

Dinámicas laborales regionales y su relevancia en el agregado nacional: Una aplicación de Clusterización de Series Temporales para Chile

FELIPE O. G. CHÁVEZ BUSTAMANTE ^a, CRISTIAN MONDACA-MARINO ^b, JULIO ROJAS-MORA ^c

^a *Universidad del Desarrollo, Facultad de Economía y Negocios, Aynavillo 456, 4030000, Concepción, Chile. E-mail: fchavezb@udd.cl*

^b *Universidad Austral de Chile, Instituto de Economía, Los Laureles n°35 interior, 5110027, Campus Isla Teja, Valdivia Chile. E-mail: cristianmondaca@uach.cl*

^c *Universidad Católica de Temuco, Departamento de Ingeniería Informática, Avenida Rudecindo Ortega 02950, 4781312, Campus San Juan Pablo II, Temuco, Chile. E-mail: jrojas@inf.uct.cl*

RESUMEN

Este trabajo tiene como objetivo analizar el comportamiento de la ocupación en el mercado del trabajo en Chile a nivel regional, determinando las diferencias y similitudes entre regiones, y de las regiones con el agregado de empleo a nivel nacional. Para la realización de dicho trabajo se han utilizado las tasas mensuales de ocupación regionales y a nivel país para el periodo 1986-2010, y utilizado métodos de Clusterización de series temporales para identificar conjuntos de regiones con comportamientos similares. Los resultados muestran la existencia de diferentes regímenes en los mercados laborales de las regiones y el relevante rol de algunas regiones en el comportamiento del agregado nacional.

Palabras clave: Mercado Laboral Regional, Clusterización de Series Temporales, Ocupación

Regional Labor Dynamics and their Relevance in the National Aggregate: A Time Series Clustering Application for Chile

ABSTRACT

The objective of this work is to analyze the employment behavior in the labor market in Chile at the regional level, determining the differences and similarities between regions, and of the regions with the aggregate employment at the national level. For this work, we studied the monthly regional employment rates during the 1986-2010 period for Chile by applying time series clustering methods to identify regions with similar dynamics. Afterward, we use that information to build a model that explains the contributions from the regional markets to the national aggregate. Results show the existence of regimes on the local labor markets and the relevance of the Metropolitan Region on the country-level behavior.

Keywords: Regional Labor Market, Time Series Clustering, Employment.

Clasificación JEL: J64, R23, J21, C32

Artículo recibido en enero de 2018 y aceptado en septiembre de 2018

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. e-36307

1. INTRODUCCIÓN Y MARCO CONCEPTUAL

En la literatura existe un especial interés por el estudio del comportamiento del mercado laboral a nivel regional, los factores que pueden incidir en las fluctuaciones del nivel de ocupación y su efecto en el agregado nacional (Decressin y Fatás, 1995). Los mercados laborales regionales pueden presentar diferencias respecto a lo que se observa a nivel país, debido a la existencia de regiones con mayor especialización en la producción de bienes y servicios fundamentales para la economía del país.

Las regiones también pueden presentar dinámicas diferentes de ajuste a las crisis económicas (sean locales, sectoriales o nacionales) puesto que su estructura y organización social puede facilitar o dificultar el proceso de adaptación de la oferta y demanda de trabajos a nivel local. Por otra parte, el componente espacial incide en la movilidad de los trabajadores, provocando que algunas regiones se adapten de forma más rápida que otras frente a ajustes en la oferta laboral por la conmutación entre regiones o entre países limítrofes. Estas diferencias, separadas o en su conjunto, pueden propiciar procesos de convergencia entre regiones, y procesos de disparidad en la dinámicas de los mercados laborales a nivel regional, y por lo tanto, los métodos de agregación para determinar la dinámica nacional pueden eventualmente estar ocultando la heterogeneidad existente a nivel local.

La identificación de las regiones que aportan significativamente al agregado de empleo nacional puede ser relevante debido a que puede ayudar a implementar de mejor manera políticas de convergencia y descentralización con especificidad territorial, y servir como insumo para estudios empíricos que determinen los factores económicos y sociales que inciden en la formación de puestos de trabajos (Decressin y Fatás., 1995).

Partiendo de la base que las divisiones administrativo-regionales convergen hacia dinámicas interregionales diferentes, nos planteamos las siguientes interrogantes: ¿existen regiones con dinámicas similares de sus mercados laborales?, ¿se pueden identificar regímenes de comportamiento característicos de los mercados laborales a nivel regional?, ¿cómo es la participación de las regiones en el agregado nacional?, ¿se puede modelar el comportamiento del agregado nacional en función de un grupo representativo de regiones?.

En este artículo buscamos dar respuesta a las preguntas mencionadas, analizando los mercados laborales de las regiones chilenas para el período 1986-2010 y utilizando como metodología la Clusterización de series temporales en el proceso de identificación de conglomerados de regiones. La principal aportación de este trabajo es utilizar nuevas herramientas de análisis de series temporales y su clasificación, que puedan contribuir a una interpretación más acabada de la información que proporcionan los mercados regionales.

Los métodos de Clusterización consisten en la división de objetos en grupos homogéneos óptimos utilizando medidas de similitud y distancias (Johnson, 1967; King, 2012; Ruspini, 1969), e identificando estructuras en un conjunto de datos al organizarlos objetivamente en grupos homogéneos, donde la similitud de los objetos y el grupo se minimiza y la disimilitud entre grupos se maximiza (Warren Liao, 2005).

El uso de herramientas de Clusterización para series temporales ha sido restringido, debido a que los métodos tradicionales se sirven de datos estáticos y en ellos, el orden de ocurrencia de los datos no es relevante. Para el caso de las series temporales, la dificultad radica en su naturaleza dinámica, ya que captura un proceso sucedido en el tiempo y el orden de ocurrencia de los valores es de máxima importancia para definir la evolución de los procesos experimentados (Li y Prakash, 2011).

