoderlein and Mihaleva 2008): prices are equal for all individuals at different points in space in a given time period (see, e.g. Moro and Sckokai 2000). However, not accounting for household variability may bias the estimation, since prices may also reflect a different composition of the (aggregate) goods.

Based on Lewbel (1989), a more appealing route is to construct Stone–Lewbel (SL) household price indexes for (aggregate) goods, using budget shares and CPIs for elementary commodities *within* each aggregate. Under this approach, it is possible to *induce* household-level variability in prices. Empirically, introducing this variability may not be trivial, since in the recording period individuals/households may not purchase some goods such that recovering households' prices of aggregate goods becomes impossible. We then need an alternative route to reconstruct SL indices when information is missing (i.e. expenditure censoring). Castellón, Boonsaeng and Carpio (2015, CBC from now onward) proposed an empirical solution, using a regression imputation method. In this article, we discuss the issues and propose a different solution comparing it with the CBC approach; further, by means of a simulation procedure, we show that our index is a valuable alternative.

II. Conceptual framework

Consider the expenditure share w_{ij} for elementary goods j entering an aggregate good i (individuals are not indexed). Lewbel (1989) showed that budget shares depend on household characteristics (i.e. demographics):

$$w_{ij} = h_{ij}(p_i, a, y_i)$$

where p_i is the vector of prices p_{ij} of elementary goods j ($j = 1,2,\ldots,n_i$) in group i, a is a vector of demographics for the individual/household and y_i is the allocated expenditure to group i. Assuming homothetic separability, then there exists some function:

CONTACT Daniele Moro daniele.moro@unicatt.it

$$v_i \equiv p_i = v_i(p_i, \bar{a})$$

where P_i is the price index for group i and $\bar{a}$ is the set of demographic characteristics for a *reference household*. We then define the SL price index for any individual/household as

$$\pi_i \equiv m_i p_i = v_i(p_i, a)$$

where m_i is the equivalence scale:

$$m_i \equiv \frac{v_i(p_i, a)}{v_i(p_i, \bar{a})}$$

In other words, variation in demographics across households will be reflected in the equivalence scales, and consequently in SL price indexes at the individual level, providing price variation.

Assume group sub-utility functions of the Cobb–Douglas type (i.e. homothetic sub-utility functions, i.e. constant within-group expenditure shares):

$$u_i(q_i, a) = \kappa_i \prod_{j=1}^{n_i} q_{ij}^{w_{ij}}$$

where q_i is the vector of quantities q_{ij} of elementary goods j (j = 1,2,. . .,n_i) in group i, and κ_i a scaling factor, defined as

$$\kappa_i = \prod_{j=1}^{n_i} \bar{w}_{ij}^{-\bar{w}_{ij}}$$

depending only on the budget shares $\bar{w}_{ij}$ of a reference household. Then, SL price indexes for the aggregate good i can be computed as

$$v_i(p_i, a) = \frac{1}{\kappa_i} \prod_{j=1}^{n_i} \left(\frac{p_{ij}}{w_{ij}}\right)^{w_{ij}} \quad (1)$$

In empirical applications, the average budget shares are usually taken for the reference household, and consequently, the computation of SL price indexes is straightforward, provided that expenditures for within-group elementary goods are nonzero. Due to many reasons, including the fact that the recording period may be short (e.g. a 2-week period), it is often the case that households do not purchase all goods, and therefore, many within-group expenditures are zero, making the computation of the SL index unfeasible.

III. Computing SL price indexes with within-group zero expenditures

Castellón et al. (2015) proposed a route for *estimating* a SL price index when within-group expenditure shares are zero (SL-CBC), based on the rationale that differences across households' SL price indexes are attributable only to differences in demographics, since prices p_{ij} are the same for all households. The SL price index in Equation (1) is first computed using *only* uncensored data (i.e. the subsample of households with nonzero expenditures for all goods j in group i). Then, its logarithm is regressed on the set of individual/household demographic variables, and estimated parameters are used to recover SL price indexes for *censored* households, conditional on their demographic characteristics. They applied their procedure using monthly CPIs, annual CPIs and no prices.[1] However, their approach does not properly account for the time dimension of the data. For example, using monthly data, changes in the logarithms of the computed indexes in the uncensored sample may also depend on changes in elementary prices through time.

