A NOTE ON THE STOCHASTIC APPROACH
TO INDEX NUMBERS

by

E. Antony Selvanathan

Department of Economics
The University of Western Australia

Discussion Paper 89.07
(A revised version of 88.15)
May 1989

ISSN 0811-6067
ISBN 0-864-22-676-4

A Note on the Stochastic Approach to Index Numbers

by

E. Antony Selvanathan
Department of Economics
The University of Western Australia
Nedlands, Western Australia 6009
AUSTRALIA

Abstract:

In a recent article Clements and Izan (1987) used the stochastic approach to index number theory to estimate the rate of inflation and its standard error. Selvanathan (1988) extends their approach to the prices of groups of goods and to prices within groups. In this note we apply the within-group results to the U.K. alcohol data. Simulation results show that the estimates are unbiased, but the asymptotic standard errors understate the true sampling variability of the estimates. To overcome this problem we applied the bootstrap technique to obtain alternate standard errors.

KEY WORDS: Simulations, Bootstrap, Alternate Standard Errors, Alcohol.

1. INTRODUCTION

In a recent article Clements and Izan (1987) used the stochastic approach to index number theory to estimate the rate of inflation and its standard error using the prices of all n goods unconditionally. Selvanathan (1988) extended their approach to the prices of groups of goods and to prices within groups conditionally and shows that all the unconditional results can be carried over to the conditional case with n replaced by n_g , where n_g is the number of goods in a group S_g . In this note we present an application of the extensions to the alcoholic beverages group (consisting of beer, wine and spirits) using data for the U.K. We also apply simulation techniques to evaluate the performance of the estimators and their asymptotic standard errors. The results show that the estimators are unbiased, but the asymptotic standard errors understate the true sampling variability of the estimates. We also find that the bootstrap standard errors can be used as alternate standard errors for the estimators.

2. WITHIN GROUP RESULTS

Selvanathan (1988) considers the price log-change of a commodity $i \in S_g$, Dp_{it} ($= \log p_{it} - \log p_{i,t-1}$), in the form

$$Dp_{it} = \alpha_t^g + \beta_i^g + \zeta_{it}^g, \quad i \in S_g, \quad g=1, \dots, G, \quad t=1, \dots, T, \quad (1)$$

where T is the number of observations. The coefficient α_t^g is interpreted as the common trend in prices, while β_i^g is the systematic change in the i^{th} relative price, both within the group S_g . The coefficients β_i^g 's satisfy the constraint $\sum_{i \in S_g} \bar{w}_i' \beta_i^g = 0$, where $\bar{w}_i'$ is the sample mean of $\bar{w}_{it}'$ [$= \frac{1}{2}(w_{it}' + w_{i,t-1}')$] and w_{it}' is the budget share of commodity i within group S_g . We assume that the errors ζ_{it}^g 's are independent over commodities and time, and that their variances are inversely proportional to the

corresponding $\bar{w}_i$. That is, $\text{cov}[\frac{z_{it}^g}{\bar{w}_i}, \frac{z_{js}^g}{\bar{w}_j}] = \eta_{gt}^2 \delta_{ij} \delta_{ts} / \bar{w}_i$, where δ is the Kronecker delta and η_{gt}^2 is a constant with respect to commodities in S_g and can be estimated consistently by $\theta_{gt}^2 / (n_g - 1)$, with θ_{gt}^2 defined below. Selvanathan (1988) weights the observations by $(\bar{w}_i')^{1/2} / \theta_{gt}$ and show that the WLS estimators of (1) constrained by $\sum_{i \in S_g} \bar{w}_i' \beta_i^g = 0$ are

$$\alpha_t^{g*} = \sum_{i \in S_g} \bar{w}_i' Dp_{it}, \quad \beta_i^{g*} = \sum_{t=1}^T \phi_t^g (Dp_{it} - \alpha_t^{g*}), \quad (2)$$

where

$$\phi_t^g = (1/\theta_{gt}^2) \sum_{\tau=1}^T (1/\theta_{g\tau}^2) \text{ and } \theta_{gt}^2 = \sum_{i \in S_g} \bar{w}_i' \left[Dp_{it} - \alpha_t^{g*} - \sum_{\tau=1}^T (Dp_{i\tau} - \alpha_\tau^{g*}) / T \right]^2. \quad (3)$$

The above θ_{gt}^2 has the interpretation of a Divisia second moment (see, Barnett and Serletis, forthcoming, for applications of such indices).

