axioms

MDPI

Article

Equation of Finite Change and Structural Analysis of Mean Value

Stan Lipovetsky

Independent Researcher, Minneapolis, MN 55305, USA; stan.lipovetsky@gmail.com

Abstract: This paper describes a problem of finding the contributions of multiple variables to a change in their function. Such a problem is well known in economics, for example, in the decomposition of a change in the mean price via the varying in time prices and volumes of multiple products. Commonly, it is considered by the tools of index analysis, the formulae of which present rather heuristic constructs. As shown in this work, the multivariate version of the Lagrange mean value theorem can be seen as an equation of the function's finite change and solved with respect to an interior point whose value is used in the estimation of the contribution of the independent variables. Consideration is performed on the example of the weighted mean value function, which is the main characteristic of statistical estimation in various fields. The solution for this function can be obtained in the closed form, which helps in the analysis of results. Numerical examples include the cases of Simpson's paradox, and practical applications are discussed.

Keywords: Lagrange mean value theorem; multiple variables; finite change formula; contributions of variables; weighted mean value; mean price; Simpson's paradox

MSC: 26A24; 26B05; 62A99

check for updates

Citation: Lipovetsky, S. Equation of Finite Change and Structural Analysis of Mean Value. *Axioms* **2023**, *12*, 962. https://doi.org/10.3390/axioms12100962

Academic Editors: Eva T. López Sanjuán and María Isabel Parra Arévalo

Received: 18 September 2023
Revised: 6 October 2023
Accepted: 10 October 2023
Published: 12 October 2023

1. Introduction

This paper considers the finite change formula as an extension of the Lagrange mean value theorem to the multivariate version [1–4]. This formula is employed for finding a finite change in a function presented as the total of contributions from the increments of the variables. Such problems appear in various applications where the influences of the independent variables are investigated and their contributions to the increment of the outcome variable are estimated. For example, characteristics of growth and rates in economics can be described with the help of the decomposition of a change in the mean price caused by the varying partial prices and structure of volumes of multiple products. These problems are commonly considered in economics and social sciences via the so-called index analysis [5–9], using rather heuristic formulae of Laspeyres, Paasche, Fisher and other indices [10–15]. A detail review of various index forms is given in [16], and a description of the R software packages for index analysis can be found in [17]. The line integral approach for decomposition of a function's change due to alternation of its different variables was suggested by F. Divisia [18] and developed by many authors [19–23]. Variational analysis for finding a geodesic curve with integration by this trajectory is considered in [24], the ideal index formulae are presented in [25,26], and application to the incremental analysis in nonlinear regression models is described in [27].

In contrast to the line integration by continuous trajectories, the Lagrange mean value theorem in its multivariate version can be expressed as an equation of a finite change in the outcome dependent variable. After solving this equation with respect to an interior point, its value is employed for the estimation of the impact of the variables' modification onto a transformation of their function. Consideration is performed on the example of weighted mean value function, which is one of the main characteristics in any statistical estimation. The solution for this function can be obtained in the closed form, useful in the analysis of the outcome decomposition by changes in the partial values and in the structure of the weighting, which can be particularly helpful in sensibility analysis. Numerical examples also include special cases of the so-called Simpson's paradox [28–32], in which each particular value increases but their mean value decreases, or vice versa—the particular values decline but their mean value grows. The suggested approach helps to interpret such results via data restructuring.

The paper is organized as follows: Section 2 focuses on the Lagrange theorem for decomposition of finite change in the function due to finite increments of its variables, Section 3 describes the application to the decomposition of the weighted mean value function (with Appendix A), Section 4 presents numerical illustrations, and Section 5 summarizes the results.

2. Lagrange Mean Value Theorem and Finite Change Equation

Consider a continuous function $F(x, y, \ldots, z)$ of many real variables $x, y, \ldots, z$. Suppose all the variables are known in the initial $x_0, y_0, \ldots, z_0$ and final $x_1, y_1, \ldots, z_1$ moments in time (or it could be two compared states of a process, two compared objects, etc.), with the two corresponding function values and their difference ΔF defined as follows:

$$\Delta F = F_1 - F_0 = F(x_1,\ y_1,\ \ldots,\ z_1) - F(x_0,\ y_0,\ \ldots,\ z_0). \tag{1}$$

The aim of the problem consists of decomposition of the increment ΔF into a sum of items representing contributions of a change in each particular variable into the total change ΔF in the function:

$$\Delta F = \Delta F(\Delta x) + \Delta F(\Delta y) + \ldots + \Delta F(\Delta z). \tag{2}$$

Such a decomposition (2) shows the relative impact of different variables in the function's alternation. The changes in variables can be parameterized as follows:

$$x(t) = x_0 + t\Delta x, \tag{3}$$

where $\Delta x = x_1 - x_0$. With the parameter t changing on the closed interval [0, 1], the variable x transforms from the initial x_0 to the final x_1 state, and similarly with all other variables.

For solving the problem of decomposition (2), the Lagrange mean value theorem can be applied. For one variable, this classic theorem can be formulated as follows: for a continuous differentiable function $F(x)$, there exists a point x^* on the interval $(x_{0,}\ x_1)$ such that the tangent at this interior point x^* equals the slope of the segment between the endpoints, which can be written as

$$F'(x^*) = (F(x_{\ 1}) - F(x_0))/(x_1 - x_0). \tag{4}$$

The relation (4) can be also rewritten as follows:

$$F(x_{\ 1}) - F(x_0) = F'(x^*)(x_1 - x_0), \tag{5}$$

which states that the finite change in the function is defined by the derivative of the function in the interior point $F'(x^*)$ multiplied by the finite change in the variable x at its endpoints.

