INDEX NUMBER THEORY AND MEASUREMENT ECONOMICS

By W.E. Diewert. February 2015.

CHAPTER 9: Two Stage Aggregation and Homogeneous Weak Separability

1. Introduction

Most statistical agencies use the Laspeyres formula to aggregate prices in two stages. At the first stage of aggregation, the Laspeyres formula is used to aggregate components of the overall index (e.g., food, clothing, services, etc.) and then at the second stage of aggregation, these component subindexes are further combined into the overall index. The following question then naturally arises: does the index computed in two stages coincide with the index computed in a single stage? We will address this question in section 3 below.[1]

However, before answering the above question, we will first ask a more fundamental question: namely, what conditions on consumer's preferences justify a two stage aggregation procedure? We address this section in section 2 below.

2. The Assumption of Homogeneous Separability of Preferences

It turns out that the assumption of homogeneous separability is one of the simplest ways of justifying aggregation over commodities[2] in such a way that the commodity aggregate has an aggregate price that behaves just as if it were a "true" microeconomic price. Essentially, this assumption allows us to apply microeconomic theory to aggregates!

The assumption of *homogeneous separability* works in the following manner. Suppose that a household or consumer has preferences over two groups of commodities where there are N_1 commodities $q^1 \equiv [q_1^1,...,q_{N_1}^1]$ in the first group and N_2 commodities $q^2 \equiv [q_1^2,...,q_{N_2}^2]$ in the second group. Let the consumer's preferences over all of the commodities be represented by the nonnegative, continuous, quasiconcave and increasing utility function $U(q^1,q^2)$ for $q^1 \geq 0_{N_1}$ and $q^2 \geq 0_{N_2}$. The consumer's preferences are homogeneously separable in the two groups if there exists a "macro" utility function $F(Q^1,Q^2)$ that is nonnegative, continuous, quasiconcave and increasing in its two

[1] Much of the material in section 3 is adapted from Diewert (1978) and Alterman, Diewert and Feenstra (1999). See also Balk (1996) for a discussion of alternative definitions for the two stage aggregation concept and references to the literature on this topic.

[2] The other simple way is through the use of Hicks' Aggregation Theorem; i.e., if the prices in a group of commodities vary in strict proportion over time, then the factor of proportionality can be taken as the price of the group and the deflated group expenditures will obey the usual properties of a microeconomic commodity. "Thus we have demonstrated mathematically the very important principle, used extensively in the text, that if the prices of a group of goods change in the same proportion, that group of goods behaves just as if it were a single commodity." J.R. Hicks (1946; 312-313).

nonnegative arguments, Q^1 and Q^2, and there exist two linearly homogeneous, quasiconcave and nondecreasing "micro" utility functions, $f^1(q^1)$ and $f^2(q^2)$, such that

(1) $U(q^1,q^2) = F[f^1(q^1),f^2(q^2)]$.

Let $p^1 \gg 0_{N_1}$ and $p^2 \gg 0_{N_2}$ be two positive vectors of commodity prices facing the consumer in a particular period. Since the micro utility functions f^1 and f^2 are linearly homogeneous, we know that their corresponding cost functions have the following form:

(2) $\min_q \{p^{1T}q^1 : f^1(q^1) \geq Q^1\} = c^1(p^1)Q^1$;
(3) $\min_q \{p^{2T}q^2 : f^2(q^2) \geq Q^2\} = c^2(p^2)Q^2$

where $c^1(p^1)$ is the *unit cost function* that is dual to f^1 and $c^2(p^2)$ is the *unit cost function* that is dual to f^2.

Now consider the cost minimization problem for a consumer that has separable preferences of the form defined by (1) above: for $p^1 \gg 0_{N_1}$ and $p^2 \gg 0_{N_2}$ and $u > 0$, we define the minimum cost of achieving the utility level u as follows:

(4) $C(u,p^1,p^2) \equiv \min_{q\text{'s}} \{ p^{1T}q^1 + p^{2T}q^2 : F[f^1(q^1),f^2(q^2)] \geq u\}$
$= \min_{q\text{'s},\, Q\text{'s}} \{ p^{1T}q^1 + p^{2T}q^2 : F[Q^1,Q^2] \geq u,\ Q^1 \equiv f^1(q^1),\ Q^2 \equiv f^2(q^2)\}$
where we added two extra constraints to the cost minimization problem by defining $Q^1 \equiv f^1(q^1)$ and $Q^2 \equiv f^2(q^2)$
$= \min_{Q\text{'s}} \{c^1(p^1)Q^1 + c^2(p^2)Q^2 : F[Q^1,Q^2] \geq u\}$ using (2) and (3)
$\equiv C^*[u, c^1(p^1),c^2(p^2)]$

