

REVISTA DE ESTUDIOS REGIONALES

I.S.S.N.: 0213-7585

2ª EPOCA Enero-Abril 2021



120

SUMARIO

Manuel César Vila, Xesús Pereira López y Rosa María Verdugo Matés. Análisis comparativo de las fuentes estadísticas para la proyección de series temporales de migraciones regionales clasificadas por niveles educativos

Montserrat Lira Raggio. El tratamiento de los *Business Angels* en las disposiciones de la Unión Europea. El impacto territorial

Eva M. de la Torre, Fernando Casani y Carmen Perez-Esparrells. Measuring universities' engagement: a revision of the European research projects and the actual use of the so-called 'third mission' indicators

Jesús Artero López, Rosario Gómez-Álvarez Díaz y David Patiño Rodríguez. El impacto redistributivo de un sistema de renta básica universal en Andalucía

Antonio Sánchez González. Cartografía y litigio territorial en los confines de Aragón y raya de Castilla: la pila bautismal que separa dos reinos

Mª Ángeles Rodríguez Domenech y Isabel Rodríguez Domenech. Brand image in intermedium size cities. Identifying the cities' first-generation effect with high rail speed in Ciudad Real (Spain)

I. Artículos

Análisis comparativo de las fuentes estadísticas para la proyección de series temporales de migraciones regionales clasificadas por niveles educativos

Comparative analysis of statistical sources for the projection of time series of regional migrations classified by educational levels

Manuel César Vila

Xesús Pereira López

Rosa María Verdugo Matés

Universidad de Santiago de Compostela

Recibido, Abril de 2018; Versión final aceptada, Septiembre de 2019.

PALABRAS CLAVE: Censo, EPA, EVR, Migraciones interiores, Nivel educativo, Iterative proportional fitting.

KEY WORDS: Population census, Labour Force Survey (EPA), Residential Variation Statistics (EVR), Internal migrations, education level, Iterative proportional fitting.

Clasificación JEL: C67, O15, R15

RESUMEN

En el presente trabajo se analiza la robustez estadística de los censos de población y de la Encuesta de Población Activa con el objetivo de estimar matrices regionales de flujos migratorios clasificados por niveles educativos, mediante variantes del procedimiento iterative proportional fitting, a partir de la información contenida en ellos.

Dado que las estadísticas oficiales sobre migraciones no permiten clasificar a los migrantes en función de su nivel formativo, es necesario utilizar fuentes alternativas para modelizar los movimientos poblacionales entre comunidades autónomas teniendo en cuenta su formación.

Recurrir a estas fuentes alternativas obliga a realizar un análisis previo de las mismas, para la elección de la que permita la mejor estimación de flujos migratorios interregionales para los diferentes niveles educativos.

Con este fin se analiza la robustez de la información migratoria de los censos y de la Encuesta de Población Activa, utilizando diferentes indicadores y procedimientos.

Se realiza una primera comparación de ambas fuentes con los datos globales de la Estadística de Variaciones Residenciales; análisis que es complementado con los

resultados de las ecuaciones compensatorias para el conjunto español y para cada una de sus comunidades autónomas, a partir de la información de los dos últimos censos de población, calculando a continuación los errores relativos medios (ERM) para los totales poblacionales y para las personas con estudios de segundo y de tercer grado, de cada una de esas regiones.

La conclusión de estos análisis es que la variable movilidad de los censos de población se encuentra afectada por la falta de respuesta y por el método de imputación utilizado por el propio INE para corregirla, explicando la infravaloración de los datos de los censos con respecto a la EVR.

A partir de los informes de evaluación de la calidad de los datos del Censo de Población 2001 y de la Encuesta de Población Activa para los años 2006 a 2017, se resumen los resultados de varios indicadores como son el porcentaje de idénticamente clasificados, la tasa de diferencia neta, el índice de cambio neto o la tasa de diferencia bruta; pudiendo concluir que una mayor robustez para los datos de la EPA que para los del Censo de 2001.

Por último, utilizando la Estadística de Variaciones Residenciales del año 2001 que ofrece datos migratorios interiores clasificados por titulación académica, son calculados indicadores de igualdad y relación como el estadístico de desigualdad U-Theil, el coeficiente de variación de Pearson y el de correlación, entre esta fuente y el censo y la EPA de ese mismo año. Igual que sucedía con las comparaciones previas, el análisis de los estadísticos anteriores reporta unos resultados mejores para los datos de la EPA con respecto a la EVR, que los datos del censo con respecto a esa fuente.

Una vez escogida la EPA como fuente estadística para la proyección de flujos migratorios regionales considerando los niveles formativos, se emplea el *iterative proportional fitting* (IPF) para obtener las matrices migratorias.

El procedimiento para la estimación de los flujos migratorios a partir de la información de los microdatos de la EPA es verificado en varias fases. Comenzando por la matriz generada anualmente a partir de los cuatro ficheros de microdatos trimestrales, filtrando la población de 16 y más años que el año anterior cambió de residencia y agregando la información en una matriz de 20x5 filas y 19 columnas, que muestra la estructura de flujos migratorios interregionales y de los inmigrantes procedentes del exterior clasificados en cinco niveles de estudios. A partir de los márgenes totales de esa matriz de microdatos, son calculados unos nuevos por elevación, convergiendo con los valores agregados del módulo de la EPA "Variables de submuestra: personas que han cambiado de residencia hace un año" utilizando un procedimiento de ajuste biproporcional. De esta manera, las migraciones clasificadas por niveles educativos de las NUTS-2 proporcionadas por la muestra de microdatos son elevadas a los valores de los movimientos migratorios de las NUTS-1 ofrecidos por las "Variables de la submuestra".

Antes de esta estimación se testa la correlación de los flujos migratorios agregados de la EPA con los de la EVR. Para esto los datos de la EVR son agregados en NUTS-1, desde los datos originales para NUTS-2. La correlación entre ambas fuentes es muy grande con valores superiores al 95 por cien para la mayoría de los casos, no bajando

del 82 por cien. Cada elemento genérico de la matriz de microdatos es denotado por $mij/k(t)$, donde i sería la comunidad de origen (17 comunidades autónomas, 2 ciudades autónomas y el extranjero), j denota la comunidad de destino (17 comunidades autónomas y 2 ciudades autónomas), k el nivel educativo (analfabetos, sin estudios primarios, estudios de primaria, estudios de secundaria y estudios superiores) y t el año concreto de la encuesta (de 2000 a 2017).

ABSTRACT

This paper analyses the statistical strength of population censuses and the Labour Force Survey (EPA), with the objective of estimating regional matrixes of migratory flows classified according to education level, by means of variations of the iterative proportional fitting procedure, based on the information from such sources.

Official migration statistics do not classify migrants according to their qualifications, which makes it necessary to use alternative sources in order to model population movements among autonomous regions, taking their education into account.

The use of these alternative sources obliges us to carry out a prior analysis of such, with the objective of selecting the optimum choice enabling the best estimation of inter-regional migratory flows, taking into account the migrant population's different education levels.

We therefore analyse the strength of migratory information from censuses and the Labour Force Survey, using different indicators and procedures, in Section 2 of this paper. This is preceded by an introduction, while Section 1 reviews the literature on qualified migrations.

In this Section 2, we analyse a first comparison of both sources with the global data from the Residential Variation Statistics (EVR). This analysis is supplemented by the results of compensatory equations for Spain as a whole and for each of its autonomous regions, based on information from the last two population censuses, calculating thereafter the mean relative errors (ERM) for the population totals and for people with second- and third-cycle studies, in each of these territories.

The conclusion of these analyses is that the mobility variable of population censuses is affected by the lack of answers from the interviewed population and by the imputation method used by the National Institute of Statistics (INE) to correct it, which explains the underestimation of census data compared to the Residential Variation Statistics.

Starting from evaluation reports on the quality of the 2001 Population Census and of the Labour Force Survey for the years 2006 to 2017, we extracted the results of several indicators such as the percentage of identically classified, the net difference rate, the net change index or the gross difference rate, concluding that in all cases there is greater strength for the Labour Force Survey data than for that of the 2001 Census.

Finally, using the Residential Variation Statistics of 2001, which includes interior migratory data classified according to academic qualifications, we calculated equality and relation indicators such as Theil's U index of inequality, Pearson's coefficient of variation and that of correlation, between this source and the census and this same source

and the Labour Force Survey of that year. As in the case of the prior comparisons, the analysis of the previous statistics produces better results for the Labour Force Survey data compared to that of the Residential Variation Statistics, than the census data with regard to this source.

After selecting the Labour Force Survey as the statistical source for projecting regional migratory flows, taking into account education levels, in Section 3 of this paper we use the iterative proportional fitting to obtain migratory matrixes.

The procedure for estimating migratory flows based on the information from the of the Labour Force Survey micro-data is verified in several phases. Beginning with the matrix generated annually from the four files of three-monthly micro-data, filtering the population ages 16 and over that changed residence the previous year and adding the information in a matrix with 20×5 rows and 19 columns, which shows the structure of inter-regional migratory flows and of immigrants from abroad classified in five education levels. The total margins of this matrix of micro-data are used to calculate new ones by elevation, converging with the values added from the Labour Force Survey module “Subsample variables: persons that have changed residence a year ago,” using a bi-proportional adjustment procedure. In this way, the migrations classified according to educational level of NUTS-2 provided by the micro-data sample are elevated to the values of the migration movements of NUTS-1 supplied by the “Subsample variables.”

Before this estimation, the correlation of the migratory flows added from the Labour Force Survey with those of the Residential Variation Statistics are tested for all the years analysed. To that end, the data of the Residential Variation Statistics is added in NUTS-1, from the original data for NUTS-2. The correlation between both sources is very high, with values exceeding 95 per cent in most cases and never lower than 82 per cent. Each generic element of the micro-data matrix is denoted by $m_{ijk}(t)$, where i is the region of origin (17 autonomous regions, 2 autonomous cities and abroad), j denotes the destination region (17 autonomous regions and 2 autonomous cities), k the education level (illiterate, without primary education, primary education, secondary education and higher education) and t the specific year of the survey (from 2000 to 2017). These matrixes reflect an inter-regional migratory structure, but the absolute data is limited to the interviewed sample, which requires an elevation adjustment using the data of the “Labour Force Survey subsample: Persons that have changed residence a year ago. Persons ages 16 and over that have changed residence a year ago due to education level attained and place of origin/destination,” whose results are encompassed in 8 NUTS-1 for the places of origin and 7 for those of destination. This elevation adjustment enables us to calculate the $r_{i/k}(t)$ and $s_{j/k}(t)$ margins, which will make it possible to initiate the iterative proportional fitting procedure in order to obtain the 18 annual matrixes of the series (2000-2017).

After adjusting the margins of the micro-data matrix to the absolute levels of the Labour Force Survey subsample, the iterative proportional fitting procedure can be used to calculate the intermediate elements of this matrix, making use of the previous margins. The proposed methodology enables us to attain convergence in the eighteen estimated annual matrixes.

The obtained results favour a broader vision of the relations among regions, with regard to this type of migrations, and enable subsequent univariant or multivariant analyses. They could also be examined using input-output methodology or even under the theory of networks, as proposed in the last section of this paper along with the conclusions.

1. INTRODUCCIÓN

Tal y como han puesto de manifiesto algunas investigaciones, en las migraciones interiores españolas el nivel educativo de los emigrantes es una variable muy correlacionada con la propensión a emigrar (Gil y Jimeno, 1993; Antolín y Bover, 1997; Recaño, 2014). Además, esta decisión individual modifica los stocks de capital humano de los distintos territorios y repercute en diferentes variables económicas.

Sin embargo, ni la Estadística de Variaciones Residenciales (EVR), ni la Estadística de Migraciones (EM), principales fuentes de información sobre los fenómenos migratorios, proveen datos sobre características de los migrantes, salvo su edad (año de nacimiento), sexo o territorios de origen y de destino. Esta falta de datos sobre las características personales de los migrantes no es exclusiva del Estado español, sino un fenómeno que se extiende a los institutos oficiales de estadística de diferentes países, que no suelen facilitar esta información, cuando resulta difícil su anonimización.