For multiple variables, the expression (5) can be generalized in the expression

$$\Delta F = F'_x[x(t), \ldots, z(t)]x'(t)\Delta t + \ldots + F'_z[x(t), \ldots, z(t)]z'(t)\Delta t, \tag{6}$$

in which F'_x and F'_z are the function's derivatives by x or z, and similarly with the other variables. The variables are defined in the parametric form (3) as $x(t), \ldots, z(t)$, and the notations $x'(t), \ldots, z'(t)$ are used for the derivatives by the parameter t, so the relation

(6) can be simplified to the so-called finite-change Formula (3) (Chapter 5), or the finite-increment Formula (4):

$$\Delta F = F'_x[x(t), \ldots, z(t)]\Delta x + \ldots + F'_z[x(t), \ldots, z(t)]\Delta z, \tag{7}$$

in which Δx is defined in (3), and similarly with other variables.

Likewise the Lagrange mean value theorem (5), the relation (7) states that for a given finite change ΔF of the function, there exists at least one point $t = t^*$ such that the total differential at the right-hand side (7) at this point equals this finite change in the left-hand side (7).

Each item in (7) corresponds to a change in the function due to the change in each one variable, which is directly related to the problem (2). For a given ΔF, the expression (7) can be considered as an equation of a finite change and solved with respect to the unknown interior point t^*, whose value can then be used in the estimation of the contribution of each variable's change in the transformation of their function (2).

3. Decomposition of Weighted Mean Value

Let us apply the described approach to the problem of decomposition of the mean value by the variables of influence. The arithmetic mean value m in a general form of the weighted values of the variable x is presented by the well-known formula

$$m = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}, \tag{8}$$

in which x_i are all i-th observations ($i = 1, 2, \ldots, k$, where k is the total number of different observations) and n_i are the counts with which the values x_i are observed. If all n_i are equal, the weighted mean reduces to the simple arithmetic mean value.

Depending on a specific problem, the variables x and n can have various meanings. For example, in studies on consumer purchases, x_i and n_i could denote the prices and amounts in a set of k products, and then the cost of each product is $x_i n_i$, and the total cost divided by total amount in (8) defines the mean price of the product unit. For a clearer exposition of the results, let us use these connotations, but of course, the terms can differ for another problem. Keeping this in mind, let us consider a problem of change in the mean price (8) for the current period of time compared with a basic period of time (denoted by 1 and 0 subindices, respectively), when the mean price change can be presented as the difference:

$$\Delta m = m_1 - m_0 = \frac{\sum_{i=1}^{k} x_{1i} n_{1i}}{\sum_{i=1}^{k} n_{1i}} - \frac{\sum_{i=1}^{k} x_{0i} n_{0i}}{\sum_{i=1}^{k} n_{0i}}. \tag{9}$$

The problem is similar to that formulated in the expression (2)—how can we decompose the total increment Δm (9) of the mean price into a sum of contributions from a change in each particular price x_i and amount n_i? For this aim, let us denote each variable change as

$$\Delta x_i = x_{1i} - x_{0i}\ ,\ \Delta n_i = n_{1i} - n_{0i}, \tag{10}$$

and with them, the changes in variables can be parameterized similarly to (3) as follows:

$$x_i(t) = x_{0i} + t\Delta x_i\ ,\ n_i(t) = n_{0i} + t\Delta n_i \tag{11}$$

The parameter t varies within the interval [0, 1], and accordingly, all variables (11) change the values from the initial to the final state. Depending on the problem, the variables $x_i(t)$ and $n_i(t)$ can be continuous or discrete numbers, but it is possible for approximate estimation to consider all of them as continuous variables. Then, the expression of finite change (7) for the mean value function (8) can be written as:

$$\Delta m = \sum_{i=1}^{k} \frac{\partial m}{\partial x_i} \Delta x_i + \sum_{i=1}^{k} \frac{\partial m}{\partial n_i} \Delta n_i. \tag{12}$$

In (12), taking derivatives of m (8) by all $2k$ variables (11) yields:

$$\Delta m = \frac{\sum_{i=1}^{k} n_i(t)\Delta x_i}{\sum_{i=1}^{k} n_i(t)} + \frac{\sum_{i=1}^{k} x_i(t)\Delta n_i}{\sum_{i=1}^{k} n_i(t)} - \frac{\sum_{i=1}^{k} x_i(t) n_i(t)}{\left(\sum_{i=1}^{k} n_i(t)\right)^2} \sum_{i=1}^{k} \Delta n_i. \quad (13)$$

To simplify notations, let us denote the total of amounts as follows:

$$N_0 = \sum_{i=1}^{k} n_{0i},\ N_1 = \sum_{i=1}^{k} n_{1i}\ ,\ \Delta N = N_1 - N_0 = \sum_{i=1}^{k} \Delta n_i. \quad (14)$$

Using (11) and (14), the relation (13) can be represented in explicit dependence on the parameter t:

$$\Delta m = \frac{\sum_{i=1}^{k} (n_{0i}\Delta x_i + \Delta n_i \Delta x_i t)}{N_0 + t\Delta N} + \frac{\sum_{i=1}^{k} (x_{0i}\Delta n_i + \Delta n_i \Delta x_i t)}{N_0 + t\Delta N} - \frac{\sum_{i=1}^{k} (x_{0i} n_{0i} + (x_{0i}\Delta n_i + n_{0i}\Delta x_i)t + \Delta n_i \Delta x_i t^2)}{(N_0 + t\Delta N)^2}\Delta N. \quad (15)$$

This expression presents the equation of finite change (7) for the function of mean value (8), and it is a rational quadratic form by the parameter t. For a given value of the function change (9), the Equation (15) can be solved for finding the internal point t^*, with which the contributions from each variable change Δx_i and Δn_i in the total change (12) can be identified. The following result can be proved.