where $C^*(u,P^1,P^2)$ is the *macro cost function* that is dual to the macro utility function $F(Q^1,Q^2)$. Looking at (4), it can be seen that *the unit cost functions*, $c^1(p^1)$ *and* $c^2(p^2)$, *act like microeconomic prices for the quantity aggregates,* $Q^1 = f^1(q^1)$ *and* $Q^2 = f^2(q^2)$. This is what is powerful about the assumption of homogeneous separability.[3] For certain functional forms for $c^m(p^m)$ or $f^m(q^m)$ for m = 1,2, listed in Pollak (1983) or Diewert (1976), we can use exact index number formulae to calculate these price and quantity aggregates using the assumption of optimizing behavior along with observed price and quantity data for two periods.

Problem

1. Repeat the analysis around equations (1) to (4) assuming 3 homogeneous aggregates instead of 2.

[3] This model of aggregation dates back to Leontief (1947) but it is most clearly explained by Shephard (1953; 61-71) (1970; 145-146). See also Arrow (1974) (in the producer context), Samuelson and Swamy (1974), Diewert (1980; 438-42) and Blackorby, Primont and Russell (1978) for additional references and exposition.

3. Two Stage Aggregation of Index Number Formulae

We return to the question that was raised in the introduction; namely, does a Laspeyres index computed in two stages coincide with a Laspeyres index computed in a single stage?

Suppose that the price and quantity data for period t, p^t and q^t, can be written in terms of M subvectors as follows:

(5) $p^t = [p^{t1},p^{t2},...,p^{tM}]$; $q^t = [q^{t1},q^{t2},...,q^{tM}]$; t = 0,1

where the dimensionality of the subvectors p^{tm} and q^{tm} is N_m for m = 1,2,…,M with the sum of the dimensions N_m equal to N. These subvectors correspond to the price and quantity data for subcomponents of the consumer price index for period t. Now construct subindices for each of these components going from period 0 to 1. For the base period, set the price for each of these subcomponents, say $P_m^{\,0}$ for m = 1,2,…M, equal to 1 and set the corresponding base period subcomponent quantities, say $Q_m^{\,0}$ for m = 1,2,…,M, equal to the base period value of consumption for that subcomponent for m = 1,2,…,M:

$$(6)\ \ P_m^0 \equiv 1;\ Q_m^0 \equiv \sum_{i=1}^{N_m} p_i^{0m} q_i^{0m} \ \text{ for } m = 1,2,...,M.$$

Now use the Laspeyres formula in order to construct a period 1 price for each subcomponent, say $P_m^{\,1}$ for m = 1,2,…,M, of the consumer price index. Since the dimensionality of the subcomponent vectors, p^{tm} and q^{tm}, differs from the dimensionality of the complete period t vectors of prices and quantities, p^t and q^t, it is necessary to use different symbols for these subcomponent Laspeyres indexes, say $P_L^{\,m}$ for m = 1,2,…M. Thus the period 1 subcomponent prices are defined as follows:

$$(7)\ \ P_m^1 \equiv P_L^m(p^{0m}, p^{1m}, q^{0m}, q^{1m}) \equiv \frac{\sum_{i=1}^{N_m} p_i^{1m} q_i^{0m}}{\sum_{i=1}^{N_m} p_i^{0m} q_i^{0m}} \ \ \textit{for } m = 1,2,...,M.$$

Once the period 1 prices for the M subindexes have been defined by (7), then corresponding subcomponent period 1 quantities $Q_m^{\,1}$ for m = 1,2,…,M can be defined by deflating the period 1 subcomponent values $\sum_{i=1}^{Nm} p_i^{\,1m} q_i^{\,1m}$ by the prices $P_m^{\,1}$:

$$(8)\ \ Q_m^1 \equiv \frac{\sum_{i=1}^{N_m} p_i^{1m} q_i^{1m}}{P_m^1} \ \ \textit{for } m = 1,2,...,M.$$

Now define subcomponent price and quantity vectors for each period t = 0,1 using equations (6) to (8) above. Thus define the period 0 and 1 subcomponent price vectors P^0 and P^1 as follows:

(9) $P^0 = (P_1^0, P_2^0, ..., P_M^0) \equiv 1_M$; $P^1 = (P_1^1, P_2^1, ..., P_M^1)$

where 1_M denotes a vector of ones of dimension M and the components of P^1 are defined by (7). The period 0 and 1 subcomponent quantity vectors Q^0 and Q^1 are defined as follows:

(10) $Q^0 \equiv [Q_1^0, Q_2^0, \ldots, Q_M^0]$; $Q^1 \equiv [Q_1^1, Q_2^1, \ldots, Q_M^1]$

where the components of Q^0 are defined in (6) and the components of Q^1 are defined by (8). The price and quantity vectors in (9) and (10) represent the results of the first stage aggregation. Now use these vectors as inputs into the second stage aggregation problem; i.e., apply the Laspeyres price index formula using the information in (9) and (10) as inputs into the index number formula. Since the price and quantity vectors that are inputs into this second stage aggregation problem have dimension M instead of the single stage formula which utilized vectors of dimension N, a different symbol is required for the new Laspeyres index which we choose to be P_L^*. Thus the Laspeyres price index computed in two stages is denoted as $P_L^*(P^0,P^1,Q^0,Q^1)$. Now ask whether this two stage Laspeyres index equals the corresponding single stage index P_L; i.e., ask whether

(11) $P_L^*(P^0,P^1,Q^0,Q^1) = P_L(p^0,p^1,q^0,q^1)$.

If the Laspeyres formula is used at each stage of each aggregation, the answer to the above question is yes: straightforward calculations show that the Laspeyres index calculated in two stages equals the Laspeyres index calculated in one stage.

Problems

2. Verify that the Laspeyres index calculated in two stages equals the Laspeyres index calculated in one stage.

3. Is it true that the Paasche index calculated in two stages equals the Paasche index calculated in one stage?

Now suppose that the Fisher or Törnqvist formula is used at each stage of the aggregation; i.e., in equations (7), suppose that the Laspeyres formula $P_L^m(p^{0m},p^{1m},q^{0m},q^{1m})$ is replaced by the Fisher formula $P_F^m(p^{0m},p^{1m},q^{0m},q^{1m})$ (or by the Törnqvist formula $P_T^m(p^{0m},p^{1m},q^{0m},q^{1m})$) and in equation (11), $P_L^*(P^0,P^1,Q^0,Q^1)$ is replaced by P_F^* (or by P_T^*) and $P_L(p^0,p^1,q^0,q^1)$ is replaced by P_F (or by P_T). Then is it the case that counterparts are obtained to the two stage aggregation result for the Laspeyres formula, (11)? The answer is no; it can be shown that, in general,

(12) $P_F^*(P^0,P^1,Q^0,Q^1) \neq P_F(p^0,p^1,q^0,q^1)$ and $P_T^*(P^0,P^1,Q^0,Q^1) \neq P_T(p^0,p^1,q^0,q^1)$.

Similarly, it can be shown that the quadratic mean of order r index number formula P^r defined and the implicit quadratic mean of order r index number formula P^{r*} defined in chapter 5 are also not consistent in aggregation.

However, even though the Fisher and Törnqvist formulae are not *exactly* consistent in aggregation, it can be shown that these formulae are *approximately* consistent in aggregation. More specifically, it can be shown that the two stage Fisher formula P_F^* and the single stage Fisher formula P_F in (12), both regarded as functions of the 4N variables in the vectors p^0,p^1,q^0,q^1, approximate each other to the second order around a point where the two price vectors are equal (so that $p^0 = p^1$) and where the two quantity vectors are equal (so that $q^0 = q^1$) and a similar result holds for the two stage and single stage Törnqvist indexes in (12).[4] As was seen in chapter 5, the single stage Fisher and Törnqvist indexes have a similar approximation property so all four indexes in (12) approximate each other to the second order around an equal (or proportional) price and quantity point. Thus for normal time series data, single stage and two stage Fisher and Törnqvist indexes will usually be numerically very close.[5] This result will be illustrated in chapter 11 for an artificial data set.

Similar approximate consistency in aggregation results (to the results for the Fisher and Törnqvist formulae explained in the previous paragraph) can be derived for the *quadratic mean of order r indexes*, P^r, and for the implicit quadratic mean of order r indexes, P^{r*}; see Diewert (1978; 889). However, the results of Hill (2006) again imply that *the second order approximation property of the single stage quadratic mean of order r index P^r to its two stage counterpart will break down as r approaches either plus or minus infinity*. To see this, consider a simple example where there are only four commodities in total. Let the first price ratio p_1^1/p_1^0 be equal to the positive number a, let the second two price ratios p_i^1/p_i^0 equal the b and let the last price ratio p_4^1/p_4^1 equal c where we assume a < c and a ≤ b ≤ c. Using the properties of means of order r, the limiting value of the single stage index is:

$$(13)\quad \lim_{r\to+\infty} P^r(p^0,p^1,q^0,q^1) = \lim_{r\to-\infty} P^r(p^0,p^1,q^0,q^1) = \sqrt{\min_i\left(\frac{p_i^1}{p_i^0}\right)\max_i\left(\frac{p_i^1}{p_i^0}\right)} = \sqrt{ac}.$$