Por otra parte, existen fuentes estadísticas diferentes a las anteriores que sin tener como objetivo principal ofrecer información sobre movilidad poblacional, compilan datos sobre esta materia. Una de estas fuentes es la EPA que, a través de los microdatos, disponibles por trimestres desde 1999, posibilita conocer el lugar de residencia anterior un año antes al del período de referencia, pudiendo filtrar estos movimientos en función de diversas características personales, como la edad y el nivel educativo, entre otras. Los censos de población también permiten la obtención de datos migratorios clasificados por características personales. Elaborados decenalmente reúnen información sobre movimientos migratorios correspondientes a los años anteriores al de su publicación. En ambos casos sería precisa la utilización de alguna técnica de estimación de datos con estructura matricial para disponer de una serie temporal que abarcara todas las entidades territoriales objeto de análisis.

El único trabajo existente para el Estado español, que estima las migraciones interregionales clasificadas por niveles educativos, utilizando el Censo de 2001 es el de Cabrer, Serrano y Simarro (2009). Sin embargo, estos au-

tores no consideraron la utilización de la otra fuente alternativa –EPA– para realizar esas estimaciones, y si ésta presentaba una mayor o menor calidad de la información sobre migrantes clasificados por niveles educativos que los censos. Constatar la calidad estadística de los datos ofrecidos por estas dos fuentes se constituye en el objetivo principal del presente trabajo.

A continuación de esta introducción se realiza un repaso del marco teórico de las migraciones cualificadas y de los principales trabajos sobre migraciones interiores en España considerando características personales.

En el segundo apartado se analiza la robustez de los datos migratorios interiores clasificados por niveles formativos de los últimos censos y de la EPA. Para esto se comparan estos datos con los de la fuente migratoria oficial con una serie temporal más alargada –EVR– para contar con una referencia temporal igual al del período objeto de estudio. A mayores del contraste mediante diversos indicadores estadísticos, serán utilizados estudios específicos elaborados por el propio INE para analizar la validez de sus fuentes.

Una vez elegida la EPA como base para la estimación de los flujos migratorios anuales interiores en España durante el período 2000-2017, en el tercer apartado serán proyectados estos flujos mediante técnicas de ajuste proporcional. Dentro de estas técnicas y en el ámbito de la Economía destaca el algoritmo RAS (Norman, 1999; Wong, 1992); pero su aplicación, con diferentes nombres, es ampliamente conocida como destacan (Lomax, Norman, Rees y Stillwell, 2013), que consideran que la elección del *iterative proportional fitting* (IPF) es preferida por muchos investigadores para la estimación de datos inexistentes en tablas migratorias. Los trabajos en el ámbito de la Geografía que utilizan esta técnica son numerosos¹.

El trabajo finaliza con un apartado donde se resumen las principales conclusiones obtenidas.

2. MARCO TEÓRICO DE LAS MIGRACIONES CUALIFICADAS

La literatura económica sobre migraciones de personas con cualificación se desarrolló fundamentalmente sobre la teoría de la fuga de cerebros,

1 Chilton y Poet (1973), Philipov (1981), Willekens (1977), Willekens, Pór y Raquillet (1981), Willekens (1982), Nair (1985), Rees y Duke-Williams (1997), van Imhoff, van der Gaag, van Wissen y Rees (1997) y Lomax et al. (2013).

basada en los modelos neoclásicos de la teoría del valor de Hicks-Samuelson (Bhagwati y Hamada, 1974) a mediados de la década de los años sesenta, centrando la atención en los aspectos cualitativos de estos movimientos por encima de los cuantitativos.

El comienzo de esta teoría de la fuga de cerebros podría situarse en el trabajo de Grubel y Scott (1966), prolongando su desarrollo durante casi 30 años hasta llegar al artículo de Haque y Kim (1994). En medio de este período, la teoría de la fuga de cerebros va elaborando diferentes modelos. Así aparece un conjunto de trabajos con formulaciones estáticas de competencia perfecta, como son los de Berry y Soligo (1969), Tobin (1973), Kenen (1971), Johnson (1979) y Grubel (1975).

Contemporáneos a estos trabajos se encuentran los modelos estáticos que introducen distorsiones en la simplificada situación de competencia perfecta como son los de Bhagwati y Hamada (1974), McCulloch y Yellen (1975a), McCulloch y Yellen (1975b), Hamada y Bhagwati (1975) y Harris y Todaro (1970). Simultáneamente aparecen modelos dinámicos como los de Mishan y Needleman (1968), McCulloch y Yellen (1974), Rodríguez (1975a) y Rodríguez (1975b).

A finales del siglo XX aparecen estudios que contradicen las propuestas anteriores, propugnando la coexistencia de crecimiento económico endógeno y de fuga de cerebros benéfica, constituyendo lo que se conoce como teoría del *brain gain*. Esta literatura localiza la fuga de cerebros en modelos explícitamente dinámicos, alcanzando resultados significativamente más optimistas que los logrados por la teoría de la fuga de cerebros. Su proposición central versa sobre la posibilidad de que la emigración anime a las personas a adquirir más formación, aumentando de esta manera el capital humano del país de origen en lugar de minorarlo, como defendía la teoría precedente del *brain drain*, dadas las oportunidades de movilidad o de trabajo en el extranjero. Además, esa acumulación de competencias tiene efectos beneficiosos, más allá de las ganancias estrictamente privadas experimentadas por los adquirentes de la cualificación, pudiendo afectar a toda la economía. Ejemplos de tales ganancias serían la mayor transmisión intergeneracional de cualificación y de educación (Vidal, 1998) y los *spillovers* entre trabajadores cualificados (Mountford, 1997).

Estos últimos modelos comparten dos características importantes. La primera es la naturaleza del beneficio social resultante del aumento de la cualificación, para el cual son posibles varias formulaciones. En la forma

más sencilla, Stark, Helmenstein y Prskawetz (1997), Stark, Helmenstein y Prskawetz (1998) y Stark y Wang (2002) asumen que el aumento del nivel de cualificación medio de la economía de origen es deseable. Por otra parte, Mountford (1997) postula una externalidad de la producción en que la productividad del trabajo actual depende positivamente de la participación de la población que obtiene la educación en el período anterior. Beine, Docquier y Rapoport (2001) formalizan ésta, permitiendo que la cualificación media de una generación sea transferida directamente a la siguiente, pudiendo construirse sobre ella, al tener como fundamento esa educación. En todos esos casos, la emigración tiene un efecto negativo directo a través de la fuga de cerebros desde la economía de origen (efecto fuga), y un efecto potencialmente beneficioso al promover la formación de capital humano (efecto ganancia).

Sin embargo, en la literatura sobre migraciones interiores en España no se prodigan los trabajos que apliquen este cuadro teórico al análisis de las mismas. Este hecho se explica fácilmente por la falta de información estadística que compile datos del nivel educativo de los emigrantes. Con todo, existe una vasta lista de estudios que analizan las migraciones interiores teniendo en cuenta otros factores explicativos.

Siguiendo la teoría de los factores *push* y *pull* de Ravenstein (1885) para explicar las migraciones podrían clasificarse los estudios recientes sobre este tópico para los movimientos poblacionales en el interior de España. Dentro de los factores de expulsión aparecen los costes de desplazamiento y la incertidumbre, el desempleo, la estructura productiva volcada en el sector primario, los impuestos y el precio de la vivienda. Por su parte, los salarios, las oportunidades de empleo, la estructura productiva basada en la industria y los servicios, el gasto público y el clima, constituyen los principales factores de atracción de los migrantes. Además de estos factores de expulsión y de atracción, existen otros como determinadas características de los territorios y de su población, y las propias características personales de los migrantes que también explicarían los movimientos interregionales. Si bien los estudios que tienen en cuenta factores de expulsión y de atracción para explicar la dinámica migratoria interior en España son mayoritarios, los que centran su atención en aspectos personales de los migrantes se reducen enormemente. Entre los escasos trabajos de este tipo destacan los de Ródenas (1994a), Antolin y Bover (1997), García y Stillwell (1999), Bover y Velilla (1999) y Recaño (2014).

La primera autora tiene en cuenta características personales de los migrantes en su libro sobre emigración y economía. Aunque este trabajo monográfico comienza repasando la emigración al extranjero desde 1882, los movimientos migratorios interiores serán estudiados a partir de la Primera Guerra Mundial, período donde se produce una intensificación de los mismos, con el éxodo rural hacia las regiones más industrializadas, como consecuencia de la caída en términos reales de los salarios agrícolas, pérdida muy superior a la sufrida por sus homónimos industriales.

Casi 40 años después se produce una segunda etapa de intensificación de las migraciones interiores, durante el período desarrollista 1959-1973, donde se alcanzan los mayores volúmenes migratorios interiores, explicados por la concentración regional de la actividad económica, requiriendo grandes stocks de mano de obra para no paralizar el crecimiento los sectores secundario y terciario, básicamente de Cataluña, Madrid, la Comunidad Valenciana y el País Vasco que concentran casi el 80% de los emigrantes. De esta manera, las desigualdades regionales de renta y de cualificación no hicieron más que aumentar.

El siguiente período analizado es el de la crisis económica de los años 1973-1985. Al igual que en el período anterior, las características de los migrantes son idénticas: hombres y mujeres emigran casi a partes iguales, casi la mitad de ellos (46,44%) tienen entre 25 y 64 años y se trata de una migración de carácter familiar, donde más de la mitad de los migrantes (63,12%) no pertenecen a la población activa. El destino regional de estas migraciones coincide también con el de la etapa anterior, aunque Madrid y la Comunidad Valenciana, al sufrir en menor medida la crisis económica, aguantan mejor el descenso de llegadas que Cataluña y el País Vasco.

Por último, la autora centra su análisis en la etapa de recuperación económica del quinquenio 1985-1989. Del mismo modo que en las etapas anteriores, tampoco se detecta un cambio en la distribución por sexo y edad de los migrantes, aunque esta última refleja el envejecimiento global de la población española. De cualquier manera, casi las tres cuartas partes de los emigrantes tenían menos de 34 años. Con respecto a las características de estos movimientos, se experimenta un incremento, a pesar de la reducción en los diferenciales regionales de renta y producto, repitiendo el esquema tradicional de regiones receptoras y expulsoras, aunque sin las grandes concentraciones del pasado.

Antolin y Bover (1997) utilizando datos individuales de la Estadística de Migraciones, incluida en la Encuesta de Población Activa, analizan la probabilidad de emigrar considerando factores de atracción, junto con determinadas características de una muestra de hombres entre 16 y 70 años, como son la edad, el nivel educativo, el estado civil y la situación familiar, tal como que el cónyuge esté trabajando, tener hijos o convivir con otros familiares; y otros datos sobre su situación profesional como estar o no registrado en el servicio público de empleo o ser autónomo.

Los autores consideran que es la interacción de factores regionales y características personales la que explica los patrones migratorios interiores, resultando los atributos personales determinantes a la hora de tomar la decisión de migrar o no, a pesar de existir componentes regionales para incentivarla. De esta manera, se concluye que las migraciones interiores no funcionan como mecanismo corrector de los diferenciales regionales en las tasas de paro y por lo tanto no ayudan a la reducción de las altas tasas de desempleo existentes.

Utilizando una fuente de datos diferente a la del trabajo precedente, García y Stillwell (1999) analizan los cambios en los patrones migratorios interprovinciales durante la década de los ochenta para una muestra de población del Censo de 1991. Fueron calculadas las tasas migratorias brutas por grupos de edad teniendo en cuenta su peso relativo sobre el total estatal. La relación entre tasas de inmigración y emigración para cuatro grupos de edad muestra una gran variabilidad según los años de experiencia de los migrantes.

Bover y Velilla (1999) en un trabajo donde repasan la historia y las tendencias actuales de las migraciones españolas dedican el apartado tercero a las migraciones interiores, donde en el subapartado segundo realizan un repaso de las migraciones interregionales en el período 1983-1995, basado en el artículo anteriormente citado de Antolin y Bover (1997). La escasez de estudios migratorios centrados en características personales se hace evidente en el subapartado anterior a este donde se describen las migraciones interiores en el período 1960-1982. Ninguno de los dos trabajos citados (Santillana, 1981 y Ródenas, 1994b) analiza este tipo de características, recurriendo a factores de atracción y de expulsión.