A pesar de la variedad de estudios que se han hecho con foco en la Clusterización de series temporales (Warren Liao, 2005), aún no existe consenso sobre un método de disimilitud óptimo para ciertos tipos de datos, y por tanto, sigue siendo un tema de controversia, cuál de los métodos funciona mejor dentro de la amplia gama de combinaciones algoritmo-distancia. Por ejemplo, Zhang *et al.* (2006) proponen la utilización de la transformada de Wavelet ortogonal, Alonso *et al.* (2006) presentan un método basado en la densidad probabilística de predicciones; Xiong y Yeung (2004) demuestran que los modelos ARMA son más efectivos que otras alternativas al trabajar con series temporales de distinta longitud, entre otros artículos. Además, una gran dificultad para los investigadores es seleccionar una medida de distancia, debido a la gran variedad de medidas de distancias posibles de utilizar, y por lo tanto, puede transformarse en un proceso complejo el decidir el método más adecuado (Shirkhorshidi *et al.*, 2015).

A pesar de estas dificultades, existen investigaciones que han utilizado métodos de Clusterización en indicadores económicos de nivel regional; por ejemplo, en Barón (2002) se identifican las regiones económicas colombianas versus las geográficas a través de un análisis de clúster. En otro trabajo similar, Galbraith *et al.* (2004) analizan sectores industriales y la desigualdad de las rentas salariales a lo largo del tiempo con métodos de clúster temporales. En Aroca y Bosch (2000) se sugiere la existencia de distintos clústeres a nivel regional chileno en términos de producto, configurándose un clúster de mayor producto, regiones I y II y otro con bajo producto en el sur de Chile - VIII, IX y X. Artículos similares de clustering espacial para variables económicas regionales pueden consultarse en Albert y Mateu, (2004); Aroca y Bosch, (2000); Brida *et al.* (2011); Brida *et al.* (2015); Sastré Gutiérrez y Rey (2008).

Al igual que la Clusterización de datos estáticos, la Clusterización de series temporales requiere un algoritmo que permita formar los grupos de clústeres en

las series de datos a partir de la medida de distancia elegida, cuya elección dependerá de los tipos de datos disponibles y el propósito particular de la aplicación.

Existen diversos métodos de clusterización para series temporales, y el lector puede encontrar una amplia revisión en Warren Liao (2005) y Aghabozorgi *et al.* (2015). De los métodos más utilizados en análisis de series temporales, se encuentra el Dynamic Time Warping (de ahora en adelante DTW), que encuentra un alineamiento óptimo para dos series temporales bajo ciertas restricciones. La DTW fue propuesta inicialmente por Sankoff y Kruskal, (1983) y aplicada a encontrar patrones en series temporales por Berndt y Clifford, (1994). Diversos trabajos mencionan su utilidad en la identificación de series temporales similares a pesar de presentar desfase en el eje temporal (Hui y Yunfa, 2005; Keogh y Pazzani, 1999; Keogh y Ratanamahatana, 2004; Müller, 2005). Esta distancia permite reconocer formas similares incluso en la presencia de transformaciones de señales como desplazamientos o escalamientos (Montero y Vilar, 2014). Su utilidad se origina en que la mayoría de algoritmos usan la distancia Euclidiana (o alguna variación de ella), la cual presenta cierta fragilidad como medida de distancia (Keogh y Pazzani, n.d).

2. METODOLOGÍA

Los datos corresponden a las tasas de ocupación mensual para el período 1986-2010 disponibles en el Instituto Nacional de Estadística de Chile (Encuesta Nacional de Empleo). Debido a la disponibilidad de datos, se utiliza la muestra de trece regiones correspondientes a la anterior organización administrativa del país que rigió hasta el año 2007; estas están enumeradas secuencialmente de norte al sur del país. En la siguiente tabla se describe brevemente las principales actividades por regiones:

Tabla 1
Resumen por Región (Sector y PIB)

Región	Principales sectores económicos	PIB Regional (Millones de Pesos)	
		1986	2007
I – de Tarapacá	Comercio, Restaurantes y Hotelería Manufactura	112.383	2.247.383
II – de Antofagasta	Minería	212.375	4.026.283
III – de Atacama	Minería Agropecuario, Silvícola y Pesca	50.079	1.132.528
IV – de Coquimbo	Agropecuario, Silvícola y Pesca	79.172	1.399.550
V – de Valparaíso	Transporte y Telecomunicaciones	300.820	4.873.998
RM – Región Metropolitana	Comercio, Restaurantes y Hoteles Servicios Financieros Manufactura	1.298.379	27.010.760

Tabla 1 (Continuación)
Resumen por Región (Sector y PIB)

Región	Principales sectores económicos	PIB Regional (Millones de Pesos)	
		1986	2007
VI – del Libertador General Bernardo O'Higgins	Agropecuaria Silvícola y Pesca Minería	184.698	2.260.947
VII – del Maule	Agropecuaria, Silvícola y Pesca Industria Manufacturera	131.449	2.125.466
VIII – del Biobío	Manufactura Transporte y Telecomunicaciones	344.164	5.741.048
IX – de la Araucanía	Comercio, Restaurantes y Hotelería Manufactura	71.735	1.518.246
X – de los Lagos	Agropecuaria, Silvícola y Pesca	122.891	2.758.379
XI – Aisén del General Carlos Ibáñez del Campo	Servicios Comunes y Sociales Agropecuaria, Silvícola y Pesca	16.392	389.892
XII – de Magallanes y de la Antártica Chilena	Manufactura Comercio, Restaurantes y Hotelería Servicios Comunes y Sociales	100.767	847.400

Fuente: Díaz, Pardo y Meller (2002) y Banco Central de Chile.