Theorem. *The equation of finite change for the mean value function (15) has only one feasible root:*

$$t^* = \frac{1}{1 + \sqrt{N_1/N_0}}\ . \quad (16)$$

Proof of Theorem. The proof of this theorem is given in Appendix A.

With only one solution for the internal point t^* (16), the decomposition (2) for the mean value function (8) by the variables of impact is also unique. This point identifies the values in trajectories (11):

$$x_i(t^*) = \frac{x_{1i}\sqrt{N_0} + x_{0i}\sqrt{N_1}}{\sqrt{N_0} + \sqrt{N_1}},\ n_i(t^*) = \frac{n_{1i}\sqrt{N_0} + n_{0i}\sqrt{N_1}}{\sqrt{N_0} + \sqrt{N_1}}. \quad (17)$$

Let us consider the first quotient in (13), which defines the change Δm occurring due to the changes in the x-variables. Using the second relation (17) in the first quotient (13) yields:

$$\Delta m(\Delta x) = \frac{\sum_{i=1}^{k} n_i(t^*)\Delta x_i}{\sum_{i=1}^{k} n_i(t^*)} = \frac{\sum_{i=1}^{k} \left(n_{1i}\sqrt{N_0} + n_{0i}\sqrt{N_1}\right)\Delta x_i}{\sum_{i=1}^{k}\left(n_{1i}\sqrt{N_0} + n_{0i}\sqrt{N_1}\right)} = \sum_{i=1}^{k} w_i \Delta x_i, \quad (18)$$

in which the weights w_i are defined as:

$$w_i = \frac{n_{1i}\sqrt{N_0} + n_{0i}\sqrt{N_1}}{\sum_{i=1}^{k}\left(n_{1i}\sqrt{N_0} + n_{0i}\sqrt{N_1}\right)} = \frac{n_{1i}/\sqrt{N_1} + n_{0i}/\sqrt{N_0}}{\sum_{i=1}^{k}\left(n_{1i}/\sqrt{N_1} + n_{0i}\sqrt{N_0}\right)} = \frac{n_{1i}/\sqrt{N_1} + n_{0i}/\sqrt{N_0}}{\sqrt{N_1} + \sqrt{N_0}}, \quad (19)$$

with their total equal to one:

$$\sum_{i=1}^{k} w_i = 1. \quad (20)$$

It is useful to mention that both of the relations (A11) could be equal to zero only when $n_{1i} = n_{0i}$ by all i, which corresponds to the trivial case when only x-s vary, so the total change in the mean value is defined by the same Formula (18) with weights (19) reduced to the values $w_i = n_{0i}/N_0$. Thus, such a special case is also covered by the general solution (18)–(19).

The last two quotients (13) are related to the change in the mean value Δm because of the variations in the n-variables, which can be presented as follows:

$$\Delta m(\Delta n) = \sum_{i=1}^{k} \frac{x_i(t^*)\Delta n_i}{\sum_{j=1}^{k} n_j(t^*)} - \frac{\Delta N}{\sum_{i=1}^{k} n_i(t^*)} \cdot \frac{\sum_{j=1}^{k} x_j(t^*) n_j(t^*)}{\sum_{j=1}^{k} n_j(t^*)} \,. \tag{21}$$

The last quotient in (21) is the mean value (8) taken in the internal point (16):

$$m(t^*) = \frac{\sum_{j=1}^{k} x_j(t^*) n_j(t^*)}{\sum_{j=1}^{k} n_j(t^*)}. \tag{22}$$

With (22), the total change Δm due to changes Δn (21) is defined as the weighted sum of the deviations of $x_i(t^*)\Delta n_i$ from the mean $m(t^*)\Delta N$:

$$\Delta m(\Delta n) = \sum_{i=1}^{k} \frac{1}{\sum_{j=1}^{k} n_j(t^*)} \{x_i(t^*)\Delta n_i - m(t^*)\Delta N\} \,. \tag{23}$$

Using both relations (17), and also the equality

$$\sum_{j=1}^{k} \left(\frac{n_{1j}}{\sqrt{N_1}} + \frac{n_{0j}}{\sqrt{N_0}} \right) = \sqrt{N_1} + \sqrt{N_0}, \tag{24}$$

we can transform (23) to the expression:

$$\Delta m(\Delta n) = \frac{1}{\sqrt{N_1} + \sqrt{N_0}} \sum_{i=1}^{k} \left(\frac{x_{1i}}{\sqrt{N_1}} + \frac{x_{0i}}{\sqrt{N_0}} \right) (\Delta n_i - w_i \Delta N). \tag{25}$$

It is the explicit form for the formulae (21) and (23), and it contains deviations of the changes in n_i from the total change in N weighted by values w_j (19). In a special case of the constant quantities by both periods of time, when $\Delta n_i = 0$ by all i, the change $\Delta m(\Delta n)$ in (25) equals zero, so a change in m can occur only due to a change in the x-variables (18). If for some quantities $\Delta n_i \neq 0$ but the total quantity is constant, $\Delta N = 0$, then the last item in (25) disappears, and this expression becomes similar to the form (18). The compact expression (23) and the explicit Formula (25) are convenient for the interpretation and calculations as well.

