PRECIOS

Documentos de trabajo

Volatility analysis and aggregate index variability decomposition: a methodological and practical framework

Autores:
Ignacio Agustín Galará
Ronald Lineros González
Marcelo Montes Muñoz

Nº.2, agosto de 2018

Instituto Nacional de Estadísticas • Chile

INSTITUTO NACIONAL DE ESTADÍSTICAS
Morandé 801, piso 22, Santiago, Chile
Teléfono: (56) 23246 1137
Correo: ine@ine.cl
Facebook: @ChileINE
Twitter: @INE_Chile
www.ine.cl

Ignacio Agustín Galará
Ronald Lineros González
Marcelo Montes Muñoz
Subdirección Técnica

Los autores agradecen la colaboración y comentarios de cada uno de los integrantes del Departamento de Estadísticas de Precios por sus valiosos aportes, observaciones y comentarios.

Los Documentos de Trabajo del INE están dirigidos a investigadores, académicos, estudiantes y público especializado en materias económicas, y tienen como objetivo proporcionar un análisis exhaustivo sobre aspectos conceptuales, analíticos y metodológicos claves de los productos estadísticos que elabora la institución y, de esta forma, contribuir al intercambio de ideas entre los distintos componentes del Sistema Estadístico Nacional.

Las interpretaciones y opiniones que se expresan en los Documentos de Trabajo pertenecen en forma exclusiva a los autores y colaboradores y no reflejan necesariamente el punto de vista oficial del INE ni de la institución a la que pertenecen los colaboradores de los documentos.

Volatility analysis and aggregate index variability decomposition: a methodological and practical framework

Abstract

By connecting a set of revised methodologies consistently and coherently, this paper provides a methodological and practical framework for identifying the most volatile components of an aggregate index[1] and the tools for decomposing its variability. The volatility of the aggregate index and its components are measured through the standard deviations of their monthly and annual variations. To complement this analysis, we discuss additional formal statistical tests for identifying components with the highest likelihood of influencing the volatility of the aggregate index. Once identified, the components are analyzed to determine whether they have sufficient impact to influence the volatility of the general index. In addition, we present an exercise in the decomposition of aggregate-index variations in order to illustrate the two elements of the general index: aggregate and idiosyncratic shocks. Aggregate shocks result either from macroeconomic effects that have an impact on most components or from the propagation of the shock among these components. Idiosyncratic shocks occur within specific components, mainly those with high relative weights. The proposed framework can be applied to the practical analysis of the Consumer Price Index (CPI) and other indices.

Keywords: aggregate index, Consumer Price Index, volatility, standard deviation, components.

[1] The calculation of an aggregate index is based on weighted components (sectors).

* The opinions and interpretations expressed in this working paper belong exclusively to the authors and do not necessarily represent the official point of view of National Statistics Institute of Chile.

1. Introduction

Economic theories seek to understand and explain how economic phenomena behave. Robust macroeconomic indicators are needed to contrast different theories with reality. Gross Domestic Product (GDP), Unemployment Rates (UR), and the Consumer Price Index (CPI) are among the national indicators used by analysts and in financial markets and the academy.

The study of the causes of the volatility of these indices is fundamental because of the impact they have on decision-making in economic policy, investment, and consumer behavior. One of the main tasks of offices of national statistics is to ensure the correct measurement of these phenomena, an is an essential task in providing these statistics and indices.

The CPI is particularly important because it measures the price variation of a basket of goods and services consumed by a household sample (representative at national level) with which inflation is calculated.

The volatility of these indices is a crucial issue. For example, the uncertainty in the movements of interest rates affects investment decisions, causing economic agents to dedicate resources to guard against this uncertainty and therefore harming productive activities. In this way, the growing volatility of inflation (sustained over time) makes the economy less efficient by introducing frictions into the markets; the markets then create a gap between the relative prices prevailing in the economy and prices determined only by market forces in the absence of inflation volatility (Friedman, 1977). For this reason, it is essential to note the degree of susceptibility of the CPI or any other aggregate index to specific elements of its composition.