El primer paso para el análisis de datos, corresponde a la aplicación de un tratamiento de desestacionalización de series temporales conocido como X-12. Este método ha sido desarrollado por el *U.S Census Bureau* y se ha utilizado ampliamente con el fin de separar las componentes: ciclo tendencia, estacional, irregular y otras asociadas al efecto calendario (Findley *et al.*, 2012.; Gallardo y Rubio, 2009; Ladiray y Quenneville, 2000). La importancia de realizar este ajuste permite mejorar notablemente las posteriores estimaciones (Rabanal y Baronio, 2010). Para revisar con más detalle la importancia de los componentes ciclo y tendencia en los análisis temporales, se recomienda al lector consultar el estudio de McLaren y Zhang (2010).

Luego, estas series de dato temporales desestacionalizadas corresponden a la materia prima del análisis, a las que posteriormente se aplicarán dos métodos de Clusterización: *Dynamic time Warping (DTW)* y *Fuzzy c-means*.

Primero, aplicamos el método DTW presente en los paquetes estadísticos de R: “*TSClust*”, propuesto por Montero y Vilar, (2014) y el “*dtwclus*” presentado en Alexis *et al.* (2016). Este método funciona alineando dos series de tal forma que su diferencia sea minimizada (Warren Liao, 2005).

Una vez que se obtienen las distancias entre los puntos, todo método de Clusterización requiere de un algoritmo que ordene estos datos (matriz de distancia) en aglomeraciones. En nuestro caso, partimos con la Clusterización jerárquica, que es conocida por su representación gráfica en forma de dendograma

(ver Figura 1). Su uso para datos de series temporales puede verificarse en Zhang *et al.* (2011). En este método, cada objeto es asignado inicialmente a su propio clúster y luego el algoritmo procede a iterar. Luego, en cada etapa se van agrupando los dos clústeres más parecidos hasta que sólo exista un único clúster (Johnson, 1967; Montero y Vilar, 2014).

El algoritmo de Clusterización por jerarquización ha sido ampliamente utilizado en distintos campos, como la bioinformática, bioestadística, finanzas y otros. Algunas referencias de aplicaciones a la extracción de información de datos en el área financiera-económica pueden encontrarse en Ahlquist y Breunig, (n.d.); Micciche' *et al.* (2005); Řezanková, (n.d); Tumminello *et al.* (2010); Xavier *et al.* (2009); Zhegao y Jianjunyang (2014).

Por otra parte, utilizamos un segundo algoritmo para el cálculo de distancias: el *fuzzy c-means*. Este método es similar al clustering particional: *algoritmo estocástico que crea una partición dura de los datos en k-clústeres donde cada uno tiene un centroide*. Para el caso particular de series temporales, el centroide también es una serie de tiempo - con la particularidad que la serie no pertenece exclusivamente a un solo clúster, sino que a todos en un grado determinado. Para cada serie, el grado total de pertenencia entre clústeres debe sumar uno. En este procedimiento, una función objetivo definida en términos de distancias cuadradas se minimiza (Alexis *et al.*, 2016). Su uso puede revisarse por ejemplo, en Chiu, (1994); Izakian *et al.* (2015) y Řezanková, (n.d).

Una vez identificados agrupamiento de regiones con comportamientos similares, se puede realizar el ejercicio de intentar predecir el comportamiento del agregado nacional en función de algunas regiones, las cuales se puede monitorear de una forma más intensa con las encuestas de empleo de nivel regional para disponer de información temprana respecto del comportamiento del empleo a nivel nacional

Para esto se propone un modelo de regresión lineal simple (ver ecuación 1) que estima la participación de los mercados regionales en el agregado nacional, descartando las regiones que se identificaron como parecidas entre sí por el método de clusterización.

$$oc_{i,t} = \alpha_i + \beta_i oc_{n,t} + u_{i,t} \quad (1)$$

Donde oc_{it} es el porcentaje de ocupación de la región i -ésima, oc_{nt} es el porcentaje de ocupación a nivel nacional, y u es una componente aleatoria normal.

3. RESULTADOS

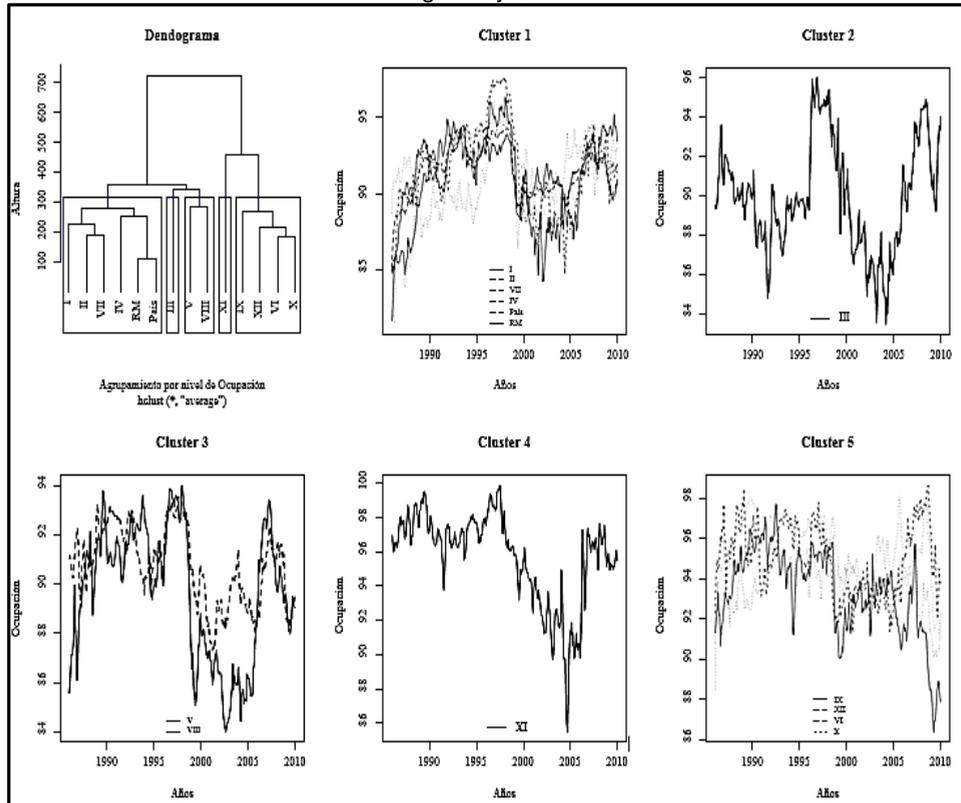
Las distancias computadas por el DTW se presentan en la siguiente tabla:

Tabla 2
Matriz de Distancias (DTW)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	RM
II	222												
III	332	324											
IV	282	255	386										
V	336	306	315	373									
VI	478	425	651	468	700								
VII	236	192	275	319	355	308							
VIII	400	353	376	513	288	486	275						
IX	544	413	537	624	493	227	327	483					
X	598	484	738	662	1115	184	367	889	249				
XI	883	936	1020	1468	1313	634	870	1303	573	333			
XII	665	662	931	880	1172	249	475	965	334	187	292		
RM	305	293	471	235	327	424	292	329	476	715	1414	791	
País	311	281	452	267	400	318	198	248	371	666	1324	714	114

Fuente: Elaboración propia.

Figura 1
Dendrograma y Clústeres



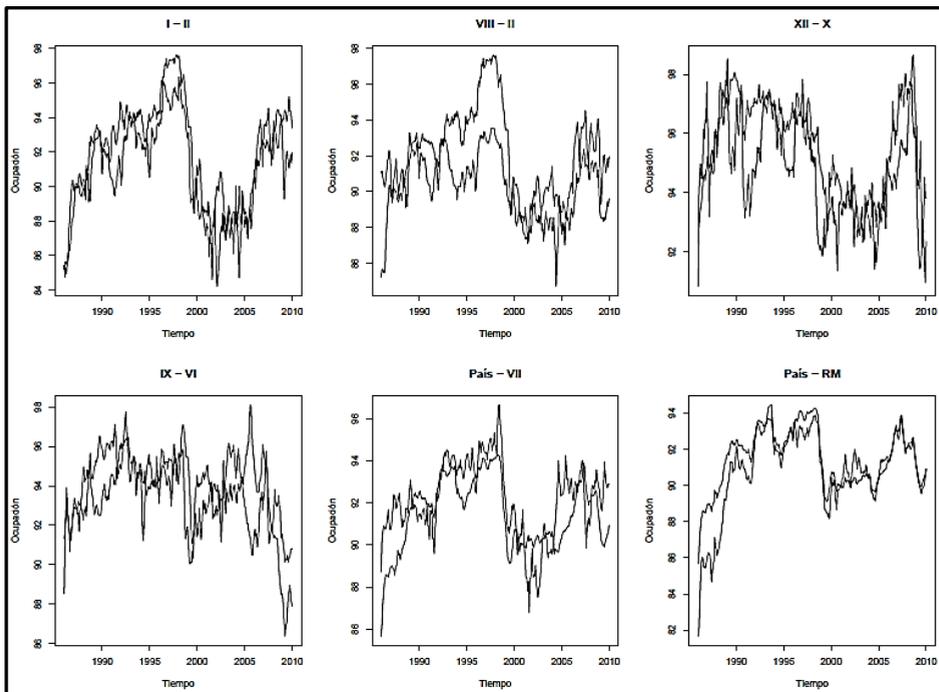
Fuente: Encuesta Nacional de Empleo, INE.

Con estas medidas, el método de jerarquización resultó en el siguiente dendograma, para el cual seleccionamos un corte de 5 clústeres¹.

Sobre la base de la figura anterior, el algoritmo determinó el dendograma presentado en el que se observa una primera agrupación de la I, II, VII, IV, RM y País, seguido de la III clusterizada individualmente, luego la V y VIII, a continuación la región XI sin agrupamiento y finalmente un clúster con cuatro regiones: la IX, XII, VI y X. Junto al dendograma adjuntamos los gráficos de los distintos clústeres obtenidos con sus respectivas series.

A continuación presentamos las principales similitudes y disimilitudes computadas por el DTW presentado en la matriz de distancias anterior. Esto, con el fin de poder observar que efectivamente las series que tienen menores distancias sean similares y aquellas con mayores distancias presenten diferencias considerables en la tendencia de la serie temporal.

Figura 2
Principales Similitudes



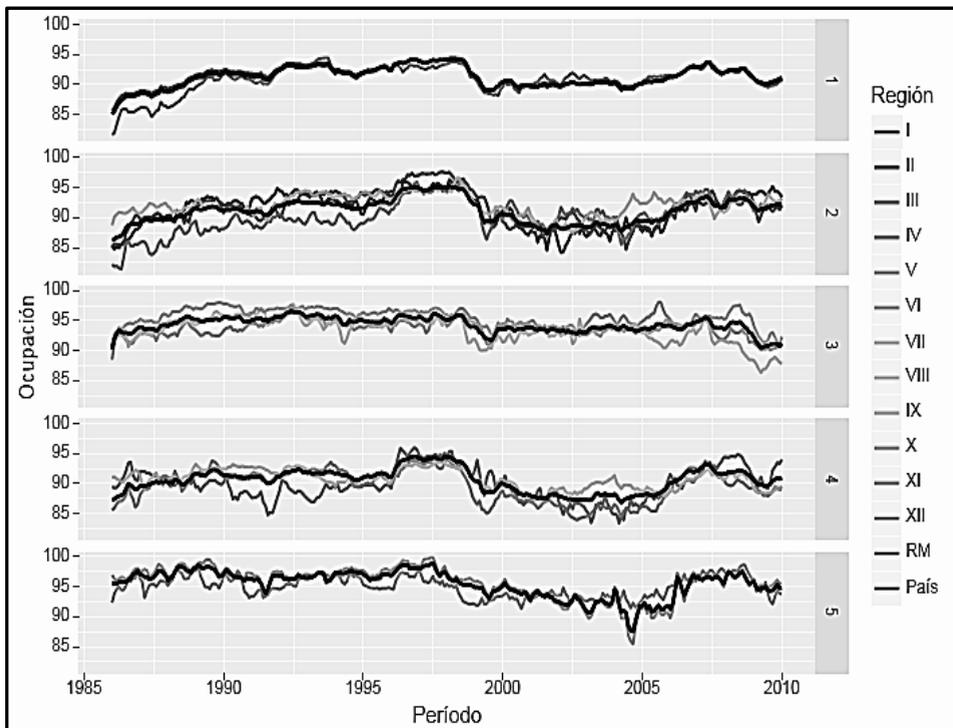
Fuente: Encuesta Nacional de Empleo, INE.

