ARGENTINE REGIONS BASED ON DYNAMIC CRITERIA*

Sigal, Facundo** Camusso, Jorge*** Navarro, Ana Inés****

ABSTRACT

In this paper, we use the Stock and Watson methodology to estimate economic coincident indexes for each of the twenty-four Argentine provinces. We extract business cycle components from these indexes using the Christiano-Fitzgerald filter and then we group Argentine provinces into regions according to the dynamic behavior of their economies, by applying a Ward-like hierarchical clustering algorithm under different scenarios. We found very varied results, but with a certain regularity that can be highlighted, since there are some groups of provinces that are clustered together in every scenario. However, neither scenario produce any regionalization similar to the statistical regions determined by the Argentine Institute of Statistics and Censuses (INDEC). When we assign equal weights to the contiguity and business cycle dimensions in the clustering process, the resulting clusters are very similar to what we can expect from a economic regionalization, that is, complete contiguity, business cycle similarities and a relatively balanced size. Another particularly interesting result is that the provinces that concentrate the country's agro-industrial production and exports (*Córdoba* and *Santa Fe*) appear together in almost all the clustering procedures As a whole, the results show that regionalization based on static criteria may not be the most appropriate approach when dynamics matter.

JEL Classification: *C19, E32, R11*

Keywords: *Business Cycles, Coincident Indicators, Clusters, Regional Economic Activity*

* The authors want thank Inés Buchara for proofreading this paper.
** Departamento de Economía, Facultad de Ciencias Empresariales, Universidad Austral, Paraguay 1950, S2000FZF Rosario, Argentina. E-mail address: fsigal@austral.edu.ar.
*** Departamento de Economía, Facultad de Ciencias Empresariales, Universidad Austral, Paraguay 1950, S2000FZF Rosario, Argentina. E-mail address: jcamusso@austral.edu.ar.
**** Departamento de Economía, Facultad de Ciencias Empresariales, Universidad Austral, Paraguay 1950, S2000FZF Rosario, Argentina. E-mail address: anavarro@austral.edu.ar.

1. INTRODUCTION

The subnational heterogeneity among the regions in which Argentine provinces are generally classified has been studied in several dimensions such as unemployment (Féliz, Panigo and Pérez, 2000; Figueras, Díaz Cafferata, Arrufat, Descalzi and Rubio, 2002; Galiani, Lamarche, Porto, and Sosa Escudero, 1999, 2005), returns to education (Paz, 2009) and distributive inequalities (Zacaria and Zoloa, 2005). For the most part, these subnational studies are limited to the analysis of a province or a group of them, which make up a certain region. One of the exceptions is the study by Blanco, Elosegui, Izaguirre, and Montes-Rojas (2019), in which the authors analyze the possible asymmetric effect of monetary policy on employment, both at the provincial level and for all economic regions. Another dimension of the analysis of subnational economies is the study of their economic cycles, trying to link these local cycles with the national ones. Garegnani and Di Gresia (1999) analyze whether the provincial cycles are correlated with the national one, Lapelle (2015) shows the degree of synchronization between the economic cycle of the urban area of Rosario with the provincial and national cycle, and Muñoz and Trombetta (2014) seek to construct a matching national index and its counterpart for each of the twenty-four Argentine provinces.

On the other hand, the estimation of coincident synthetic indexes at the provincial level is quite widespread in Argentina. Based on the pioneering studies by Jorrat (2003, 2005) for the Province of *Tucumán*, there have been several applications to other provinces and geographical areas, including *Córdoba* (Michel Rivero, 2007), *Santa Fe* (D´Jorge, Cohan, Henderson and Sagua, 2007), *Misiones* (Heredia and Alvarez, 2017), and *Rosario* (Lapelle, 2015), which are based on the traditional methodology of the National Economic Research Office (NBER). In Berardi, Navarro and Uría (2010), a coincident synthetic index for the Province of *Santa Fe* was presented but following the Stock and Watson (SW from now on) methodology (Clayton-Matthews and Stock, 1998/1999; Stock and Watson, 1989, 1991). Then, in Navarro and Sigal (2012), several modifications were introduced to this index seeking to adapt it to changes in available information. In this line, some provincial statistical institutes -in collaboration with other institutions- estimate economic synthetic indicators for their provinces by using the SW methodology, like ISAE[1] and ISAEER[2] for the provinces of *Santa Fe* and *Entre Ríos*, respectively. However, there have been no attempts to measure regional cycles, using the SW methodology, except one incipient program that is being developed by a group of researchers

[1] *Indicador Sintético de Actividad Económica.*
[2] *Indicador Sintético de Actividad Económica de la Provincia de Entre Ríos.*

at *Universidad Austral* for the provinces grouped in a politically defined region called "*Región Centro*" (*Santa Fe, Córdoba* and *Entre Ríos*) due to its location at the center of the country.

Besides these studies, the Argentine Institute of Statistics and Censuses (INDEC) groups the Argentine provinces into five or six regions.[3] This approach resembles the Bureau of Economic Analysis (BEA) one, which since the 1950s has grouped the fifty states of the United States into eight regions based primarily on cross-sectional similarities in their socioeconomic characteristics. As Crone (2005) points out, economists have tended to use the BEA regions to examine various regional trends and cycles assuming that the regions are properly defined for this purpose. [4]

However, when the analysis focuses on business cycle phenomena, multistate regions based on similarities at a given moment may not be the appropriate set of observations. For this reason, Crone (1998/1999, 2005) groups the U.S states into regions based on similarities in their dynamic economic behavior, using a set of coincident indexes -estimated from a SW-type model- and the business cycle components extracted from them. Likewise, Miller and Sabbaresse (2012) estimate a state space model for Georgia and test if it produces reasonable forecasts for the real growth rates of the gross domestic products of Alabama, Florida, Georgia, North Carolina, South Carolina, and Tennessee, for 1991 through 2008. Brida, Garrido and London (2012) study the economic performance of the Argentine provinces –using real per capita Gross Domestic Product (GDP)- during the period 1961-2000. Applying hierarchical clustering techniques, they detect groups of provinces with similar performance, but their regions do not correspond with INDEC's regions at all. In the same line, Sigal, Camusso and Navarro (2021) apply clustering methodologies to group Argentine provinces into regions according to the dynamic behavior of their economies. The authors estimate coincident indicators for each province by using the SW model and then they implement three clustering techniques: k-means procedure with and without geographic contiguity constraint in the grouped provinces, and agglomerative hierarchical grouping.