To do this, the sources of volatility of the indices must be identified during their construction. When atypical variations or discrepancies between comparative series are observed, we must focus on the series or specific periods in which they occur. With a proposed battery of statistical tests, we can discard these variations and discrepancies if the volatility of the index is due to problems of measurement or calculation methodology. If not, we advance to a comprehensive analysis of the interaction of intrinsic and exogenous economic dynamics that alter the measured phenomenon and consequently affect the behaviour of the index.

The main purpose of this paper is to provide a methodological and practical framework for analyzing the dynamics and behavior of the volatility of any aggregate index. Standard deviation, as used in this paper, is a measure of dispersion or volatility and is widely used in the literature for studying different degrees of volatility[2].

Based on the methodology presented in this paper, frequent and informative analyses may be made. In essence, we propose a parsimonious analysis loaded with strong information, with a sunk cost at the beginning but with zero marginal cost after the analysis has been automated.

To achieve this (following recommendations raised in a different context by McKenzie and Sasin), a holistic (sensitive) analysis compares several indicators and specifications (monthly, quarterly, and annually) of different levels or components of the aggregated indices, formal statistical tests, and time periods, in order to decompose the index volatility in a consistent manner (internal validity) (McKenzie and Sasin, 2007). A diagram and basic statistical measurements[3] at various levels of the index are added to the main analysis. Finally, the economic context is considered (external validity).

In section II, we briefly present the methodology used to identify components that have a high likelihood of affecting the volatility of the general index. In section III, we discuss the sources of variability of index numbers. Our analysis differentiates between aggregate and idiosyncratic shocks. We describe the limitations in section IV and conclude in section V.

II. Identification of volatile components in an aggregate index

The volatility of an index is measured by monthly or annual variations of standard deviation. For any index whose construction is based on the weighted aggregation of many lower-level components of lower levels, the components normally have different degrees of influence on the aggregate index, either because the intrinsic volatility of a particular group of components is affecting the index or because its weights are important. Knowing these patterns of influence is fundamental to the study of index behavior.

The methodology discussed in this section is based on Atuk et al in which the authors analyze the impact of seasonal products on the CPI in Turkey (Atuk et al, 2013). This study can be extrapolated to any aggregate index because it is straightforward in application and very informative in results.

---

[2] Detmeister and Hulseman (2017), Bowdler and Malik (2017), and Karras (2015, 2017), among others.

[3] Mean, median, standard deviation, and maximum and minimum, among others.

First of all, we need to know the contribution of the various components of an index. The monthly contribution[4] of component m to aggregate index AI in month t is defined as follows:

$$C_{AI}^{m} = \frac{m_{t-1} * w_{m,t}}{AI_{t-1}} * \Delta m_t \qquad (1)$$

Where Δm_t is the monthly variation of component m, and $w_{m,t}$ is the weight of m. With this definition, we can calculate the monthly variation of the aggregate index AI as the sum of all monthly contributions[5]. By analyzing the contributions of a given set of components of the aggregate index, the impact of each component on the volatility of an aggregate index can be observed.

Contributions are the focus of the analysis because they measure the weight or impact that a component, or set of components, has on the total variation of an aggregate index, enabling us to recognize which factors are disturbing the variation of the general index in a particular month (monthly contribution) or year (annual contribution). The annual contribution reflects the accumulated monthly changes of a range of specific components on the annual variation of the general index.

Once all the contributions for all components of the index have been determined, the volatility (standard deviation) between monthly and annual contributions is contrasted. Through a test of statistical variance, we can assess the probability, associated with the null hypothesis, of the equality of variance between annual and monthly contributions. In this exercise, we conjecture that if a component of the index is of high volatility with respect to its monthly contribution and that, in addition, this volatility is shared by its annual contribution, then it is highly likely that the component is introducing volatility into the general index. By ranking the components by the standard deviations of their monthly contribution and selecting all components with volatility above the 95th percentile (or 90th percentile for a broader interval), we can identify the components with the most substantial influence on the index.

Additionally, the influence of volatility of any component on the index can be displayed graphically by creating a scatter plot that contrasts the monthly or annual standard deviations of the variations of the component with the standard deviation of the annual

---

[4] The annual contribution has the same definition considering t as a particular year and annual variations instead of monthly variations.

[5] The same applies to annual variation – it is the sum of all annual contributions.

variation of the general index. This tool enables us to visualize the tendencies in the volatility of particular components. These tendencies may be transferred to the general index to facilitate an analysis of the behavior of any aggregate index.