Recaño (2014) centra su estudio en las migraciones interiores para el período que va desde 1971 a 2011, utilizando microdatos de los cuatro últimos censos elaborados por el INE, mediante modelos de regresión logística

multinomial. Son consideradas características personales de los migrantes como sexo, edad, estado civil, nivel de estudios, CC AA de residencia 10 años antes y lugar de nacimiento, analizando las migraciones intrarregionales, las interregionales, las de retorno y las de nacidos en el extranjero.

Con respecto a las dos primeras, los principales resultados serían que en el grupo de los jóvenes, las mujeres tienen una menor propensión a migrar, sobre todo a nivel interregional. De igual manera, entre todos los estados civiles, los solteros se mueven menos. Con respecto al nivel educativo, la propensión migratoria aumenta con éste, resultando mayor para las migraciones interregionales.

A nivel territorial, Cataluña, Madrid, el País Vasco y Navarra son las comunidades más dinámicas en las migraciones intrarregional para los jóvenes; resultando Extremadura, Castilla-La Mancha y Castilla y León las que poseen mayor poder de expulsión de su población joven a otras regiones. Comunidad Valenciana, Cataluña y Madrid serían las que cuentan con mayor capacidad para retener su población.

Algunos de estos resultados también se repiten para la población de mayor edad analizada, como acontece con las mujeres. Sin embargo, en este grupo, serían los casados los que presentan menor propensión a migrar. También se repite un resultado similar al de la población joven, en lo referente al nivel educativo.

Cuanto a las migraciones de retorno que tienen lugar entre diferentes regiones, no existen diferencias destacables entre sexos de la muestra de población joven, resultando mayor cuando existe una ruptura de pareja, y también aumentando con el nivel de formación. Sin embargo, la otra muestra de población de mayor edad, cuenta con comportamientos diferenciados por sexo, al existir una propensión menor en el caso de los hombres, aumentando en ambos casos en la situación de soltería, no resultando relevante el nivel formativo.

Por último, las migraciones de los nacidos en el extranjero, están caracterizadas por la masculinización de las mismas, no resultando relevante el estado civil y aumentando junto con el nivel de estudios.

Una propuesta diferente a la descrita anteriormente, es presentada en Cabrer et al. (2009) y Pereira, César y Verdugo (2015), donde mediante técnicas de ajuste proporcional son elaboradas matrices de flujos migratorios interregionales clasificados por niveles educativos a partir de los censos de población. Así, en el primero de ellos (Cabrer et al., 2009) estiman los flujos

migratorios de capital humano entre las regiones españolas a partir de un conjunto de hipótesis. Los resultados indican un aumento de la intensidad de los flujos brutos de migrantes con niveles de instrucción de segundo y tercer grado. Se constata que en España existen tres polos de atracción y emisión de capital humano centrados en Madrid, País Vasco y Cataluña, y actuando de forma independiente. Asimismo, se obtiene una alta adecuación entre el nivel educativo de los inmigrantes y los requerimientos de cualificación de la mano de obra de cada región. Sin embargo, se constata que los años medios de estudio de la población potencialmente activa han aumentado frente a la estabilidad de los de la población migrante, por lo que no cabe señalar a la inmigración como un factor relevante para reducir las disparidades interregionales en capital humano.

Pereira et al. (2015) analizan los flujos migratorios regionales de España por nivel educativo mediante una técnica para proyectar las migraciones intra e interregionales –por años y desagregadas por niveles formativos– a partir de los censos. Para ello utilizan una extensión del *iterative proportional fitting* (IPF) aprovechando no solamente los vectores de márgenes anuales, sino también los flujos migratorios totales entre comunidades autónomas, reformulando el IPF básico para poder efectuar ajustes de carácter triproporcional, motivados precisamente por la información disponible.

Con las matrices obtenidas son calculados los coeficientes de Chenery y Watanabe y los de Streit para identificar las migraciones interiores de la población con estudios de tercer grado en España. Los resultados obtenidos favorecen una visión más amplia de las relaciones entre comunidades, en lo que se refiere a este tipo de migraciones, detectando aquéllas capaces de crear mayores tensiones de movilidad.

3. ROBUSTEZ DE LAS FUENTES ESTADÍSTICAS PARA LA ESTIMACIÓN DE FLUJOS MIGRATORIOS CLASIFICADOS POR NIVELES EDUCATIVOS.

Analizar la robustez de la información migratoria de los censos y de la EPA constituye el principal objetivo de este apartado y del trabajo en su conjunto. Para ello será utilizada la EVR como fuente de contraste de los datos agregados de la EPA y de los censos, sobre todo para testar los valores absolutos de las migraciones, de manera similar a la utilizada por

Ródenas y Martí (2009), que constataron la infravaloración de los flujos migratorios del Censo de 2001 con respecto a la EVR. A pesar de reconocer que la EVR puede infravalorar o sobrevalorar los inmigrantes extranjeros, y que puede contabilizar migraciones repetidas en el mismo año, así como “falsas migraciones”, como mucho, estas carencias apenas explicarían el 11,8% de la divergencia existente. Por lo tanto, y después de un análisis comparativo de la EVR con el Censo de 2001, estas autoras concluyen que las novedades metodológicas y la influencia de la falta de respuesta parcial para la variable migratoria del censo explicarían la mayor parte de estas discrepancias. Estas novedades metodológicas con respecto al censo anterior de 1991 afectan a la manera de formular las preguntas migratorias, al permitir anteriormente la auto clasificación de las personas como migrantes, evitando la falta de respuesta parcial y la presencia de inconsistencias. Desafortunadamente, en 2001 el formulario de la encuesta provoca falta de respuesta cuando el lugar de residencia es diferente del de nacimiento y no se declara movilidad, apareciendo un problema de datos incompletos y el procedimiento, utilizado para resolver esto, repercute en las estimaciones censales de la movilidad. Teniendo como objetivo testar este fenómeno, se analiza la coherencia interna de los datos de los censos, realizando un cotejo con otras fuentes estadísticas, sabiendo que tanto las migraciones como la natalidad deben hacer cumplir determinadas igualdades contables demográficas. Así, la desigualdad migratoria presenta inconsistencias que únicamente pueden ser explicadas por la falta de respuesta parcial. De otra manera, si el problema fuese la falta de respuesta total, la desigualdad contable relativa a los nacidos también dejaría de confirmarse y no es el caso. Por lo tanto, sería la variable movilidad la que se encuentra afectada por la falta de respuesta y por el método de imputación utilizado por el propio INE para corregirla, explicando la infravaloración de los datos migratorios entre 1991 y 2001 del censo con respecto a la EVR, como señalan las autoras anteriormente citadas.

Prolongando el análisis hasta el último censo (2011) mediante la ecuación compensadora [1], la robustez de los datos migratorios recogidos en el mismo puede ser testada. El método de los componentes de la ecuación de equilibrio demográfico entre los censos de 2001 y de 2011 para estimar el saldo migratorio del período ($SM_{2001-11}$) sería:

$$SM_{2001-11} = Pob\ Censo_{2011} - Pob\ Censo_{2001} - (Nacimientos_{2001-11} - Muertes_{2001-11}) \quad [1]$$

El saldo migratorio ($SM_{2001-11}$) fue calculado de acuerdo con las poblaciones totales de los dos últimos censos y el movimiento natural (nacidos y muertos) procedentes de la estadística de *Movimiento Natural de la Población* (MNP) del INE. Mediante la opción “Lugar de residencia en 2001” el Censo de 2011 permite obtener datos de la relación entre la residencia en aquel año y la actual, utilizando “Residía en el extranjero” a nivel estatal para contabilizar los inmigrantes, y a nivel de comunidad autónoma aumentando al valor anterior el obtenido en el apartado “Otra comunidad” para alcanzar los inmigrantes, tanto interiores, como exteriores que llegaron a la comunidad autónoma durante el período intercensal. Resulta evidente la imposibilidad de obtener un saldo migratorio a partir del Censo de 2011 al carecer de datos sobre los emigrantes al exterior. Los resultados de las ecuaciones compensadoras de cada región aparecen en la Cuadro 1, así como los saldos totales (interiores y exteriores) de la EVR y los saldos interiores más los inmigrantes procedentes del exterior según el Censo de 2011 en la década anterior. A nivel estatal el censo recoge un saldo interior más los inmigrantes exteriores de 2.830.380, cifra muy inferior a la esperada para alcanzar el saldo teórico resultante de la ecuación compensadora². No obstante, la EVR aproxima bastante ese resultado, al contabilizar un saldo exterior de 4.825.489 personas (alrededor del 95% del saldo teórico). A nivel autonómico los saldos migratorios teóricos también se encuentran muy próximos a los contabilizados por la EVR para ese período. A su vez, los datos del censo se encuentran muy alejados de los saldos teóricos. Calculando los emigrantes al extranjero³ –que el censo no facilita– mediante el saldo teórico y los saldos interiores y los inmigrantes exteriores del censo, serían obtenidas cifras incongruentes de conformidad con las esperadas, mostradas en la última columna de la Cuadro 1. Con la excepción de Melilla, los resultados son negativos, indicando que, en lugar de emigrantes, serían inmigrantes desde el exterior, los necesarios para garantizar los saldos teóricos del período, no existiendo ninguna emigración al exterior, situación totalmente irreal.

- 2 Si los emigrantes al exterior alcanzaron una proporción similar a la de la EVR sobre unos inmigrantes exteriores de 2.830.380, los primeros deberían ser 792.506 (28% de los inmigrantes), devolviendo un saldo de 2.037.874, apenas un 40% del saldo calculado por la ecuación compensadora para el período.
- 3 Los saldos migratorios teóricos (SM_t) obtenidos de las ecuaciones compensadoras tendrían que ser iguales a la suma de los saldos migratorios interior (SM_i) y exterior (SM_e) del censo, calculados restando a los inmigrantes (I), los emigrantes (E). Así los emigrantes al extranjero del censo (E_c) podrían ser calculados a partir de la siguiente ecuación: $SM_t = SM_i + (I_e - E_c)$, es decir, $E_c = SM_i + I_e - SM_t$.

CUADRO 1
**SALDOS MIGRATORIOS DE LA ECUACIÓN COMPENSADORA, DE
 LA EVR Y SALDO MIGRATORIO INTERIOR MÁS INMIGRANTES
 EXTERIORES DEL CENSO DE 2011, 2001-2011**

	Censo 2011	Censo 2001	Nacimientos	Muertes	Saldo natural	SM teórico	SM EVR	SM _t + I _e censo año a año	Emigrantes al extranjero
TOTAL	46.574.725	40.595.861	4.719.587	3.806.159	913.428	5.065.437	4.825.489	2.830.380	-2.235.057
Andalucía	8.343.655	7.325.866	919.634	646.353	273.280	744.509	697.574	401.765	-342.744
Aragón	1.331.190	1.191.636	121.806	133.838	-12.032	151.586	136.985	96.830	-54.756
Asturias	1.069.275	1.056.298	76.819	126.932	-50.113	63.090	51.961	31.345	-31.745
Baleares	1.096.905	837.094	113.601	74.307	39.293	220.518	181.342	127.025	-93.493
Canarias	2.078.280	1.686.928	193.203	128.531	64.673	326.679	282.843	135.765	-190.914
Cantabria	589.175	530.942	51.753	54.652	-2.900	61.133	49.657	33.705	-27.428
Castilla y León	2.515.755	2.427.685	196.721	270.618	-73.897	161.967	116.773	79.520	-82.447
Castilla-La Mancha	2.092.395	1.747.368	193.930	178.625	15.304	329.723	316.812	209.545	-120.178
Cataluña	7.472.935	6.304.366	804.008	598.956	205.052	963.517	793.601	549.620	-413.897
C. Valenciana	4.990.345	4.145.087	508.114	398.022	110.092	735.166	806.852	453.470	-281.696
Extremadura	1.097.695	1.051.032	101.426	107.641	-6.216	52.879	29.704	24.320	-28.559
Galicia	2.759.890	2.681.025	214.370	296.476	-82.106	160.971	137.726	85.015	-75.956
Madrid	6.387.250	5.394.140	720.810	409.904	310.906	682.204	803.607	345.825	-336.979
Murcia	1.458.250	1.192.462	175.780	99.753	76.028	189.760	211.621	132.665	-57.095
Navarra	635.175	548.166	65.144	51.360	13.784	73.225	55.767	40.305	-32.920
País Vasco	2.173.265	2.065.476	203.735	192.826	10.908	96.881	100.572	45.275	-51.606
La Rioja	319.460	274.028	30.787	27.831	2.955	42.477	44.685	28.745	-13.732
Ceuta	83.185	71.060	13.958	5.125	8.833	3.292	1.278	3.110	-182
Melilla	80.655	65.202	13.990	4.408	9.582	5.871	6.129	6.530	659

Fuente: EP a partir de los Censo de 2001 y 2011, MNP y EVR.