This study continues the analysis of Sigal, Camusso and Navarro (2021) but using the business cycle component of the coincident indexes to cluster the provinces into regions. Besides, here the clustering algorithm implemented allow us to control the relative weights given to the economic and geographical contiguity dimensions, which offers a more flexible approach to

---

[3] *Área Metropolitana de Buenos Aires* (AMBA), *Región Cuyo*, *Región Noreste Argentino* (NEA), *Región Noroeste Argentino* (NOA), *Región Pampeana* and *Región Patagonia*.
[4] See Crone (2005) for a complete list of studies.

determine the composition of the regions. The main objective of our work is to determine which groups of provinces have a similar economic behavior adopting a dynamic criteria and if it changes when adding contiguity restrictions. Particularly, we are interested in analyzing if the provinces that make up the so-called “*Región Centro*” have a common business cycle, since most of the country's agro-industrial production and exports are concentrated there. We also consider interesting doing a comparison between our results and INDEC’s regionalization.

This study uses information from a set of series at the state and metropolitan area level for the twenty-four Argentine provinces obtained from different official sources and sectoral chambers that have been compiled and seasonally adjusted. Based on Crone (1998/1999, 2005), composite indexes are estimated for each province following the SW methodology, which is especially suitable for constructing synthetic indicators for provinces and regional economies because it does not require having a timely series of regional GDP, due to the assumption that the "state of the economy" underlying is unobservable. We also extract the cyclical component of the indexes, by using the Christiano-Fitzgerald (CF from now on) filter. Then, we group the provinces into five regions by using, separately, both indicators and applying a Ward-like hierarchical clustering algorithm.

Broadly this paper contributes to the field of empirical analysis that studies regional economies performance and, specifically, into the empirical study of the regions’ composition from a dynamic perspective of economic similarity of the provinces that comprise them. One contribution of this paper is the application of clustering methodology to select the provinces that make up the different regions according to the dynamic behavior of their economies. In turn, when applying this analysis to Argentina, it allows us to observe whether the regional disparity found in developed countries between the definitions at a point in time and those that arise from analyzing cyclical behavior of the provinces that make up regions, is also verified in an emerging economy. An additional contribution consists of obtaining coincident indexes for the twenty-four provinces, with a statistically precise methodology, which until now has been used exclusively to calculate indexes in a few bunches of provinces.

We found very varied results, but with a certain regularity that can be highlighted, since there are some groups of provinces that are clustered together in every scenario. However, neither scenario produce any regionalization similar to the statistical regions determined by INDEC, which illustrates that regionalization based on static criteria may not be the most appropriate approach when dynamics matter. Another particularly interesting result is that the provinces that concentrate the country's agro-industrial production and exports (*Córdoba* and *Santa Fe*) appear together in almost all the clustering procedures.

The outline of the paper is as follows. In Section 2 data and methodology for building coincident indexes for each province and the procedures to combine them in regions based on the similarity of their dynamic economic behavior are presented. Section 3 show the results of our cluster analysis and the comparison of these respect to the official regionalization of INDEC, with particular interest in the analysis of *Región Centro*. Finally, the study closes in Section 4 with a global evaluation of the results.

2. DATA AND METHODOLOGY

2.1. Data

When subnational coincident indexes are used to compare the state of the economy across the whole country, they must have a certain degree of consistency. In particular, the indexes should be constructed from the same set of indicators for each jurisdiction (Crone and Clayton-Matthews, 2005). In addition to this reasonable methodological justification, the use of the same series to build up the coincident indexes for the different provinces has operational advantages, since it restricts the data search for candidate indicators to a unique common set. The data collection is also facilitated by the fact that most of empirical studies use SW methodology to extract the state of the economy from a bunch of few “standard” economic series, like employment, tax revenues, among others (Clayton-Matthews and Stock, 1998/1999; Crone and Clayton-Matthews, 2005). Once identified this series set, it is necessary to apply some additional economic and statistical criteria to select the indicators that finally will be components of the coincident index. Several criteria have been proposed in the literature to select the variables better suited for that purpose.[5]

The data used herein comes from the database named *Indicadores Regionales* built by IDIED-*Universidad Austral* which compiles public and private economic information for the twenty-four provinces of Argentina. For each province we have a set of series that could be classified in eight broad categories: employment, consumption, investment, sector production, energy, prices and wages, public finance, and financial services. On these, for selecting the series to be incorporated in the coincident index we apply the standard economic and statistical criteria.

[5] See, for example, Crone (1998/1999) and Muñoz and Trombetta (2014).

First, we verify that the length of the series reaches the minimum amount of one hundred twenty months for each indicator at provincial level and that these variables have economic significance. For the rest of criteria, we do the tests by using the series at the national level in order to simplify the procedure and to ensure a common set of variables for the estimation of the provincial coincident indexes. Thus, secondly, to check for smoothness, we construct the boxplots for each of the series that pass the first selection stage. Third, we analyze the business cycle properties of the series. Since a key notion of the business cycle is that fluctuations are common across all sectors, the series should exhibit a consistent timing pattern over time as a coincident indicator and fit well with the national business cycle. Following Stock and Watson (1989), for each variable we analyze both the basic univariate characteristics and the co-motion properties with the aggregate activity. For this purpose, we need a reference series for the behavior of the entire national economy. Ideally, the reference series would be the Gross State Product (GSP), but this is reported on an annual basis and with a lag of several years. Thus, we use EMAE[6], a monthly estimate of Argentine economic activity produced by INDEC. The criterion of consistent timing is tested by calculating the co-movements -through contemporaneous correlations- between each series and EMAE. Since we are looking for strong indicators, we prefer the series with high correlation, notwithstanding, *a priori* we do not discard any of the series based on a lower level of correlation.

There are eighteen series that best accomplished all the tests for being included in the coincident index. All these series were included in the estimation of successive versions of the index. The model finally selected was the one that produced the best index in terms of smoothness and conformity. This means that it satisfies the required assumptions of SW model. The series finally included in the best model are “registered employees”, “large users of electric power provided”, “motorcycle sales”, “grade 2 gasoil provided”, “cement sales” and “gasoline sales”.[7]

2.2. Methodology

2.2.1. Background in the empirical literature

For a frequent and timely monitoring of the “state of the economy”, the empirical literature usually estimates a composite index of coincident indicators since, by aggregating the movements of several key economic indicators, it represents a single summary statistic that

[6] *Estimador Mensual de Actividad Económica.*

[7] A great advantage of the six selected variables is that all of them are published almost simultaneously. Another advantage is that these series are published by the national offices of each area, so we can trust that the series for each state are calculated with the same methodology.

tracks the economy's current state. The National Bureau of Economic Research (NBER) based in Mitchell (1927) and Burns and Mitchell (1938, 1946), developed one of the first methodologies to estimate this type of composite indexes, which has been widely used since then.[8] The NBER approach it is easy to understand at conceptual level and the calculations it involves are not complex. However, since business cycle is not precisely defined, it is not entirely clear what the index is really measuring. Moreover, the vector of weights used to aggregate the set of data series and construct the index is exogenous and arbitrary, so it is not based in a statistical or economic optimization process.[9]

