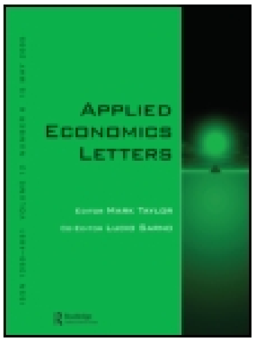

Applied Economics Letters

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

A note on commodity price aggregation bias without separability

Camilo Sarmiento & Richard E. Just

Applied Economics Letters, 2006, **13**, 365–368

A note on commodity price aggregation bias without separability

Camilo Sarmiento[a,*] and Richard E. Just[b]

[a]*Pacific Institute for Research and Evaluation, Calverton, USA*
[b]*University of Maryland, College Park, USA*

This study identifies sources of price aggregation bias when separability restrictions do not apply. It shows that even though the assumption of the generalized composite commodity theorem guarantees aggregate integrability, it does not guarantee consistent price aggregation except in the homothetic translog model.

I. Introduction

The generalized commodity composite theorem (GCT) (Lewbel, 1996) under its assumption implies for a variety of flexible forms that, if a disaggregate demand system satisfies integrability conditions, then so does its form under commodity price aggregation.[1] Davis (2003) generalizes linear tests of the GCT for consistent price aggregation to the nonlinear case.

Contrary to common belief, this paper shows that the assumption of the GCT allows consistent price aggregation in terms of an uncorrelated residual to included regressors only if preferences conform to the homothetic translog model (Silberberg, 1990, p. 411). This homothetic model contains restrictions that are generally binding for composite demand analysis and especially for commodities with low income elasticities.

II. The GCT Assumption

Assume the aggregate commodity demand for group k is

$$X_{kt} = f_k(\mathbf{p}_t, M_t) \tag{1}$$

where the share of commodity group k in total expenditures at time t is X_{kt}; the price vector partitioned by w commodity groups is $\mathbf{p}_t = (\mathbf{p}_{1t}, \ldots, \mathbf{p}_{wt})$; the logarithm of income at time t is M_t; and the price vector of each commodity group is $\mathbf{p}_{it} = (\mathbf{p}_{ilt}, \ldots, \mathbf{p}_{in,t})$ where the logarithm of price j in commodity group i at time t is p_{ijt} and the number of goods in commodity group k is n_k.[2]

Because direct regression with many individual prices requires many observations and is typically intractable, most demand equations are estimated

*Corresponding author. E-mail: csarmiento@pire.org

[1] The GCT applies for a variety of flexible forms including all homothetic utility functions, the almost ideal demand system, and the translog demand system.

[2] Although not shown explicitly, the analytical implications derived in this study for share demand equations apply also for systems that use quantities rather than shares. Similarly, the analytical implications extend to cases where the explanatory variables are not necessarily defined in logarithms. Logarithms of explanatory variables are used to be consistent with Lewbel (1996). However, the conditions of independence derived below for absence of bias would be different in non-logarithmic cases.

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13504850500395290

as a function of linear aggregate indexes of prices,

$$X_{kt} = f_k(R_{1t}, \ldots, R_{wt}, M_t) \tag{2}$$

where the traditional linear price index is $R_{it} = \ln[\sum_{j=1}^{n_i} s_{ijt} \exp(p_{ijt})]$ and s_{ijt} is the share of commodity j in commodity group i at time t. Lewbel (1996) illustrates that the source of omitted information from linearly aggregating prices, i.e., $\mathbf{p}_{1t} - \mathbf{e}_1 R_{1t}, \ldots, \mathbf{p}_{wt} - \mathbf{e}_w R_{wt}$, where $\mathbf{e}_i$ is an $n_i \times 1$ vector of ones, is not constant over time in Equation 2 and, hence, does not satisfy the Hicks–Leontief composite theorem (Hicks, 1936; Leontief, 1936).

As a more plausible alternative, Lewbel (1996) formulates the GCT using the following assumption.