Estos resultados de la EVR son coherentes con los alcanzados por Ródenas y Martí (2009), y también con los de otro trabajo anterior, Ródenas y Martí (2005), donde las migraciones captadas por el censo eran espectacularmente inferiores a las declaradas en la EVR, presentando importantes problemas de consistencia y de no-respuesta. Por no tener los migrantes la posibilidad de clasificarse de esa manera en los censos de 2001 y 2011 al no existir ninguna pregunta al respecto –debiendo ser el propio INE quien los caracterice a partir de preguntas sobre el año de residencia en la vivienda, municipio, comunidad autónoma y país– se hace muy evidente que estén repitiendo los problemas descritos anteriormente. Utilizando un esquema similar al de las autoras anteriores, se comparan los cuestionarios censales y los resultados migratorios obtenidos a partir de esas preguntas (Cuadro 2). La movilidad puede ser estimada a partir de diferentes preguntas, la residencia hace determinado período de tiempo, normalmente con referencia al censo inmediatamente anterior (pregunta 1.1); la residencia habitual cinco años antes (pregunta 1.2); la residencia un año antes (pregunta 1.3); y por el año de llegada a la residencia actual (pregunta 2). Comparando los resultados de la pregunta 1.1 y los de la pregunta 2, sumando los resultados para los

CUADRO 2
COMPARACIÓN ENTRE LAS PREGUNTAS MIGRATORIAS DE LOS
CENSOS 1991, 2001 Y 2011

Censo 1991 (marzo)	Censo 2001 (noviembre)	Censo 2011 (noviembre)
Pregunta 1: lugar de residencia anterior hace...	Pregunta 1: lugar de residencia anterior hace...	Pregunta 1: lugar de residencia anterior hace...
Lugar de residencia habitual marzo 1981 (10 años).	Lugar de residencia habitual noviembre 1991 (10 años y 8 meses).	Lugar de residencia habitual noviembre 2001 (10 años).
Estimación a partir de la Pregunta 1.1.:	Estimación a partir de la Pregunta 1.1.:	Estimación a partir de la Pregunta 1.1.:
Personas residentes en viviendas familiares 10 y más años	Personas residentes en viviendas familiares 10 y más años	Personas residentes en viviendas familiares 10 y más años
Total 4.017.599	Total 4.778.821	Total 9.076.755
Procedentes del interior	Procedentes del interior	Procedentes del interior
3.614.209	3.800.409	5.658.755
Procedentes del exterior	Procedentes del exterior	Procedentes del exterior
403.390	978.412	3.418.000
Lugar de residencia habitual abril 1986 (5 años).	(no es realizada).	(no es realizada).
Lugar de residencia habitual marzo 1990 (1 año).	(no es realizada).	Lugar de residencia habitual noviembre 2010 (1 año).
Pregunta 2: año de llegada a la actual residencia	Pregunta 2: año de llegada a la actual residencia	Pregunta 2: año de llegada a la actual residencia
Solo las personas que en los últimos 10 años residieron en otro municipio: año en que estableció por última vez su residencia en este municipio y lugar de procedencia.	Desde que año reside (aunque sea desde el nacimiento) en: España, esta comunidad, este municipio (si antes residía en otro, señalar).	Desde que año reside (aunque sea desde el nacimiento) en: España, esta comunidad, este municipio (si antes residía en otro municipio, provincia, país, señalar).
Estimación a partir de la Pregunta 2 (suma dato año-a-año de 1981-91):	Estimación a partir de la Pregunta 2 (suma dato año-a-año de 1991-01):	Estimación a partir de la Pregunta 2 (suma dato año-a-año de 2001-11):
Personas residentes en viviendas familiares y colectivas	Personas residentes en viviendas familiares	Personas residentes en viviendas familiares
Todas las edades	Todas las edades	Todas las edades
Total 4.434.777	Total 6.048.287	Total 10.279.905
Procedentes del interior	Procedentes del interior	Procedentes del interior
3.997.886	4.872.136	7.449.645
Procedentes del exterior	Procedentes del exterior	Procedentes del exterior
436.891	1.176.151	2.830.260
Diferencia Pr 2 – Pr 1.1.:	Diferencia Pr 2 – Pr 1.1.:	Diferencia Pr 2 – Pr 1.1.:
417.178	1.269.466	1.203.150

Fuente: EP en línea con Ródenas y Martí (2005) y Ródenas y Martí (2009) a partir de los censos de 1991, 2001 y 2011.

diez años anteriores, estos últimos siempre son superiores, tal como indicaron Ródenas y Martí (2005). Esto puede ser debido a estar computando los menores de 10 años con movilidad, pudiendo incluir algunas de las migraciones intermedias, como las de ida y vuelta, producidas en el período entre censos. Para los dos primeros censos los resultados migratorios de las dos preguntas experimentan crecimientos muy heterogéneos (19% para la pregunta 1.1 y 36% para la 2), y en los dos últimos las cifras resultan más similares (90 y 70%, respectivamente). Una hipótesis explicativa del aumento de la diferencia entre ambas estimaciones (pregunta 1.1 y 2) podría estar en la población menor de 10 años que la pregunta 2 recogería, pero no la 1.1. Para el Censo de 1991 estas migraciones alcanzarían a 366.250 menores de 10 años, mientras que para el de 2001 ya serían de 480.943 y para el de 2011 de 658.295.

A partir de estos datos puede ser deducida la movilidad intermedia, capturada por cada uno de los censos: 50.928 migraciones en el de 1991, 788.523 en el de 2001 y 544.855 en el último; constituiría multiplicar la movilidad intermedia capturada por los dos primeros por más de quince, pero reduciría la capturada entre los dos últimos, más de un 30%. Admitiendo que se tuvieran dado de alta en los registros municipales todos estos cambios de residencia, este descubrimiento encajaría perfectamente con el hecho de que la EVR recoge mayores niveles de movilidad en estos últimos veinte años.

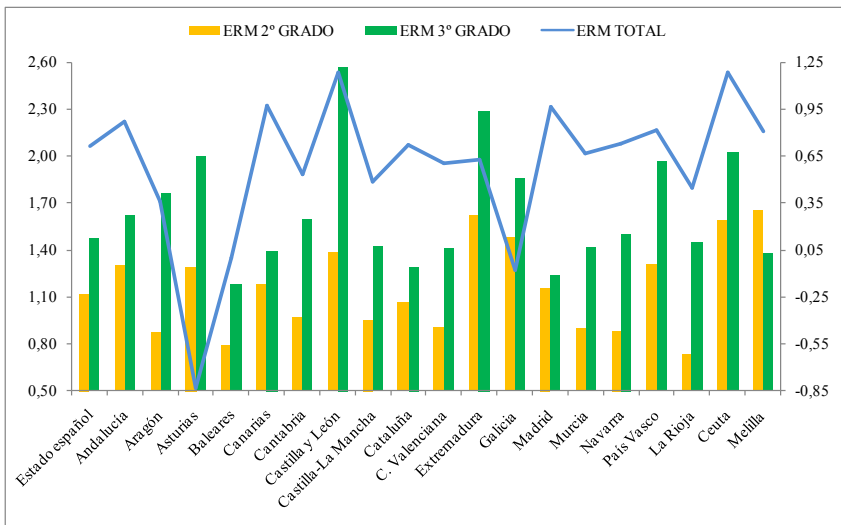
Continuando con los resultados parciales de las ecuaciones compensadoras y centrandó la atención en las diferencias poblacionales de los dos últimos censos⁴, es posible detectar los desfases entre territorios y para distintos niveles educativos, concretamente para los dos niveles superiores por resultar los más interesantes para este trabajo. Mediante los errores relativos medios (ERM)⁵, calculados para los totales poblacionales y para las personas con estudios de segundo y de tercer grado, se hace evidente que solamente Asturias y Galicia obtienen valores por debajo de cero con respecto a la población total, resultado que haría plausible la ecuación compensadora, al considerar los emigrantes al

4 Las ecuaciones compensadoras utilizadas anteriormente para los totales sin desagregar por nivel educativo se encuentran, cuando se intenta analizar los saldos intercensales considerando esta variable, con la ausencia de desagregación del movimiento natural. Esta carencia obliga a trabajar con las diferencias entre los datos censales.

5 Es calculado como el cociente entre la diferencia de los saldos censales del período 2001-2011 y la suma del saldo migratorio interior más los inmigrantes al extranjero, obtenidos año-a-año del Censo de 2011, dividida por la media de ambas cantidades.

exterior y los saldos naturales negativos del período, pero siendo insuficientes para cubrir la diferencia. Por el contrario, cuando son vistos los errores para los migrantes con estudios de segundo y de tercer grado, los valores obtenidos se sitúan muy por encima de estos resultados. La Figura 1 muestra en el eje de ordenadas derecho los valores del ERM total y en el izquierdo los relativos a los migrantes con estudios de segundo y de tercer grado. La evolución de los estadísticos totales con los clasificados por estudios, para estas dos comunidades diverge, al contrario de lo que sucede para el resto de territorios. Por lo tanto, el censo no está únicamente infravalorando los datos migratorios globales, sino que en ciertas comunidades y para determinados niveles educativos el sesgo podría ser mayor.

FIGURA 1
ERM DE LA VARIACIÓN CENSAL Y LAS MIGRACIONES INTERIORES MÁS LOS INMIGRANTES EXTERIORES SEGÚN EL CENSO DE 2011 PARA EL PERÍODO 2001-2011



Fuente: EP a partir de los censos de 2001 y 2011.

El propio INE en el análisis realizado en 2007 sobre la calidad de los datos del Censo de 2001 (INE, 2007a), al evaluar los errores en determinadas características de la población recogidas en los cuestionarios censales,

originados al no clasificar las unidades correctamente, detecta sesgo en el sentido indicado anteriormente. El análisis del INE centra la atención en diversas características analizadas en el censo comparando sus resultados con los alcanzados por la EPA. Los distintos métodos de entrevista utilizados en ambas encuestas pueden dar origen a respuestas diferentes a la misma pregunta, obteniendo clasificaciones diversas de una misma persona con respecto a una determinada característica. La referencia considerada es la EPA por ser elaborada por entrevistadores especialmente entrenados para obtener la información y las familias elegidas colaboran durante seis trimestres consecutivos, lo que también ayuda a mejorar la calidad de la encuesta.

Cruzando la información disponible en ambas fuentes para la misma persona, se confeccionan tablas de errores de contenido y de indicadores de calidad. A partir de la tabla de concordancia se pueden obtener distintos indicadores de calidad para cada modalidad:

- **Porcentaje de idénticamente clasificados (PIC):** Es un indicador de la estabilidad de respuesta, varía entre 0 y 100. Su valor óptimo (100) indica que todas las personas pertenecientes, según la EPA a una modalidad, fueron clasificadas de igual forma en el censo.
- **Tasa de diferencia neta (TDN):** Indicador del sesgo de respuesta. Puede ser positiva o negativa. En el primer caso indica que el censo tiene un sesgo de ese signo en la contabilización de personas con esa modalidad.
- **Índice de cambio neto (ICN):** Complementa la información suministrada por la TDN, al reflejar la importancia de las variaciones netas de respuesta con respecto del total de personas clasificadas, el ICN refleja la cuantía de esas variaciones con respecto del número de personas poseedoras de esa característica según la encuesta de control (EPA).
- **Tasa de diferencia bruta (TDB):** Indicador de la varianza de respuesta. Su valor puede ser nulo o positivo. Refleja el porcentaje de personas que discrepan en ambas encuestas acerca de la clasificación o no en la modalidad.
- **Índice de cambio bruto (ICB):** Representa el porcentaje de personas clasificadas de distinta manera en ambas encuestas con respecto de las clasificadas en la modalidad según la EPA. Complementa la información facilitada por la TDB.

