

Efectos sobre las Comunidades Autónomas de los cambios metodológicos en la EPA-2002. Un análisis mediante la técnica de escalados multidimensionales

1.- Introducción.

La Encuesta de Población Activa (EPA) elaborada por el Instituto Nacional de Estadística (INE) desde 1964 es la principal fuente estadística empleada por los analistas e investigadores para el estudio del mercado de trabajo. El análisis de sus datos permite el conocimiento de los factores demográficos y socioeconómicos que afectan al tamaño y composición de la fuerza de trabajo y sirve de base para el diseño y evaluación de las políticas económicas orientadas a la promoción y creación de empleo.

A lo largo de su vida esta encuesta ha experimentado numerosos y variados cambios: en la periodicidad de su publicación, en su metodología o en las definiciones de las variables consideradas, que han obligado a la reconstrucción, en la medida de lo posible, de las series para conseguir una cierta homogeneidad en las mismas.

Por el último de ellos, tres son los cambios conceptuales y metodológicos que incorporan los datos de la EPA correspondientes al primer trimestre de 2002:

1º) El empleo de nuevas proyecciones de población: En el tercer trimestre de 2001, el INE revisó las proyecciones de población que venía utilizando, basadas en el Censo de Población del año 1991, y que no recogían el notable incremento de la población inmigrante experimentado por nuestro país en los últimos años del pasado siglo. Esta revisión provocó un aumento del valor absoluto de todas las categorías del mercado laboral (activos, ocupados, parados,...).

2º) Se efectúa una reponderación de los factores de elevación para adecuar la estructura demográfica de la población por sexo y edad resultante de la EPA con la obtenida de la información ofrecida por el Padrón Continuo. Derivada de las dificultades de realización de las entrevistas, la estructura demográfica ofrecida por la EPA subestimaba a la población relativamente joven (25-50 años) y sobre estimaba a la población de mayor edad (mayores de 65 años).

3º) Las nuevas instrucciones para considerar como “activos” los métodos de búsqueda de empleo que permiten clasificar a una persona como parada, derivados de la entrada en vigor del nuevo Reglamento 1897/2000 sobre encuestas de población activa aprobado por la Comisión Europea en septiembre de 2000. Esta modificación afecta notablemente a los datos de paro ofrecidos por la EPA, produciendo una disminución de las cifras de desempleo tanto absolutas como relativas.

El efecto combinado de estos cambios no afecta de igual forma a las diversas Comunidades Autónomas de nuestro país. El INE ha facilitado datos relativos a 2000 y 2001 con los nuevos factores, así como información sobre los efectos de la “nueva definición de parado”, dado que en 2001 incluyó en los cuestionarios de la EPA preguntas que permitiesen analizar estos efectos. Con ayuda de esta información, tomando como referencia los datos correspondientes al segundo trimestre del año 2001 y mediante el empleo de la técnica de escalados multidimensionales, trataremos de visualizar, en el presente trabajo, las diferencias

en las posiciones relativas que, con respecto a las variables relacionadas con el empleo, el desempleo y la actividad, muestran las diferentes Comunidades Autónomas españolas.

El trabajo continúa detallando en la sección dos la metodología seguida en el análisis. La sección tres muestra los resultados, que se comentan en la cuarta sección, y en la quinta y última sección se extraen las conclusiones oportunas.

2.- Metodología.

Para estudiar cómo han afectado los cambios metodológicos de la EPA a las posiciones relativas que ocupan las Comunidades Autónomas en el mercado laboral se han considerado, con la definición empleada por la propia encuesta, las tres tasas, a nuestro juicio, más significativas del panorama laboral:

- Tasa de actividad: Cociente entre el total de activos y la población de 16 o más años.
- Tasa de ocupación o empleo: Cociente entre el total de ocupados y la población de 16 o más años.
- Tasa de desempleo o paro: Es el cociente entre el número de parados y el de activos.

Con objeto de introducir en el análisis la variable género, se han empleado también las tasas correspondientes a la población femenina y masculina, definidas de igual forma pero restringidas al colectivo de la población en cuestión.

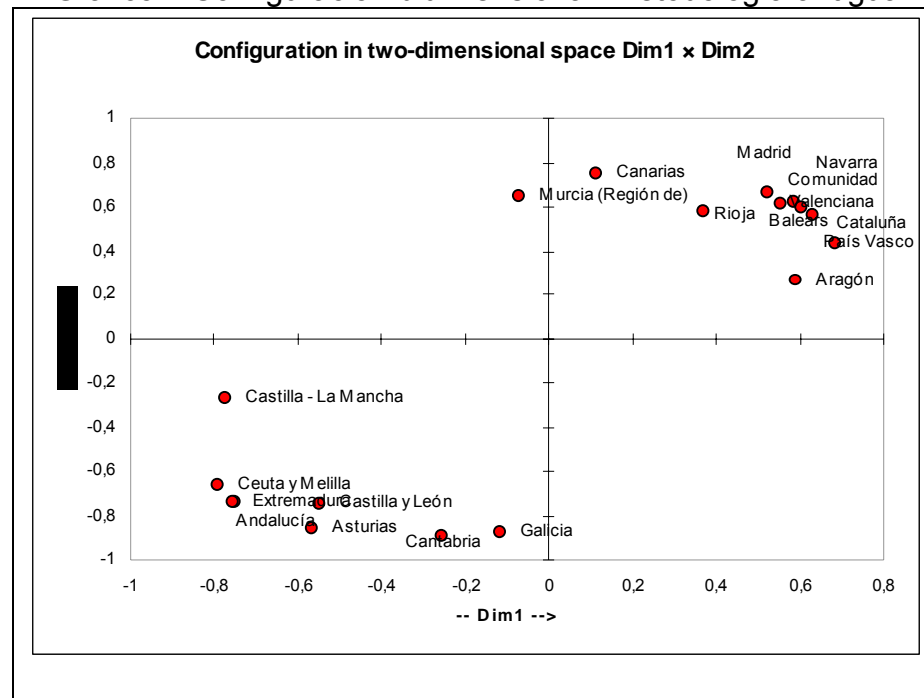
Con esta información se dispone de una tabla de datos que recoge nueve variables por Comunidad Autónoma. La situación de las mismas respecto a este conjunto de datos vendría representada en un espacio 9-dimensional, de difícil interpretación. Por ello, nuestro primer objetivo es reducir la dimensionalidad de este espacio, de modo que, sin perder demasiada información, podamos visualizar la posición relativa de las Comunidades Autónomas. Para ello hemos empleado una técnica de análisis de datos multivariantes, conocida como escalados multidimensionales (ver apéndice), que produce una información gráfica fácilmente interpretable.