Formulas (18) and (25) for an i-th item from their totals identify an impact of the change in each particular x_i and n_i variables, which can be presented as

$$m(\Delta x_i) = \frac{\Delta x_i}{\sqrt{N_1} + \sqrt{N_0}} \left(\frac{n_{1i}}{\sqrt{N_1}} + \frac{n_{0i}}{\sqrt{N_0}} \right) , \tag{26}$$

and the second one is

$$\Delta m(\Delta n_i) = \frac{1}{\sqrt{N_1} + \sqrt{N_0}} \left(\frac{x_{1i}}{\sqrt{N_1}} + \frac{x_{0i}}{\sqrt{N_0}} \right) \left\{ \Delta n_i - \frac{\Delta N}{\sqrt{N_1} + \sqrt{N_0}} \left(\frac{n_{1j}}{\sqrt{N_1}} + \frac{n_{0j}}{\sqrt{N_0}} \right) \right\}, \tag{27}$$

in which the weights w_j are defined in (19). The sum of the contributions (26) and (27) is:

$$\begin{aligned} \Delta m(\Delta x_i) + \Delta m(\Delta n_i) &= \tfrac{1}{\sqrt{N_1}+\sqrt{N_0}} \left(\Delta x_i \left(\tfrac{n_{1i}}{\sqrt{N_1}} + \tfrac{n_{0i}}{\sqrt{N_0}} \right) + \Delta n_i \left(\tfrac{x_{1i}}{\sqrt{N_1}} + \tfrac{x_{0i}}{\sqrt{N_0}} \right) \right) \\ - \tfrac{\sqrt{N_1}-\sqrt{N_0}}{\sqrt{N_1}+\sqrt{N_0}} \left(\tfrac{x_{1i}}{\sqrt{N_1}} + \tfrac{x_{0i}}{\sqrt{N_0}} \right) \left(\tfrac{n_{1i}}{\sqrt{N_1}} + \tfrac{n_{0i}}{\sqrt{N_0}} \right) &= \tfrac{x_{1i} n_{1i}}{\sqrt{N_1}+\sqrt{N_0}} \left(\tfrac{1}{\sqrt{N_1}} + \tfrac{\sqrt{N_0}}{N_1} \right) \\ - \tfrac{x_{0i} n_{0i}}{\sqrt{N_1}+\sqrt{N_0}} \left(\tfrac{1}{\sqrt{N_0}} + \tfrac{\sqrt{N_1}}{N_0} \right) &= \tfrac{x_{1i} n_{1i}}{N_1} - \tfrac{x_{0i} n_{0i}}{N_0} = m_{1i} - m_{0i}. \end{aligned} \tag{28}$$

Thus, the total of the i-th contributions from the variables' change equals the change in the i-th component of the mean value (8). Summing the relation (28) by all i-th contributions, yields the change in the mean value (9):

$$\sum_{i=1}^{k} \Delta m(\Delta x_i) + \sum_{i=1}^{k} \Delta m(\Delta n_i) = \sum_{i=1}^{k} m_{1i} - \sum_{i=1}^{k} m_{0i} = \Delta m, \quad (29)$$

which proves that the obtained formulae for all inputs are correct. It is also useful to note that if one of contributions is already calculated, then to facilitate the calculations, the complementary one can be found as its difference from the mean's change by the relation (28). For example, when we know $\Delta m(\Delta x_i)$, then it is possible to estimate another contribution by the i-th product as:

$$\Delta m(\Delta n_i) = m_{1i} - m_{0i} - \Delta m(\Delta x_i). \quad (30)$$

Summing (30) with respect to i yields a similar relation for the total values, which can be easily obtained from (30) by omitting the i-subindex.

In interpretation of the results, we should take into account the following features of the obtained formulae. Each i-th contribution into the total change $\Delta m(\Delta x)$ (18) depends on the signs of Δx, so a positive change $\Delta x_i > 0$ increases the outcome Δm, while a negative change $\Delta x_i < 0$ diminishes the outcome. In contrast to it, an i-th contribution into the total change $\Delta m(\Delta n)$ (25) because of the change Δn_i, is more complicated: a positive contribution to the change in the mean value is given when $\Delta n_i - w_i \Delta N > 0$, if the change Δn_i is above the weighted total change $w_i \Delta N$. Similarly, for the opposite case $\Delta n_i - w_i \Delta N < 0$, the contribution of the change Δn_i to the change in the mean value is negative. This complicated impact of the quantities n_i onto the mean value and its change (8)–(9) leads to the possibility of encountering the so-called Simpson's paradox when the increased value in each item produces a decrease in their mean outcome, and vice versa, when a decrease in each item yields an unexpected increase in the mean value.

The described approach based on the Lagrange mean value theorem was demonstrated for decomposition of the mean value function. Actually, many functions can be transformed to the structure similar to the mean value. It especially concerns the statistical functions with summing by the data observations. For example, in the pair regression model $y = a + bz$, the slope coefficient equals the quotient of the sample covariance to the variance of the predictor, which can be transformed to the following form:

$$b = \frac{\sum_{i=1}^{k} (z_i - m_z)(y_i - m_y)}{\sum_{j=1}^{k} (z_j - m_z)^2} = \sum_{i=1}^{k} \frac{(z_i - m_z)^2}{\sum_{j=1}^{k} (z_j - m_z)^2} \cdot \frac{y_i - m_y}{z_i - m_z} = \sum_{i=1}^{k} v_i b_i \,, \quad (31)$$

in which v_i denotes the weights of squared deviations $(z_i - m_z)^2$ in their total, and b_i denotes the partial slope coefficients in each i-th observation. The obtained function (31) has a structure of the weighted mean value, and its change can be studied in the described approach. Another example of the functions of mean value structure can be found in the readability indices, in average number of elements per word or words per sentence [33]. Changes in values of those functions can be investigated via decomposition by the factors of influence as well.

4. Numerical Examples

To illustrate the described approach in numerical estimations, let us consider a set of ten products sold at a market in two consecutive time periods. For example, it can be a car dealership with ten models of trucks. The prices and quantities of the basic period x_0 and n_0, and of the current period x_1 and n_1, are shown in the first columns of Table 1, together with their total values given there in the last row. The total quantity diminishes from $N_0 = 48$ to $N_1 = 37$, so by $\Delta N = -11$, and the internal point (16) value equals $t^* = 0.532$.

Table 1. Change in the mean price (dataset-1).