¹ Cabe destacar, que con el método de jerarquización no existe un número “ideal” u “óptimo” de clústeres, sino que es decisión del investigador de acuerdo con la interpretación y características de los datos. Esto configura una gran diferencia con el conocido método *k-means* en el que se deben escoger de antemano el número de clústeres requeridos.

En esta sección es importante destacar la asociación País-Región Metropolitana, la cual es interesante de obtener ya que confirma nuestra presunción inicial que gran parte del agregado nacional es explicado por el comportamiento del mercado laboral capitalino, por concentrar los niveles más altos de población y de empleo. Como se esperaba, la Región Metropolitana se encuentra muy relacionada con la dinámica a nivel País.

La segunda parte de nuestro estudio consistió en computar la Clusterización para los mismos datos, pero esta vez utilizando el algoritmo *fuzzy-c means* (FCM) para series temporales. La aplicación de esta metodología concluyó con la Figura 3 en la que se observan los clústeres formados y sus respectivas formaciones internas.

Figura 3
Fuzzy c-means Clustering



Fuente: Elaboración propia.

Al ser una metodología estocástica, el *fuzzy clustering* permite, a diferencia del *hierarchical clustering*, calcular la probabilidad que una región pertenezca a un determinado clúster. El FCM parte del supuesto que cada set de datos (región) pertenece a todos los clústeres en distintos grados. La asignación de

cada uno a su respectivo clúster se obtiene para aquella en que la probabilidad de pertenencia sea mayor. Esto nos permite un mayor nivel de detalle de los resultados obtenidos, ya que es posible analizar si una región pertenece con mayor certeza a un determinado clúster o si la asignación es un proceso aleatorio. La elección nuevamente de 5 clústeres, en este caso se justifica en que se logra la maximización de las diferencias entre clústeres y la minimización de diferencias intra-clústeres. A continuación presentamos la tabla que resume las probabilidades de pertenencia obtenidas por el FCM:

Tabla 3
Probabilidades pertenencia Región- Clúster

	Clúster 1	Clúster 2	Clúster 3	Clúster 4	Clúster 5
I	0,285	0,084	0,322	0,272	0,038
II	0,284	0,078	0,303	0,310	0,025
III	0,276	0,125	0,292	0,262	0,045
IV	0,270	0,066	0,312	0,341	0,011
V	0,324	0,050	0,335	0,269	0,021
VI	0,101	0,618	0,100	0,119	0,062
VII	0,298	0,087	0,298	0,301	0,016
VIII	0,357	0,051	0,303	0,276	0,013
IX	0,137	0,470	0,138	0,150	0,105
X	0,033	0,615	0,035	0,036	0,281
XI	0,010	0,069	0,010	0,010	0,902
XII	0,040	0,403	0,041	0,043	0,473
RM	0,306	0,035	0,248	0,370	0,005
Pais	0,322	0,028	0,243	0,405	0,003

Fuente: Elaboración propia.

A partir de la tabla, se originan algunas consideraciones de importancia. Por una parte, encontramos que ciertas regiones tienen probabilidades muy similares de pertenencia en más de un clúster. Por ejemplo, la VII región es asignada al Clúster 4 pero con una diferencia de 0,003 y 0,004 de pertenecer al Clúster 1 y 3, respectivamente. La II región, es asignada al Clúster 4, pero tiene una diferencia de probabilidad de 0,007 de ser parte del Clúster 3. Estos ejemplos implican una mayor dificultad al momento de caracterizar o dar una explicación empírica de la asociación Región-Clúster, ya que dichas series temporales presentan un comportamiento que no logra ser determinante a la hora de definir su agrupamiento.

La siguiente tabla muestra un comparativo de las agrupaciones de regiones obtenidas por ambos métodos de Clusterización

Tabla 4
Resumen de Clusters por Método

Clúster	Fuzzy c-means	Hierarchical
1	I – II – IV - VII	I – II – VII – IV – RM – País
2	XI – XII	III
3	III – V – VIII	V – VIII
4	RM - País	XI
5	VI – IX - X	VI – IX – X - XII

Fuente: Elaboración propia.

En caso contrario, existen regiones con una probabilidad mayor de asignación, lo que implicaría esta vez que sus dinámicas son relativamente únicas o de difícil comparación con otras. Ejemplo de esta situación encontramos las regiones más australes del país: XI, IX, la X y la XII.

En la Tabla 3 se presentan las regresiones lineales individuales respecto del agregado nacional para cada una de las regiones de Chile. En este ejercicio, se puede apreciar que las regiones con mayor similitud al agregado nacional son las regiones: RM, II, I y V, que presentan un valor alto del coeficiente de bondad de ajuste de la regresión. En el otro extremo, las regiones con menor relación con el agregado nacional son: III, XI, VI, IX, XII, las cuales presentan un bajo valor del coeficiente de bondad de ajuste de la regresión.