Puesto que el presente estudio trata de escenificar el efecto que, los cambios metodológicos de la EPA, ha supuesto en las posiciones relativas de las Comunidades Autónomas, una vez elegida la técnica que nos permite una ordenación de las mismas, hemos seleccionado el período temporal en el que aplicarla: el segundo trimestre del año 2001. Con datos de dicho período y mediante la técnica del escalado multidimensional, hemos procedido a representar las distintas Comunidades Autónomas según los valores de las nueve variables consideradas en el estudio, medidas con la antigua metodología. A continuación, y en el mismo plano, se han representado las Comunidades Autónomas atendiendo a los valores que toman dichas variables según la metodología aplicada a partir del año 2002. Esto nos permitirá comparar las posiciones resultantes. Los resultados se muestran en el siguiente apartado.

3.- Resultados

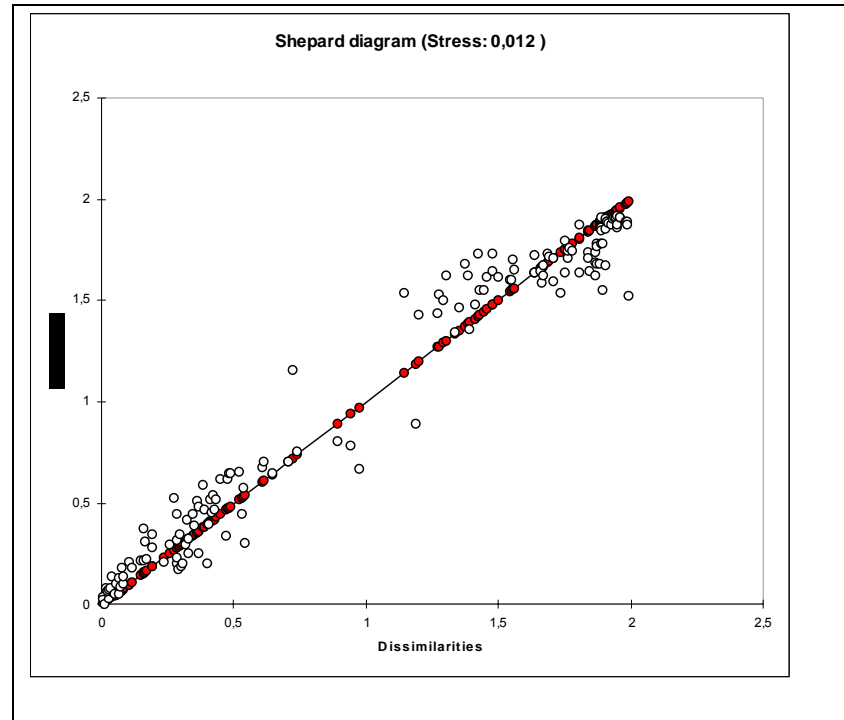
En el gráfico 1 presentamos las posiciones que ocupan las Comunidades Autónomas en un espacio bidimensional construido a partir de los datos correspondientes al segundo trimestre del año 2001, con anterioridad a los cambios metodológicos. La bondad del ajuste podemos analizarla a través del gráfico de Shepard (gráfico 2), donde observamos como las disimilaridades que presentan las observaciones están próximas a las distancias estimadas, ya que los puntos están próximos a la bisectriz. El valor para la función de Stress correspondiente a esta configuración es $S=0'012$, lo que, según la terminología de Kruskal, se considera muy bueno¹.

Gráfico 1: Configuración bidimensional. Metodología antigua



¹ Para una dimensión el valor de la función de Stress sube hasta 0'031. Para un espacio tridimensional el valor de la función de Stress es 0'011, lo que mejora sólo levemente el correspondiente a dos dimensiones y, a cambio añade complejidad al gráfico. Por ello, hemos optado por dos dimensiones.

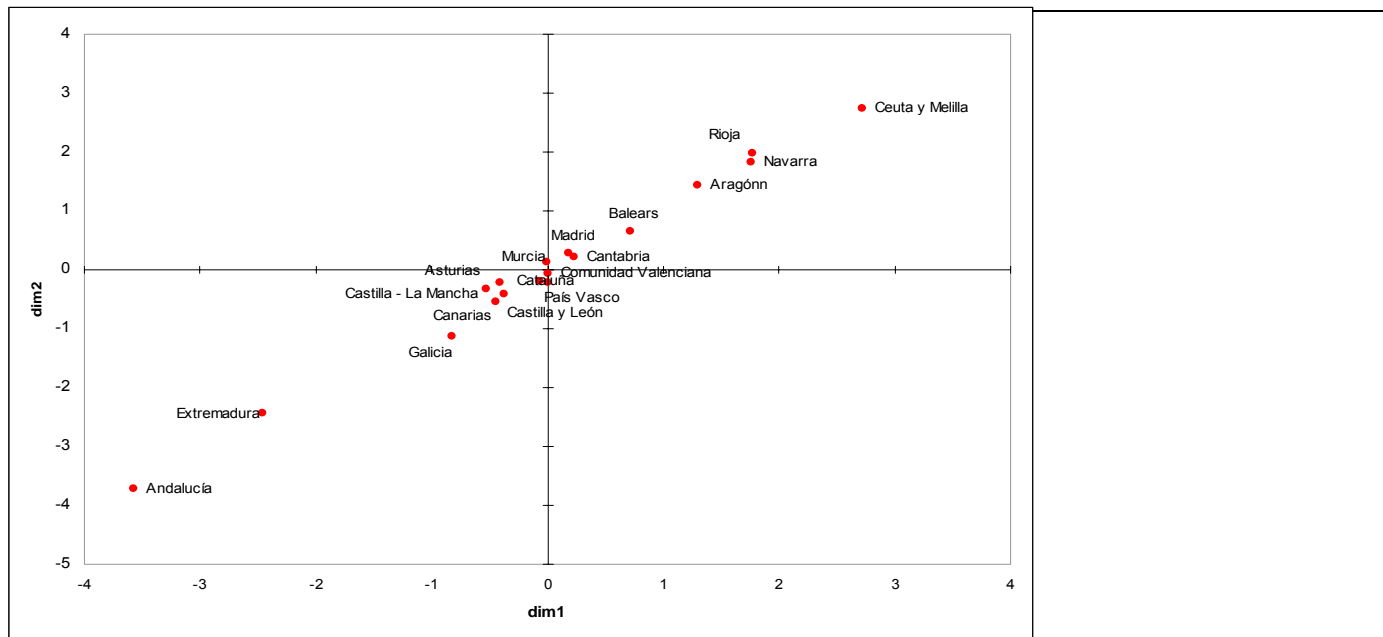
Gráfico 2: Diagrama de Shepard



Con objeto de poder representar a las Comunidades Autónomas con los datos relativos a la EPA según la nueva metodología, para el mismo ámbito temporal, hemos regresado las dos dimensiones sobre las nueve variables. Los resultados obtenidos se ofrecen en el apéndice, aquí sólo señalaremos que ambas regresiones presenta coeficientes de determinación superiores a 0'9.