III. Sources of variability of aggregate-index numbers

In this section, we study a range of mechanisms that affect the variability of aggregate-index numbers, which are constructed from the aggregation of specific micro-indices that have a specific structure of weights.

The two general sources of variation are aggregate and idiosyncratic (i.e., sector-specific) shocks. Aggregate shocks, such as the application of a minimum wage law, are associated with the macroeconomic structure and its dynamics and affect all sectors[6] or components and therefore the general average. Idiosyncratic shocks give rise to two cases of interest. The first is a shock to sectors that have sufficient weight in the index to affect the general average. The second is a shock to a sector whose connection with the remaining sectors of the economy facilitates its spread and creates the probability of becoming an aggregate shock.

A technical mechanism that allows us to observe these sources of variation is the Gabaix disaggregation methodology (Gabaix, 2011) as used by Foerster, Sarte, and Watson in their structural analysis of the US Industrial Production Index (IPI) (Foerster, Sarte, and Watson, 2011). This decomposition uses mathematical operations that do not alter the aggregate calculation and that are applicable in general to any aggregate index of the following form:

$$AI_t = \sum_{i=1}^{N} m_{it} * w_{it} \qquad (2)$$

Where AI_t is the index in period t[7], m_{it} is the micro-index at the fundamental-aggregation level i in the same period, and w_{it} is the weight of the aggregate i in period t. In general, AI_t is composed of '*N*' sectors or components. On the other hand, the monthly variation of the general index is defined in the following manner[8]:

[6] In this paper, the definition of "sectors" is used as a synonym for any economic unit, the fundamental unit of calculation of the specific index, such as prices, firms, activity of economic sectors, etc.

[7] This period can be monthly, quarterly, or yearly.

[8] This is the sum of the contributions of all components.

$$\Delta AI_t = \sum_{i=1}^{N} \frac{m_{i,t-1} * w_{i,t}}{AI_{t-1}} * \Delta m_{i,t} \qquad (3)$$

The monthly variation of the index can be explained not only by the variations of all components of the index but also by the variation of the relative level of the micro-index with respect to the aggregate index and by the variations of the relative weight of the micro-index. To simplify the previous calculation, the variation of the aggregate index can be rewritten in the following manner:

$$\Delta \mathrm{AI}_t = \sum_{i=1}^{N} \Delta m_{i,t} * \widehat{w}_{i,t} \qquad (4)$$

Where, $\widehat{w}_{i,t}$ is called "conjunctural weight" in reference to the variability of its structure resulting from the historical interaction between the micro-index and the aggregate index.

To simplify notation, we denote G_t and g_{it} as the rate of variation of the general index and of the micro-index, respectively. Thus, expression (4) can be written as the follows:

$$G_t = \sum_{i=1}^{N} g_{i,t} * \widehat{w}_{i,t} \qquad (5)$$

Operating mathematically (5), we obtain the **Gabaix equation** for an aggregate index:

$$G_t = \underbrace{\frac{1}{N}\sum_{i=1}^{N} g_{i,t}}_{\boldsymbol{A}} + \underbrace{\sum_{i=1}^{N}\left(\overline{\widehat{w}}_i - \frac{1}{N}\right) g_{i,t}}_{\boldsymbol{C}} + \underbrace{\sum_{i=1}^{N}\left(\widehat{w}_{i,t} - \overline{\widehat{w}}_i\right) g_{i,t}}_{\boldsymbol{D}} \qquad (6)$$