Tabla 5
Modelos de participación del empleo regional en el agregado nacional (1986-2010)

Regiones	α_i	β_i	R^2	p-valor (β)
I	-34,0	1,37	0,63	2,19e-63 ***
II	-42,5	1,47	0,70	1,42e76 ***
III	30,2	0,65	0,15	5,21e-12 ***
IV	-17,4	1,18	0,47	1,80e-41 ***
V	-26,1	1,27	0,60	1,92e-59 ***
RM	-21,5	1,23	0,80	5,7e-102 ***
VI	55,4	0,42	0,21	1,03e-16 ***
VII	23,9	0,75	0,46	6,44e-40 ***
VIII	37,8	0,58	0,38	2,33e-31 ***
IX	38,3	0,60	0,22	3,28e-17 ***
X	34,9	0,66	0,45	3,34e-39 ***
XI	32,3	0,69	0,19	1,77e-14 ***
XII	49,5	0,50	0,23	1,09e-17 ***

Fuente: Elaboración propia.

Considerando estos resultados y los obtenidos por el método de Clusterización, podemos proponer un modelo de regresión lineal que integre ambos análisis, considerando la selección de regiones según agrupamiento

identificado por el método DTW y los casos de más alto nivel de correlación entre las series regionales y el agregado nacional y.

En este caso, la selección considera las regiones que presentan un alto grado de correspondencia con el comportamiento del agregado nacional y descartar aquellas regiones que tienen comportamientos similares a nivel temporal, con lo cual se proponen las siguientes regiones como candidatas a ser incluidas en un modelo de regresión lineal que modela el comportamiento del agregado nacional, como son las regiones RM, III, V, y X (ver ecuación 2).

$$\text{ocu}_{n,t} = 10,83 + 0,50 \cdot \text{ocu}_{RM,t} + 0,045 \cdot \text{ocu}_{III,t} + 0,19 \cdot \text{ocu}_{V,t} + 0,15 \cdot \text{ocu}_{X,t} \quad (2)$$

(1,10)
(0,007)
(0,007)
(0,0106)
(0,0146)

$$N=289, R^2=0,978$$

De la ecuación 2 se puede apreciar que las regiones propuestas permiten modelar el agregado nacional, con un alto grado de ajuste, y que desde una perspectiva territorial permiten modelar el agregado nacional monitoreando sólo cuatro regiones, localizadas en la zonas: Norte (III), Centro (RM y V) y Sur (X) de Chile, regiones representativas de las diferentes zonas geográficas que se pueden encontrar en el país, las cuales tienen diferentes vocaciones productivas vinculadas a la Minería (III), Servicios e industria (RM y V) y Silvoagropecuario y pesca, sectores de gran importancia en la economía nacional.

Del análisis anterior, se han identificado distintos patrones de comportamiento en los mercados laborales a nivel regional y su peso relativo sobre el agregado nacional, con cinco grupos de regiones en el que la Región Metropolitana comparte lugar con el agregado nacional. Además, las regiones más australes configuran comportamientos completamente distintos a las del centro y norte del país, que son las que más inciden con el comportamiento agregado nacional. Sin embargo, es importante notar que aun cuando las regiones I y II están alejadas geográficamente de la RM- que mejor refleja el comportamiento nacional- siguen siendo significativas a la hora de explicar la caracterización de empleo a nivel país.

En términos generales, la importancia de nuestro estudio tiene que ver con enfatizar que las dinámicas regionales importan. Quah (1996) cuestiona la hipótesis que caricaturiza la irrelevancia de la economía geográfica, y sus resultados refuerzan la idea que las divisiones administrativas regionales y sus localizaciones geográficas importan más aún que los factores macro. En consecuencia, nuestro trabajo permite extender lo planteado por Quah (1996) en términos que las visiones regionales importan en términos geográficos; pero también en la dimensión temporal.

En este contexto, los métodos de Clusterización de series temporales efectivamente pueden ayudar a identificar grupos afines, y con esto obtener información útil para la toma de decisiones (Barragan *et al.*, 2016). En particular, se debe tener en cuenta que las iniciativas públicas pueden influenciar al

fortalecimiento económico de los territorios a través de coordinación y planificación estratégica específica a cada zona, con sus distintas particularidades (Arufe y Carrión, 2005). Además la clusterización ha ayudado a identificar grupos de regiones con un comportamiento histórico similar en términos de ocupación y dan cuenta de la diversidad que plantea la longitud geográfica del territorio chileno y las distintas concentraciones de población y actividad industrial en que cada una está especializada.

Que existan grupos de regiones con rasgos similares, no debería ser una sorpresa debido a la cercanía geográfica; sin embargo, también se han identificado regiones vecinas con comportamiento diferentes. Tal como concluye Barón (2002), incluso si las zonas geográficas estudiadas son vecinas, no implica que las dinámicas internas sean necesariamente parecidas.

Estos resultados pueden ser muy útiles en el ámbito del diseño de Política de Empleo que se puede desarrollar en Chile, debido a que la existencia de diferentes regímenes de comportamiento de los mercados laborales en las regiones, justificaría la creación de políticas de empleo, no solamente descentralizadas y flexibles, sino que además desarrollar mecanismos que activan los subsidios y ayudas, para funcionar en el momento más adecuado de acuerdo a las necesidades de las regiones. Ya que dependiendo del ciclo económico regional, puede suceder que el comportamiento de la ocupación a nivel nacional no amerite la intervención de programas de empleo, pero si la intervención en algunas regiones que estén experimentando dificultades, aprovechando la identificación de grupos para seleccionar las regiones a ser beneficiadas, y con esto aprovechar de mejor forma los recursos públicos en aquellas regiones que efectivamente lo necesitan, en el momento adecuado.

Finalmente, mencionar que otro resultado interesante de destacar es la posibilidad de monitorear el comportamiento del empleo a nivel nacional, realizando un esfuerzo estadístico focalizado en solo algunas regiones del país, de tal forma de evaluar en forma rápida el comportamiento del agregado nacional como una especie de indicador temprano del comportamiento de la ocupación.

4. DISCUSIÓN Y CONCLUSIONES

En este artículo se ha analizado el comportamiento de la ocupación en el mercado del trabajo en Chile a nivel regional, utilizado métodos de Clusterización de series temporales para identificar conjuntos de regiones con comportamientos similares, y de las regiones con el nivel de ocupación a nivel nacional en el período 1985-2010. Además se analizaron las correlaciones lineales entre las regiones y el agregado nacional para identificar un modelo que ayude a monitorear el agregado nacional utilizando el menor número posible de regiones.