Empleando ambas ecuaciones de regresión hemos estimado para cada Comunidad Autónoma, sus coordenadas respecto a las dimensiones 1 y 2. En el gráfico 3 se muestran las nuevas posiciones.

Gráfico 3: Configuración bidimensional. Nueva metodología



Asimismo, para resaltar los cambios de posición de las diversas CCAA, se presentan en un mismo gráfico ambas ordenaciones. Para distinguir unas de otras hemos presentado en color verde y seguidas de una n las Comunidades con nueva metodología.

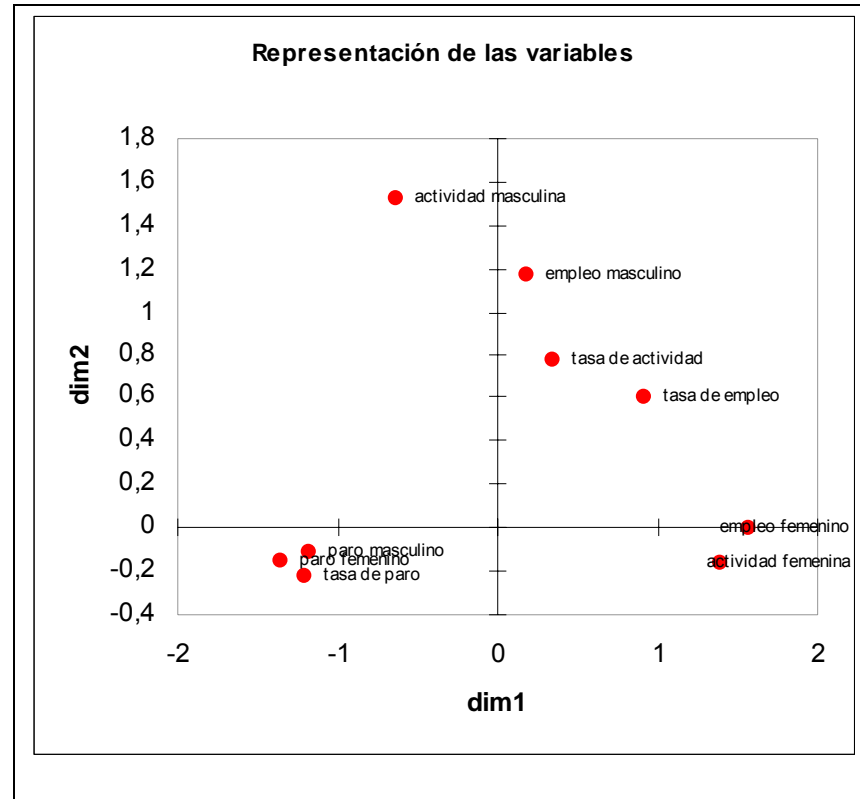
Gráfico 4: Configuración bidimensional. Antigua y nueva metodología

El significado, en términos de las variables que estamos manejando, de estas posiciones y distancias exige la interpretación de los ejes respecto a los cuales hemos representado las CCA². Para ello hemos regresado las nueve variables respecto a los dos ejes o dimensiones. Los aceptables resultados de estas regresiones, así como el análisis Profit que nos proporciona los cosenos directores de las variables con las dimensiones consideradas se ofrecen en el apéndice adjunto. En estos resultados observamos como las variables tasa de empleo femenino, tasa de paro femenino y masculino y tasa de actividad femenina, son las más relacionadas con el eje o dimensión 1, mientras que la tasa de actividad, tasa de actividad masculina y tasa de empleo femenino, son las más relacionadas con la dimensión 2.

La representación gráfica de estos resultados se ofrecen en el siguiente gráfico:

Gráfico 5: Representación de las variables

² En el algoritmo de Shepard-Kruskal que venimos manejando las posiciones de los ejes son arbitrarias, lo que dificulta su interpretación



Tras su observación, nos parece razonable asociar la dimensión 1 con la situación en el mercado laboral femenino y la dimensión 2 con el mercado laboral masculino. Cuanto más positiva sea la dimensión 1 de una Comunidad Autónoma, mejores serán sus tasas femeninas. Cuanto más positiva sea su dimensión 2, mayores son sus tasas de actividad y empleo masculino.

4.- Comentario a los resultados

En las tablas 1 y 2 del apéndice podemos observar los valores originales de las diversas tasas empleadas en nuestro análisis para ambas metodologías. Podemos comprobar como el paso de una a otra metodología ha mantenido constante el valor

de las tasas de ocupación y ha rebajado, en diverso grado, las tasas de actividad y paro de las diferentes CCAA. Pero si bien esta información nos permite apreciar la evolución, debida al cambio metodológico, seguida por dichas CCAA en cuanto a dichas tasas, no permite situar comparativamente el estado del mercado laboral de cada comunidad con los del resto de comunidades, ni apreciar los efectos producidos por el cambio metodológico en dicha situación.

Una aproximación de esta comparación entre la situación de los mercados laborales de las CCAA podemos obtenerla con los valores normalizados de las variables empleadas en este trabajo que se ofrecen en las tablas 3 y 4 del apéndice. Pero el empleo de la cantidad de información numérica contenida en dichas tablas hace bastante engorroso el análisis de la comparación deseada. El atractivo de la técnica del escalamiento multidimensional que hemos empleado reside, fundamentalmente, en presentar gráfica y sintéticamente toda esa información de la forma fácilmente interpretable que se nos muestra en el anterior apartado de resultados.