Given the limitations of the traditional approach, in the late 1980s NBER's economists developed new methodologies to construct coincident and leading indexes (Stock and Watson, 1989, 1991). The SW methodology, mathematically based in probabilistic state space models, rest on the hypothesis that the observed co-movements in indicator series can be captured by a single unobserved variable that represents the unknown "state of the economy". This methodology assumes that each series has a component attributable to the unique variable unobserved and a particular or idiosyncratic component. The problem to be solved in this approach is to estimate the current state of the economy, namely the common element in the fluctuations of each time series. The main contribution of their research was the use of a statistical technique called the Kalman filter for the estimation of the optimal weights on the component indicators. In contrast to the traditional composite index methodology which applies equal weights once the volatility in each series is standardized, in SW approach the weights of the component series are estimated, by using a maximum likelihood procedure, in such a way that best identifies the single underlying factor, which is time dependent and best represents the co-movement in the components. Hence, from a statistical point of view this approach is more rigorous than the former NBER's methodology (Orr et al., 1999) and the index also provides a better definition of the underlying state of the economy from a mathematical perspective. This methodology is currently being used for several estimations both at state or regional level (e.g., Crone, 1998/1999; Crone and Clayton Matthews, 2005; Méndez, 2007; Orr et al., 1999; Orr et

[8] In Argentina, this traditional approach was widely applied in academic works for the measurement of the evolution of subnational and national economic activity (Arredondo *et al.* 2009; Jorrat, 2003, 2005; Michel Rivero, 2007).

[9] Burns and Mitchell (1938, 1946) define the coincident indicators as coincident with the "reference cycle," that is, with the broad-based swings in economic activity known as the business cycle. This definition is intuitively appealing but, as Burns and Mitchell (1946) recognized, lacks precise mathematical content, so is unclear what conclusions one should draw from swings in the index.

al., 2001; Tebaldi and Kelly, 2012) and at national level (e.g., Dias, 1993; Fukuda and Onodera, 2001; Hall and Zonzilos, 2003; Reklaite, 2011).

2.2.2. Stock and Watson's Model

The underlying assumption of SW`s methodology is that in addition to stochastic component that represents an idiosyncratic movement, each macroeconomic variable has a common unobserved component, called "state of the economy". The authors formulate a linear model in the unobserved variable for estimating the common component using the Kalman filter to build the likelihood function and to obtain the maximum likelihood estimators of the parameters of the model. Following Clayton-Matthews and Stock (1998/1999), the structure of the model as applied here is:

$$\Delta \mathbf{x_t} = \boldsymbol{\beta} + \gamma(\mathrm{L})\Delta \mathbf{c_t} + \boldsymbol{\mu}_\mathbf{t} \tag{1}$$

$$\mathrm{D}(\mathrm{L})\boldsymbol{\mu}_\mathbf{t} = \boldsymbol{\varepsilon}_\mathbf{t} \tag{2}$$

$$\varphi(\mathrm{L})\Delta \mathbf{c_t} = \delta + \eta_t \tag{3}$$

where $\Delta \mathbf{x_t}$ is a Gx1 vector of observable series in first-difference log form to achieve stationarity and $\mathrm{var}(\eta_t) = 1$. A scalar latent stationary series that is common to the G observable series is captured by $\boldsymbol{\Delta c_t}$ which, in this context, can be interpreted as deviations from the average growth rate of the economy or, alternatively, as the growth rate of the unobserved state of economy. This component follows an autoregressive moving average (ARMA) process and enters in the equation (1) with different lags and weights. The $\boldsymbol{\mu}_\mathbf{t}$ vector, called the idiosyncratic portion, consists of G mutually uncorrelated, mean zero, stationary ARMA processes. The Gx1 vector $\boldsymbol{\varepsilon}_\mathbf{t}$ and the scalar η_t comprise Gx1 mutually uncorrelated white noise processes. The symbol L is the lag operator, i.e., $\mathrm{L^k x_t = x_{t-k}}$. The lag polynomial matrix $\mathrm{D(L)}$ is assumed to be diagonal, hence $\boldsymbol{\mu}_\mathbf{t}$ in each of the G series in equation (2) are serially uncorrelated among them. The parameters of the model can be expressed as follows:

$$\gamma(\mathrm{L}) = [\gamma_1(\mathrm{L}),\gamma_2(\mathrm{L}),...,\gamma_G(\mathrm{L})]`, \text{ where } \gamma_g(\mathrm{L}) = \gamma_{g0} + \gamma_{g1}\mathrm{L} + \gamma_{g2}\mathrm{L}^2 + \cdots,\ \ g = 1,\dots,G. \tag{4}$$

$$\mathrm{D(L)} = \mathrm{diag}[(\mathrm{d_1(L)},\dots,\mathrm{d_G(L)}]`, \text{ where } \mathrm{d_g(L)} = 1 - \mathrm{d_{g1}L} - \mathrm{d_{g2}L^2} - \cdots,\ \ g = 1,\dots,G. \tag{5}$$

$$\varphi(\mathrm{L}) = 1 - \varphi_1\mathrm{L} - \varphi_2\mathrm{L}^2 - \cdots \tag{6}$$

$$\sum = \mathrm{cov}(\,|\boldsymbol{\varepsilon}_\mathbf{t}, \eta_t|) = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_G^2, \sigma_\eta^2) \tag{7}$$

The system of equations (1)-(3) is estimated by maximum likelihood procedure. It is accomplished by representing the system in state space form, using the Kalman filter. Following the literature, we normalize each of the coincident series x_{it} by subtracting its mean difference and dividing by the standard deviation of its differences. This identifying restriction constrains the constant $\boldsymbol{\beta}$ to be zero. Therefore, the obtained Kalman filter, denoted by $\boldsymbol{\Delta c_{t/t}}$, can be considered as a composite coincident index of the economic activity, constructed using data on the coincident variables available through time t.

Finally, following Crone and Clayton-Matthews (2005), we perform a calibration of the coincident index of each province transforming its first two moments of growth -the average growth rate and the average deviation tendency- in the first two moments of a selected indicator. Ideally, we should calibrate each coincident index with the corresponding GSP series, but, given the data limitations about GSP mentioned above, we use EMAE. By doing this, we ensure a certain consistency in the comparison of the different estimated indexes.

2.2.3. Business cycle extraction

Since, in the spirit of Crone (2005), we will group provinces not only based on their SW coincident index, but also based on their business cycle component, we need to implement a cycle extraction procedure. This author uses the bandpass filter of Baxter and King (1999), BK from now on, which is a linear filter that eliminates trend components and high frequency irregular components, but retains the intermediate cyclical components, specifying the business cycle as fluctuations within specific ranges. Indeed, BK filter is an approximation of an ideal bandpass filter (Larsson and Vasi, 2012). However, we decided to use the bandpass filter of Christiano and Fitzgerald (2003) to extract cyclical component from coincident indexes. This filter is based on the same principles of BK method, but it has the advantage of work well on a larger class of time series, converges in the long run to the optimal filter, and in real time applications outperforms the BK filter. Also, CF filter uses the whole time series for the calculation of each filtered data point (Nilsson and Gyomai, 2011).