Assumption 1: *The vectors* $\mathbf{p}_{1t} - \mathbf{e}_1 R_{1t}, \ldots, \mathbf{p}_{wt} - \mathbf{e}_w R_{wt}$ *are statistically independent of* $(R_{1t}, \ldots, R_{wt}, M_t)$.

Assumption 1 treats each variable as a stochastic process and has important implications under stationarity. In particular, following Hamilton (1994), stationarity of R_{it} implies $E(R_{it}) = c_i$ where c_i is a group specific constant (p. 54) and $\text{Cov}(R_{it}, \mathbf{p}_{ht} - \mathbf{e}_1 R_{ht}) = 0$ if R_{it} and $\mathbf{p}_{ht} - \mathbf{e}_h R_{ht}$ are uncorrelated across time (pp. 264–66).

III. Price Aggregation Bias under the Assumption of the GCT

The generalized composite theorem (GCT) implies under its assumption for a variety of flexible forms that if a disaggregate demand system satisfies integrability conditions, then so does its form under commodity price aggregation (Lewbel, 1996).[3] This note demonstrates the restrictiveness of this assumption when analysed in terms of price aggregation bias. To do so, the underpinning of the Stone–Weierstrass theorem is used next.

By the Stone–Weierstrass theorem (Simon and Blume, 1994), any continuous function can be represented by a polynomial and, therefore, any nonlinear continuous function contains a subset of the sources of price aggregation bias found in a quadratic polynomial.[4] To demonstrate that Assumption 1 is not sufficient for unbiased estimates of elasticities and consumer surplus under linear price aggregation, thus the quadratic form is used,

$$X_{kt} = \gamma k + \sum_{i=1}^{w} \mathbf{A}_{ki}\mathbf{p}_{it} + b_k M_t + \sum_{i=1}^{w} \mathbf{C}_{ki}\mathbf{p}_{it} M_t + \sum_{j=1}^{w} \mathbf{p}'_{it}\mathbf{B}_{kij}\mathbf{p}_{jt} + c_k M_t^2 \tag{3}$$

where an arbitrary quadratic element of Equation 3, $p_{ijt}p_{hvt}$, is alternatively represented as

$$(p_{ijt} - R_{it})(p_{hvt} - R_{ht}) + R_{it}R_{ht} + R_{ht}(p_{ijt} - R_{it}) + R_{it}(p_{hvt} - R_{ht}) \tag{4}$$

In a regression with commodity group price indexes, the arbitrary quadratic element would be represented by $R_{it}R_{ht}$. The existence of a nonzero correlation between a regressor R_{it} in Equation 2 and $R_{ht}(p_{ijt} - R_{it})$ in Equation 4 when Assumption 1 holds would imply that the price aggregation error is not distributed independently of the price indexes included in Equation 2 and, thus, the projection of the aggregation error on to Equation 2 is generally nonzero regardless of the GCT.[5]

To consider the correlation between R_{it} and $R_{ht}(p_{ijt} - R_{it})$, note that

$$\text{Cov}[R_{it}, R_{ht}(p_{ijt} - R_{it})] = E[R_{it}R_{ht}(p_{ijt} - R_{it})] - E(R_{it})E[R_{ht}(p_{ijt} - R_{it})]$$

and, under Assumption 1,

$$\begin{aligned}\text{Cov}[R_{it}, R_{ht}(p_{ijt} - R_{it})] &= E(R_{it}R_{ht})E(p_{ijt} - R_{it}) - E(R_{it})E(R_{ht})E(p_{ijt} - R_{it}) \\ &= [E(R_{it}R_{ht}) - E(R_{it})E(R_{ht})]E(p_{ijt} - R_{it})\end{aligned} \tag{5}$$

In particular, $\text{Cov}[R_{it}, R_{ht}(p_{ijt} - R_{it})] = 0$ only if in addition to Assumption 1 the stochastic processes of R_{it}, R_{ht} and $(p_{ijt} - R_{it})$ satisfy at least one of two conditions: $E(R_{it}R_{ht}) = E(R_{it})E(R_{ht})$ or $E(p_{ijt} - R_{it}) = 0$.