Desde el punto de vista de este apartado y al no contener información sobre migraciones, el interés se centraría en los principales resultados referidos a la característica “Nivel de formación mayor completado”. Estos resultados obtienen los peores indicadores de calidad, como ya viene siendo habitual en otras operaciones estadísticas. El PIC alcanza su valor máximo, próximo al 70%, en la modalidad Tercer Grado, Licenciatura; siendo el menor para la modalidad Analfabetos, con el 42%. La modalidad con mayor sesgo es la de Tercer Grado, Doctorado, con casi el 100%. Esta última, además de ser la menos importante cuantitativamente, presenta considerables transferencias de población con la modalidad de Tercer Grado, Licenciatura. El ICB exhibe valores bastante elevados en todas las modalidades, sobresaliendo Tercer Grado, Doctorado, que alcanza el 183%, y Sin estudios y Segundo Grado, FP Grado Medio alrededor del 115%.

Las transferencias poblacionales entre niveles educativos adyacentes son usuales. Por ejemplo, el 40% de los analfabetos de la EPA aparecen en el censo como sin estudios. Igualmente, el 39% de las personas sin estudios en la EPA son clasificadas en el censo con estudios de primer grado. A su vez, el 29% de los clasificados en la EPA con estudios de primer grado aparecen en el censo con estudios de segundo grado, ESO, EGB, Bachillerato Elemental. Similar porcentaje aparece en la EPA clasificados como doctorados y como licenciados en el censo. Estas discrepancias pueden ser debidas en muchos casos a respuestas no bien especificadas, llevando a clasificar de manera diferente la misma persona, o a la dificultad de conocer la equivalencia entre los estudios y títulos antiguos y los actuales.

Estos resultados son confirmados a través del menor ICG observado en todas las características analizadas, donde nuevamente el “Nivel de formación mayor completado” alcanza apenas el 53,33%, cuando la variable “Edad” llega hasta el 98,09% o para la “Nacionalidad” al 99,67%; advirtiendo del sesgo que los datos del censo poseen en la categoría sobre formación, lo que obliga a actuar con cautela cuando se toman estos datos del Censo de 2011. Desgraciadamente el INE no realiza ninguna evaluación de la calidad de los datos de este último censo y tampoco es posible, con la información ofrecida por este instituto, una comparación similar a la anterior. Por todo esto, deberá ser efectuado un contraste alternativo de la EPA y del Censo de 2011 con respecto a los movimientos migratorios clasificados por niveles educativos.

CUADRO 3
**ERROR RELATIVO MEDIO¹ ENTRE LOS MOVIMIENTOS MIGRATORIOS DE LA EPA Y DEL
 CENSO 2011 POR TERRITORIOS DE PROCEDENCIA Y NIVELES EDUCATIVOS, 2002-2011**

Territorios	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
ESTADO ESPAÑOL	0,17	0,08	-0,19	-0,37	-0,26	-0,19	-0,10	-0,24	-0,12	-0,37
Galicia, Asturias y Cantabria	-0,23	-0,04	-0,30	-0,97	-0,79	-0,77	-0,37	-0,49	-0,54	-0,68
País Vasco, Navarra, La Rioja y Aragón	0,22	-0,15	0,26	-0,39	-0,41	-0,43	-0,25	-0,40	-0,20	-0,66
Madrid	0,85	0,55	-0,09	-0,28	0,01	-0,20	0,00	0,16	-0,04	-0,37
Castilla y León, Castilla-La Mancha y Extremadura	-0,02	-0,23	-0,49	-0,85	-0,42	-0,37	-0,20	-0,11	0,26	-0,52
Cataluña, C. Valenciana y Baleares	-0,15	-0,30	-0,46	-0,67	-0,83	-0,83	-0,55	-0,76	-0,62	-0,85
Andalucía, Murcia, Ceuta y Melilla	0,06	0,53	-0,21	-0,87	-0,58	-0,59	-0,54	-0,72	-0,59	-0,85
Canarias	-0,04	-0,58	-0,43	-1,09	-1,00	-0,28	-0,33	-0,66	-1,18	-1,11
Extranjero	0,24	0,11	-0,07	0,22	0,23	0,45	0,28	0,24	0,68	0,41
Niveles educativos										
Analfabetos	0,82	-0,18	-0,09	-0,23	-0,21	0,26	0,69	-0,18	0,89	0,39
Estudios primarios incompletos										
1º grado	-0,13	-0,34	-0,70	-0,65	-0,59	-0,41	-0,51	-0,65	-0,56	-0,46
2º grado	0,29	0,22	0,00	-0,14	0,00	-0,08	-0,01	-0,11	-0,11	-0,18
3º grado	-0,20	-0,35	-0,45	-0,85	-0,77	-0,51	-0,29	-0,42	-0,12	-0,71

Fuente: EP a partir de EPA y Censo 2011.

1 Cociente entre la diferencia de los migrantes de la EPA y del censo, dividida por la media de ambas cantidades, (Cabrer et al., 2009).

CUADRO 4
**ERROR RELATIVO MEDIO ENTRE LOS MOVIMIENTOS MIGRATORIOS DE LA EPA Y DEL
 CENSO 2011 POR TERRITORIOS DE DESTINO Y NIVELES EDUCATIVOS, 2002-2011**

Territorios	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
ESTADO ESPAÑOL	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Galea, Asturias y Cantabria	-0,40	-0,07	-0,11	-0,59	-0,53	-0,58	-0,27	-0,25	-0,42	-0,31
País Vasco, Navarra, La Rioja y Aragón	0,05	-0,19	0,45	-0,02	-0,15	-0,24	-0,15	-0,16	-0,08	-0,29
Madrid	0,69	0,51	0,10	0,09	0,27	-0,01	0,10	0,39	0,08	0,00
Castilla y León, Castilla-La Mancha y Extremadura	-0,19	-0,27	-0,30	-0,48	-0,16	-0,19	-0,10	0,13	0,38	-0,15
Cataluña, C. Valenciana y Baleares	-0,31	-0,33	-0,27	-0,30	-0,57	-0,65	-0,45	-0,52	-0,50	-0,47
Andalucía, Murcia, Ceuta y Méjilla	-0,10	0,49	-0,03	-0,49	-0,32	-0,40	-0,44	-0,48	-0,47	-0,48
Canarias	-0,21	-0,61	-0,24	-0,71	-0,74	-0,10	-0,23	-0,42	-1,06	-0,74
Extranjero	0,08	0,08	0,12	0,60	0,49	0,64	0,39	0,48	0,80	0,78
Niveles educativos										
Analfabetos	0,65	-0,21	0,10	0,15	0,05	0,45	0,80	0,06	1,01	0,77
Estudios primarios incompletos	-0,30	-0,37	-0,51	-0,27	-0,33	-0,22	-0,40	-0,41	-0,44	-0,08
1º grado	0,12	0,19	0,19	0,23	0,26	0,11	0,10	0,13	0,01	0,19
2º grado	-0,36	-0,38	-0,26	-0,48	-0,51	-0,32	-0,19	-0,18	0,00	-0,34
3º grado										

Fuente: EP a partir de EPA y Censo 2011.

El Censo de 2011, tal como sucede con el de 2001, ofrece información sobre la población que cambió de residencia en el año anterior, permitiendo construir tablas de contingencia de 20x5 filas (comunidades (NUTS-2), ciudades autónomas y extranjero para cinco niveles educativos) y 19 columnas (comunidades y ciudades autónomas). En las primeras son representados los territorios de origen de los migrantes, mientras que en las segundas aparecen los territorios de destino. Por la contra, la EPA no permite un nivel de desagregación tan amplio, con datos globales. Los archivos de microdatos de esta encuesta ofrecen información con una estructura similar a la anterior para la población en edad de trabajar que cambió de residencia el año anterior. También los resultados anuales de la EPA en el apartado de variables de la sub-muestra, "Personas que cambiaron de residencia hace un año" ofrecen datos sobre movilidad de la fuerza de trabajo en sus diferentes categorías. Esta información aparece organizada para 8 territorios de procedencia y 7 de destino (NUTS-1)⁶, ofreciendo apenas datos para los totales, obligando de esta manera a presentar los datos censales con un esquema idéntico para realizar un análisis comparativo.

Se constatan diferencias en los valores absolutos, siendo globalmente mayores para el censo hasta 2003 e invirtiendo ese resultado en adelante, existiendo una gran heterogeneidad entre los distintos grupos territoriales y niveles educativos. El Cuadro 3 y el Cuadro 4 presentan los errores relativos medios para los valores absolutos totales por NUTS-1 de origen y de destino y para los niveles educativos del total estatal. La distribución de estos errores usando como referencia el total estatal muestra una evolución temporal razonablemente similar con la única excepción de los territorios de origen Extranjero, Canarias y en mucho menor grado Castilla y León, Castilla-La Mancha y Extremadura (NUT-1 Centro). En relación a los territorios de destino, la mayor discrepancia aparece nuevamente en Canarias y

6 Los niveles educativos fueron agrupados en cinco: analfabetos, estudios primarios incompletos, primer grado, segundo grado y tercer grado. La EPA con alteraciones mínimas dependiendo del año suele utilizar los niveles de analfabetos, educación primaria, educación secundaria primera etapa y formación e inserción laboral correspondiente, educación secundaria segunda etapa y formación e inserción laboral correspondiente, formación e inserción laboral con título de secundaria segunda etapa y educación superior (incluye doctorado). Para el año 2014 utiliza los niveles de: analfabetos, estudios primarios incompletos, educación primaria, primera etapa de educación secundaria, segunda etapa de educación secundaria, con orientación general, segunda etapa de educación secundaria, con orientación profesional (incluye educación post-secundaria en el superior y educación superior).

la NUT-1 Centro. Con respecto al nivel de estudios, las diferencias mayores se presentan en la población que no sabe leer ni escribir, al no considerar la categoría de estudios primarios incompletos, al carecer la EPA de datos.

Otro aspecto destacable, que se mantiene para toda la serie, es la infravaloración de los migrantes con estudios de tercer grado del censo con respecto de la EPA. Esta cuestión no es anodina al condicionar los resultados finales del presente trabajo. Un análisis gráfico⁷ permite observar los diferenciales en los porcentajes de migrantes con estudios de tercer grado entre la EPA y los censos de 2001 y 2011 para la serie temporal 2000-2011. Con alguna excepción para años concretos los porcentajes de los inmigrantes y de los emigrantes con estudios de tercer grado son mayores en la EPA con respecto de los dos últimos censos. Estos resultados, que evidencian una reducida robustez de los datos censales, obligan a procurar una evaluación de los datos de la EPA como contrapartida de fuente migratoria por niveles educativos.

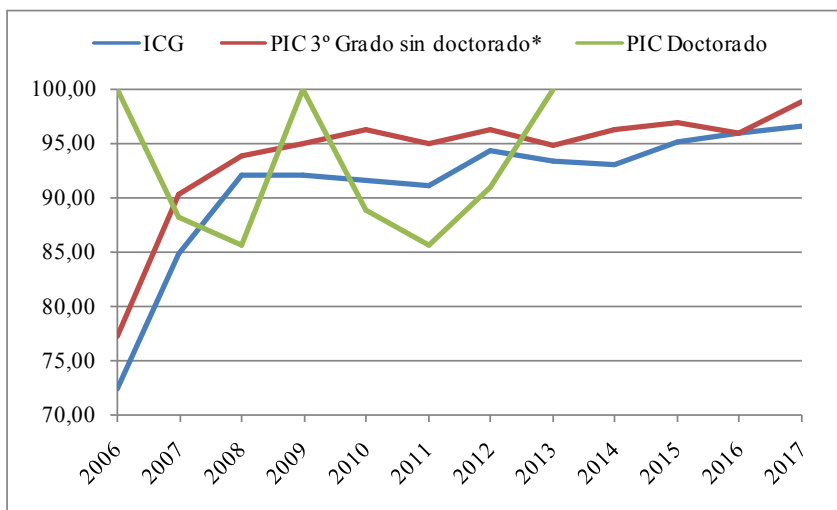
El INE tiene evaluado anualmente la calidad de la EPA durante el período 2006-2017 construyendo tablas de concordancia y calculando diversos indicadores como los descritos anteriormente para la evaluación del Censo de 2001. En esta ocasión el instituto estadístico realiza dos encuestas independientes (original y repetida) para analizar los errores de muestreo y también los ajenos a la misma. Así, poco tiempo después de tener realizada la entrevista original, a una parte de las unidades entrevistadas, se la repite nuevamente. Mediante la comparación de los datos recogidos en ambas entrevistas para las mismas unidades se puede estimar la calidad de los resultados y calcular índices numéricos sobre ella. Este procedimiento se basa en el modelo elaborado por Hansen, Hurwitz y Bershad aplicado por la Oficina de Censos de Estados Unidos (INE, 2007b).