Item	Basic Period		Current Period		Mean Price		Change in Variables			Change in Mean Value		
i	x_0	n_0	x_1	n_1	m_0	m_1	Δx	Δn	w_i	$\Delta m(\Delta x)$	$\Delta m(\Delta n)$	Δm
1	30	2	40	2	1.250	2.162	10	0	0.047	0.475	0.438	0.912
2	50	4	40	3	4.167	3.243	−10	−1	0.082	−0.823	−0.101	−0.923
3	40	3	60	5	2.500	8.108	20	2	0.096	1.929	3.679	5.608
4	70	5	80	1	7.292	2.162	10	−4	0.068	0.681	−5.811	−5.130
5	15	1	70	4	0.313	7.568	55	3	0.062	3.390	3.865	7.255
6	50	7	45	6	7.292	7.297	−5	−1	0.153	−0.767	0.773	0.006
7	60	6	30	5	7.500	4.054	−30	−1	0.130	−3.892	0.446	−3.446
8	40	4	50	4	3.333	5.405	10	0	0.095	0.949	1.123	2.072
9	80	3	40	3	5.000	3.243	−40	0	0.071	−2.847	1.091	−1.757
10	20	13	100	4	5.417	10.811	80	−9	0.195	15.581	−10.186	5.394
total	455	48	555	37	44.063	54.054	100	−11	1	14.675	−4.683	9.992

The next columns in Table 1 contain the corresponding total costs divided by the total quantities, $x_{0i}n_{0i}/N_0$ and $x_{1i}n_{1i}/N_1$, which define the i-th items and their total m_0 and m_1 in the mean prices (8)–(9) of each period. The changes in the i-th prices and quantities (10) are given in the next two columns, and then the column w_i presents the weights (19)–(20). After this, the next two columns show the change in the mean price due to the changes (18) in the partial prices and due to the changes (25) in the quantities. The sum of these two columns in Table 1 yields the last column of the total change in the mean price for each i-th product (28) and by all of them in total (29), which equals:

$$\Delta m = \Delta m(\Delta x) + \Delta m(\Delta n) = 14.675 - 4.683 = 9.992. \tag{32}$$

Thus, the changes in the particular prices led to the increment $\Delta m(\Delta x)$ in the mean price, but restructuring according to the changes in the amounts $\Delta m(\Delta n)$ decreased the total mean price Δm. The change (32) in the mean price equals the difference (9) of the mean prices m_1 and m_0 in Table 1.

Table 1 also demonstrates that the signs of difference in all i-th contributions $\Delta m(\Delta x_i)$ coincide with the direction of changes in the partial prices Δx_i, as follows from the Formula (18). However, the signs of the contributions $\Delta m(\Delta n_i)$ and the signs of changes in the amounts Δn_i themselves, due to (25), can vary in different directions. For example, for the products with i = 3, 5, the quantities grow, $\Delta n_i > 0$, and the contribution to the mean price is positive, $\Delta m(\Delta n_i) > 0$; for the products with i = 1, 8, 9, there is no change in quantities, $\Delta n_i = 0$, but their impact onto the mean price is positive, $\Delta m(\Delta n_i) > 0$; for the rest of products with i = 2, 4, 6, 7, 10, there is a reduction in quantities, $\Delta n_i < 0$, and the input into the mean price is negative, $\Delta m(\Delta n_i) < 0$ for the products i = 2, 4, 10, but the input is positive, $\Delta m(\Delta n_i) > 0$, for the products i = 6 and 7. Therefore, a redistribution of the amounts n can yield various results depending on the structure of the weights and total amounts, as expressed in the Formula (25).

In Table 1, all ten prices go up and the mean price also grows, which seems natural and is not surprising. However, the complex structure of the amounts n_i and their evolution Δn_i can influence the mean price so that it would change in the opposite direction, which produces the famous Simpson's paradox. Let us consider it in the next example presented in Table 2 organized similarly to the previous table.

Table 2. Change in the mean price and Simpson's paradox (dataset-2).

Item	Basic Period		Current Period		Mean Price		Change in Variables			Change in Mean Value		
i	x_0	n_0	x_1	n_1	m_0	m_1	Δx	Δn	w_i	$\Delta m(\Delta x)$	$\Delta m(\Delta n)$	Δm
1	30	5	36	8	2.419	7.200	6	3	0.134	0.803	3.978	4.781
2	50	8	54	3	6.452	4.050	4	−5	0.105	0.420	−2.821	−2.402
3	40	10	45	1	6.452	1.125	5	−9	0.101	0.503	−5.830	−5.327
4	70	9	73	1	10.161	1.825	3	−8	0.092	0.275	−8.611	−8.336
5	15	3	18	4	0.726	1.800	3	1	0.071	0.214	0.860	1.074
6	40	7	45	8	4.516	9.000	5	1	0.152	0.758	3.725	4.484
7	20	6	21	8	1.935	4.200	1	2	0.143	0.143	2.122	2.265
8	40	5	43	2	3.226	2.150	3	−3	0.067	0.201	−1.277	−1.076
9	30	3	33	4	1.452	3.300	3	1	0.071	0.214	1.634	1.848
10	20	6	22	1	1.935	0.550	2	−5	0.065	0.130	−1.515	−1.385
total	355	62	390	40	39.274	35.200	35	−22	1	3.661	−7.735	−4.074

The prices and quantities of the basic and current periods in Table 2 seem to be very similar to those in Table 1. The total quantity become N_0 = 62 in the basic and N_1 = 40 in the current periods, so the change is also negative, $\Delta N = -22$, and the internal point (16) is t^* = 0.555. The price of each product increases, so all $\Delta x_i > 0$, and all contributions to the change in the mean price are positive $\Delta m(\Delta x_i) > 0$ as well. However, the total change in the mean price is negative, so it diminishes:

$$\Delta m = \Delta m(\Delta x) + \Delta m(\Delta n) = 3.661 - 7.735 = -4.074. \tag{33}$$