In fact, suppose for the moment that no price information about within-group commodities are available; without loss of generality, we may arbitrarily set elementary prices equal to 1 in one period and the SL price index becomes

$$v_i(p_i, a) = \frac{\prod_{j=1}^{n_i} w_{ij}^{-w_{ij}}}{\prod_{j=1}^{n_i} \bar{w}_{ij}^{-\bar{w}_{ij}}}$$

that is, only differences in budget shares are responsible for index variability across individuals in that period. This expression holds for every time period only under stringent conditions, otherwise differences in the SL price index in other periods depend also on price dynamics:

$$v_i(p_i, a) = \frac{\prod_{j=1}^{n_i} w_{ij}^{-w_{ij}}}{\prod_{j=1}^{n_i} \bar{w}_{ij}^{-\bar{w}_{ij}}} \prod_{j=1}^{n_i} p_{ij}^{w_{ij}}$$

To compute the SL index, we propose an alternative procedure (SL-MCS) that is free from price dynamics bias. We first construct price indexes for the uncensored individuals, *setting prices equal to* 1 *in each time period* (i.e. implicitly in each month, we

[1]In their article, Castellón et al. (2015) found that different specifications of the SL index do not have a significant impact on demand estimation. Although limited to the data set used in their paper, this finding suggests that demand estimation may not suffer from the absence of any price information.

scale all prices to be equal to 1), obtaining for each period a scaling factor relating the price of each individual to that of the reference household.[2] Then, for the uncensored subsample, the logarithm of this scaling factor is regressed for all individuals and time periods on the set of demographics, and estimated parameters are used to compute *fitted* scaling factors for the whole sample. The SL-MCS price index can be computed using the following approximation of the original SL price index formula in Equation (1),

$$v_i(p_i, a) = \left(\frac{\widehat{\prod_{j=1}^{n_i} w_{ij}^{-w_{ij}}}}{\prod_{j=1}^{n_i} \bar{w}_{ij}^{-\bar{w}_{ij}}} \right) \prod_{j=1}^{n_i} p_{ij}^{\bar{w}_{ij}} \qquad (1a)$$

Equation (1a) is introduced to account for price dynamics: the term in brackets provides an estimated measure of the scaling between the price index of each individual and that of the reference household, while the Cobb–Douglas price index computed at the reference household budget shares takes into account only the dynamics of the elementary prices within the group.[3]

IV. Empirical evidence and concluding remarks

To compare the two procedures, we use data from the Italian Household Budget Survey, conducted by ISTAT (National Bureau of Statistics), for a 3-year period (2011–2013). We use a sample of 66,771 households, treated as a time series of cross-sections. We group food expenditures in 14 aggregates (Table 1); each aggregate contains two or more elementary goods. Monthly CPI*s* are also taken from ISTAT.

As it is evident from Table 1, the issue of zero expenditures (i.e. at least one subgroup elementary expenditure is zero) is relevant, ranging from a minimum of 63.6% for group 2 to a maximum of 98.3% in group 5; this is mainly due to the 2-week recording period used in the survey. Households characteristics have been selected given the available information: geographic area (dummies for North, Centre and South), household composition (number of household members), age of the household head (dummies for young, adult and elderly), presence of kids (dummies for presence or not) and level of education (five dummies).

Table 1. List of commodity groups and their composition.

Commodity groups	No. of subgroup goods	Share of zero expenditures
Grains and grain-based products	4	0.775
Vegetables and vegetable products	3	0.636
Starchy roots, tubers, legumes, nuts and oilseeds	3	0.909
Fruit, fruit products and fruit and vegetable juices	3	0.921
Beef, veal and lamb	3	0.983
Poultry, eggs and other fresh meat	3	0.935
Pork + processed and other cooked meats	3	0.953
Fish and other seafood	2	0.775
Milk, dairy products and milk product imitates	3	0.831
Sugar and confectionary and prepared desserts + soft drinks	5	0.966
Animal fats	2	0.705
Plant based fats	2	0.876
Tea, coffee, cocoa and drinking water	3	0.848
Alcoholic beverages	3	0.972