The sampling variances of these estimators are

$$\begin{aligned} \text{cov}[\alpha_t^{g*}, \alpha_s^{g*}] &= \theta_{gt}^2 \delta_{st} / (n_g - 1), \\ \text{cov}[\beta_i^{g*}, \beta_j^{g*}] &= \left[(\delta_{ij} / \bar{w}_i') - 1 \right] / (n_g - 1) \sum_{t=1}^T (1/\theta_{gt}^2), \quad i, j \in S_g \end{aligned} \quad (4)$$

The mean of α_t^{g*} over all T periods and its sampling variance are

$$\bar{\alpha}^{g*} = \sum_{t=1}^T \alpha_t^{g*} / T, \quad \text{var } \bar{\alpha}^{g*} = \sum_{t=1}^T \theta_{gt}^2 / (n_g - 1) T^2. \quad (5)$$

3. AN APPLICATION

We now apply the results reported in the last section to the prices of the alcoholic beverages (namely, beer, wine and spirits) in the U.K., 1955-1975. We also analyse the alcohol quantity data by using the same methodology. This simply involves replacing the price log-changes Dp_{it} with the corresponding quantity changes Dq_{it} and appropriate re-interpretations

of θ_{gt}^2 , α_t^g and β_i^g .

Columns 2-3 and 8-9 of Table 1 present the estimates of common trends in prices and quantities of beer, wine and spirits and their asymptotic standard errors, computed using equations (2)-(4). The underlying data are from Selvanathan (1987). As can be seen, most of the estimates are significantly different from zero. In the last row we give the means and their asymptotic standard errors (ASE), computed using (5). On average, alcohol prices and consumption increased by 5.1 and 3.4 percent per annum, respectively. The standard errors of these averages are .4 and .5, respectively, indicating that they are significantly different from zero.

Columns 2-3 of Table 2 give the estimates of β_i^g and their asymptotic standard errors. All except one of these estimates are significantly different from zero. The entries in rows 1-3 of column 2 show that the estimated relative price of beer increased by .49% per annum while relative price of wine declined by 1.53% per annum. These two changes are significantly different from zero. On the other hand the relative price of spirits fell by .28%, but it is insignificantly different from zero. The estimates for quantities given in rows 4-6 of column 2 can be interpreted in the same manner.

4. HOW RELIABLE ARE THE ASYMPTOTIC STANDARD ERRORS?

As we have used asymptotic theory to derive the standard errors (see above equation (2)), it is proper to ask, how does this theory perform with this data set where there are only 20 observations? In this section we analyse this issue by using Monte-Carlo simulation experiments.

Briefly, the simulation procedure works as follows. As $\theta_{gt}^2/(n_g-1)$ is an asymptotically unbiased estimator of η_{gt}^2 , it follows that $\text{var } \zeta_{it}^g = \theta_{gt}^2/(n_g-1)\bar{w}_i$, $i=1,2,3$; $t=1,\dots,20$. We generate pseudo normal random variates $\hat{\zeta}_{it}^g$ ($i=1,2,3$; $t=1,\dots,20$) each with zero mean and

Table 1. Data-Based Estimates and Monte-Carlo Simulation Results for Common Trends in Prices and Quantities for Alcohol: United Kingdom, 1955-1975.