Así, en el gráfico 1, para los datos de la antigua metodología, las Comunidades Autónomas aparecen distribuidas en dos cluster o grupos perfectamente delimitados. El primero de estos cluster se sitúa, con la excepción de Murcia que presenta un valor ligeramente negativo respecto a la dimensión 1, en el primer cuadrante. Agrupa, salvo excepciones como Canarias, Aragón y Murcia, a las Comunidades Autónomas con mayores niveles de renta (o mayor grado de desarrollo económico): Baleares, Cataluña, Comunidad Valenciana, La Rioja, Madrid, Navarra y País Vasco. Sus puntos presentan coordenadas positivas respecto a ambos ejes lo que adjudica a dichas regiones una buena situación relativa en cuanto a empleo y paro. El segundo cluster, que agrupa, con la excepción cántabra, las Comunidades Autónomas con menores niveles de renta, se sitúa en el tercer cuadrante - ambas coordenadas negativas-, indicando con ello la mala situación relativa, en cuanto a tasas de actividad, empleo y paro de estas comunidades.

Los datos calculados con la nueva metodología aparecen en el gráfico 3, y conjuntamente con los datos obtenidos con la anterior metodología en el gráfico 4. Lo primero que podemos apreciar de la observación de estos gráficos es que los cambios metodológicos producen dos reacciones que podíamos calificar de contradictorias: De una parte aumenta la dispersión de la ordenación y, con ello, las distancias entre las comunidades con mejor y peor situación relativa, y de otra, una tendencia hacia el centro de la ordenación de la mayor parte de las comunidades. Así, las Comunidades con mejores y peores situaciones relativas se encuentran más alejadas entre sí y del centro del gráfico y, al mismo tiempo, con una mejora o un empeoramiento de su posición relativa, el grueso de comunidades acercan sus posiciones.

Así, en el primero de los cluster mencionados, Aragón, la Rioja y Navarra mejoran sustancialmente su ya buena posición relativa y pasan a comandar la ordenación. Por contra, el resto de comunidades del cluster empeoran su posición relativa. En algunos casos, como Cataluña, Comunidad Valenciana, País Vasco y, espectacularmente, la comunidad canaria, hasta el extremo

de pasarlas al grupo enmarcado en el tercer cuadrante, representativo de una mala situación relativa en cuanto a las tasas laborales.

En cuanto al segundo cluster, aparece un comportamiento similar de las comunidades integrantes del mismo. Así, Las Comunidades Autónomas de Andalucía, Galicia y Extremadura, extreman su mala posición relativa en la primera de las ordenaciones y cierran la nueva ordenación tras los cambios metodológicos. Las demás comunidades de este grupo mejoran posiciones, destacando Cantabria y Ceuta y Melilla. La primera, con un comportamiento análogo pero de sentido inverso al canario, pasa en la nueva ordenación al primer cluster, un lugar más acorde con su buena situación económica. En cuanto al cambio de Ceuta y Melilla, que de una mala posición relativa de su mercado laboral, pasa a liderar la nueva ordenación con la mejor posición relativa en cuanto a sus tasas de actividad, empleo y paro, conviene señalar que, dada la escasa entidad y significatividad de sus cifras absolutas en la muestra de la EPA, cualquier cambio en las mismas producen variaciones espectaculares en sus cifras relativas. Todo ello permite considerarla como outlier y justifica su no inclusión en la mayoría de análisis de este tipo.

Llegados a este punto, convendría preguntarnos por las causas que han motivado este comportamiento de las diversas Comunidades Autónomas ante los cambios metodológicos. En la respuesta que ofrecemos aquí a esta pregunta cabe adelantar algunas consideraciones.

En primer lugar, señalaremos que algunas de las variaciones en la ordenación relativa de las Comunidades Autónomas con respecto a las variables consideradas, pueden ser debidas a la estacionalidad de los datos empleados (trimestrales), así como a factores cíclicos. Igualmente, las diversas estructuras demográficas de las distintas comunidades, así como las peculiares características de los diferentes mercados regionales también aportarían causas explicativas a los cambios acaecidos. Pero dadas las limitaciones y objetivos del presente trabajo, centraremos nuestra atención exclusivamente en las causas directamente derivadas de los cambios metodológicos.

En cuanto al primero de ellos, la utilización de nuevas proyecciones de población en el cálculo de los factores de elevación de la encuesta³, no produce cambios en las tasas. Las anteriores proyecciones, basadas en el Censo de 1991, subestimaban la

³ El factor de elevación es el coeficiente utilizado para elevar los resultados muestrales a poblacionales. Se obtiene mediante el cociente entre la población total y el total de efectivos de la muestra. A partir de los datos de la encuesta, la estimación del número de efectivos en la población de las diversas categorías laborales, por ejemplo el nº de ocupados, se calcula del siguiente modo:

$$\text{Estimación del nº de ocupados} = \text{nº de ocupados de la muestra} \times \frac{\text{Pobl. Total}}{\text{Total de la muestra}}$$

Por ello, si aumenta las proyecciones de población, es decir, la población total, también lo hará todas las estimaciones (de activos, ocupados, parados...) de la EPA.

inmigración realmente producida. La aplicación a la encuesta de las nuevas proyecciones provoca un aumento del valor absoluto de los efectivos de todas las categorías del mercado de trabajo, sin embargo, las tasas prácticamente no deben variar al incrementarse en proporción similar numerador y denominador de las mismas.

Como consecuencia de este cambio, la población española en edad de trabajar aumentó, en el 4º trimestre de 2001⁴, un 2'5% respecto a la proyección anteriormente vigente. Este aumento de población se repartió de forma muy desigual entre las diferentes comunidades. Generalmente, los mayores incrementos se dieron en las regiones con mayor recepción inmigratoria. Así, según datos de estudios recientes (IEA, 2002), para el 4º trimestre de 2001 los mayores incrementos porcentuales de la población de 16 o más años se dieron en Madrid y las comunidades autónomas del arco mediterráneo (Baleares, C. Valenciana, Cataluña y Murcia) que son las que sufren mayor presión inmigratoria. Todas ellas junto con La Rioja presentan, en las nuevas proyecciones para dicho trimestre, incrementos porcentuales de población superiores a la media española. El resto de comunidades crecen por debajo de la media nacional, y en algunos casos como Aragón, Extremadura y Castilla-la Mancha no se perciben prácticamente diferencias con las nuevas proyecciones.

Puesto que el colectivo de inmigrantes suele presentar tasas de actividad y de paro superiores a los de la población autóctona, es razonable suponer que el factor migratorio tenga bastante que decir sobre los cambios en las posiciones relativas de las comunidades.