The CF filter assumes that the time series y_t follows a random walk without drift. The method formulates the detrending and smoothing problem in the frequency domain (Larsson and Vasi, 2012; Nilsson and Gyomai, 2011), with a cyclical component c_t that is estimated as follows:

$$c_t = b_0 y_t + b_1 y_{t+1} + \cdots + b_{T-1-t} y_{T-1} + \tilde{b}_{T-t} y_T + b_1 y_{t-1} + \cdots + b_2 y_{t-2} + \tilde{b}_{T-1} y_1$$

where:

$t = 3,4, \ldots, T-2$

$$b_j = \frac{\sin(jc) - \sin(ja)}{\pi j}, j \geq 1$$

$$b_0 = \frac{c - a}{\pi}, a = \frac{2\pi}{p_h}, c = \frac{2\pi}{p_l}$$

$$\tilde{b}_k = -\frac{1}{2} b_0 - \sum_{j=1}^{k-1} b_j$$

p_l and p_h are the cut off lengths of the cycle. This implies that cycles that are longer than p_l but shorter than p_h are preserved in the cyclical component c_t. With monthly data, $p_l = 18$ and $p_h = 96$. It is important to note that the CF filter assign a different weight on each observation, so it is not symmetric.

2.2.4. Cluster analysis

There are several techniques for grouping a set of n objects or observations into K clusters, using the information contained in a set of variables of interest. In a study similar to ours, Crone (2005) applies k-means clustering procedure to group U.S states into regions based on similarities in their business cycles components. However, this clustering approach does not impose ex-ante geographical contiguity constraints for the provinces that make up a cluster, but contiguity is introduced ad-hoc through the inclusion of proximity variables. Although there are modifications of k-means method in which the contiguity of observations is an explicit constraint (Costanzo, 2001), it generally does not allow to control the relative weight that this constraint has in clustering procedure.

Instead, in our study we use the method of Chavent et al. (2018), consisting of a Ward-like hierarchical clustering algorithm including spatial/geographical constraints. Basically, the method consists of using two dissimilarity matrices, D_0 and D_1, which represent the dissimilarities in the space of characteristics (in our case, the state of the economy of the provinces or their business cycle, i.e., an economic dimension) and in the space of the restriction (geographical contiguity). To quantify the relative weight that these two dimensions have in the clustering process, a parameter $\alpha \in [0,1]$ is used, whose value can be arbitrarily selected by the user, or one can use the value that generates the same loss of information in both matrices ("optimal value" of the parameter). The basic aspects of the methodology of Chavent et al. (2018) are explained below.

Consider a set of n observations, for which w_i represents the weight of the i-th observation ($i = 1,2,\dots,n$).[10] Define by $D_0 = [d_{0,ij}]_{n\times n}$ and $D_1 = [d_{1,ij}]_{n\times n}$ two dissimilarity matrices, which represents, respectively, distances in the economic and geographic dimensions. Thus defined these matrices, the parameter α allows to control the relative importance of the spatial constraint in the clustering process.

Let $I_\alpha(C_k^\alpha)$ be the mixed pseudo inertia[11] (or simply mixed inertia) of the cluster C_k^α, defined as:

$$I_\alpha(C_k^\alpha) = (1-\alpha)\sum_{i\in C_k^\alpha}\sum_{j\in C_k^\alpha}\frac{w_i w_j}{2\mu_k^\alpha}d_{0,ij}^2 + \alpha\sum_{i\in C_k^\alpha}\sum_{j\in C_k^\alpha}\frac{w_i w_j}{2\mu_k^\alpha}d_{1,ij}^2$$

Where $\mu_k^\alpha = \sum_{i\in C_k^\alpha} w_i$ is the weight of C_k^α, while $d_{0,ij}$ y $d_{1,ij}$ are the normalized dissimilarities between the observations i and j. [12] Note that $I_\alpha(C_k^\alpha)$ is a convex combination of the pseudo inertia of the economic and geographical dimensions. The smaller $I_\alpha(C_k^\alpha)$ is, the more homogeneous are the observations of the cluster.

On the other hand, the mixed pseudo within-cluster inertia of a partition in K clusters $P_K^\alpha = (C_1^\alpha, C_2^\alpha, \dots, C_K^\alpha)$ is the sum of the mixed inertia of its clusters:

$$W_\alpha(P_K^\alpha) = \sum_{k=1}^{K} I_\alpha(C_k^\alpha)$$

Given a value of α, the clustering procedure is a sequential process. First, the initial number of clusters is $K = n$. Then, in each subsequent step two clusters A and B are grouped such that the new partition has a minimum mixed within-cluster inertia. That is, the optimization problem to be solved is:

$$\arg\min_{A,B\in P_{K+1}^\alpha} I_\alpha(A\cup B) - I_\alpha(A) - I_\alpha(B)$$

At each step, the algorithm groups two clusters so that $\delta_\alpha(A,B) \equiv W_\alpha(P_{K+1}^\alpha) - W_\alpha(P_K^\alpha) = I_\alpha(A\cup B) - I_\alpha(A) - I_\alpha(B)$ it is minimal. That is, the difference in mixed within-cluster inertia between two successive partitions is minimized. This sequential procedure generates a hierarchical set of partitions $\{P_n, \dots, P_K^\alpha, \dots, P_1\}$[13] that is represented graphically by a dendogram

[10] Since a priori we do not have reasons to give different weights to the observations, we apply uniform weights equal to $\frac{1}{n}$.
[11] Pseudo inertia generalizes the inertia to the case of dissimilarity data (Euclidean or not). The formula presented correspond to the general case, but we use Euclidean distances.
[12] Dissimilarities are normalized so that they vary between 0 and 1.
[13] Note that the initial and final partition do not depend on the value of α.

in which the height of the cluster $A \cup B$ is given by $\delta_{\alpha}(A, B)$. The process ends when K or one cluster is formed. In our case, since we will compare our clustering results with INDEC's regionalization, we select $K = 5$ as the cut-off point. [14]

Finally, it is important to note that a key point in the algorithm is the selection of the value of α. While the user can determine the value of this parameter arbitrarily, based on the relative weight they want to give to the spatial dimension, Chavent et al. (2018) also propose a procedure that, given a value for K, best compromises between loss of economic and loss of geographical homogeneity. Applied to our case, this implies determine the value α that increases the geographical homogeneity of a partition in K clusters without negatively affecting homogeneity in terms of the state of the economy or business cycle.[15] Likewise, although we will use the "optimal" value obtained α in this way, we will also arbitrarily choose values to give it greater weight relative to geographical contiguity or economic dimension.