Year (1)	Prices, a_t^g						Quantities, a_t^q					
	Data-Based Values		Simulated Values, 500 Trials				Data-Based Values		Simulated Values, 500 Trials			
	Point Estimate (2)	ASE (3)	Mean (4)	RMSE (5)	RMSASE (6)	RMSASE/RMSE (7)	Point Estimate (8)	ASE (9)	Mean (10)	RMSE (11)	RMSASE (12)	RMSASE/RMSE (13)
1956	.82	.71	.81	.74	.79	1.07	2.89	.90	2.86	.91	.99	1.09
1957	2.77	.65	2.77	.65	.74	1.13	1.33	.63	1.31	.65	.60	1.23
1958	1.00	.84	.99	.82	.92	1.12	-.99	.65	-.95	.62	.60	1.29
1959	-4.57	2.49	-4.51	2.52	2.30	.91	5.52	.19	5.51	.20	.55	2.84
1960	-2.14	1.88	-2.07	1.84	1.77	.96	5.95	1.41	5.92	1.55	1.41	.91
1961	5.55	1.64	5.51	1.65	1.55	.94	4.77	.99	4.80	.99	1.09	1.11
1962	6.55	2.42	6.57	2.43	2.37	.97	.00	.74	.05	.70	.65	1.22
1963	2.93	4.16	2.65	4.09	3.95	.97	2.75	1.55	2.77	1.61	1.63	1.01
1964	5.30	.64	5.39	.64	.72	1.12	5.95	.93	5.94	.92	1.02	1.11
1965	9.71	.44	9.77	.47	.57	1.20	-2.72	3.59	-2.66	3.56	3.41	.95
1966	6.03	.43	6.02	.43	.54	1.25	1.58	.74	1.57	.74	.88	1.19
1967	4.15	.57	4.15	.55	.66	1.19	2.02	1.49	2.15	1.49	1.54	1.04
1968	3.85	1.09	3.87	1.10	1.08	.98	3.24	1.15	3.23	1.08	1.24	1.14
1969	7.95	1.31	7.91	1.27	1.29	1.01	-.25	4.22	-.55	4.35	4.15	.95
1970	7.10	1.56	7.21	1.60	1.51	.94	5.13	3.09	4.82	3.11	3.00	.95
1971	6.14	1.01	6.15	.95	.99	1.04	5.84	1.81	5.87	1.85	1.74	.94
1972	3.88	1.27	3.90	1.25	1.22	.97	6.95	2.79	7.12	2.86	2.72	.95
1973	2.90	.27	2.88	.29	.45	1.57	12.31	5.17	12.42	4.83	4.94	1.02
1974	8.74	2.54	8.72	2.39	2.52	1.05	4.97	1.69	5.11	1.72	1.68	.97
1975	22.90	1.00	22.98	1.02	1.04	1.02	-.17	4.24	.04	4.25	4.05	.95
Mean	5.08	.37	5.08	.38	.36	.94	3.35	.53	3.37	.54	.52	.95

NOTE: The year 1956, for example, refers to the transition from 1955 to 1956. All entries in columns 2-6 and 8-12 are to be divided by 100.

Table 2. Data-Based Estimates and Simulation Results for Relative Prices and Quantity Changes for Alcoholic Beverages: United Kingdom, 1955-1975.

Beverage (1)	Data-Based Values		Monte-Carlo Simulations				Bootstrap Outer-Loop Simulations				Bootstrap Inner-Loop Simulations			
	Point Estimate (2)	ASE (3)	Mean (4)	RMSE (5)	RMSASE (6)	(6)/(5) (7)	Mean (8)	RMSE (9)	RMSASE (10)	(10)/(9) (11)	Mean (12)	RMSE (13)	RMSASE (14)	(13)/(9) (15)
Prices, β_1^g														
1. Beer	.49	.13	.49	.25	.09	.36	.57	.33	.14	.44	.58	.32	.15	.97
2. Wine	-1.53	.40	-1.52	.72	.27	.38	-1.47	.55	.43	.79	-1.40	.72	.46	1.31
3. Spirits	-.28	.25	-.29	.50	.17	.35	-.49	.68	.28	.40	-.53	.68	.29	1.00
Quantities, β_1^q														
4. Beer	-1.48	.13	-1.48	.33	.13	.38	-1.41	.52	.16	.31	-1.43	.51	.19	.98
5. Wine	3.62	.39	3.66	1.04	.38	.37	3.80	.95	.49	.51	3.88	1.08	.57	1.14
6. Spirits	1.37	.25	1.34	.63	.24	.38	1.13	.91	.31	.34	1.12	.86	.36	.95

NOTE: All entries in columns 2-6, 8-10 and 12-14 are to be divided by 100.