En cuanto a las reponderaciones, los resultados de las mismas corrigen de forma sustancial la estructura por sexo y edad de la población. En relación al sexo y a escala nacional, el efecto conjunto de las reponderaciones y el aumento poblacional debido a las nuevas proyecciones hace crecer la población masculina y femenina, pero en mayor grado la primera⁵. Ello parece confirmar que las estimaciones de la antigua EPA subestimaba ambas poblaciones pero, especialmente, la masculina.

Con referencia a la estructura por edades de la población, el grupo de “adultos jóvenes” más afectado por la reponderación son el de 25-44 años para los hombres (incrementos del 14'8%, en el 2º trimestre de 2000) y 25-39 para las mujeres (14'1%). En contrapartida, las generaciones mayores corrigen a la baja sus efectivos. Por efecto de la reponderación, los grupos de edad de 25 a 54 años y 55 y más años pasan del 49'3% y 35'3% de la población, al 52'7% y 31'7%, respectivamente según información del INE para el 4º trimestre de 2001. El resto de grupos permaneció invariable.

Por tanto, así como el incremento de población de las nuevas proyecciones conlleva el aumento de efectivos de todas las categorías laborales, las correcciones de las reponderaciones en cuanto a la estructura de edades tenderán a elevar las cifras de actividad, puesto que incrementan fundamentalmente los grupos de población más jóvenes, atribuyendo mayor peso en la estructura poblacional a los grupos de edades más activos. Ahora bien, este aumento de la proporción de personas jóvenes, más

⁴ Último trimestre que puede ser empleado para la comparación. A partir del primer trimestre de 2002 sólo se publican los datos con la nueva metodología.

⁵ En algunas comunidades (Andalucía) y para el 2º trimestre de 2000, la reponderación reduce la población femenina.

activas desde el punto de vista económico, no tiene por que afectar, significativamente, a que dicho incremento se reparta de alguna forma entre ocupados y parados, por lo que la reponderación, en cierto modo, podemos considerarla neutral respecto a las tasas de paro.

Por otra parte, la tasa de actividad es la suma de la tasa de ocupación y la tasa absoluta de paro:

$$(\text{Activos}/\text{Pob.}>16 \text{ años}) = (\text{Ocupados}/ \text{Pob.}>16 \text{ años}) + (\text{Parados}/ \text{Pob.}>16 \text{ años})$$

y, asimismo, la tasa de paro puede expresarse como el cociente de la tasa absoluta de paro y la tasa de actividad:

$$(\text{Parados}/\text{Activos}) = (\text{Parados}/ \text{Pob.}>16 \text{ años}) : (\text{Activos}/\text{Pob.}>16 \text{ años})$$

Estas relaciones muestran la complejidad del análisis de los efectos de la reponderación en las tasas. En la tasa de actividad es suma de los efectos de estas reponderaciones sobre las tasas de ocupación y la tasa absoluta de paro, pero como los efectos sobre la tasa de actividad influyen a su vez sobre la tasa de paro, los efectos sobre ésta de la reponderación dependerá de la importancia relativa de los efectos de dicha reponderación sobre la tasa de ocupación y la tasa absoluta de paro.

Lo anteriormente expresado indica la dificultad que supone adentrarnos más en el análisis de la incidencia sobre las tasas de las reponderaciones de los factores de elevación, pero pone de manifiesto la importancia que sobre las variaciones en las ordenaciones relativas de las comunidades tiene el tercero de los cambios metodológicos, al incidir directamente sobre la tasa absoluta de paro.

Dicho cambio, motivado por la nueva interpretación que da al criterio de búsqueda activa de empleo la aplicación del reglamento comunitario 1897/2000, afecta directamente sólo a las cifras de parados (rebajándolas), pero por las relaciones antes señaladas traslada su incidencia sobre las tasas de actividad y paro.

El nuevo reglamento cambia el concepto de búsqueda activa de empleo, eliminando algunos métodos de búsqueda considerados hasta entonces válidos y añadiendo la exigencia de un “contacto activo” (es decir, para encontrar trabajo) con una oficina de empleo de la Administración en las cuatro semanas anteriores a la realización de la encuesta, a los que señalan como método de búsqueda de empleo “la inscripción en las oficinas de empleo” (85’4% de los parados, el 30’5% de ellos lo señalan como único método de búsqueda).

La aplicación del nuevo reglamento produce importantes exclusiones (20’6% en promedio para el año 2001) en las cifras de parados. Pero del total de exclusiones, la mayor parte de ellas están relacionadas con los que mencionan la inscripción en las oficinas de empleo como forma de búsqueda activa y no tienen un “contacto activo”. Así, para el 2º trimestre de 2001 (período de nuestro análisis), de las 504.600 exclusiones por la aplicación del nuevo reglamento de las cifras de paro, 476.400 (94’4%) están relacionadas con la inscripción, 10.200 (2%) por haber encontrado empleo y 18.100 (3’6%) por utilizar métodos de búsqueda de empleo que ya no se consideran activos.

Además, como se señala en el estudio (IEA, 2002) de donde provienen las anteriores cifras, para el 4º trimestre de 2001, en la Rioja, Asturias y Aragón, los porcentajes de parados que señalan la inscripción como único método de búsqueda superan ampliamente el 60%, mientras que no llegan al 20% en el País Vasco, Comunidad Valenciana y Canarias, y ni siquiera alcanza el 5% en Cataluña. Asimismo, se observan importantes diferencias en el porcentaje de exclusiones motivadas por la ausencia de “contacto activo” entre las diferentes comunidades. Así, mientras en Baleares apenas se dan exclusiones⁶ por esta causa, en Ceuta y Melilla, Extremadura, Asturias, Canarias, Castilla-La Mancha y Cantabria, superan el 70%. El resto de comunidades, como la andaluza, presenta valores intermedios entre los señalados para los parámetros analizados.

A la vista de lo expuesto, podemos señalar este cambio metodológico como el causante de algunos de los cambios de posición observados en los gráficos anteriores.

5.- Conclusiones

Con las salvedades apuntadas al comentar los resultados:

- Efectos contrapuestos de los cambios metodológicos en las ordenaciones relativas de las Comunidades Autónomas.
- Supuesto razonable: factor inmigratorio como causa explicativa de algunos cambios.
- Complejidad del análisis de la influencia de las reponderaciones sobre las diferentes tasas.
- Importancia del endurecimiento del concepto de búsqueda activa de empleo, que reduce las cifras y las tasas de paro, y que presenta una influencia geográfica muy diversa, como causa explicativa de los cambios en las ordenaciones relativas de las Comunidades Autónomas respecto a las variables consideradas.