Similarly to the previous data results (32), the change in the mean price has the positive impact of changes in the prices and negative impact of changes in the amounts. However, with respect to the absolute value, there is a relation $|\Delta m(\Delta x)| > |\Delta m(\Delta n)|$ in (32), while there is the opposite inequality $|\Delta m(\Delta x)| < |\Delta m(\Delta n)|$ in (33). Thus, in spite of the increases in all the prices, the mean price decreases, which occurs because the negative changes in the amounts $\Delta m(\Delta n)$ overcome the positive impact $\Delta m(\Delta x)$. By referring to Table 2, we can identify which products give a negative impact: those with i = 2, 3, 4, 8, 10, with contributions $\Delta m(\Delta n_i) < 0$. If to change places for the data of the basic and the current periods, the results in Table 2 receive the opposite signs. It would correspond to another situation when a decrease in all prices produces the mean price growth. Such an ambiguity could distort an adequate understanding of the results presented in some statistical reports, which should be considered with attention and caution.

Let us also compare the newly developed technique based on the Lagrange mean theorem and one of the common techniques of the logarithmic decomposition of the total increment, described in [24,25] and also [26] (Formula (11)) and [32] (Formula (10)). Table 3 at first presents results for dataset-1: Lagrange-based decomposition for the share $\Delta m(\Delta x)$, repeated from Table 1 for the sake of comparison with its logarithmic estimation, and their ratio. The last three columns in Table 3 show similar results for dataset-2 from Table 2. We can see that within each dataset, the results via both methods are very close with respect to any i-th product, mostly within several precent, and the mean difference shown in the last row is about 4–8%. Thus, these methods support the results of each other, and are open for further investigation.

It is important to note that the decomposition of a function change due to the changes in its independent variables presents a special kind of descriptive analysis which can indicate in which directions researchers and managers can find how to improve the outcome values. With some products of positive and others of negative contributions into the change in the mean price, the best and worst players can be identified. Of course, it is difficult to predict an actual rate of the mean price change with improvement in the product characteristics because many factors play their role in the market. For example, some products can be complementary, others substitutional, the market context has its effects, and other conditions can influence the consumers decisions [34,35]. It can also be useful to build a spreadsheet calculator for performing the described decompositions and considering various "what-if" scenarios according to the obtained results.

Table 3. Comparison of the newly developed and standard techniques.

	Dataset-1			Dataset-2		
i	Lagrange Based, from Table 1 a=$\Delta m(\Delta x)$	Logarithmic Method b=$\Delta m(\Delta x)$	Their Ratio b/a	Lagrange Based, from Table 2 c=$\Delta m(\Delta x)$	Logarithmic Method d=$\Delta m(\Delta x)$	Their Ratio d/c
1	0.475	0.479	1.009	0.803	0.799	0.995
2	−0.823	−0.822	1.000	0.420	0.397	0.945
3	1.929	1.933	1.002	0.503	0.359	0.714
4	0.681	0.563	0.827	0.275	0.204	0.741
5	3.390	3.507	1.034	0.214	0.216	1.007
6	−0.767	−0.769	1.002	0.758	0.766	1.010
7	−3.892	−3.883	0.998	0.143	0.143	0.999
8	0.949	0.956	1.008	0.201	0.192	0.954
9	−2.847	−2.813	0.988	0.214	0.215	1.002
10	15.581	12.563	0.806	0.130	0.105	0.810
mean			0.967			0.918

5. Summary

This paper described the generalized Lagrange mean value theorem for multiple variables in application to the decomposition of a function's change and presented it as the sum of contributions from the change in each independent variable. The multivariate version of the Lagrange mean value theorem is considered as a finite change equation that can be solved with respect to an interior point, whose value is used for the estimation of the contribution of the independent variables. The derivation and analysis are performed on the example of the weighted mean value function, which is one of the main characteristics of statistical description practically in all areas of research. The solution for this function is obtained in the closed form, which is helpful in the analysis of results. Numerical examples include also the cases of the Simpson's paradox. The described possibilities of the finite change equation can be useful in practical applications when a researcher or manager needs to identify which components of the characteristic of mean value give the main positive or negative impact, because these items can be considered as the main drivers for reaching an increase or decrease in the mean price value. The suggested approach can be implemented for finding a structure of change for other functions as well. It can enrich the possibilities of data analysis and serve various practical applications.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: I am grateful to the three reviewers for their comments and suggestions which helped to improve the paper.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

The Proof of Theorem is as follows. Multiplying the equation of finite change (15) by the common denominator yields:

$$(N_0+t\Delta N)^2\Delta m = (N_0+t\Delta N)\sum_{i=1}^{k}\left((n_{0i}\Delta x_i + x_{0i}\Delta n_i) + 2\Delta n_i\Delta x_i t\right) - \Delta N\sum_{i=1}^{k}\left(x_{0i}n_{0i} + (x_{0i}\Delta n_i + n_{0i}\Delta x_i)t + \Delta n_i\Delta x_i t^2\right) \quad \text{(A1)}$$

Opening parentheses and gathering items by the power of parameter t produces the following quadratic equation:

$$t^2\Delta N\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right)+2tN_0\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right)+\left(\Delta m{N_0}^2-N_0\sum_{i=1}^{k}(x_{0i}\Delta n_i+n_{0i}\Delta x_i)+\Delta N\sum_{i=1}^{k}x_{0i}n_{0i}\right)=0. \quad \text{(A2)}$$

To simplify the last expression of the intercept, we notice the equality

$$x_1n_1=(x_0+\Delta x)(n_0+\Delta n)=x_0n_0+\Delta x\Delta n+x_0\Delta n+n_0\Delta x\,. \quad \text{(A3)}$$