3. RESULTS

In this section, we firstly present the results obtained from the clustering process and the resultant regions based on the coincident composite indicators estimated for each province using the SW methodology. In the second place, we show the results from clustering provinces based on their business cycle component, extracted from the coincident index by using the CF filter. In both approaches, we consider three scenarios, according to the relative weight that we set for the contiguity dimension in the clustering process.

The results from the first clustering approach are shown in Figure 1. The first clustering scenario (left panel) groups provinces only by the economic dimension, i.e., the estimated coincident index. The second uses an "optimum" mixture of economic similarities and geographic proximity and the last one combines equally the coincident index and contiguity between provinces. As expected, the scenario that only takes in account the state of the economy ($\alpha = 0$), generates groupings of provinces that are not necessarily geographically close. For example, *Tierra del Fuego* is grouped with some provinces from the center and north of the country, as *San Luis*, *La Pampa*, *Santa Fe* and *Chaco*. As the weight of the coincident index in clustering decreases, the groups become increasingly compact, but without reaching complete contiguous regions. When we use optimum alpha ($\alpha = 0.375$), some contiguous regions are built. There is a compact Northern Region including *Salta*, *Chaco* and *Formosa*. *Santa Fe*, *Córdoba* and

[14] INDEC*'s* statistical regions are six, but one of them is *Área Metropolitana de Buenos Aires* (AMBA). Because of this regionalization, the selected number of clusters is five.
[15] For more technical details on determining the "optimal" value of α, see Chavent et al. (2018).

Buenos Aires belongs to a compact Central Region, among other provinces. But some isolated provinces can be found, as *Jujuy, Tierra del Fuego* and *Ciudad de Buenos Aires* that are clustered together. In the scenario that gives equal weights to economic and geographical dimensions, *Ciudad Autónoma de Buenos Aires* and the Province of *Jujuy* are clustered with the three southernmost provinces of the country. This scenario produces similar results as the second one. Both scenarios present four contiguous cluster and a scattered one, but the only cluster that has no complete contiguity in the third scenario is divided in three compact regions, while in the second scenario this cluster is divided in five compact groups of provinces. Also, if we focus on the *Región Centro*, it is interesting to note that, in both scenarios, the provinces of *Córdoba* and *Santa Fe* belong to the same cluster

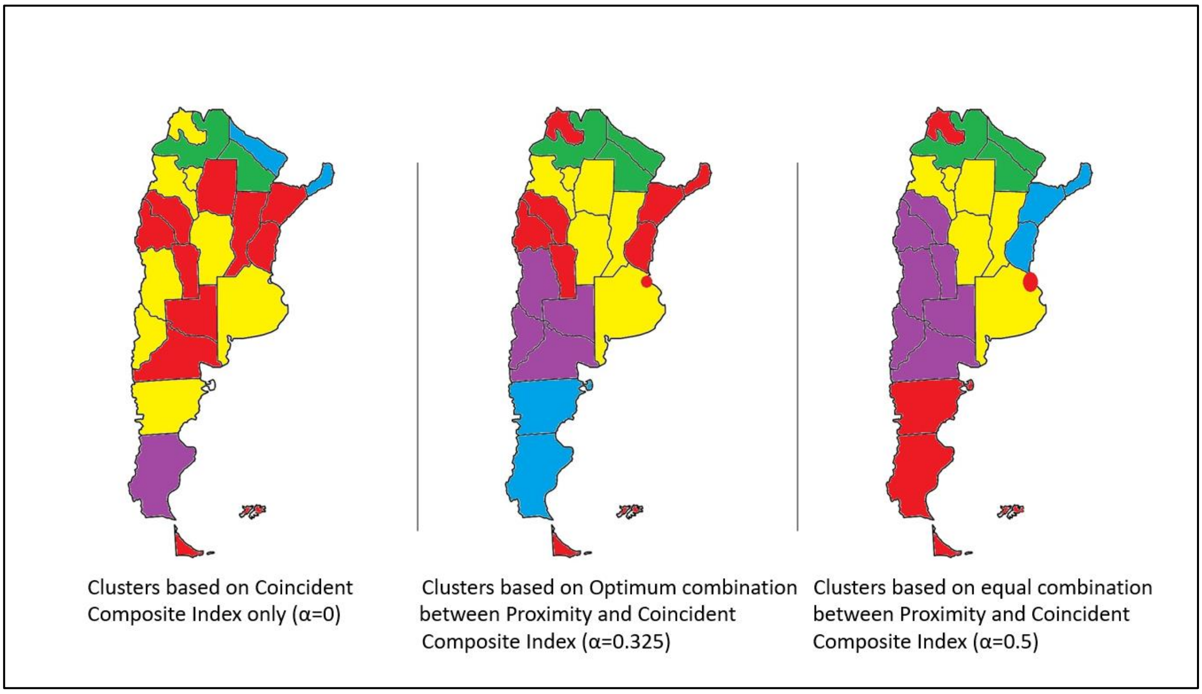

Figure 1: Resulting clusters applying different combinations of contiguity and economy.

<u>Source:</u> own elaboration with data from different official sources and sectoral chambers.

As a second approach, business cycles components were extracted from the coincident indexes by applying the CF filter. Then, using these resulting cycles, the provinces were grouped using the same clustering technique and the three scenarios described in the previous paragraphs. The results are shown in Figure 2. The application of clustering technique based on business cycle and contiguity generates groupings of provinces that are different from the regions shown in Figure 1. Indeed, the clusters of Figure 2 seem to be more compact. Some regularities can be found, since *Santa Fe* and *Córdoba* are clustered together in the three scenarios, as well as Central Western provinces as *Mendoza, San Juan, San Luis* and *La Rioja*. Also, it is interesting to note that the equally combined clustering is the only one that produces five compact regions without isolated provinces. These clusters are similar to what we can expect from economic regionalization, that is, complete contiguity and business cycle similarities. We can also find balanced regions based on their size, since -except for one region- most of them are made up of three or four provinces.