[3] The GCT applies for a variety of flexible forms including all homothetic utility functions, the almost ideal demand system, and the translog demand system.

[4] As in Equation 2, all variables are defined in logarithms.

[5] Of course, a weaker condition for unbiasedness is where the sum of correlations of R_{it} with $(p_{ijt} - R_{it})(p_{hvt} - R_{ht})$, $R_{ht}(p_{ijt} - R_{it})$, and $R_{it}(p_{hvt} - R_{ht})$ in Equation 4 is zero. Under the assumption of the GCT, the sum the covariances of R_{it} with the stochastic processes of $R_{ht}(p_{ijt} - R_{it})$ and $R_{it}(p_{hvt} - R_{ht})$ is larger than the case in which individual stochastic processes are considered if $E(p_{ijt}) > E(R_{it})$, $E(p_{hvt}) > E(R_{ht})$, and $\text{Cov}(R_{it}, R_{ht}) > 0$; or if $E(p_{ijt}) > E(R_{it})$, $E(p_{hvt}) < E(R_{ht})$, and $\text{Cov}(R_{it}, R_{ht}) < 0$. Correlations could partially cancel one another if $E(p_{ijt}) < E(R_{it})$, $E(p_{hvt}) > E(R_{ht})$, and $\text{Cov}(R_{it}, R_{ht}) > 0$; or if $E(p_{ijt}) > E(R_{it})$, $E(p_{hvt}) < E(R_{ht})$, and $\text{Cov}(R_{it}, R_{ht}) > 0$. However, no plausible underlying force is apparent that would cause these correlations to cancel one another except by chance. Thus, such possibilities are ignored for the analytical purposes of this study.

With respect to the first condition, $E(R_{it}R_{ht}) \neq E(R_{it})E(R_{ht})$ unless R_{it} and R_{ht} are uncorrelated stochastic processes, which if $i=h$ requires that $\mathrm{Var}(R_{it})=0$, i.e., that R_{it} is a degenerate stochastic process. Thus, without degeneracy, the condition that R_{it} and R_{ht} are uncorrelated stochastic processes in addition to Assumption 1 eliminates some, but not all, of the sources of bias in a quadratic form. That is, under Assumption 1, $\mathrm{Cov}[R_{ht}, R_{ht}(p_{ijt}-R_{it})]=0$ only if $\mathrm{Var}(R_{ht})=0$ or $E(p_{ijt}-R_{it})=0$.

With respect to the second condition, note that $E(p_{ijt}-R_{it}) \neq 0$ unless $(p_{ijt}-R_{it})$ is a zero mean stationary process.[6] This condition for all i and j together with the independence in Assumption 1 implies that

$$\begin{aligned} &E[f_k(R_{1t},\ldots,R_{wt},M_t)R_{ht}(p_{ijt}-R_{it})] \\ &\quad = E[f_k(R_{1t},\ldots,R_{wt},M_t)R_{ht}]E(p_{ijt}-R_{it}) = 0 \end{aligned}$$

and, thus,

$$\begin{aligned} &\mathrm{Cov}[f_k(R_{1t},\ldots,R_{wt},M_t),(p_{ijt}-R_{it})(p_{hvt}-R_{ht}) \\ &\quad + R_{ht}(p_{ijt}-R_{it}) + R_{it}(p_{hvt}-R_{ht})] = 0 \end{aligned}$$

which verifies unbiasedness of Equation 2 if Equation 1 is quadratic. Therefore, if Assumption 1 holds and $(p_{ijt}-R_{it})$ is a zero mean stationary process, then the projection of the aggregation error on to Equation 2 has a zero mean when Equation 1 has a quadratic form.

Integrability conditions do not eliminate the bias noted in Equation 5. For example, if in Equation 3 $A_{ki}=(a_{k1},\ldots,a_{kw})$ and B_{kij} is a diagonal matrix, $B_{kji}=\mathrm{diag}(b_{kij1},\ldots,b_{kij}n_i)$, then Equation 5 is equal to zero only if $[\Sigma_i a_{ki}]b_{kijh}=0$. Integrability conditions in the almost ideal model (Deaton and Muellbauer, 1980), which is a special case of Equation 3, do not imply that either $[\Sigma_i a_{ki}]=0$ or $b_{kijh}=0$.