Now aggregate commodities 1 and 2 into a subaggregate and commodities 3 and 4 into another subaggregate. Using the properties of means of order r again, it is found that the limiting price index for the first subaggregate is $[ab]^{1/2}$ and the limiting price index for the second subaggregate is $[bc]^{1/2}$. Now apply the second stage of aggregation and use the properties of means of order r once again to conclude that the limiting value of the two

---

[4] See Diewert (1978; 889). In fact, these derivative equalities are still true provided that $p^1 = \lambda p^0$ and $q^1 = \mu q^0$ for any numbers $\lambda > 0$ and $\mu > 0$.

[5] For an empirical comparison of the four indexes, see Diewert (1978; 894-895). For the Canadian consumer data considered there, the chained two stage Fisher in 1971 was 2.3228 and the corresponding chained two stage Törnqvist was 2.3230, the same values as for the corresponding single stage indexes.

stage aggregation using P^r as the index number formula is $[ab^2c]^{1/4}$. Thus the limiting value, as r tends to plus or minus infinity, of the single stage aggregate over the two stage aggregate is $[ac]^{1/2}/[ab^2c]^{1/4} = [ac/b^2]^{1/4}$. Now b can take on any value between a and c and so the ratio of the single stage limiting P^r to its two stage counterpart can take on any value between $[c/a]^{1/4}$ and $[a/c]^{1/4}$. Since c/a is less than 1 and a/c is greater than 1, it can be seen that the ratio of the single stage to the two stage index can be arbitrarily far from 1 as r becomes large in magnitude with an appropriate choice of the numbers a, b and c.

The results in the previous paragraph show that some caution is required in assuming that *all* superlative indexes will be approximately consistent in aggregation. However, for the three most commonly used superlative indexes (the Fisher ideal P_F, the Törnqvist-Theil P_T and the Walsh P_W), the available empirical evidence indicates that these indexes satisfy the consistency in aggregation property to a sufficiently high enough degree of approximation that users will not be unduly troubled by any inconsistencies.[6]

Problem

4. Use the properties of means of order r to establish the results in (13) above.

References

Alterman, W.F., W.E. Diewert and R.C. Feenstra, (1999), *International Trade Price Indexes and Seasonal Commodities*, Bureau of Labor Statistics, Washington D.C.

Arrow, K.J. (1974), "The Measurement of Real Value Added", in *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz,* P. David and M. Reder (eds.), New York: Academic Press.

Balk, B.M. (1996), "Consistency in Aggregation and Stuvel Indices", *The Review of Income and Wealth* 42, 353-363.

Blackorby, C., D. Primont and R.R. Russell (1978), *Duality, Separability and Functional Structure: Theory and Economic Applications*, New York: North-Holland.

Diewert, W.E. (1976), "Exact and Superlative Index Numbers", *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (1978), "Superlative Index Numbers and Consistency in Aggregation", *Econometrica* 46, 883-900.

Diewert, W. E. (1980), "Aggregation Problems in the Measurement of Capital", pp. 433-528 in *The Measurement of Capital*, Dan Usher (ed.), University of Chicago Press, Chicago.

---

[6] See chapter 11 for some additional evidence on this topic.

Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: Clarendon Press.

Hill, R.J. (2006), "Superlative Indexes: Not All of Them are Super", *Journal of Econometrics* 130, 25-43.

Leontief, W. W. (1947), "Introduction to a Theory of the Internal Structure of Functional Relationships", *Econometrica* 15, 361-373.

Pollak, R.A. (1983), "The Theory of the Cost-of-Living Index", pp. 87-161 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada; reprinted as pp. 3-52 in R.A. Pollak, *The Theory of the Cost-of-Living Index*, Oxford: Oxford University Press, 1989; also reprinted as pp. 5-77 in *Price Level Measurement*, W.E. Diewert (ed.), Amsterdam: North-Holland, 1990.

Samuelson, P. A. and S. Swamy (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis", *American Economic Review* 64, 566-593.

Shephard, R. W. (1953), *Cost and Production Functions*, Princeton: Princeton University Press.

Shephard, R. W. (1970), *Theory of Cost and Production Functions*, Princeton: Princeton University Press.