Para el año 2006, los resultados más notables de la característica "Nivel de formación" para la población de 16 y más años serían el reducido número de personas que se clasifican, especialmente en la entrevista repetida (ER) en las modalidades Formación e inserción laboral con título de secundaria de 2ª etapa y Doctorado, especialmente en la primera, donde en la ER no fue clasificada ninguna persona, no devolviendo indicadores significativos.

7 Este análisis gráfico superaría la extensión permitida para la publicación del presente trabajo, pero puede ser solicitado a los autores del mismo.

Los mayores PIC corresponden, sin tener en cuenta la modalidad Doctorado, a las modalidades Educación primaria y Educación superior, excepto doctorado, que alcanzan valores del 77%. Con respecto al ICN y al ICB no muestran valores muy altos (sin tener en cuenta, nuevamente, la modalidad Doctorado).

FIGURA 2
INDICADORES DE CALIDAD DE LA EPA



* A partir de 2014 la PIC 3º Grado sin doctorado también incluiría estos estudios.

Fuente: EP a partir de INE (2007b), INE (2008), INE (2009), INE (2010), INE (2011a), INE (2012), INE (2013), INE (2014a), INE (2015), INE (2016), INE (2017) e INE (2018).

Por lo demás, el ICG es del 72,50%, valor superior al de 2004, y muy por encima del alcanzado para la evaluación del Censo de 2001 (53,33%). Conviene recordar que este índice evaluaba la calidad global de los datos de una característica. Para los siguientes años el índice aumenta hasta alcanzar valores por encima del 90% como muestra la Figura 2, incluyendo los porcentajes de idénticamente clasificados (PIC) para la población de 16 o más años con educación superior. El desempeño de este indicador también es muy bueno, con dígitos que suelen situarse por encima del 90%.

Como aparece en la Figura 2 los peores resultados serían los del primer año de la serie. El propio informe del INE señala, que de la misma forma que con los resultados del Censo de 2001, existen transferencias de población importantes entre determinados niveles educativos. Así el 44% de las personas clasificadas como analfabetos en la ER, fueron clasificadas en la modalidad de Educación primaria de la entrevista original (EO). También son significativas las transferencias entre las modalidades Educación secundaria, primera etapa y Educación primaria, debidas fundamentalmente a los sucesivos cambios en los planes académicos, haciendo difícil la determinación del nivel de estudios y complicando su codificación. Igualmente existe una tendencia en las personas a aumentar su status social declarando un nivel de formación mayor en la ER que en la EO. De cualquier manera, estos indicadores resultan mucho mejores que los alcanzados en el análisis similar realizado por el INE para el Censo de 2001. Esta mayor robustez en los datos de la EPA condiciona la preferencia de esta base de datos por delante de los censos.

Aún es posible realizar otro test entre la EPA y los censos, en este caso para el año 2001. La EVR ofrece datos migratorios interiores clasificados por titulación académica para algunos años⁸. El análisis de las diferencias relativas de los migrantes por niveles de estudio entre las tres fuentes resulta viable, ya que tanto el Censo de 2001, como la EPA de 2000 también permiten obtener los movimientos entre comunidades realizados en ese año clasificados con una nomenclatura parecida⁹.

- 8 La EVR dispone de esta variable entre 1988 y 2000, éste último año será el utilizado para comparar los datos entre las dos fuentes, por ser el último disponible. Los migrantes son clasificados según su nivel de estudios en: No saben leer, ni escribir; Título inferior a graduado escolar; Graduado escolar o equivalente; Bachillerato y títulos equivalentes o superiores y No clasificable y no correctamente definido.
- 9 Analfabetos; Sin estudios; Primer grado; ESO, EGB, Bachillerato elemental; Bachillerato superior; FP Grado medio; FP Grado superior; Diplomatura, Licenciatura y Doctorado. Igualando No saben leer, ni escribir a Analfabetos; Título inferior a graduado escolar a Sin estudios y Primer grado; Graduado escolar o equivalente a ESO, EGB, Bachillerato elemental y FP Grado medio y Bachillerato y títulos equivalentes o superiores a Bachillerato superior, FP Grado superior, Diplomatura, Licenciatura y Doctorado, se puede realizar una comparación entre ambas fuentes para el año 2000.

CUADRO 5
INDICADORES DE IGUALDAD ENTRE LOS MOVIMIENTOS
MIGRATORIOS DE LA EVR-CENSO, EVR-EPA POR NIVELES DE
FORMACIÓN DE LOS MIGRANTES, 2000

Comunidades de origen	Total	Analfabetos	Sin estudios	Graduado escolar	Bachillerato o superior
EVR-Censo					
U Theil	0,78	0,97	0,78	0,77	0,72
Coefficiente R ² Pearson	0,99	0,98	0,99	0,98	0,99
Coefficiente de correlación	0,99	0,99	0,99	0,99	0,99
<i>p-valor</i>	0,00000	0,00000	0,00000	0,00000	0,00000
EVR-EPA					
U Theil	0,71	0,94	0,88	0,84	0,65
Coefficiente R ² Pearson	1,00	0,96		0,99	0,99
Coefficiente de correlación	1,00	0,98		0,99	1,00
<i>p-valor</i>	0,00000	0,00000	0,00000	0,00000	0,00000
Comunidades de destino	Total	Analfabetos	Sin estudios	Graduado escolar	Bachillerato o superior
EVR-Censo					
U Theil	0,78	0,97	0,79	0,77	0,72
Coefficiente R ² Pearson	0,99	0,99	0,99	0,98	0,98
Coefficiente de correlación	0,99	1,00	0,99	0,99	0,99
<i>p-valor</i>	0,00000	0,00000	0,00000	0,00000	0,00000
EVR-EPA					
U Theil	0,71	0,92	1,00	0,61	0,56
Coefficiente R ² Pearson	0,98	0,86		0,98	0,97
Coefficiente de correlación	0,99	0,93		0,99	0,99
<i>p-valor</i>	0,00000	0,00000	0,00000	0,00000	0,00000

Fuente: EP a partir de EVR 2000, de Censo 2001 y de EPA 2000.

Los censos en general y la EPA en particular muestran valores absolutos menores a los de la EVR. En el caso del censo, además de otras carencias, cuenta con el problema de la falta de respuesta parcial (Ródenas y Martí, 2009) al no permitir sus preguntas sobre movilidad, identificar a la persona entrevistada como migrante, debiendo utilizar otros métodos de imputación indirectos, como ya quedó indicado anteriormente. Para la finalidad de este trabajo, resulta importante ver si el método de imputación utilizado distribuye los posibles errores de manera aleatoria entre los diferentes niveles formativos. Al tratarse de magnitudes con valores absolutos muy dispares son calculados indicadores de igualdad y relación como el estadístico de desigualdad U-Theil, el coeficiente de variación de Pearson

y el de correlación utilizado anteriormente, mostrando los resultados en el Cuadro 5, donde son verificados U-Theil menores para el nivel de estudios superior en todas las comparaciones, y también para este mismo nivel en el par EVR-EPA con respecto de EVR-Censo. Sin embargo, el estadístico U-Theil permanece próximo a la unidad en la mayoría de los casos, denotando desigualdad entre el par de fuentes, conviene destacar los mejores resultados para el nivel de formación superior tanto para emigrantes como para inmigrantes con respecto de los otros niveles educativos. Por oposición con este indicador, los coeficientes de Pearson y de correlación muestran una gran relación entre los datos de cada par de fuentes, con p-valores muy cercanos a cero para el coeficiente de correlación, lo que hace rechazar la hipótesis nula de incorrelación.

Nuevamente, el análisis comparativo lleva a escoger la EPA por delante del censo como fuente de datos migratorios clasificados por niveles educativos.

También resulta interesante comparar la evolución de la población entre 16 y 64 años de la EPA, al encontrarse incluida en los cálculos de la ecuación compensadora¹⁰. Los resultados para los años 2001 a 2013 están fuertemente correlacionados con los saldos totales de la EVR, con la excepción de los años 2008 a 2010, y también para la serie completa para la mayoría de los territorios considerados individualmente. En el Cuadro 6 aparecen los resultados de estas ecuaciones compensadoras y los coeficientes de correlación con los saldos migratorios totales de la EVR para la población entre 16 y 64 años.

La no correlación para los años 2008 a 2010 es debida a los saldos de la EVR que no muestran la emigración exterior que se produjo durante los primeros años de la crisis económica. La causa de este fenómeno, como indica la *Metodología* de la EVR (INE, 2014b), sería que debida a que a partir del año 2004 son incluidas las altas por omisión y las bajas por inclusión indebida de extranjeros, siendo consideradas, respectivamente, Altas por variación residencial procedente del exterior en las que no consta el país de procedencia y Bajas por variación residencial con destino al extranjero que

10 Se calcularon las defunciones por grupos de edades según las tasas de mortalidad del *Movimiento Natural de la Población: Defunciones* y los nacimientos según la población de 15 años del *Padrón Municipal*. El resultado sería el saldo migratorio total del Estado o de la comunidad autónoma.

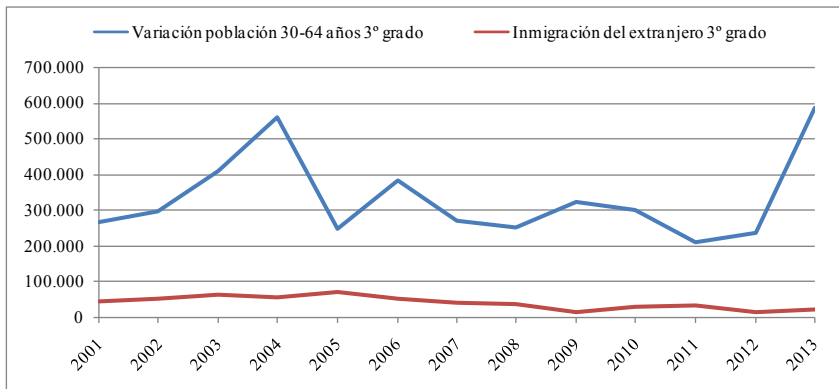
CUADRO 6
RESULTADOS DE LAS ECUACIONES COMPENSADORAS DE LA EPA Y CORRELACIÓN CON
LOS SALDOS DE LA EVR