Then, the last two items in (A3) can be represented via the other items, and summing such relation by all i yields:

$$\sum_{i=1}^{k}(x_{0i}\Delta n_i+n_{0i}\Delta x_i)=\sum_{i=1}^{k}x_{1i}n_{1i}-\sum_{i=1}^{k}x_{0i}n_{0i}-\sum_{i=1}^{k}\Delta x_i\Delta n_i \quad \text{(A4)}$$

Taking into account the definitions (9) and (14), the relation (A4) can be simplified to the expression:

$$\sum_{i=1}^{k}(x_{0i}\Delta n_i+n_{0i}\Delta x_i)=N_1m_1-N_0m_0-\sum_{i=1}^{k}\Delta x_i\Delta n_i. \quad \text{(A5)}$$

Substituting (A5) in place of the intercept in (A2) leads to such an expression for it:

$$\Delta m{N_0}^2-N_0\sum_{i=1}^{k}(x_{0i}\Delta n_i+n_{0i}\Delta x_i)+\Delta N\sum_{i=1}^{k}x_{0i}n_{0i}=N_0\left(\Delta mN_0-N_1m_1+N_0m_0+\Delta Nm_0+\sum_{i=1}^{k}\Delta x_i\Delta n_i=(-N_0)\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right).\right. \quad \text{(A6)}$$

Using the obtained result (A6) for the intercept in the Formula (A2) yields the quadratic equation of the following form:

$$t^2\Delta N\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right)+2tN_0\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right)-N_0\left(\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i\right)=0\,. \quad \text{(A7)}$$

This equation contains the same term in parentheses with each of its items by power t. If this multiplier does not equal zero, then it can be cancelled from the Equation (A7).

In a general case of changed quantities, when $\Delta n_i\neq 0$ at least for some i, this multiplier differs from zero—indeed, using definitions (9), (10) and (14) lets us transform it as follows:

$$\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i=\left(\frac{\sum_{i=1}^{k}x_{1i}n_{1i}}{\sum_{i=1}^{k}n_{1i}}-\frac{\sum_{i=1}^{k}x_{0i}n_{0i}}{\sum_{i=1}^{k}n_{0i}}\right)\left(\sum_{i=1}^{k}n_{1i}-\sum_{i=1}^{k}n_{0i}\right)-\left(\sum_{i=1}^{k}x_{1i}n_{1i}-\sum_{i=1}^{k}x_{1i}n_{0i}-\sum_{i=1}^{k}x_{0i}n_{1i}+\sum_{i=1}^{k}x_{0i}n_{0i}\right). \quad \text{(A8)}$$

Opening parentheses in (A8) simplifies this expression further to the following one:

$$\Delta m\Delta N-\sum_{i=1}^{k}\Delta x_i\Delta n_i=\sum_{i=1}^{k}n_{0i}\left\{x_{1i}-\frac{\sum_{j=1}^{k}x_{1j}n_{1j}}{\sum_{j=1}^{k}n_{1j}}\right\}+\sum_{i=1}^{k}n_{1i}\left\{x_{0i}-\frac{\sum_{j=1}^{k}x_{0j}n_{0j}}{\sum_{j=1}^{k}n_{0j}}\right\}=\sum_{i=1}^{k}n_{0i}\{x_{1i}-m_1\}+\sum_{i=1}^{k}n_{1i}\{x_{0i}-m_0\}. \quad \text{(A9)}$$

A sum of the weighted deviations of x_i from their weighted mean value equals zero only if the weights are the same as used in the definition of the mean value, for example:

$$\sum_{i=1}^{k}n_{1i}\left\{x_{1i}-\frac{\sum_{j=1}^{k}x_{1j}n_{1j}}{\sum_{j=1}^{k}n_{1j}}\right\}=0\,,\ \sum_{i=1}^{k}n_{0i}\left\{x_{0i}-\frac{\sum_{j=1}^{k}x_{0j}n_{0j}}{\sum_{j=1}^{k}n_{0j}}\right\}=0\,. \quad \text{(A10)}$$

However, the result in (A9) differ from the equalities (A10)—the deviations from the mean values in (A9) are taken with the counts n_i of the other set, so these totals differ from zero:

$$\sum_{i=1}^{k} n_{0i}\{x_{1i} - m_1\} \neq 0\,,\ \sum_{i=1}^{k} n_{1i}\{x_{0i} - m_0\} \neq 0\,. \quad \text{(A11)}$$

Therefore, the same term in three parentheses in (A7) is not of zero value and can be canceled, yielding the following simple quadratic equation:

$$t^2 \Delta N + 2tN_0 - N_0 = 0\,. \quad \text{(A12)}$$

For the same total amounts $\Delta N = 0$ (although with some $\Delta n_i \neq 0$), (A12) becomes the linear equation with the solution $t^* = 1/2$. For the general case of different N_0 and N_1, the quadratic Equation (A12) has the following solutions:

$$t_{1,2} = \frac{-2N_0 \pm \sqrt{4{N_0}^2 + 4N_0 \Delta N}}{2\Delta N} = \frac{-N_0 \pm \sqrt{N_0 N_1}}{\Delta N}\,. \quad \text{(A13)}$$

Taking the positive solution from (A13), feasible for the definitions in (11), yields the result:

$$t^* = \frac{\sqrt{N_0}\left(\sqrt{N_1} - \sqrt{N_0}\right)}{\Delta N} = \frac{\sqrt{N_0}\left(\sqrt{N_1} - \sqrt{N_0}\right)}{N_1 - N_0} = \frac{\sqrt{N_0}}{\sqrt{N_0} + \sqrt{N_1}} = \frac{1}{1 + \sqrt{N_1/N_0}} \quad \text{(A14)}$$

It is the meaningful and unique solution (16) for the Equation (15) of finite change. It holds for the equal total amounts $N_0 = N_1$ as well, reducing to $t^* = 1/2$.