We have computed the two alternative (SL-CBC, and SL-MCS) indexes for the 14 groups. Dummy variables for the year of the survey have been added in the estimation, as in Castellón et al. (2015). Results are summarized in Table 2. The correlation coefficient between the two indexes ranges from a minimum of 0.495 (group 2) to a maximum of 0.987 (group 1), with an average of 0.925, indicating that differences between the two methods can be quite large. Although there is no clear evidence of a relation between sample size and differences in the two indexes, it may be worth noting that the lowest correlation coefficient (group 2) is associated with the lowest incidence of zero expenditures. Thus, it may be the case that with a high level of zero occurrences, the two methods produce similar results, while with larger samples of uncensored data, the two imputation methods may lead to strongly different results.

To provide more evidence, we have conducted a simulation exercise to compare the two approaches. Using the demographic structure of our sample, we have randomly generated 100 samples of within-group budget shares for 3 elementary goods $j = 1, 2, 3,$

[2]Note that this is somewhat similar to the 'no price' case in Castellón et al. (2015), where individual price differences are only due to demographic heterogeneity (see Lewbel 1989); however, to account for price dynamics, we introduce the Cobb–Douglas index in Equation (1a).

[3]As pointed out by a Referee, being (1a) an empirical approximation to the true index in Equation (1), alternative formulations for the Cobb–Douglas price index in Equation (1a) could be adopted, such as time-specific average budget shares. We have chosen to resort to this formulation because it uses constant budget shares, and therefore, the dynamics of the Cobb–Douglas price index is only due do the dynamics of prices within the group.

Table 2. Estimation results of the SL-CBC and SL-MCS indexes: summary statistics and correlations.

	Mean	SD	Min	Max	Corr.	Nonzero obs.	Share of zeros
SL-CBC-1	93.997	3.617	82.944	105.411	0.987	15,025	0.775
SL-MCS-1	94.013	3.667	84.061	105.578			
SL-CBC-2	89.360	4.543	76.946	99.388	0.495	24,290	0.636
SL-MCS-2	89.202	2.285	84.727	94.723			
SL-CBC-3	95.898	3.286	89.378	104.399	0.937	6,094	0.909
SL-MCS-3	95.906	3.002	91.487	102.481			
SL-CBC-4	96.040	5.035	81.988	115.030	0.824	5,246	0.921
SL-MCS-4	95.602	3.876	84.103	107.602			
SL-CBC-5	92.878	2.381	86.910	100.461	0.975	1,162	0.983
SL-MCS-5	92.895	2.346	87.241	99.631			
SL-CBC-6	93.690	3.412	86.858	101.403	0.982	4,352	0.935
SL-MCS-6	93.694	3.325	87.532	100.488			
SL-CBC-7	93.490	3.503	84.327	103.238	0.986	3114	0.953
SL-MCS-7	93.513	3.472	85.272	102.731			
SL-CBC-8	98.888	2.815	91.135	106.638	0.954	15,016	0.775
SL-MCS-8	98.917	2.721	93.071	106.249			
SL-CBC-9	92.310	2.556	85.300	97.880	0.965	11,316	0.831
SL-MCS-9	92.330	2.529	86.610	97.346			
SL-CBC-10	90.142	3.560	81.897	100.085	0.980	2,294	0.966
SL-MCS-10	90.190	3.545	82.458	100.914			
SL-CBC-11	102.298	4.120	90.703	115.173	0.968	19,719	0.705
SL-MCS-11	102.294	4.056	92.618	114.566			
SL-CBC-12	97.989	3.359	91.664	106.862	0.982	8,312	0.876
SL-MCS-12	98.007	3.426	92.306	106.691			
SL-CBC-13	99.494	3.593	89.298	107.878	0.929	10,132	0.848
SL-MCS-13	99.488	3.379	92.828	107.525			
SL-CBC-14	91.218	3.777	82.328	102.252	0.982	1,875	0.972
SL-MCS-14	91.271	3.708	83.174	101.467			

entering an aggregate i, as $w_{ij} = m_{ij}(a) + \varepsilon_{ij}$, according to the suggestion of Lewbel (1989). We have also tried different values of the SDs of the random term in the simulation, with the constraint that all budget shares were positive.