Bibliografía

Arce, C. Escalamiento Multidimensional. Ed. PPU. Barcelona 1993

Instituto de Estadística de Andalucía. Encuesta de Población Activa: Informe para Andalucía. Dirección: Toharia, L. Sevilla 2002

Kruskal, J. B.: “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis”, Psychometrika , 29, 1964^a, pp. 1-27.

⁶ El propio estudio apunta como causa la importancia del trabajo fijo discontinuo en esta comunidad: Un tipo de trabajo que facilita el contacto activo con las oficinas del INEM

Kruskal, J. B.: "Nonmetric multidimensional scaling: a numerical method", Psychometrika, 29, 1964b, pp. 115-130.
Mantha, V.: "Factor Analysis and Multidimensional Scaling", Seminar Papers, Department for Electrical and Computer Engineering, Universidad Estatal de Mississippi, 1998.
Peña, D.: Análisis de Datos Multivariantes, McGraw-Hill, Madrid, 2002.

APÉNDICE

EL ESCALADO O ESCALAMIENTO MULTIDIMENSIONAL.

Es un conjunto de técnicas o métodos matemáticos utilizados en el campo de las Ciencias Sociales para el análisis de las evaluaciones subjetivas entre pares de entidades. Mediante estas técnicas se transforman las distancias o disimilaridades entre observaciones en un espacio de menor dimensión de tal forma que se conserva la estructura subyacente de los datos. Asimismo permiten obtener una representación geométrica de los datos en un espacio de mínima dimensionalidad, de forma que sea interpretable por simple inspección visual. Esta última característica del escalado multidimensional lo hace especialmente atractivo.

El análisis de escalados multidimensionales⁷ se basa en lo siguiente: Dada n observaciones de R^p , se define una medida de proximidad que cuantifica el grado de similitud o disimilitud entre éstas. A partir de esta medida se construye la matriz de similitud o disimilitud⁸. Como medida de proximidad se puede utilizar la correlación.

La técnica parte de una matriz de distancias o disimilaridades. Existe una multitud de variantes de los escalados multidimensionales con funciones de costes ligeramente diferentes y diversos algoritmos de optimización. Los algoritmos diseñados para analizar la matriz de disimilitud y que pueden ser utilizados para reducir la dimensionalidad del conjunto de datos se dividen en dos tipos básicos: métricos, cuando se parte de una matriz de distancias⁹, y no métricos, cuando se parte de una matriz de similitudes o similitudes¹⁰.

⁷ Una descripción más detallada de esta técnica puede consultarse en Peña (2002) o Arce C. (1993).

⁸ Se dirá de disimilitud, si a mayor semejanza le corresponde mayor puntuación, y de similitud, si a mayor semejanza le corresponde menor puntuación.

⁹ Un ejemplo típico sería la matriz de distancias kilométricas entre un conjunto de ciudades dadas dos a dos y mediante esta técnica trataríamos de representar en un plano esas ciudades.

¹⁰ Una matriz de este tipo se obtendría de comparar, por ejemplo, una marca de tabaco con otras, o una marca de coche con otras.

En el supuesto del escalado métrico, la disimilaridad δ_{ij} entre dos observaciones i y j , vendrá representada en un espacio multidimensional por una distancia d_{ij} , de forma que: $\delta_{ij} = f(d_{ij})$

Aunque los modelos de escalamiento multidimensional permiten considerar diversas definiciones de distancia, la mayoría emplean la distancia euclídea:

$$d_{ij} = \left[\sum_{s=1}^p (x_{is} - x_{js})^2 \right]^{1/2}$$

y f es una función lineal de pendiente positiva¹¹.

En el escalado no métrico (Shepard, 1962; Kruskal, 1964^a, 1964b), se plantea que la función que relaciona las disimilaridades con las distancias es una función desconocida a la que sólo se le impone que sea una función monótona: creciente si es de disimilaridades o decreciente si es de similaridades.

En este caso (que es el nuestro) de escalado no métrico, el objetivo es encontrar unas coordenadas que sean capaces de reproducir las distancias a partir de la única condición de monotonía. Ello obliga a definir:

- 1.- Un criterio de bondad de ajuste que sea invariante ante transformaciones monótonas de los datos.
- 2.- Un algoritmo para obtener las coordenadas, optimizando el criterio establecido.

Habitualmente se utiliza como criterio de ajuste la denominada función de STRESS, definida como:

$$S = \frac{\sum_{ij} (d_{ij} - \hat{d}_{ij})}{\sum_{ij} d_{ij}^2}$$

siendo \hat{d}_{ij} las denominadas “disparidades”, que son unos valores ajustados a las distancias y que están en un orden lo más similar posible a los datos.

El procedimiento de cálculo sigue los siguientes pasos:

- Generar una configuración inicial de puntos en un espacio de dimensionalidad prefijada.
- Normalizar la configuración.

¹¹ Para ver como proceder a partir de la matriz de distancias puede consultarse Peña (2002).

- Calcular las distancias entre cada par de puntos.
- Ajustar las distancias mediante una función monótona obteniéndose las disparidades.
- Calcular el valor de la función de STRESS¹². Si es aceptable, la configuración de puntos de partida es la solución final, en caso contrario, hay que mover la configuración de puntos en alguna dirección y se comienza de nuevo.