For a more general non-quadratic functional form, the condition $E(p_{ijt}-R_{it})=0$ for all i and j in addition to Assumption 1 is not sufficient for unbiasedness. To illustrate this point, a polynomial of order three is assumed for Equation 1. From the cubic polynomial, consider the term $p_{ijt}^2 p_{hvt}$, which is alternatively represented as[7]

$$\begin{aligned} &R_{it}(p_{ijt}-R_{it})(p_{hvt}-R_{ht}) + R_{it}^2 R_{ht} + R_{it}R_{ht}(p_{ijt}-R_{it}) \\ &\quad + R_{it}^2(p_{hvt}-R_{ht}) + (p_{ijt}-R_{it})^2(p_{hvt}-R_{ht}) \\ &\quad + R_{it}R_{ht}(p_{ijt}-R_{it}) + R_{ht}(p_{ijt}-R_{it})^2 \\ &\quad + R_{it}(p_{ijt}-R_{it})(p_{hvt}-R_{ht}) \qquad (6) \end{aligned}$$

where $R_{ht}(p_{ijt}-R_{it})^2$, which is a subset of the aggregation error in Equation 6, can be omitted without biasing elasticity estimates in Equation 2 only if $\mathrm{Cov}[R_{it}, R_{ht}(p_{ijt}-R_{it})^2]=0$. But under Assumption 1,

$$\begin{aligned} &\mathrm{Cov}[R_{it}, R_{ht}(p_{ijt}-R_{it})^2] \\ &\quad = E[R_{it}R_{ht}(p_{ijt}-R_{it})^2] - E(R_{it})E[R_{ht}(p_{ijt}-R_{it})^2] \\ &\quad = [E(R_{it}R_{ht}) - E(R_{it})E(R_{ht})]E[(p_{ijt}-R_{it})^2] \end{aligned}$$

Hence, the bias has mean zero only if, in addition to Assumption 1, R_{it} and R_{ht} are uncorrelated stochastic processes or $(p_{ijt}-R_{it})^2$ is a zero mean stationary process. The latter implies that the stochastic processes of p_{ijt} and R_{it} are perfectly collinear.

Based on these results, three properties that contribute to eliminating price aggregation bias in Equation 2 are:

(a) Stochastic independence of R_{it} and R_{ht} for $i \neq h$;[8]
(b) Zero mean stationarity of $p_{ijt}-R_{it}$; and
(c) Assumption 1.

In empirical work, properties (a) and (b) are not likely to hold. For example, stochastic independence of R_{it} and R_{ht} for $i \neq h$ in Property (a) implies that every economic shock affects only one sector (commodity group) in the economy. This condition is intuitively unrealistic and not surprisingly rejected by actual data (see Table 1). For example, using commodity groups as classified by the National Income and Product Accounts (NIPA) and data from 1954 to 1994, the test in Table 1 rejects zero covariance between R_{it} and R_{ht}. Moreover, zero mean stationarity of $p_{ijt}-R_{it}$ for all i and j in Property (b) is also in conflict with most economic time series data. For example, for all commodity groups as classified by the National Income and Product Accounts (NIPA) and data from 1954 to 1994, Lewbel (1996, p. 235) finds that $(p_{ijt}-R_{it})$ is a nonstationary process for all i and j. Lewbel (1996, p. 235), however, finds empirical evidence in favor of (c).