Therefore,

$$G_t = A_t + C_t + D_t \quad (7)$$

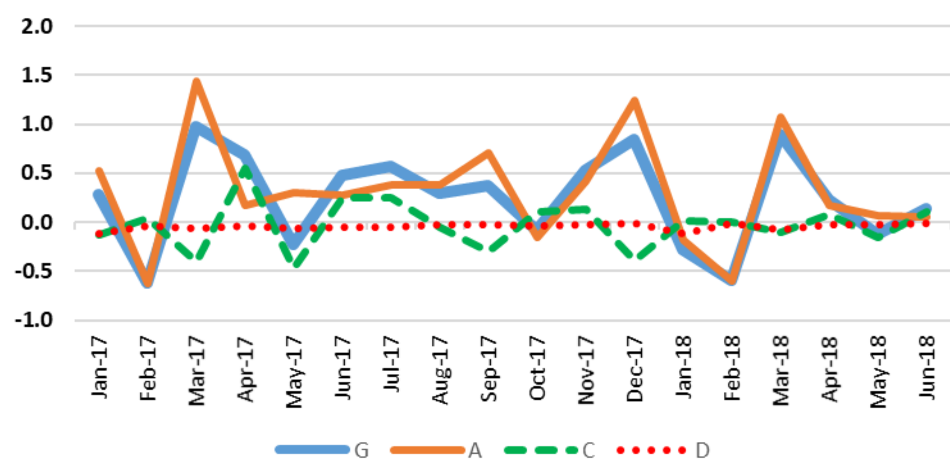
Each component allows us to analyze different phenomena. Component A_t shows the effects of aggregate shocks because the same relevance (weight = 1/N) is assigned to all variations of the sector in the construction of the index. Component C_t shows us how much the general variability is caused by sectors (components) of greatest relevance in the index (with high average conjunctural weighting in regard to 1/N) such that the idiosyncratic shocks in them (whether small or large) will alter the general variation of the aggregate index.

Component D_t is useful when the general index is constructed on the basis of aggregates that constantly change in relative importance[9]. That is, if the sectors themselves are volatile, depending on the variations of $\widehat{w}_{i,t}$, the aggregate index will be affected by them.

Once the structure of components has been established, we can proceed to an analysis based on equation (7). First, we need to identify which of the series components (A, C, D) best replicates the behavior of the general series (G) in order to evaluate whether the general dynamic is the product of the aggregate effects (A), the dynamics of the main sectors (C), or the volatility of its component sectors (D). We can do this exercise by using any of three different processes: contrasting graphs, observing differences in series, or calculating the standard deviation for each component as an indicator of its volatility and then observing which is the closest to the value reported by the general series. With these results, we can demonstrate which series component most influences the general tendency of the aggregate index and which marks the behavior of the phenomena in each specific period.

The following figure illustrates the application of this tool in the identification of dominant series in general and of specific periods.

Figure 1. Gabaix chart of component decomposition: monthly variation of Chile's Labour Cost Index (ICMO) from Jan-17 to Jun-18.



Components	Standard deviation
G	0.483
A	0.557
C	0.258
D	0.030

Source: National Statistics Institute – series decomposition of the Labour Cost Index (ICMO)

The figure above shows that series A closely replicates the dynamic of the aggregate (general) index (G) although in some months the behavior of the aggregate index can almost exclusively be attributed to component C, as can be seen from April to August 2017.

[9] For example, in the CPI we generally have two situations. First, the index is updated every year (chained index) or with certain regularity (mostly every 5 or 10 years).

This is consistent with the figures presented in the complementary table, where the general volatility of G (measured by its standard deviation) has a value closer to the volatility of component A. Nevertheless, the volatility of component C is not negligible although it is less volatile than A. Thus, the general volatility decreased during the period.

We indicated previously that idiosyncratic shocks can result in aggregated effects because of the propagation resulting from the interconnection among sectors (components)[10]. The degree of interconnection among sectors should be identified and measured to show the effect of this interaction on the dynamics.

To identify this effect, we first calculate the average covariance of the variations for a given period within each sector. However, it must be remembered that, when measuring economic phenomena, substitution and complementary relations may coexist (as in the CPI). Thus, the average aggregate of the positive links (complements) and negative links (substitutes) can be compensated, resulting in an average covariance closer to zero. We therefore recommend that the study be complemented with the calculation of the average positive and negative covariance so that the existence of these links can be observed.

To broaden the calculation and avoid the problem of substitutes and complements, Shea proposes a method that starts from the following premise: "if $\mathrm{G_t}$ is the sum of the variations of each sector, then the variance of $\mathrm{G_t}$ will be equal to the sum of the variances of each sector plus the covariance between these" (Shea, 2002). With this premise, if the variability of a series is calculated and then compared with the same result while discounting the covariance, the effect of links between sectors on this variability can be observed. To calculate deviations discounted for covariability (σ^*), use the following formula:

$$\sigma^* = \sqrt{T^{-1} \sum_t \sum_i h_{i,t}^2 \left(g_{i,t} - \bar{g}_\iota\right)^2} \quad (8)$$

Where $h_{i,t} = \widehat{w}_{it}$ for the series G_t, and $h_{i,t} = N^{-1}$ for the component A_t. Cross-comparing the four components (the standard deviations of G_t and A_t with and without covariance), we can identify the importance of not only the links between sectors or components on the variability of both series but also the sectors of greatest conjunctural weight in the variability and the relevance of the interconnections among them.