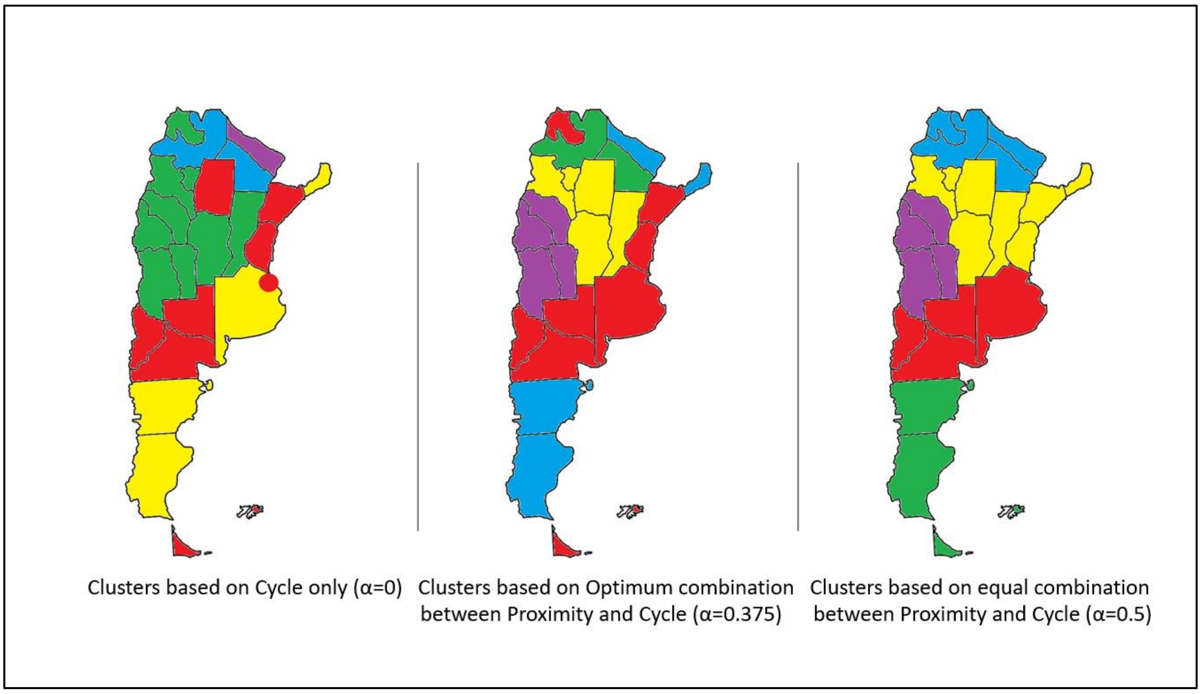

Figure 2: Resulting clusters applying different combinations of contiguity and business cycle.

Source: own elaboration with data from different official sources and sectoral chambers.

Analyzing clustering scenarios as a whole -both approaches- we could say that very varied results are found, but with a certain regularity that can be highlighted. There are some groups that are clustered together in every scenario: *Salta* and *Chaco; Catamarca*, *Córdoba* and *Tucumán; La Rioja*, *San Juan* and *San Luis;* and *Corrientes* and *Entre Ríos*. Likewise, the most extreme provinces are geographically isolated in many scenarios. *Jujuy* is clustered with no contiguous province in 5 out of 6 scenarios, and *Tierra del Fuego* in 4 out of 6. The Province and the City of *Buenos Aires* appear grouped only in half of the scenarios.

It is natural to compare our results with the official regionalization of INDEC. As mentioned before, the scenario that produces results that can be interpreted as an economic regionalization is the one that clusters provinces using an equal combination of business cycle and geographic contiguity. Neither of the five clusters is exactly the same as any of INDEC's statistical regions. This result allows us to observe that the regional disparity found by the literature in developed countries between the static and dynamic definitions of regions, is also verified in an emerging economy, like Argentina. Some of INDEC's Northeastern and Northwestern provinces are grouped together in a Northern Region (*Salta*, *Jujuy*, *Chaco* and *Formosa*). In our results, a large central region is built from three INDEC's Northwestern provinces (*Catamarca*, *Tucuman* and *Santiago del Estero*), three provinces from *Región Pampeana* (*Santa Fe*, *Cordoba* and *Entre Ríos*) and two Northeastern provinces (*Corrientes* and *Misiones*). The three southernmost provinces are grouped together, as in INDEC's *Patagonia*, but without *Neuquén* and *Río Negro*, which are clustered with *La Pampa* and both Province and City of *Buenos Aires*.

On the other hand, it is interesting to focus our attention on the provinces that concentrate agro-industrial production and exports in Argentina, that is, *Córdoba* and *Santa Fe*. These provinces, in addition to *Entre Rios*, make up the *Región Centro*, which is a regional integration bloc created from a political decision of their governors. Although, our results suggest that, beyond geographical proximity, *Córdoba* and *Santa Fe* have many similarities in their business cycles, which are stronger than those shared, separately, with the Province of *Entre Ríos*. Besides, *Córdoba* and *Santa Fe* are grouped in five of the six scenarios, the exception being only the clustering based only in coincident index. Secondly, *Entre Ríos* qualifies in the same group as *Santa Fe* twice, and only once with *Córdoba*. Thus, the grouping estimates kept all the *Región Centro* provinces in the same group only once, corresponding to the clustering with equal distribution between business cycle and contiguity. As can be seen in Figure 3, the movements in the state of the economy (SW index) are similar for *Santa Fe* and *Córdoba*, while *Entre Ríos* moves a little away from the pattern.

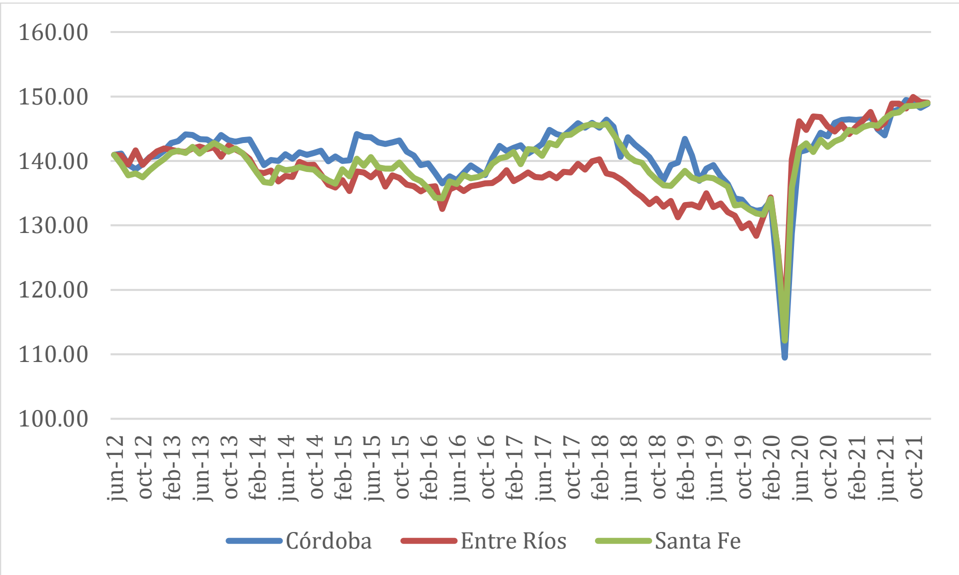

Figure 3: State of the economy (ILCE) of *Región Centro* provinces from jun-12 to dec-21.

Source: own elaboration with data from different official sources and sectoral chambers.