References

1. Fikhtengol'tz, G.M. *The Fundamentals of Mathematical Analysis*; Pergamon Press: Oxford, UK, 1965; Volume 1.
2. Cao, H.; Li, B. The Lagrange Mean Value Theorem of a Function of n Variables. 2014. Available online: https://www.researchgate.net/publication/238757054 (accessed on 11 October 2023).
3. Zając, K. Generalized Lagrange theorem. *arXiv* **2023**, arXiv:2303.09237.
4. Finite-Increments Formula—Encyclopedia of Mathematics. 2023. Available online: http://encyclopediaofmath.org/index.php?title=Finite-increments_formula&oldid=38670 (accessed on 11 October 2023).
5. Hummelbrunner, S.A.; Rak, L.J.; Fortura, P.; Taylor, P. *Contemporary Business Statistics with Canadian Applications*, 3rd ed.; Pearson Education Canada: Don Mills, ON, Canada, 2003.
6. Hill, R.J. Constructing price indexes across space and time: The case of the European Union. *Am. Econ. Rev.* **2004**, *94*, 1379–1410. [CrossRef]
7. Coelli, T.J.; Rao, D.S.P.; O'Donnell, C.J.; Battese, G.E. *An Introduction to Efficiency and Productivity Analysis*, 2nd ed.; Springer Science: New York, NY, USA, 2005.
8. O'Donnell, C.J. Nonparametric estimates of the components of productivity and profitability change in U.S. agriculture. *Am. J. Agric. Econ.* **2012**, *94*, 873–890. [CrossRef]
9. Balk, B.M. *Price and Quantity Index Numbers*; Cambridge University Press: New York, NY, USA, 2012.
10. Fisher, I. *The Making of Index Numbers*; Houghton Mifflin: Boston, MA, USA, 1922.
11. Yule, G.U.; Kendall, M.G. *An Introduction to the Theory of Statistics*; Griffin: London, UK, 1950.
12. Griliches, Z. (Ed.) *Price Indexes and Quality Change*; Harvard University Press: Cambridge, MA, USA, 1961.
13. Allen, R.G.D. *Index Numbers in Theory and Practice*; Macmillan & Co.: London, UK, 1975.
14. Vogt, A.; Barta, J. *The Making of Tests for Index Numbers: Mathematical Methods of Descriptive Statistics*; Springer: Heidelberg, Germany, 1997.
15. Ralph, J.; O'Neill, R.; Winton, J. *A Practical Introduction to Index Numbers*; Wiley Online Books: Hoboken, NJ, USA, 2015. [CrossRef]
16. Turvey, R. *Consumer Price Index Manual: Theory and Practice*; Consumer Price Index Manual: Theory and Practice (ilo.org); Oxford Academic: Oxford, UK, 2005.
17. Zhou, N.B. Simple Index and Weight Index Examples in R. 2021. Available online: mssqltips.com (accessed on 11 October 2023).
18. Divisia, F. L'indice monetaire et la theorie de la monnaie. *Rev. D'economie Polit.* **1925**, *39*, 842–861.
19. Montgomery, J.K. *The Mathematical Problem of the Price Index*; P.S. King: London, UK, 1937.
20. Hulten, C.R. Divisia index numbers. *Econometrica* **1973**, *41*, 1017–1025. [CrossRef]
21. Diewert, W.E. Superlative index numbers and consistency in aggregation. *Econometrica* **1978**, *46*, 980–1008. [CrossRef]
22. Banerjee, K.S. *On the Factorial Approach Providing the True Index of Cost of Living*; Vandenhoeck & Ruprecht: Gottingen, Germany, 1980.

23. Barnett, W.A.; Offenbacher, E.K.; Spindt, P.A. The new Divisia monetary aggregates. *J. Political Econ.* **1984**, *92*, 1049–1085. [CrossRef]
24. Lipovetsky, S. Variational analysis of the breakdown of the increase between factors. *Matekon* **1983**, *20*, 93–103.
25. Vartia, Y.O. Ideal log-change index numbers. *Scand. J. Stat.* **1976**, *3*, 121–126.
26. Sato, K. The ideal log-change index number. *Rev. Econ. Stat.* **1976**, *58*, 223–228. [CrossRef]
27. Lipovetsky, S. Extraction of increments in multifactor models. *Ind. Lab.* **1984**, *50*, 280–283.
28. Simpson, E.H. The interpretation of interaction in contingency tables. *J. R. Stat. Soc. Ser. B* **1951**, *13*, 238–241. [CrossRef]
29. Gurland, J.; Sethuraman, J. How pooling failure data may reverse increasing failure rates. *J. Am. Stat. Assoc.* **1995**, *90*, 1416–1423. [CrossRef]
30. Kocik, J. Proof without words: Simpson's paradox. *Math. Mag.* **2001**, *74*, 399. [CrossRef]
31. Curley, S.P.; Browne, G.J. Normative and descriptive analyses of Simpson's paradox in decision making. *Organ. Behav. Hum. Decis. Process.* **2001**, *84*, 308–333. [CrossRef] [PubMed]
32. Lipovetsky, S.; Conklin, M. Data aggregation and Simpson's paradox gauged by index numbers. *Eur. J. Oper. Res.* **2006**, *172*, 334–351. [CrossRef]
33. Lipovetsky, S. Readability Indices Structure and Optimal Features. *Axioms* **2023**, *12*, 421. [CrossRef]
34. Tversky, A.; Simonson, I. Context-dependent preferences. *Manag. Sci.* **1993**, *39*, 1179–1189. [CrossRef]
35. Lipovetsky, S.; Conklin, M. Finding items cannibalization and synergy by BWS data. *J. Choice Model.* **2014**, *12*, 1–9. [CrossRef]