variance as above with $n_g = 3$. These variances are computed using the data-based estimates of θ_{gt}^2 and the $\bar{w}_i$. The estimates of θ_{gt}^2 are obtained using equation (3) and the estimates of α_t^g presented in column 2 of Table 1. We use these generated values of $\hat{\zeta}_{it}^g$ together with the data-based estimates α_t^{g*} and β_i^{g*} given in column 2 of Tables 1 and 2 to obtain generated values of $Dp_{it} (= \alpha_t^{g*} + \beta_i^{g*} + \hat{\zeta}_{it}^g)$. We repeat this procedure 500 times to obtain 500 different sets of price data. We then estimate model (1) 500 times to obtain 500 new sets of estimates of α_t^g , β_i^g ($i=1,2,3$; $t=1,\dots,20$) and their variances using equations (2)-(4). We then compare the means over the 500 trials of the new estimates with the original estimates which were used to generate the data. If the asymptotic procedure is satisfactory, then the mean of the new estimates should be quite close to the original estimates. We use exactly the same procedure for quantities.

Columns 4-7 and 10-13 of Table 1 present summary measures of the 500 trials. As can be seen, the means of the 500 trials are very close to their true values, indicating that the estimates are unbiased. In columns 5-6 and 11-12 of the table we give the root-mean-squared errors (RMSEs) computed around the true parameter values (i.e., the data-based values) and the root-mean-squared asymptotic standard errors (RMSASEs) of the estimates. The RMSASE of each parameter is obtained by summing over the 500 trials the asymptotic variance, dividing by 500 and then taking the square root. The ratios RMSASE/RMSE, given in columns 7 and 13 of the table, mainly fluctuate around unity. This indicates that on average the asymptotic standard errors give a realistic picture of the true sampling variability of the estimates.

We give in columns 4-7 of Table 2 the simulation results for the coefficients β_i^g . Column 4 gives the means over the 500 trials. As can be seen, the means are very close to the data-based estimates. However, all the RMSASE/RMSE ratios given in column 7 are very low, indicating that the RMSASEs severely understate the true sampling variability of the estimates. In the next two sections we attempt to resolve this problem.

5. BOOTSTRAP ESTIMATES

As the asymptotic standard errors of the β_i^{g*} 's are not satisfactory, the question arises whether alternatives are available which have superior qualities. In this section we use Efron's (1979) bootstrap technique to obtain alternative standard errors for the β_i^{g*} 's. In Section 6 we show that these standard errors are more satisfactory.

First we describe how to bootstrap model (1). We assume that the disturbance terms ζ_{it}^g have zero means and are statistically independent and identically distributed in time. We generate a new set of data for the dependent variables Dp_{it} ($i=1,2,3$; $t=1,\dots,20$) in the following steps. Step 1: Assign mass $1/20$ to each residual vector $\hat{\zeta}_t^g$, $t=1,\dots,20$, where $\hat{\zeta}_t^g = [\hat{\zeta}_{1t}^g \ \hat{\zeta}_{2t}^g \ \hat{\zeta}_{3t}^g]'$ and $\hat{\zeta}_{it}^g = Dp_{it} - \alpha_t^{g*} - \beta_i^{g*}$ for $i=1,2,3$. We denote bootstrap residuals, values of the dependent variables and estimators by the superscript '**'. We draw 20 uniform random numbers with replacement in the range 1 to 20. Let the set of drawn random numbers be $\{t_1, \dots, t_{20}\}$. Step 2: Define the bootstrap error vectors as $[\zeta_1^{g**} \ \dots \ \zeta_{20}^{g**}] = [\hat{\zeta}_{t_1}^g \ \dots \ \hat{\zeta}_{t_{20}}^g]$. Step 3: Generate data for the dependent variables as $Dp_{it}^{**} = \alpha_t^{g*} + \beta_i^{g*} + \zeta_{it}^{g**}$, $i=1,2,3$; $t=1,\dots,20$, where ζ_{it}^{g**} is the i^{th} element in the bootstrap error vector ζ_t^{g**} .