---

[10] Changes in the price of some commodities can have an impact on other prices of the economy, affecting the CPI dynamics. Gasoline and oil, for example, are used by other agents of the economy as an input for productive processes.

The speed of propagation is an important consideration in the identification of links. In the economy, the effects within certain sectors can spread to others; however, this usually does not happen immediately. Therefore, the time window with which data is measured and studied plays a fundamental role in identifying spillover effects across components. It is, for example, harder to capture this effect in monthly variables than in quarterly variables. On the other hand, large time windows (such as annual series) make clear observations of the propagation phenomenon impracticable because other components of macroeconomic dynamics (associated with aggregate shocks) affect a large part of the sectors equally, leading to spurious correlations.

The previous analysis provides an approximate notion of the existence of propagation; however, the direction of the correlation can only be determined if we have a structural model that describes the aggregate phenomenon from its components (sectors).

Although the analysis of dominant dynamics and interconnections is useful for the economic study of the phenomenon, the identification of volatile or atypical structures is more relevant for statistical offices.

Regardless of whether the component series replicate the behavior of the general series of the index, the presence of unexpected volatility in certain periods should be studied in depth. We can thus identify the causes of volatility and contrast the volatility with information of the market to verify whether the variations are consistent or whether they are due to information errors, operational problems associated with the collection or upload of data, or methodological problems associated with the calculation.

When atypical variations or discrepancies between comparative series are identified, we must focus on the series or specific periods in which they occur. To do this, we propose the development of a ranking system based on the intertemporal variation of each sector and its specific weight. Such as system would improve the interpretation of results[11] by identifying those sectors with the greatest impact on volatility[12].

---

[11] The weights assigned to the component series C and D arise from differences between parameters, so these weights may have negative values, causing some added variations to change sign. Thus, what seems to be a significant fall in D, for example, may be the product of a strong rise in a sector with low conjunctural weight with respect to its average.

[12] The ranking system should consider the most relevant positive and negative variations, the weight of their sectors, and the aggregate result (the product of weight and variation), in order to identify the causes of idiosyncratic shocks and to contrast them with the evidence.

Comparative analysis

In statistical analysis, the following scenarios have been identified: the comparative study of short series (contemporary series) and studies of the long series (historical series). The short series compares two or more series from the same period that measure the same phenomenon but with different methodologies, structures, or frameworks of analysis[13]. We recommend a base analysis with a cross-sectional comparison that identifies discrepancies of the general series in specific periods of time and that evaluates which of the main components (A, C, D) can best explain these discrepancies. For example, a comparison could be made of the dominant effects of Gabaix decomposition on the CPI across a set of countries. Finally, the most relevant components should be identified for the series and period.

On the other hand, studies of historical series (long series) analyze the same series over a long period for methodological or structural discrepancies while using the same framework of analysis. The general base analysis must be applied to the whole series. For comparison, we recommend generating a historical subseries based on general criteria such as economic structural breaks (crises, changes in economic systems, changes of government, and new laws, among others) or statistical structural breaks (methodological changes as modifications in the way of measuring the phenomenon). Making these distinctions, we next analyze the subseries by applying the base analysis to each of them and comparing aggregated results. By using this methodology, an informative analysis can be made with high frequency. In essence, this paper proposes a parsimonious analysis loaded with strong information, with a sunk cost at the beginning of the process but with zero marginal cost once the process has been automated[14].