Regresiones dim 1 y dim 2 respecto a las tasas de actividad empleo y paro

dim1				
Parameter	Value	Standard deviation	Student's t	Pr > t
Intercept	0,000	0,049	0,000	1,000
tasa de empleo	28,186	17,466	1,614	0,145
empleo femenino	-14,327	9,925	-1,444	0,187
empleo masculino	-21,064	10,024	-2,101	0,069
tasa de paro	15,446	10,287	1,501	0,172
paro femenino	-8,439	4,211	-2,004	0,080
paro masculino	-10,962	6,841	-1,602	0,148
tasa de actividad	-20,905	13,249	-1,578	0,153
actividad femenina	9,867	7,335	1,345	0,215
actividad masculin:	17,369	8,105	2,143	0,064

Coefficiente de determinación (R^2)=0'94 C. de determinación corregido = 0'87

¹² Kruskal (1964 a y b) considera que un valor de la función de STRESS igual a 0'01 indica una solución excelente, entre 0'01 y 0'05 buena, entre 0'05 y 0'10 pasable, entre 0'10 y 0'15 regular y mayor de 0'15 mala.

dim2

Parameter	Value	Standard deviation	Student's t	Pr > t
Intercept	0,000	0,066	0,000	1,000
tasa de empleo	20,723	23,475	0,883	0,403
empleo femenino	-10,084	13,339	-0,756	0,471
empleo masculino	-18,145	13,473	-1,347	0,215
tasa de paro	13,254	13,826	0,959	0,366
paro femenino	-7,364	5,659	-1,301	0,229
paro masculino	-10,157	9,194	-1,105	0,301
tasa de actividad	-15,991	17,807	-0,898	0,395
actividad femenina	6,852	9,859	0,695	0,507
actividad masculina	15,659	10,893	1,437	0,189

Coeficiente de determinación (R^2)=0'92 C. de determinación corregido = 0'83

Ordenación de las CCAA según sus coordenadas en cada dimensión

METODOLOGÍA ANTIGUA

observation	Dim1	observation	Dim2
País Vasco	0,684	Canarias	0,752
Cataluña	0,628	Comunidad Valenciana	0,670
Madrid (Comunidad de)	0,604	Murcia (Región de)	0,650
Aragón	0,588	Balears (Illes)	0,622
Balears (Illes)	0,583	Navarra (Comunidad For)	0,620
Navarra (Comunidad For)	0,551	Madrid (Comunidad de)	0,604
Comunidad Valenciana	0,523	Rioja (La)	0,580
Rioja (La)	0,370	Cataluña	0,569
Canarias	0,113	País Vasco	0,436
Murcia (Región de)	-0,072	Aragón	0,272
Galicia	-0,119	Castilla - La Mancha	-0,270
Cantabria	-0,257	Ceuta y Melilla	-0,663
Castilla y León	-0,548	Andalucía	-0,740
Asturias (Principado de)	-0,566	Extremadura	-0,740
Andalucía	-0,754	Castilla y León	-0,749
Extremadura	-0,757	Asturias (Principado de)	-0,854
Castilla - La Mancha	-0,775	Galicia	-0,873
Ceuta y Melilla	-0,795	Cantabria	-0,886

METODOLOGÍA NUEVA

observación	dim1	observación	dim2
Ceuta y Melilla	2,723	Ceuta y Melilla	2,732
Rioja (La)	1,771	Rioja (La)	1,978
Navarra (Comunidad For:	1,761	Navarra (Comunidad For:	1,819
Aragón	1,298	Aragón	1,433
Balears (Illes)	0,715	Balears (Illes)	0,655
Cantabria	0,232	Madrid (Comunidad de)	0,281
Madrid (Comunidad de)	0,178	Cantabria	0,218
Comunidad Valenciana	0,001	Murcia (Región de)	0,132
País Vasco	0,000	Comunidad Valenciana	-0,075
Murcia (Región de)	-0,001	Cataluña	-0,196
Cataluña	-0,060	País Vasco	-0,214
Castilla y León	-0,377	Asturias (Principado de)	-0,228
Asturias (Principado de)	-0,414	Castilla - La Mancha	-0,321
Canarias	-0,448	Castilla y León	-0,409
Castilla - La Mancha	-0,525	Canarias	-0,535
Galicia	-0,818	Galicia	-1,133
Extremadura	-2,465	Extremadura	-2,424
Andalucía	-3,573	Andalucía	-3,712

Distancias euclídeas entre las posiciones para cada CCAA en la representación bidimensional con una u otra metodología

Distancias entre Comunidades

Andalucía	4,096482281
Aragón	1,360651588
Asturias (Principad	0,644426407
Balears (Illes)	0,136648587
Canarias	1,403699169
Cantabria	1,207251169
Castilla y León	0,380336084
Castilla - La Manch:	0,25541892
Cataluña	1,028347438
Comunidad Valenci	0,909751986
Extremadura	2,398338371
Galicia	0,746303335
Madrid (Comunidad	0,533707441
Murcia (Región de)	0,522799074
Navarra (Comunida	1,703203513
País Vasco	0,943279852
Rioja (La)	1,979080795
<u>Ceuta y Melilla</u>	4,889083972

Ordenación de las CCAA de mayor o menor distancia euclídea entre sus posiciones con ambas metodologías

Ceuta y Melilla
Andalucía
Extremadura
La Rioja

Navarra
Canarias
Aragón
Cantabria
Cataluña
País Vasco
Comunidad Valenciana
Galicia
Asturias
Madrid
Castilla y León
Castilla- La Mancha
Balears

Regresiones de las variables empleadas en el análisis respecto a los ejes

T empleo = 0,913 Dim1 + 0,606 Dim2 R²=0,814 F=32,758 $\alpha \leq 0,0001$
T empleo fem = 1,570 Dim1 + 0,003 Dim2 R²=0,832 F=37,170 $\alpha \leq 0,0001$
T empleo masc = 0,177 Dim1 + 1,178 Dim2 R²=0,796 F= 29,248 $\alpha \leq 0,0001$
T paro = -1,212 Dim1 – 0,218 Dim2 R²=0,689 F= 16,594 $\alpha \leq 0,001$
T paro fem = -1,363 Dim1 – 0,147 Dim2 R²=0,766 F= 24,559 $\alpha \leq 0,0001$
T paro masc = -1,187 Dim1 – 0,112 Dim2 R²=0,567 F=9,83 $\alpha \leq 0,002$
T actividad = 0,335 Dim1 + 0,780 Dim2 R²=0,493 F=7,291 $\alpha \leq 0,006$
T actividad fem = 1,392 Dim1 – 0,16 Dim2 R²=0,518 F=8,05 $\alpha \leq 0,004$
T actividad masc = -0,64 Dim1 + 1,53 Dim2 R²=0,591 F= 10,826 $\alpha \leq 0,001$

Análisis Profit
Cosenos directores de las variables con las dimensiones consideradas

Variables	con dim1	con dim2
Tasa de empleo	0'830	0'550
Tasa de empleo femenino	1'000	0'000
Tasa de empleo masculino	0'150	0'988
Tasa de paro	0'980	0'177
Tasa de paro femenino	0'990	0'107
Tasa de paro masculino	0'990	0'094
Tasa de actividad	0'390	0'918
Tasa de actividad femenina	0'993	0'114
Tasa de actividad masculina	0'390	0'922