De los principales resultados que se pueden mencionar son que los métodos de clusterización facilitaron la identificación de diferentes regímenes en los

mercados laborales de las regiones y detectar el relevante rol de algunas regiones en el comportamiento del agregado nacional.

Además los métodos de clusterización y el análisis de correlaciones lineales entre las regiones y el agregado nacional permitieron identificar cuatro regiones que pueden ser utilizadas para monitorear el comportamiento del agregado nacional, con el consiguiente beneficio de economía y rapidez que esto involucra.

Estos resultados pueden ser útiles para fundamentar una política de empleo diseñada con enfoque descentralizado y con programas flexibles que sean capaces de activarse en diferentes momento temporales, de acuerdo a los regímenes observados a nivel regional y no necesariamente de acuerdo a la coyuntura nacional.

Para futuras investigaciones, puede ser interesante analizar el comportamiento dinámico de los clústeres, es decir, observar durante el tiempo cómo las regiones transitan entre distintos clústeres. También puede ser útil analizar la relación de las dinámicas laborales con el crecimiento del producto regional, para predecir el comportamiento de la ocupación en relación al crecimiento económico regional.

REFERENCIAS BIBLIOGRÁFICAS

- AGHABOZORGI, S., SHIRKHORSHIDI, A. S., y WAH, T. Y. (2015). "Time-series clustering—A decade review. *Information Systems*", 53, pp. 16-38.
- AHLQUIST, J. S., y BREUNIG, C. (n.d.). *Country Clustering in Comparative Political Economy*.
- ALBERT, J.M. Y MATEU, J. (2004). *Concentración versus dispersión: Un análisis espacial de la localización de la actividad económica en la U.E*. Mimeo.
- ALEXIS, A., MAINTAINER, S.-E., y SARDA, A. (2016). Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance.
- ALONSO, A. M., BERRENDERO, J. R., HERNÁNDEZ, A., y JUSTEL, A. (2006). Time series clustering based on forecast densities. *Computational Statistics y Data Analysis*, 51(2), pp. 762-776. <http://doi.org/10.1016/j.csda.2006.04.035>
- AROCA, P., y BOSCH, M. (2000). "Crecimiento, convergencia y espacio en las regiones chilenas: 1960-1998". *Estudios de Economía*, ISSN-E 0304-2758, Vol. 27, No. 2, 2000, pp. 199-224, 27(2), 199-224.
- ARUFE, J. E. F., y CARRIÓN, J. J. J. (2005). "Sistemas productivos locales en Castilla y León: identificación y rasgos básicos". *Estudios de Economía Aplicada*, 23(2), pp. 423-450.
- BARÓN, J. (2002). *Las regiones económicas de Colombia: un análisis de clusters*. Documentos de Trabajo sobre Economía Regional y Urbana n°3125, Banco de la República, Colombia
- BARRAGAN, J. F., FONTES, C. H., y EMBIRUÇU, M. (2016). "A wavelet-based clustering of multivariate time series using a Multiscale SPCA approach". *Computers y Industrial Engineering*, 95, pp. 144-155. <http://doi.org/10.1016/j.cie.2016.03.003>

- BEMDT, D. J., y CLIFFORD, J. (1994). *Using Dynamic Time Warping to Find Patterns in Time Series*. AAAI-94 workshop on knowledge discovery in databases, pp 229-248.
- BEZDEK, J. C. (1984). "Fcm: The Fuzzy C-Means Clustering Algorithm". *Computers y Geosciences*, 10(3), pp. 191-203.
- BRIDA, J. G., GARRIDO, N., y MATESANZ GÓMEZ, D. (2015). "Análisis jerárquico de la dinámica económica de las comunidades españolas en el periodo 1955-2009". *Investigaciones Regionales: Journal of Regional Research*, ISSN 1695-7253, No. 31, 2015, pp. 121-141, (31), 121-141.
- BRIDA, J. G., PEREYRA, J. S., ANYUL, M. P., y RISSO, W. A. (2011). Documentos de Trabajo Regímenes de desempeño económico y dualismo estructural en la dinámica de las entidades federativas de México, 1970 -2006.
- CHIU, S. L. (1994). "Fuzzy Model Identification Based on Cluster Estimation". *Journal of Intelligent y Fuzzy Systems*, 2(3), pp. 267-278. <http://doi.org/10.3233/IFS-1994-2306>
- DECRESSIN, J., y FATÁS, A. (1995). "Regional labor market dynamics in Europe". *European Economic Review*, 39(9), pp. 1627-1655. [http://doi.org/10.1016/0014-2921\(94\)00102-2](http://doi.org/10.1016/0014-2921(94)00102-2)
- DÍAZ, R., PARDO, A., y MELLER, P. (2002). *Análisis económico-descriptivo de las regiones chilenas*. Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.
- FINDLEY, D. F., MONSELL, B. C., BELL, W. R., OTTO, M. C., y CHEN, B.-C. (2012). "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program". *Journal of Business y Economic Statistics*. 16 (2), pp 127-152. <http://doi.org/10.2307/1392565>
- GALBRAITH, J. K., BERNER, M., y CÁMARA IZQUIERDO, S. (2004). *Desigualdad y cambio industrial: una perspectiva global*. *Desigualdad y cambio industrial: una perspectiva global, 2004* (pp. 289-296). Madrid: Akal Ediciones.
- GALLARDO, M., y RUBIO, H. (2009). "Estudios económicos estadísticos Diagnóstico de estacionalidad con x-12-arima". *Studies in Economic Statistics Central Bank of Chile*.
- HUI, X., y YUNFA, H. (2005). "Data Mining Based on Segmented Time Warping Distance in Time Series Database". *Journal of Computer Research and Development*, 42 (1), pp. 72-78.
- IZAKIAN, H., PEDRYCZ, W., y JAMAL, I. (2015). "Fuzzy clustering of time series data using dynamic time warping distance". *Engineering Applications of Artificial Intelligence*, 39, pp. 235-244. <http://doi.org/10.1016/j.engappai.2014.12.015>
- JOHNSON, S. C. (1967). "Hierarchical clustering schemes". *Psychometrika*, 32(3), pp. 241-254. <http://doi.org/10.1007/BF02289588>
- KEOGH, E. J., y PAZZANI, M. J. (N.D.). *Scaling up Dynamic Time Warping for Datamining Applications*. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '00)*. ACM, New York, NY, USA, 285-289. DOI=<http://dx.doi.org/10.1145/347090.347153>
- KEOGH, E. J., y PAZZANI, M. J. (1999). *Scaling up Dynamic Time Warping to Massive Datasets* (pp. 1-11). Springer: Berlin Heidelberg. http://doi.org/10.1007/978-3-540-48247-5_1
- KEOGH, E., y RATANAMAHATANA, A. (2004). *Exact indexing of dynamic time warping*. *Knowledge and Information Systems*. <http://doi.org/10.1007/s10115-004-0154-9>
- KING, B. (2012). "Step-Wise Clustering Procedures". *Journal of the American Statistical Association*. 69, pp. 86-101.