Those differences are more obvious when we make focus on business cycles (Figure 4). The cycles extracted by CF filter show that *Entre Ríos* has a steeper cycle than *Córdoba* and *Santa Fe*. In *Entre Ríos*, early 2018 peak is sharper than in the other provinces. The latest business cycle trough appears one year later in *Entre Ríos* and in a steeper way than in *Córdoba* and *Santa Fe*. Given the latter results, the case of *Región Centro* is a useful and illustrative precedent to show that regionalizing the country ignoring business cycles and only considering political or geographical issues may not be the most appropriate choice when formulating targeted economic policies.

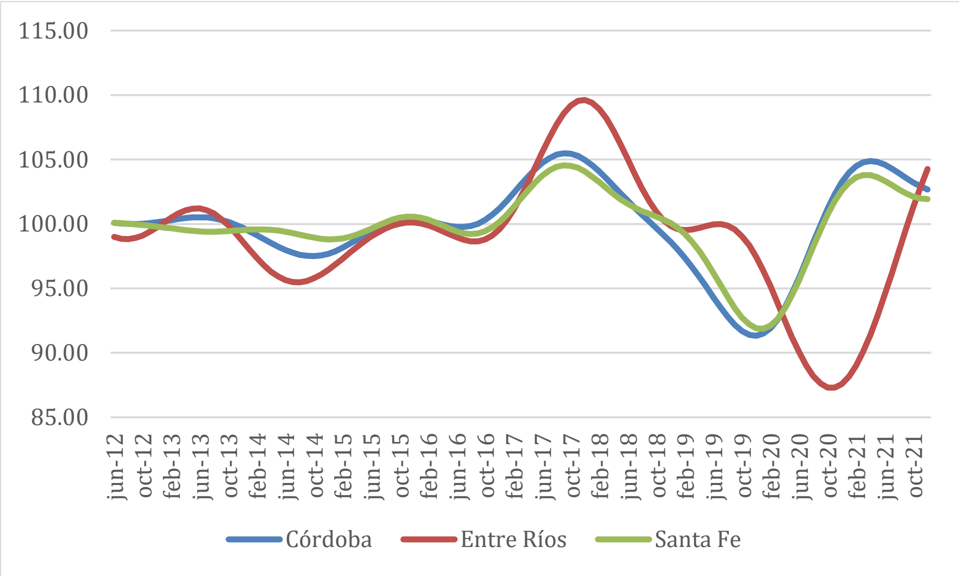

Figure 4: Business Cycle extracted from ILCE by Cristiano-Fitzgerald filter of *Región Centro* provinces from jun-12 to dec-21.

Source: own elaboration with data from different official sources and sectoral chambers.

4. CONCLUDING REMARKS

In this paper, we applied the SW methodology to estimate coincident composite indexes for the twenty-four Argentine provinces as a measure of the state of their economy. This is a statistically precise procedure which until now has been used exclusively to calculate indexes in a few bunches of Argentine provinces. We also extracted the business cycle components of these coincident indexes by applying the CF filter. Then, by implementing a Ward-like hierarchical clustering algorithm in different conditions, we grouped the provinces whose economical behaviors are similar among them and different from the rest. In this sense, our work makes a valuable contribution to the empirical literature on regional economic dynamics, a field for which there are few studies of the state of the Argentine economy.

Provinces were clustered using differently weighted combinations of both geographic contiguity and economic similarities, applying two approaches. The first one used the coincident indexes, while the second one used the business cycle components extracted from these indexes. The application of different clustering scenarios generates different groupings of provinces. As the weight of the economic similarities in clustering decreases, the groups become increasingly compact.

Beyond the clustering scenarios, one of the main results of our work is that grouping provinces according to their business cycle presents some regularities among them. Certain groups of provinces as *Salta* and *Chaco* are grouped in every scenario. The same happens with *La Rioja, San Juan* and *San Luis*, and also with *Catamarca, Córdoba* and *Tucumán*. Even in some cases, the results show isolated provinces, surrounded by others with noticeable differences in their business cycles, such as *Jujuy*, that is only grouped with neighboring provinces once, and *Tierra del Fuego*, that is clustered with contiguous provinces only twice over six scenarios. Neither scenario produce any regionalization similar to the statistical regions determined by INDEC, but when the clustering is made using an equal combination of contiguity and business cycle, the resulting clusters are very similar to what we can expect from economic regionalization, that is, complete contiguity, business cycle similarities and a relatively balanced size in terms of the number of provinces.

Another particularly interesting result of our work is that *Córdoba* and *Santa Fe* -provinces that concentrate the country's agro-industrial production and exports- appear together in almost all the clustering procedures. The same is not the case with the Province of *Entre Ríos*, also a member of the -politically defined- *Región Centro*, since only in one scenario the three provinces belong to the same cluster. These results are logical, according to the similarities in coincident indexes and business cycles between *Córdoba* and *Santa Fe.*

Finally, the results as a whole show that, at least in those cases where the purpose is to analyze the economic performance of a region over time, the impact of a national economic policy or the effects of an external shock on the performance of the economic activity, it seems that groupings based on static criteria may not be the most appropriate approach.

References

Arredondo, F., D´Jorge, M. L., Cohan, P., and Sagua, C. (2009). Metodología y Proceso de Construcción del Índice Compuesto Coincidente de Actividad Económica de la Provincia de Santa Fe. Centro de Estudios y Servicios Bolsa de Comercio de Santa Fe.

Baxter, M., & King, R. G. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics*, *81*(4), 575-593.

Berardi, M. V., Navarro, A. I., and Uría, M. C. (2010). An Application of the Stock and Watson Index Methodology to the Santa Fe Economy. *Anales XLV Reunión Anual de la Asociación Argentina de Economía Política*.

Blanco, E., Elosegui, P., Izaguirre, A., and Montes-Rojas, G. (2019). Regional and state heterogeneity of monetary shocks in Argentina, *The Journal of Economic Asymmetries*, Elsevier, vol. 20(C).

Brida, J. G., Garrido, N., and London, S. (2012). Estudio del desempeño económico regional: el caso argentino. *Cuadernos de Economía*, 32(60), 399-427.

Burns, A. F., and Mitchell W. C. (1938). Statistical Indicators of Cyclical Revivals. New York: NBER.

Burns, A. F., and Mitchell, W. C. (1946.). Measuring Business Cycles. New York: NBER.

Chavent, M., Kuentz-Simonet, V., Labenne, A., & Saracco, J. (2018). ClustGeo: an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, *33*(4), 1799-1822.

Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *International Economic Review*, *44*(2), 435-465.

Clayton-Matthews A., and Stock J. H. (1998/1999). An Application of the Stock/Watson Index Methodology to the Massachusetts Economy, *Journal of economic and Social Measurement*, 25, (3/4), 183–233.