Tabla 1. Tasas de actividad, ocupación y paro totales y según género, por CC.AA. Metodología ANTIGUA

2º trim 2001	tasa de empleo	empleo femer	empleo masci	tasa de paro	paro femer	paro masculin	tasa de activic	actividad femer	actividad masi
Andalucía	40,661%	26,285%	55,822%	21,809%	31,686%	15,760%	52,002%	38,477%	66,266%
Aragón	47,190%	33,825%	61,097%	8,306%	13,661%	4,876%	51,465%	39,177%	64,251%
Asturias (Principado de)	39,130%	26,815%	52,755%	15,018%	23,031%	9,743%	46,045%	34,839%	58,424%
Balears (Illes)	55,872%	43,145%	69,072%	6,438%	9,165%	4,582%	59,717%	47,498%	72,389%
Canarias	48,496%	34,441%	62,763%	12,654%	17,744%	9,537%	55,522%	41,870%	69,380%
Cantabria	45,228%	32,480%	58,888%	13,051%	20,502%	8,048%	52,017%	40,802%	63,993%
Castilla y León	43,406%	29,252%	58,077%	12,018%	20,168%	7,082%	49,335%	36,630%	62,504%
Castilla - La Mancha	44,325%	27,952%	61,072%	12,731%	22,481%	7,272%	50,792%	36,059%	65,861%
Cataluña	52,861%	41,154%	65,262%	8,584%	11,728%	6,357%	57,825%	46,622%	69,692%
Comunidad Valenciana	49,685%	35,712%	64,444%	11,204%	17,444%	7,084%	55,953%	43,258%	69,364%
Extremadura	40,925%	25,775%	56,577%	21,958%	34,094%	14,544%	52,425%	39,108%	66,206%
Galicia	44,694%	33,782%	56,638%	14,468%	20,972%	9,596%	52,249%	42,747%	62,650%
Madrid (Comunidad de)	51,373%	39,054%	64,925%	9,506%	13,247%	6,842%	56,770%	45,017%	69,694%
Murcia (Región de)	48,057%	32,250%	64,455%	11,454%	18,575%	7,277%	54,273%	39,581%	69,514%
Navarra (Comunidad Foral de)	51,982%	38,371%	65,865%	6,014%	8,884%	4,213%	55,308%	42,160%	68,761%
País Vasco	48,851%	36,322%	62,120%	10,918%	16,483%	7,098%	54,838%	43,491%	66,867%
Rioja (La)	48,078%	33,770%	62,757%	7,332%	10,829%	5,277%	51,834%	37,871%	66,159%
Ceuta y Melilla	43,015%	26,415%	58,602%	20,812%	33,649%	13,684%	54,320%	39,811%	68,100%

Fuente: EPA. INE

Tabla 2. Tasas de actividad, ocupación y paro totales y según género, por CC.AA. Metodología NUEVA

2ºtrim2001nueva	tasa de empleo	empleo femenino	empleo masculino	tasa de paro	paro femenino	paro masculino	tasa de actividad	actividad femenina	actividad masculina
Andalucía	40,661%	26,285%	55,822%	18,301%	26,523%	13,493%	49,769%	35,778%	64,528%
Aragón	47,190%	33,825%	61,097%	4,835%	7,729%	3,085%	49,599%	36,659%	63,043%
Asturias (Principado de)	39,130%	26,815%	52,755%	8,409%	12,810%	5,736%	42,723%	30,755%	55,965%
Balears (Illes)	55,872%	43,145%	69,072%	5,818%	8,422%	4,053%	59,340%	47,113%	71,990%
Canarias	48,496%	34,441%	62,763%	10,455%	14,835%	7,835%	54,166%	40,440%	68,099%
Cantabria	45,228%	32,480%	58,888%	8,147%	13,341%	4,790%	49,240%	37,431%	61,851%
Castilla y León	43,406%	29,252%	58,077%	9,623%	16,231%	5,759%	48,028%	34,909%	61,626%
Castilla - La Mancha	44,325%	27,952%	61,072%	9,141%	15,966%	5,551%	48,784%	33,263%	64,646%
Cataluña	52,861%	41,154%	65,262%	8,364%	11,318%	6,273%	57,686%	46,411%	69,629%
Comunidad Valenciana	49,685%	35,712%	64,444%	9,623%	14,997%	6,150%	54,975%	42,013%	68,667%
Extremadura	40,925%	25,775%	56,577%	14,041%	21,680%	9,872%	47,609%	32,910%	62,774%
Galicia	44,694%	33,782%	56,638%	11,434%	16,620%	7,686%	50,464%	40,516%	61,353%
Madrid (Comunidad de)	51,373%	39,054%	64,925%	7,180%	9,918%	5,282%	55,350%	43,354%	68,546%
Murcia (Región de)	48,057%	32,250%	64,455%	9,083%	14,601%	5,926%	52,858%	37,764%	68,516%
Navarra (Comunidad Foral de)	51,982%	38,371%	65,865%	4,261%	5,466%	3,462%	54,273%	40,636%	68,226%
Pais Vasco	48,851%	36,322%	62,120%	9,587%	14,619%	6,158%	54,031%	42,554%	66,197%
Rioja (La)	48,078%	33,770%	62,757%	3,972%	5,839%	2,909%	50,066%	35,864%	64,637%
Ceuta y Melilla	43,015%	26,415%	58,602%	5,455%	7,895%	4,373%	45,496%	28,679%	61,470%

Fuente: EPA. INE

María José Vázquez Cueto y Vicente Rodríguez Sosa

Universidad de Sevilla

e-mail: pepi@us.es ; vsosa@us.es ; Fax: 954551667

Resumen:

Los cambios metodológicos incorporados a la EPA a partir del primer trimestre de 2002 han supuesto modificaciones en los valores absolutos de las variables más relevantes del mercado de trabajo, respecto a los valores ofrecidos con la anterior metodología. Dichas modificaciones no han afectado por igual a las diversas Comunidades Autónomas. El objetivo del presente trabajo es dejar constancia del distinto impacto de estos cambios en las diferentes regiones, así como reflejar su incidencia en las posiciones relativas de las Comunidades Autónomas respecto a las tasas más significativas de los mercados laborales: tasas de actividad, empleo y paro. Para ello, se hace uso de una técnica empleada con frecuencia en el campo de las ciencias sociales para el análisis de evaluaciones subjetivas entre pares de entidades: el escalado multidimensional.