Scatter plots

Scatter plots are a good complement for these analyses. To construct a scatter plot, first, the variance of each component variation for a specific period is calculated. Second, a scatter plot with the variances and their respective structural weight is drawn. In the graph, the x-axis represents relevance (weight) and the y-axis represents volatility (variance). Next, the means of both measurements are included in the scatter plot; we

---

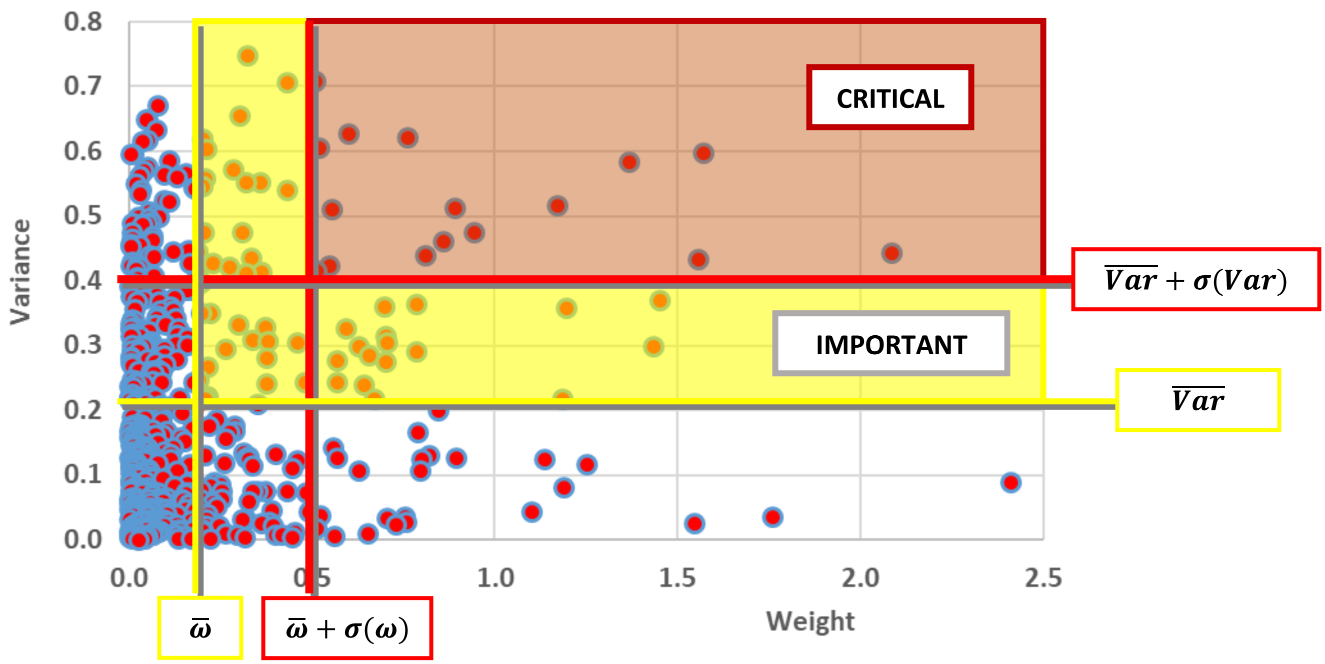
[13] The following are some examples: the comparison of referential vs official series in the period in which the base year for the CPI changes, the comparison between series (measured by different organisms or the entities) of the same phenomenon in a specific economy, and the comparison of the same phenomenon among different economies.

[14] Many types of software for statistical analysis can be used for automation.

recommend a second set of thresholds, defined by these same means plus one standard deviation. For consistency, we also recommend that analyses have several values for an interval.

With this method, two types of components can be identified: *critical components* (located above the upper threshold for both measures) are the components with greater weight and volatility and, therefore, with greater potential impact on the general index; *important components* (located between the thresholds) still have substantial weight and volatility, but less than critical components. An example of a scatter plot is provided in the figure below.

Figure 2. Scatter plot - identification of critical and important components of an index



Note: the example is based on random data.

Source: National Statistics Institute

IV. Limitation

It is important that the analyses described in the previous sections be carried out entirely with the statistical information generated for the calculation of the aggregate indices of interest. This can be an advantage because all requisite information is readily available within the statistical office that conducts the study. However, if no other calculation is made regarding the phenomena or if administrative records are insufficient[15], it may

[15] This results from either the scope or the definition of the metadata not being consistent with or not completely encompassing the phenomenon to be measured.

difficult or even impossible to contrast the information and the dynamics of these indices with empirical evidence, and thus the quality of the measurements is usually questionable, except for those measurements for which the methodology was designed.