Costanzo, G. D. (2001). A Constrained k-means clustering algorithm for classifying spatial units. *Statistical Methods and Applications*, 10(1), 237-256.

Crone, T. M. (1998/ 1999). Using State Indexes to Define Economic Regions in the US, *Journal of Economic and Social Measurement* 25, 259-275.

Crone, T. M. (2005). An Alternative Definition of Economic Regions in the United States Based on Similarities in State Business Cycle. *The Review of Economics and Statistics*, 87(4), 617-626.

Crone, T. M., & Clayton-Matthews, A. (2005). Consistent economic indexes for the 50 states. *Review of Economics and Statistics*, *87*(4), 593-603.

D´Jorge, M. L., Cohan, P. P., Henderson, S. J., and Sagua, C. E. (2007). Proceso de construcción del Índice Compuesto Coincidente Mensual de Actividad Económica de la Provincia de Santa Fe (ICASFe). *Anales XLII Reunión Anual de la Asociación Argentina de Economía Política.*

Dias, F. (1993). A composite coincident indicator for the Portuguese economy. Working Paper No 18/93. Banco de Portugal.

Féliz, M., Panigo, D., and Pérez, p. E. (2000). Determinantes del desempleo en el ámbito regional y su influencia sobre la implementación de políticas de empleo, *Anales XXXV Reunión Anual de la Asociación Argentina de Economía Política.*

Figueras, A. J., Díaz Cafferata, A., Arrufat, J.L., Descalzi, R., and Rubio, A. (2002). Evolución del desempleo regional: análisis a través de casos. *Anales XXXVII Reunión Anual de la Asociación Argentina de Economía Política.*

Fukuda, S., and Onodera, T. (2001). A new composite index of coincident economic indicators in Japan: How can we improve forecast performances? *International Journal of Forecasting*, 17(3), 483-498.

Galiani, S, Lamarche, C., Porto, A., and Sosa Escudero, W. (1999). Regional unemployment in Argentina. *Económica*, 45(4), p. 137-165.

Galiani, S., Lamarche, C., Porto, A., and Sosa Escudero, W. (2005). Persistence and regional disparities in unemployment (Argentina 1980-1997). *Regional Science and Urban Economics*, 35: 375-394.

Garegnani, M. L., and Di Gresia, L. M. (1999). Relación entre ciclo económico regional y nacional: análisis del período 1961-1995. *Económica*, 45(3), 261-281.

Hall, S. G., and Zonzilos, N. G. (2003). An Indicator Measuring Underlying Economic Activity in Greece. Working Paper No. 4. Bank of Greece.

Heredia, J., and Alvarez, N. (2017). Indicador Compuesto de Actividad Económica para la provincia de Misiones. *Anales LII Reunión Anual de la Asociación Argentina de Economía Política.*

Jorrat, J. M. (2003). Indicador Económico Regional: El Índice Mensual de Actividad Económica de Tucumán (IMAT). *Anales XXXVIII Reunión Anual de la Asociación Argentina de Economía Política.*

Jorrat, J. M. (2005). Construcción de Índices Compuestos mensuales Coincidentes y Líder de Argentina. *Progresos en econometría*, 43-100. Asociación Argentina de Economía Política.

Lapelle, H. C. (2015). El ciclo económico de la Región Rosario: Su relación con el nacional y provincial. *Anales L Reunión Anual de la Asociación Argentina de Economía Política.*

Larsson, G., & Vasi, T. (2012). *Comparison of Detrending Methods*. (Doctoral dissertation, Uppsala University).

Méndez, F. (2007). Aplicación de los modelos de espacio de estados para la estimación de un indicador agregado de la actividad económica provincial. Universidad Nacional de Rosario.

Michel Rivero, A. (2007). Indicador Económico Regional: el índice compuesto coincidente mensual de la actividad económica de Córdoba (ICA-COR). 1994-2005. *Revista de Economía*,Vol. XLV - N° 1 - (2007) - pp. 31-73

Miller, T. W., and Sabbarese, D. (2012). An economic indicator for the state of the economy in the southeastern U.S. *Journal of Regional Analysis and Policy* 42(1): 1-27.

Mitchell, W. C. (1927). *Business Cycles. The Problem and Its Setting*. New York: NBER.

Muñoz, F., and Trombetta, M. (2014). Indicador sintético de actividad provincial (ISAP): Un aporte al análisis de las economías regionales. *Anales XLIX Reunión Anual de la Asociación Argentina de Economía Política.*

Navarro, A. I. and Sigal, F. (2012). Measuring Regional Economy Cycle in Argentina: An Application of the Stock/Watson Index Methodology to the Santa Fe Economy. 13th. Annual International Business Conference, University of North Florida.

Nilsson, R., & Gyomai, G. (2011). Cycle extraction: A comparison of the phase-average trend method, the Hodrick-Prescott and Christiano-Fitzgerald filters.

Orr, J., Rich, R., and Rosen, R. (1999). Two New Indexes Offer a Broad View of Economic Activity in the New York-New Jersey Region. Federal Reserve Bank of New York, 5(14).

Orr, J., Rich, R., and Rosen, R. (2001). Leading Economic Indexes for New York State and New Jersey. *Economic Policy Review*, 7(1), 73-94.

Paz, J. (2009). Retornos a la educación en Argentina. Estructura Regional. Documentos de Trabajo, Facultad de Ciencias Económicas, Jurídicas y Sociales, Universidad Nacional de Salta, Argentina.

Reklaite, A. (2011). Coincident, leading and recession indexes for the Lithuanian economy. *Baltic Journal of Economics*, 11(1), 91-108.

Sigal, F., Camusso, J., & Navarro, A. I. (2021). Beyond the political or administrative definition of the Argentinean economic regions, do their provinces have a common economic cycle?", *13th World Congress of Regional Science Association International (RSAI).* https://www.regionalscience.org/images/PDF/PROCEEDINGS_2021RSAICongress_V16_07.pdf [Accessed August 31, 2022].

Stock, J. H. and Watson, M. W. (1989). New Indexes of coincident and leading economic indicators. In *NBER Book Series Macroeconomics Annual* (pp. 351-394). Cambridge, Mass.: MIT Press.

Stock, J. H. and Watson, M. W. (1991). A Probability Model of the Coincident Economic Indicators. In *The Leading Economic Indicators: New Approaches and Forecasting Records* (pp. 63–90.). Cambridge University Press.

Tebaldi, E., and Kelly, L. J. (2012). Measuring economic conditions: an extension of the Stock/Watson methodology. *Applied Economics Letters*, 19(18), 1865-1869.

Zacaria, H., and Zoloa, J. I. (2005). Desigualdad y pobreza entre las regiones argentinas: Un análisis de microdescomposiciones. *Anales XL Reunión Anual de la Asociación Argentina de Economía Política*.