- LADIRAY, D., y QUENNEVILLE, B. (2000). *Desestacionalizar con el método x-11*. Ed. Español. Metodologica, Revue des Techniques, Méthodes et Instrumens de Recherche en Sciences Humaines, Université Libre de Bruxelles, Belgique.
- LI, L., y PRAKASH, B. A. (2011, July). "Time series clustering: Complex is simpler!". In *ICML* pp. 185-192.
- MCLAREN, C. H., y ZHANG, X. M. (2010). "The Importance of Trend-Cycle Analysis for National Statistics Institutes". *Estudios de economía aplicada*, 28(3), 607-624.
- MICCICHE', S., LILLO, F., y MANTEGNA, R. N. (2005). "Correlation based hierarchical clustering in financial time series". In *Complexity, Metastability and Nonextensivity - 31st Workshop of the International School of Solid State Physics* (pp. 327-335). Singapore: World Scientific Publishing Co. Pte. Ltd. http://doi.org/10.1142/9789812701558_0037
- MONTERO, P., y VILAR, J. A. (2014). "TSclust: An R Package for Time Series Clustering". *Journal of Statistical Software*, 62(1), pp. 1-43. <http://doi.org/10.18637/jss.v062.i01>
- MÜLLER, M. (2005). *Information Retrieval for Music and Motion*. Berlin: Springer, Heidelberg
- QUAH, D. T. (1996). "Regional convergence clusters across Europe". *European Economic Review*, 40(3-5), pp. 951-958. [http://doi.org/10.1016/0014-2921\(95\)00105-0](http://doi.org/10.1016/0014-2921(95)00105-0)
- RABANAL, C., y BARONIO, A. M. (2010). "Alternativas para la modelización de tendencias y ciclos en la economía argentina: 1880-2009". *Estudios de economía aplicada*, 28(3), pp. 651-670.
- ŘEZANKOVÁ, H. (n.d.). *Cluster Analysis of Economic Data*.
- RUSPINI, E. H. (1969). "A new approach to clustering". *Information and Control*, 15(1), pp. 22-32. [http://doi.org/10.1016/S0019-9958\(69\)90591-9](http://doi.org/10.1016/S0019-9958(69)90591-9)
- SANKOFF, D., y KRUSKAL, J. B. (1983). *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Addison-Wesley Publication, 1.
- SASTRÉ GUTIÉRREZ, M. L., y REY, S. J. (2008). "Polarización espacial y dinámicas de la desigualdad interregional en México". *Problemas Del Desarrollo*, 39(155), pp. 181-204.
- SHIRKHORSHIDI, A. S., AGHABOZORGI, S., WAH, T. Y., SHIRKHORSHIDI, A., AGHABOZORGI, S., WAH, T., Medico, E. (2015). "A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data". *PLOS ONE*, 10(12), e0144059. <http://doi.org/10.1371/journal.pone.0144059>.
- TUMMINELLO, M., LILLO, F., y MANTEGNA, R. N. (2010). "Correlation, hierarchies, and networks in financial markets". *Journal of Economic Behavior y Organization*, 75(1), pp. 40-58. <http://doi.org/10.1016/j.jebo.2010.01.004>
- WARREN LIAO, T. (2005). "Clustering of time series data-a survey". *Pattern Recognition*, 38(11), pp. 1857-1874. <http://doi.org/10.1016/j.patcog.2005.01.025>
- XAVIER, A., CARVALHO, Y., HENRIQUE, P., ALBUQUERQUE, M., REZENDE, G., JUNIOR, A., y DANTAS GUIMARÃES, R. (2009). *Spatial hierarchical clustering*, 27(3), pp. 411-442.
- XIONG, Y., y YEUNG, D.-Y. (2004). "Time series clustering with ARMA mixtures". *Pattern Recognition*, 37(8), pp. 1675-1689. <http://doi.org/10.1016/j.patcog.2003.12.018>
- ZHANG, H., HO, T. B., ZHANG, Y., y LIN, M.-S. (2006). "Unsupervised Feature Extraction for Time Series Clustering Using Orthogonal Wavelet Transform". *Informatica*, 30, pp. 305-319.

- ZHANG, X., LIU, J., DU, Y., y LV, T. (2011). "A novel clustering method on time series data". *Expert Systems with Applications*, 38(9), pp. 11891-11900. <http://doi.org/10.1016/j.eswa.2011.03.081>
- ZHEGAO, y JIANJUNYANG. (2014). "Financial Time Series Forecasting with Grouped Predictors using Hierarchical Clustering and Support Vector Regression". *International Journal of Grid Distribution Computing*, 7(5), pp. 53-64. <http://doi.org/10.14257/ijgdc.2014.7.5.05>