Using the generated budget shares and monthly prices, we have computed the *true* SL index (SL-true) applying the formula in Equation (2); then, we have artificially censored the sample, using three different levels of censoring (low censoring, with 69.0% of zero, medium censoring, with a 84.1% of zero and high censoring, with 96.6% of zero), computing both the SL-MCS and the SL-CBC indexes. We have then computed the correlation coefficients between the two indexes and the 'true' index, SL-true, averaging over 100 samples. Results are presented in Table 3.

It is clear that the SL-MCS always outperforms the SL-CBC, in terms of correlations with the true index. Further, when censoring is randomly induced (i.e. it is not related to demographics), the degree of censoring does not seem to be relevant for the performance of the procedures. Of course, the variability of the within-group budget shares does affect the ability of the two approaches in replicating the true SL index.

In sum, the SL-MCS index, based on a regression imputation method that does account for price dynamics, appears as a more plausible approach, since it allows to better disentangle the role of demographics from the role of prices. Therefore, it provides a valid alternative to the use of the CBC approach to construct SL indexes with censored samples when price dynamics is strong over the sample period. Of course, both the MCS and the CBC methods are just approximations to the true SL index, and therefore, they may impact the quality of the results (i.e. estimated elasticities) in demand estimation, although Castellón et al. (2015) did not

Table 3. Simulation results: correlations between SL-true index and SL-CBC and SL-MCS indexes.

	Low censoring: % zero = 69.0%		Medium censoring: % zero = 84.1%		High censoring: % zero = 96.6%	
	SL-TRUE/SL-MCS	SL-TRUE/SL-CBC	SL-TRUE/SL-MCS	SL-TRUE/SL-CBC	SL-TRUE/SL-MCS	SL-TRUE/SL-CBC
SD = 0.010	0.966	0.920	0.965	0.920	0.964	0.917
SD = 0.015	0.935	0.890	0.934	0.890	0.934	0.889
SD = 0.020	0.894	0.850	0.892	0.851	0.892	0.849
SD = 0.025	0.846	0.807	0.848	0.808	0.844	0.803
SD = 0.030	0.794	0.757	0.791	0.755	0.789	0.751

find major differences by applying different specifications of their method in demand estimation.

Acknowledgement

This work was supported by the Daniel & Nina Carasso Foundation through the [FP7 SUSFOOD ERA-Net Research Project 'SUSDIET] – Implementing sustainable diets in Europe' (Scientific Coordinator: Louis-George Soler).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Daniel & Nina Carasso Foundation through the [FP7 SUSFOOD ERA-Net Research Project 'SUSDIET] – Implementing sustainable diets in Europe' (Scientific Coordinator: Louis-George Soler).

References

Castellón, C. E., T. Boonsaeng, and C. E. Carpio. 2015. "Demand System Estimation in the Absence of Price Data: An Application of Stone-Lewbel Price Indices." *Applied Economics* 47: 553–568. doi:10.1080/00036846.2014.975332.

Cox, T. L., and M. K. Wohlgenant. 1986. "Prices and Quality Effects in Cross-Sectional Demand Analysis." *American Journal of Agricultural Economics* 68: 908–919. doi:10.2307/1242137.

Deaton, A. 1988. "Quality, Quantity, and Spatial Variation of Price." *The American Economic Review* 78: 418–430.

Hoderlein, S., and S. Mihaleva. 2008. "Increasing the Price Variation in a Repeated Cross Section." *Journal of Econometrics* 147: 316–325. doi:10.1016/j.jeconom.2008.09.022.

Lewbel, A. 1989. "Identification and Estimation of Equivalence Scales under Weak Separability." *The Review of Economic Studies* 56: 311–316. doi:10.2307/2297464.

Moro, D., and P. Sckokai. 2000. "Heterogeneous Preferences in Household Food Consumption in Italy." *European Review of Agricultural Economics* 27: 305–323. doi:10.1093/erae/27.3.305.