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Coef. correl.
TOTAL	72.562	126.380	132.847	150.068	310.973	241.510	72.257	-257.013	-412.937	-421.735	-506.898	-686.817	-649.933	0,9758
Andalucía	-17.684	-3.533	6.873	17.160	46.555	13.283	-10.943	-46.844	-59.897	-62.208	-73.054	-96.894	-95.467	0,9453
Aragón	-2.063	469	1.779	2.743	13.631	10.159	7.888	-8.331	-15.057	-14.363	-16.468	-19.307	-17.751	0,8947
Asturias	-8.757	-7.432	-6.459	-6.106	-3.364	-2.991	-3.280	-7.228	-10.085	-10.974	-13.337	-15.899	-16.861	0,9608
Balears	13.707	13.643	10.740	11.257	7.946	20.048	14.178	3.535	-5.020	-6.584	-5.050	-5.830	-7.642	0,8133
Canarias	15.695	14.348	14.018	13.768	-23.026	14.478	9.525	-4.371	-9.967	-7.508	-6.909	-8.579	-12.568	0,5639
Cantabria	-973	-381	338	716	2.541	1.949	1.257	-1.668	-4.010	-4.944	-6.673	-8.840	-8.338	0,9598
Castilla y León	-14.605	-15.231	-13.097	-10.227	-5.032	10	-8.846	-23.706	-27.023	-27.358	-32.292	-38.309	-40.379	0,9504
Castilla-La Mancha	3.654	9.724	11.428	12.981	27.379	32.278	16.560	-4.255	-11.810	-12.680	-22.762	-33.010	-33.885	0,9895
Cataluña	16.193	47.498	49.263	52.719	110.665	50.518	29.236	-37.729	-72.726	-76.172	-97.873	-133.322	-126.425	0,9180
C. Valenciana	39.414	48.202	47.647	48.973	60.756	50.256	11.520	-43.161	-58.818	-50.967	-63.251	-85.678	-70.358	0,9879
Extremadura	-8.405	-5.942	-5.967	-6.331	-6.282	-2.520	-5.808	-8.168	-8.398	-8.371	-10.117	-12.709	-13.615	0,8892
Galia	-21.676	-15.847	-15.537	-14.509	-14.067	-13.017	-15.249	-21.101	-27.576	-29.454	-31.140	-36.205	-38.474	0,9653
Madrid	61.536	42.087	30.409	24.636	62.341	52.835	26.508	-29.269	-61.770	-64.058	-71.292	-118.813	-101.590	0,9777
Murcia	10.069	11.823	11.564	12.624	21.995	12.782	4.201	-5.538	-12.646	-15.244	-17.530	-17.098	-17.535	0,9638
Navarra	-787	1.175	973	223	3.511	4.432	2.807	-1.170	-3.785	-4.988	-7.062	-10.077	-8.773	0,8652
País Vasco	-15.165	-15.730	-13.850	-12.633	327	-7.778	-10.671	-17.716	-23.922	-25.275	-29.518	-38.910	-32.718	0,8000
La Rioja	2.147	2.296	2.373	2.394	2.701	2.729	1.180	-2.611	-3.668	-3.276	-4.327	-6.529	-6.120	0,9683
Ceuta	-910	-897	-447	-1.010	362	500	-378	718	1.154	153	-1.512	-1.653	1.642	0,3243
Melilla	-902	-799	-1.399	-989	568	192	1.138	582	726	1.369	2.157	-373	-4.002	0,6579

Fuente: EP a partir de EPA y EVR.

es desconocido el país de destino. Al contrario de lo que sucede con los españoles desplazados al extranjero que tienen la obligación de inscribirse en los Registros de Matrícula Consular que generan el *Padrón de españoles residentes en el extranjero* pudiendo contrastar con el banco de registro municipal del INE, localizando el municipio de residencia anterior y disponiendo del mismo para la variación residencial y viceversa. Por el contrario, con los extranjeros esto no es posible, de tal manera que la única forma de tener conocimiento de la variación residencial es mediante la declaración de la persona cuando cubre la hoja de inscripción municipal. Si la persona no comunica el país de procedencia, el motivo del alta en el *Padrón* es por omisión cuando en realidad es un cambio de residencia; con respecto a las bajas con destino al extranjero, por el mismo motivo, la única manera de registrar en el *Padrón* la baja como cambio de residencia al extranjero sería que la persona de nacionalidad no española solicitara la misma, hecho no muy habitual. La consecuencia de esto es un aumento en las altas procedentes del exterior de los no nacidos en España de más del 47% en el año 2004 con respecto al anterior y que continúa creciendo hasta 2008. Sin embargo, y teniendo en cuenta que después de 2006 se incluyen las bajas por caducidad como consecuencia del cambio legal en la Ley Orgánica 14/2003 sobre los ciudadanos extranjeros, que establece la obligación de los no comunitarios sin autorización de residencia permanente, de renovar su inscripción municipal cada dos años, y de no hacerlo los municipios deben declarar la caducidad de dicha inscripción; las bajas exteriores de los nacidos en el extranjero representan valores muy bajos en comparación con las altas y solamente alcanzan valores superiores a estos en 2013. Esto explicaría la discrepancia indicada con los datos de la EPA, ya que la propia EVR produce saldos exteriores negativos de los nacidos en España desde 2007. La robustez de los datos obtenidos estaría confirmada por otros registros poblacionales como el *Padrón de habitantes* que reproduce cifras casi idénticas a la EPA con coeficientes de correlación muy próximos a uno para todos los años del presente siglo. También la proporción de la varianza de estos saldos estatales de la ecuación compensadora atribuida a la varianza de los inmigrantes procedentes del exterior es mayor cuando los valores de estos flujos proceden de la EPA ($R^2 = 0,876$) que cuando lo hacen de la EVR ($R^2 = 0,670$).

FIGURA 3
VARIACIÓN ANUAL DE LA POBLACIÓN DE 30 A 64 AÑOS DE 3º GRADO E INMIGRACIÓN DEL EXTRANJERO, 2001-2013



Fuente: EP a partir de EPA.

Realizando una hipótesis simplificadora, considerando muy reducido el número de graduados una vez cumplidos los 29 años de edad¹¹, se calculan las variaciones anuales de la población entre 30 y 64 años, a partir de los datos de la EPA agrupados por niveles educativos. El resultado se ajusta sumando las muertes producidas en el período con datos por edades del *Movimiento Natural de la Población: Defunciones*. Centrando la atención en los resultados para el total estatal, de la población de 30 a 64 años con estudios de tercer grado, obtenidos del proceso anterior y teniendo en cuenta la hipótesis enunciada anteriormente, deberían ser una *proxy* al saldo migratorio exterior del total del Estado. La Figura 3 recoge estos resultados y las inmigraciones procedentes del exterior con estudios superiores para la población mayor de 15 años. Resulta obvio que en la ecuación compensadora faltan dos componentes (los graduados con estudios de tercer grado y los emigrantes al exterior con igual cualificación). De todos modos, estas componentes estarían rebajando los resultados parciales, aproximando las dos líneas al eje de abscisas, pero sin hacer posible la convergencia entre ellas, teniendo en cuenta los

11 En Ministerio de Educación, Cultura y Deporte (2014) y en Ministerio de Educación, Cultura y Deporte (2016) el porcentaje de graduados universitarios con 30 o más años no alcanza el 30% del total, que traducido en términos absolutos serían menos de 70.000 nuevos titulados por año.

previsibles nuevos graduados superiores mayores de 30 años y la emigración al exterior. Los resultados para este fenómeno, obtenidos de manera residual en las ecuaciones compensadoras anuales, fueron incongruentes con los que se debían esperar (resultados positivos, cuando deberían ser negativos).

Después de este análisis de las fuentes estadísticas, se concluye que la EPA constituye una fuente de datos más adecuada para estimar las matrices de flujos migratorios interiores entre las distintas comunidades autónomas, en lugar de los censos de población.

4. METODOLOGÍA DE LA ESTIMACIÓN DE LOS FLUJOS MIGRATORIOS MEDIANTE EL ***ITERATIVE PROPORTIONAL FITTING***

Para estimar un banco de datos completo de flujos migratorios por niveles educativos entre las comunidades autónomas con periodicidad anual entre 2000 y 2017 se utiliza el procedimiento IPF, propuesto por primera vez por Deming y Stephan (1940). En Birkin (1987), Norman (1999), Vejnárová (2003) y Založnik (2011) se pueden encontrar pormenorizadas descripciones del procedimiento, su historia, su práctica y sus utilidades.

El IPF se emplea para hacer ajustes o proyecciones matriciales en distintos ámbitos científicos, tanto en el formato simétrico –se entiende mismo número de filas que columnas– como en el rectangular. Se trata de un algoritmo en el que tienen lugar, de forma sucesiva, rectificaciones proporcionales por filas y columnas, sobre una matriz base. Además, puede formularse bien como un programa de optimización matemática con restricciones de igualdad, o bien como un proceso iterativo: algoritmo de escala.

El IPF es de vital importancia dado que aporta soluciones de carácter convergente, pero su aplicabilidad depende del conocimiento a priori de los márgenes –vectores suma por filas y columnas– de la matriz a proyectar o ajustar. Las extensiones de esta técnica son múltiples y estarán condicionadas por el tipo de información disponible.

El procedimiento para la estimación de los flujos migratorios a partir de la información de los microdatos de la EPA es verificado en varias fases. Comenzando por la matriz generada anualmente a partir de los cuatro ficheros de microdatos trimestrales, filtrando la población de 16 y más años que el año anterior cambió de residencia y agregando la información en una matriz de 20×5 filas y 19 columnas, que muestra la estructura de flujos migratorios

interregionales y de los inmigrantes procedentes del exterior clasificados en cinco niveles de estudios. A partir de los márgenes totales de esa matriz de microdatos, son calculados unos nuevos por elevación, convergiendo con los valores agregados del módulo de la EPA “Variables de submuestra: personas que han cambiado de residencia hace un año” utilizando un procedimiento de ajuste biproporcional. De esta manera, las migraciones clasificadas por niveles educativos de las NUTS-2 proporcionadas por la muestra de microdatos son elevadas a los valores de los movimientos migratorios de las NUTS-1 ofrecidos por las “Variables de la submuestra”.

Antes de esta estimación se testa la correlación de los flujos migratorios agregados de la EPA con los de la EVR. Para esto los datos de la EVR son agregados en NUTS-1, desde los datos originales para NUTS-2. La correlación entre ambas fuentes es muy grande con valores superiores al 95% para la mayoría de los casos, no bajando del 82%¹². Cada elemento genérico de la matriz de microdatos es denotado por $m_{ijk}(t)$, donde i sería la comunidad de origen (17 comunidades autónomas, 2 ciudades autónomas y el extranjero), j denota la comunidad de destino (17 comunidades autónomas y 2 ciudades autónomas), k el nivel educativo¹³ (analfabetos, sin estudios primarios, estudios de primaria, estudios de secundaria y estudios superiores) y t el año concreto de la encuesta (de 2000 a 2017). Estas matrices recogen una estructura migratoria interregional, pero los datos absolutos se limitan a la muestra entrevistada por lo que se hace necesario un ajuste por elevación utilizando los datos de la “Submuestra de la EPA: Personas que han cambiado de residencia hace un año. Personas de 16 y más años que han cambiado de residencia hace un año por nivel de formación alcanzado y zona de procedencia/destino”, cuyos resultados están agrupados en 8 NUTS-1 para las zonas de procedencia y 7 para las de destino. Este ajuste por elevación permite calcular los márgenes $r_{ijk}(t)$ y $s_{ijk}(t)$ que posibilitarán iniciar el procedimiento IFP para obtener las 18 matrices anuales de la serie (2000-2017).

12 Existe un análisis gráfico de este proceso que puede ser solicitado a los autores.

13 Los cambios metodológicos producidos en la EPA afectan a esta variable a partir del año 2013. Hasta ese período la encuesta trabajaba con 6 niveles formativos, pasando a 7 a partir de 2014. No obstante, ambas clasificaciones permiten agrupar los migrantes en los 5 niveles formativos utilizados en el presente trabajo.

Una vez ajustados los márgenes de la matriz de microdatos a los niveles absolutos de la submuestra de la EPA, mediante el procedimiento IPF pueden ser calculados los elementos intermedios de esa matriz con los márgenes anteriores, mediante la siguiente expresión

$$\tilde{m}_{ij/k}(t) = r_{i/k}(t) m_{ij/k}(t) s_{j/k}(t) \quad [2]$$

Las celdas con valores nulos de las matrices base tienen que ser susceptibles de rectificación¹⁴. La formulación del IPF es de carácter multiplicativo, por lo que es necesario introducir un mecanismo para favorecer la rectificación, de acuerdo con las restricciones. Esto es tan sencillo como transformar los ceros en valores positivos inferiores a la unidad utilizando la propia estructura de los márgenes y de los totales y al final del proceso iterativo se observará como algunas celdas serán modificadas y otras no. De este modo se propone a sustitución de los ceros por las proporciones que representan sobre los subtotales y multiplicarlos por las proporciones que representa cada territorio sobre los flujos globales, según la siguiente expresión:

$$n_{ij/k} = \frac{n_{\cdot j/k}}{n_{\cdot j}} \frac{n_{\cdot j}}{N} \frac{n_{i \cdot k}}{n_{i \cdot}} \frac{n_{i \cdot}}{N} = \frac{n_{\cdot j/k}}{N} \frac{n_{i \cdot k}}{N} = \frac{n_{i \cdot k} n_{\cdot j/k}}{N^2} \quad [3]$$

donde $n_{ij/k}$ representa el elemento cero de migrantes con origen en i y destino en j con un nivel educativo k ; $n_{j/k}$ es la suma de inmigrantes con destino a j con nivel educativo k ; $n_{i \cdot k}$ es el total marginal de emigrantes con origen en i y nivel educativo k y N es el total de migrantes en ese año. Esta propuesta es congruente con lo enunciado por Pearson (Pérez, 2012), que establece que dos atributos de una tabla de contingencia son independientes estadísticamente cuando la frecuencia relativa conjunta es igual al producto de las frecuencias marginales respectivas, es decir, $n_{ij} = \frac{n_{i \cdot} n_{\cdot j}}{N}$ para todo i y j .