Overall, from the Stone–Weierstrass theorem and from biases present under any second- or higher-order term of a polynomial representing Marshallian demands, Assumption 1 without Properties (a) and (b) is sufficient for unbiased estimation under linear price aggregation only if the structure of preferences

[6] If $p_{ijt}-R_{it}$ is a zero mean stationary process, then $E(p_{ijt}-R_{it})=0$ for all t (see Hamilton, 1994, p. 53).
[7] Equation 6 is obtained by multiplying Equation 4 by p_{ijt}, and then expressing p_{ijt} as a deviation from an aggregate price index, e.g., $p_{ijt}=(p_{ijt}-R_{it})+R_{it}$.
[8] For $i=h$, the assumption that R_{it} is a degenerate stochastic process, $\mathrm{Var}(R_{it})=0$, clearly conflicts with the nature of most economic data. Additionally, R_{it} is useless for evaluating price elasticities without variability across time.

Table 1. The *t*-statistics from regression of R_{it} on R_{ht}

Commodity groups[a]	Raw index
a_{12}[b]	−2.12**
a_{13}	−0.14
a_{14}	1.80*
a_{15}	1.69*
a_{16}	−1.78*
a_{23}	3.92***
a_{24}	−1.96*
a_{25}	0.68
a_{26}	21.37***
a_{34}	−1.05
a_{35}	4.38***
a_{36}	4.25***
a_{45}	−0.66
a_{46}	−1.77*
a_{56}	1.39

Notes: [a] Each regression contains an unreported constant term.
[b] The a_{ik} for the raw index column indicates the *t*-statistic from the regression (including a constant term) of R_{it} on to R_{kt} where food, clothing, housing, medical care, transportation, and recreation are represented by the subscripts 1, 2, 3, 4, 5 and 6, respectively. Significance is indicated at the 10%, 5%, and 1% levels by '*', '**', '***', respectively. The critical values are 1.68, 2.02, and 2.70, respectively.

follows the homothetic translog case (Silberberg, 1990, p. 411) with a linear-in-prices form such as,[9]

$$X_{kt} = \sum_{i=1}^{w} \mathbf{A}_{ki}\mathbf{p}_{it} \tag{7}$$

where $\sum_{i=1}^{w}\mathbf{A}_{ki}\mathbf{e}_i R_{it} + \sum_{i=1}^{w}\mathbf{A}_{ki}(\mathbf{p}_{it} - \mathbf{e}_i R_{it})$ is an alternative representation of Equation 7. To see why Assumption 1 is sufficient under Equation 7, note that Assumption 1 implies $\text{Cov}(p_{ijt} - R_{it}, R_{ht}) = 0$ for all i, j and h. Thus, unbiased estimates of $\mathbf{A}_{ki}\mathbf{e}_i$ can be obtained for the case where $\sum_{i=1}^{w}\mathbf{A}_{ki}(\mathbf{p}_{it} - \mathbf{e}_i R_{it})$ is omitted in the homothetic translog case if Assumption 1 holds.

IV. Conclusion

This study evaluated analytical correlations between aggregate price indexes and the omitted information from price aggregation, and showed that integrability conditions do not eliminate price aggregation bias under the GCT assumption. For example, the study showed that lack of independence across aggregate price indexes contributes to price aggregation bias. Whereas the GCT is operational for using integrability conditions under commodity price aggregation, the price aggregation bias is likely to be still important beyond the homothetic translog preferences and especially in the presence of non-stationary data.

References

Davis, G. C. (2003) The generalized composite commodity theorem: stronger support in the presence of data limitations, *Review of Economics and Statistics*, **85**, 476–80.

Deaton, A. S. and Muellbauer, J. (1980) An almost ideal demand system, *American Economic Review*, **70**, 312–26.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hicks, J. R. (1936) *The Value of Capital*, Oxford University Press, Oxford.

Leontief, W. (1936) Composite commodities and index numbers, *Econometrica*, **4**, 39–59.

Lewbel, A. (1996) Aggregation without separability: a generalized composite commodity theorem, *American Economic Review*, **86**, 524–43.

Silberberg, E. (1990) *The Structure of Economics: A Mathematical Analysis*, 2nd edn, McGraw-Hill, New York.

Simon, C. and Blume, L. (1994) *Mathematics for Economists*, W. W. Norton & Company, New York.

[9] A linear form such as in Equation 7 with prices in logarithms is consistent with integrability only if preferences are homothetic (Silberberg, 1990, p. 411).