These limitations arise from the total dependence of these exercises on the quality of the data itself. Although the proposed analyses enable us to focus on volatile or atypical components to study their dynamics and to review every stage of its calculation (from the survey to the construction of the corresponding micro-index) in detail, the quality of microdata is fundamental for a correct analysis.

Moreover, the proper design and estimation of the structure used to define the associated weights is also fundamental because all the proposed revisions depend almost entirely on these assigned weights, their temporal dynamics, and their interrelation with the variations of the micro-indices between periods.

V. Conclusion

Along with quality microdata and an appropriate methodology, having a battery of tools for statistical analysis is essential for the technical and operational study of the construction of aggregate indices, their dynamics, and the explanation of causal factors[16].

With a tool that focuses and directs the analyses to specific components and sectors, we can address problems quickly, mainly for very short-term indicators (with high frequency of publication[17] and recent reference periods) and for indices calculated with numerous inputs, including baskets of hundreds of economic units.

In technical terms, it is easier to show the contrast proposed by a hypothesis with the evidence of the more volatile components that move the index. Based on this information and from an operational point of view, we recommend focusing on the improvement of collection methods, data uploading, imputation, processing, and the development of the calculation methodology, all of which help to minimize errors and to improve the measurement of the phenomenon.

The holistic methodology, based on the literature, proposed in this paper enables us to perform a parsimonious yet also strong analysis. Although the method is at first intensive in per person hours, it costs approach zero once it has been automated. Therefore, this approach is both parsimonious and cost-efficient.

[16] This becomes even more relevant when an immediate and timely analysis is needed.

[17] For example, the CPI, Unemployment Rates, PPI, and other indices that are published monthly.

VI. Bibliography

Angulo, S. & Gutiérrez, M. A. (2017). *Inflación interanual México se acelera inesperadamente a 6.59 pct en primera mitad de noviembre*. *Reuters*. Recovered from https://mx.reuters.com/article/businessNews/idMXL1N1NT0UL.

Atuk, O. et al. (2013) *Treatment of Seasonal Products and CPI Volatility*. Central Bank of the Republic of Turkey, 13*,* 51-82.

Baldwin, J. & Macdonald, R. (2013). *Fluctuations des prix á l'échelle internacionale selon les indices des prix á la consomation*. Statistique Canada, 11-626-X au catalogue (025).

Bowdler, C., & Malik, A. (2017). Openness and inflation volatility: Panel data evidence. *The North American Journal of Economics and Finance*, *41*, 57-69. doi: 10.1016/j.najef.2017.03.008.

Detmeister, A., & Hulseman, E. (2017). Was There a Great Moderation for Inflation Volatility? *FEDS Notes*, *2017* (2011). doi: 10.17016/2380-7172.2011.

FAO (2012). Voltilité des prix á l'échelle mondiale. Food and Agriculture Organization of the United Unions (FAO). Recovered from

http://www.fao.org/fileadmin/templates/est/meetings/price_volatility/ME260F_Technical_Paper_03.pdf.

Foerster, A., Sarte, P-D. & Watson, M. (2011). Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production. *Journal of Political Economy, 2011,* 119(1). University of Chicago.

Friedman, M. (1977). Nobel Lecture: Inflation and Unemployment. *Journal of Political Economy, 85*, 451–472.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica, 79*(3).

Galará, Hueichapan, Lineros & Montes (2018). *Análisis de volatilidad y descomposición de variabilidad del IPC en Chile*. Instituto Nacional de Estadísticas de Chile (INE), documento de trabajo no 1.

Karras, G. (2015). Low Inflation vs. Stable Inflation: Evidence from the UK, 1688–2009. *Scott J Polit Econ, 62*, 505–517.

Karras, G. (2017). Is the relationship between inflation and its volatility asymmetric? US evidence, 1800–2016. *The Journal of Economic Asymmetries, 16,* 79-86. doi 10.1016/j.jeca.2017.08.002.

McKenzie and Sasin (2007), Migration, Remittances, Poverty, and Human Capital: Conceptual and empirical challenges. World Bank Policy Research, Working Paper 4272.

Shea, J. (2002). Complementarities and Comovements. *Journal of Money, Credit and Banking, 34*(2), 412-434.