Using the generated values of Dp_{it} , we re-estimate model (1) to obtain the bootstrap parameter estimates α_t^{g**} and β_i^{g**} . These estimates are obtained by applying equations (2) - (4) to the generated data. We repeat this procedure 500 times to obtain 500 sets of bootstrap estimates of α_t^g and β_i^g . Selvanathan (1988) shows that it is always true that $\alpha_t^{g**} = \alpha_t^{g*}$, whereas $\beta_i^{g**} = \beta_i^{g*}$ only under a special case. Consequently, it makes sense to consider only the bootstrap estimates of β_i^g . Columns 8-11 of Table 2 summarize the bootstrap simulations (outer-loop) results. The bootstrap standard errors of the β_i^g 's, computed as the RMSEs, are reported in column 9

of the table. The RMSASE/RMSE ratios given in column 11 are nearly the same as those of column 7, except for β_i^{g*} for wine prices. We shall come back to these results in the next section.

6. THE QUALITY OF THE BOOTSTRAP STANDARD ERRORS

The Monte-Carlo simulation results with normal errors show that the asymptotic standard errors understate the true sampling variability of the estimates of β_i^g (see column 7 of Table 2). The bootstrap results also indicate that the asymptotic standard errors are still smaller than the true variability, but there is some improvement (see column 11 of Table 2). This finding agrees with the results of Freedman and Peters (1984; columns 4 and 5 of their Table 1). In order to analyse the quality of the bootstrap standard errors, we use the inner-outerloop simulation procedure described in Freedman and Peters (1984) which involves a nested iteration.

For each of the 500 passes in the outer loop, we perform 500 passes in the inner loop and compute the means, the RMSEs and the RMSASEs of the β_i^g 's. In such case we have 500 sets of means, RMSEs and RMSASEs. Columns 12-15 of Table 2 summarizes the results. Column 12 presents the mean of the 500 means. Column 13 gives the RMS of the 500 RMSEs (obtained around the parameter values of the corresponding pass through the outer loop). For full details of the procedure, see Selvanathan (1987). Column 14 of the table presents the mean of the 500 RMSASEs obtained for each pass of the outer loop. The RMSASE for each pass of the outer loop is obtained by summing the variances obtained using (4) for each of the 500 passes of the inner loop, dividing by 500 and taking the square root.

The entries in column 12 are close to those in column 8, again indicating unbiasedness. As the entries in column 13 are reasonably close to those in column 9, we conclude that the bootstrap standard errors give a

fair picture of the true sampling variability of the estimates. A comparison of column 14 with 3 shows that the data-based asymptotic standard errors are of the same order of magnitude as the bootstrap RMSASE. These asymptotic standard errors are far below the RMSEs given in columns 9 and 13 of the table. These results confirm those of Freedman and Peters (1984), Theil (1987) and Theil, Rosalsky and Finke (1984).

ACKNOWLEDGMENTS

The author would like to thank Professor Kenneth W. Clements and, the Associate Editor and two anonymous referees of this Journal for their helpful comments.

REFERENCES

- Barnett, W.A., and Serletis, A., (forthcoming), "A Dispersion-Dependency Diagnostic Test for Aggregation Error," Journal of Econometrics.
- Clements, K.W., and Izan, H.Y., (1987), "The Measurement of Inflation: A Stochastic Approach," Journal of Business and Economic Statistics, 5, 339-50.
- Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7, 1-26.
- Freedman, D.A., and Peters, S.C., (1984), "Bootstrapping a Regression Equation: Some Empirical Results," Journal of the American Statistical Association, 79, 97-106.
- Selvanathan, E. A. (1987), Explorations in Consumer Demand, unpublished Ph.D. thesis, Department of Economics, Murdoch University.
- _____, (1988), "Stochastic Approach to Index Numbers: Extensions," Discussion Paper No: 88.15, Department of Economics, The University of Western Australia.
- Theil, H., (1987), "The Econometrics of Demand Systems," Chapter 3 in H. Theil and K.W. Clements, Applied Demand Analysis: Results from System-Wide Approaches. Cambridge, Mass.: Ballinger Publishing Company.
- _____, Rosalsky, M.C., and Finke, R., (1984), "A Comparison of Normal and Discrete Bootstraps for Standard Errors in Equation Systems," Statistics and Probability Letters, 2, 175-80; erratum, 2:255.