Se procede a verificar los resultados de las matrices estimadas agregando los datos por territorios sumando los flujos para todos los niveles

14 En el ámbito de la contabilidad nacional, en principio, este problema no se da porque las estructuras productivas apenas cambian en períodos cortos de tiempo.

educativos y así comparar estos flujos totales con los proporcionados por la EVR y por los censos de 2001 y 2011. Construyendo matrices 19×19 , más los márgenes totales, pueden ser calculados varios estadísticos de selección como los propuestos por Agresti (2011): cociente de verosimilitud D (*likelihood ratio*) e índice de discrepancia ID (*dissimilarity index*). Los ratios de error medio absoluto (Robles y Sanjuán, 2005) calculados en función de los coeficientes técnicos directos y también de distribución completan este análisis, junto con los coeficientes de correlación y de Pearson¹⁵.

Los resultados de los análisis anteriores corroboran la selección de la EPA como fuente de información migratoria por niveles educativos al estar siempre por encima de los obtenidos por la fuente estadística alternativa –los censos.

5. CONCLUSIONES

Este trabajo analiza la elección de la fuente estadística no migratoria más idónea para la estimación de los flujos migratorios interregionales clasificados por nivel de estudios dentro del Estado español mediante un procedimiento de ajuste proporcional.

La literatura sobre el tópico (Cabrer et al., 2009 y Pereira et al., 2015) utilizó uno de los posibles bancos de datos –los censos– sin embargo después de una multiplicidad de contrastes, se concluye que resultaría más fiable el empleo de la otra fuente estadística, la EPA. Entre los principales resultados de estos contrastes estaría el hecho de la menor robustez de los datos censales referidos a los niveles de formación de la población, y la infravaloración de los movimientos migratorios de la población con mayores niveles educativos.

Una vez realizada la elección de los datos origen de la información migratoria por niveles formativos, el procedimiento de ajuste propuesto adapta un método ampliamente utilizado en Geografía para estimar datos demográficos incompletos de los flujos migratorios interiores, conocido como *iterative proportional fitting*, aunque empleado en otras ramas del saber,

15 Igual que en la nota anterior, la introducción de este análisis en el presente trabajo superaría con mucho la extensión permitida para su publicación. Por este motivo, estos resultados pueden ser facilitados por los autores, bajo solicitud.

como en Economía que lo denomina método RAS. Siguiendo una serie de etapas descritas en el apartado 3 fueron estimados los flujos migratorios intra e interregionales clasificados por nivel de estudio para los años 2000 a 2017, utilizando los datos de la EPA. La metodología propuesta permite obtener la convergencia en las dieciocho matrices anuales estimadas.

Los resultados obtenidos favorecen una visión más amplia de las relaciones entre comunidades, en lo que se refiere a este tipo de migraciones y permitirían posteriores análisis de tipo univariante o multivariante. También podrían ser examinados mediante la metodología input-output o incluso bajo la teoría de redes.

BIBLIOGRAFÍA

- AGRESTI, A. (2011). *Categorical data analysis*, New Jersey: John Wiley and Sons, Inc.
- ANTOLÍN, P. y BOVER, O. (1997): "Regional migration in Spain: the effects of personal characteristics and of unemployment, wages, house price differentials using pooled cross-sections", *Oxford Bulletin of Economics and Statistics*, 59 (2), 215-235.
- BEINE, M., DOCQUIER, F. y RAPOPORT, H. (2001). "Brain drain and economic growth: Theory and evidence", *Journal of Development Economics*, 64 (1), 275-289.
- BERRY, R.A. y SOLIGO, R. (1969). "Some welfare aspects of international migration", *The Journal of Political Economy*, 77 (5), 778-794.
- BHAGWATI, J. y HAMADA, K. (1974). "The brain drain, international integration of markets for professionals and unemployment: A theoretical analysis", *Journal of Development Economics*, 1 (1), 19-42.
- BIRKIN, M. (1987). *Iterative proportional fitting (IPF): Theory, method, and examples. Computer Manual*, Leeds, School of Geography, University of Leeds.
- BOVER, O. y VELILLA, P. (1999). "Migration in Spain: Historical background and current trends", *Discussion Paper of IZA*, 8.
- CABRÉ, A. y RECAÑO, J. (1997). Migraciones y coyuntura económica en las regiones españolas: Periodos 1988-90 y 1992-94, XXII Reunión de Estudios Regionales: El desarrollo de las regiones. Nuevos escenarios y perspectivas de análisis, Pamplona-Iruña.
- CABRER, B., SERRANO, G. y SIMARRO, R. (2009). "Flujos migratorios y movilidad del capital humano", *Investigaciones Regionales*, 16, 5-42.
- CHILTON, R. y POET, R. (1973). "An entropy maximizing approach to the recovery of detailed migration patterns from aggregate census data", *Environment and Planning*, 5 (1), 135-146.
- DEMING, W.E. y STEPHAN, F.F. (1940). "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *The Annals of Mathematical Statistics*, 11 (4), 427-444.
- GARCÍA, A. y STILLWELL, J. (1999). "Inter provincial migration in Spain: Temporal trends and age specific patterns", *International Journal of Population Geography*, 5 (2), 97-115.
- GIL, L.A. y JIMENO, J. (1993). "The determinants of labour mobility in Spain: Who are the migrants?", FEDEA, Documento de trabajo, 9305.
- GRUBEL, H.B. y SCOTT, A.D. (1966). "The international flow of human capital", *American Economic Review*, 56 (1-2), 268-274.
- GRUBEL, H. (1975). Evaluating the welfare effects of the brain drain from developing countries, Bellagio Conference on Brain Drain and Income Taxation, Bellagio
- HAMADA, K. y BHAGWATI, J. (1975). "Domestic distortions, imperfect information and the brain drain", *Journal of Development Economic*, 2 (3), 265-279.
- HAQUE, N.U. y KIM, S.J. (1994). "Human Capital Flight: Impact of Migration on Income and Growth", *Staff Papers-International Monetary Fund*, 577-607.
- HARRIS, J.R., TODARO, M.P. (1970). "Migration, unemployment and development: a two-sector analysis", *American Economic Review*, 60 (1), 126-142.
- INE (2007a). Evaluación de la calidad de los datos del Censo de Población 2001. Accesible en <http://www.ine.es/censo2001/EvaluacionCenso2001VFinal3.pdf>
- INE (2007b). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2006. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa06.pdf

- INE (2008). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2007. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa07.pdf
- INE (2009). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2008. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa08.pdf
- INE (2010). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2009. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa09.pdf
- INE (2011). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2010. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa10.pdf
- INE (2012). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2011. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa11.pdf
- INE (2013). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2012. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa12.pdf
- INE (2014a). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2013. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa13.pdf
- INE (2014b). Metodología de la Estadística de Variaciones Residenciales. Accesible en <http://www.ine.es/daco/daco42/migracion/notaevr.htm>
- INE (2015). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2014. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa14.pdf
- INE (2016). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2015. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa15.pdf
- INE (2017). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2016. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa16.pdf
- INE (2018). Evaluación de la calidad de los datos de la Encuesta de Población Activa. Año 2017. Accesible en http://www.ine.es/docutrab/eval_epa/evaluacion_epa17.pdf
- JOHNSON, H.G. (1979). "Some economic aspects of the brain drain", *Issue: A Journal of Opinion*, 9, 4, 7-14.
- KENEN, P. (1971). "Migration, the terms of trade and economic welfare in the source country", en J. Bhagwati, R. Jones, R. Mundell y J. Vanek (eds.) *Trade, balance of payments and growth*, Amsterdam, North-Kolland.
- LOMAX, N., NORMAN, P., REES, P. y STILLWELL, J. (2013). "Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates", *Journal of Population Research*, 30 (3), 265-288.
- McCULLOCH, R. y YELLEN, J.L. (1974). Factor mobility and the steady state distribution of income, Discussion paper 3, Harvard Institute of Economic Research.
- McCULLOCH, R. y YELLEN, J.L. (1975a) Should the brain drain be taxed?, Bellagio Conference on Brain Drain and Income Taxation, Bellagio.
- McCULLOCH, R. y YELLEN, J.L. (1975b). "Consequences of a tax on the brain drain for unemployment and income inequality in less developed countries", *Journal of Development Economic*, 2 (3), 249-264.
- MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE (2014). *Datos básicos del sistema universitario español. Curso 2013-2014*. Accesible en https://www.educacionyfp.gob.es/dms/meecd/servicios-al-ciudadano-meecd/estadisticas/educacion/universitaria/datos-cifras/DATOS_CIFRAS_13_14.pdf
- MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE (2016). *Datos y cifras del sistema universitario español. Curso 2014-2015*. Accesible en <https://www.educacionyfp.gob.es/dms/meecd/>

- servicios-al-ciudadano-mecd/estadisticas/educacion/universitaria/datos-cifras/Datos-y-Cifras-del-SUE-Curso-2014-2015.pdf
- MISHAN, E.J. y NEEDLEMAN, L. (1968). "Immigration: Some long term economic consequences", *Economica Internazionale*, Part A, 281-300.
- MOUNTFORD, A. (1997). "Can a brain drain be good for growth in the source economy?", *Journal of Development Economic*, 53 (2), 287-303.
- NAIR, P.S. (1985). "Estimation of period-specific gross migration flows from limited data: Bi-proportional adjustment approach", *Demography*, 22 (1), 133-142.
- NORMAN, P. (1999). "Putting iterative proportional fitting on the researcher's desk", White Rose Research Online, 1-37.
- PEREIRA, X., CÉSAR, M. y VERDUGO, R.M. (2015). "Estudio del perfil educativo de los migrantes en España: Una aplicación del *iterative proportional fitting* mixto", *Anales Asepuma*, 23. Accesible en <http://www.asepuma.org/2015/2015.htm>
- PÉREZ, C. (2012). *Estadística Aplicada. Conceptos y Ejercicios a través de Excel*, Madrid, Ibergarceta Publicaciones.
- PHILIPPOV, D. (1981). "Migration and settlement in Bulgaria", *Environment and Planning A*, 1978 (10), 593-617, Laxenburg, Austria International Institute for Applied Systems Analysis [IIASA].
- RAVENSTEIN, E.G. (1885). "The law of migration", *Journal of the Royal Statistical Society*, 48, 167-227.
- RECAÑO, J. (2014). Los factores sociodemográficos y territoriales de las migraciones internas en España: 1971-2011, Cambio demográfico y socio territorial en un contexto de crisis, Sevilla, XIV Congreso Nacional de Población, AGE, 391-405.
- REES, P. y DUKE-WILLIAMS, O. (1997). "Methods for estimating missing data on migrants in the 1991 British Census", *International Journal of Population Geography*, 3 (4), 323-368.
- ROBLES, L. y SANJUÁN, J. (2005). "Análisis comparativo de las tablas input-output en el tiempo", *Estadística Española*, 47 (158), 143-178, Instituto Nacional de Estadística.
- RÓDENAS, C. (1994a). *Emigración y economía en España*, Madrid: Editorial Civitas.
- RÓDENAS, C. (1994b). "Migraciones interregionales en España (1960-1989): Cambios y barreras", *Revista de Economía Aplicada*, 4 (II), 5-36.
- RÓDENAS, C. y MARTÍ, M. (2005). Migraciones: qué describen el censo de 2001 y la estadística de variaciones residenciales, VI Jornadas de Economía Laboral, Alacant.
- RÓDENAS, C. y MARTÍ, M. (2009). "¿Son fiables los datos de migraciones del Censo de 2001?", *Revista de Economía Aplicada*, XVII (50), 97-118, Universidad de Zaragoza, Departamento de Estructura Económica y Economía Pública.
- RODRIGUEZ, C.A. (1975a). "Brain drain and economic growth: A dynamic model", *Journal of Development Economic*, 2 (3), 223-247.
- RODRIGUEZ, C.A. (1975b). "On the welfare aspects of international migration", *The Journal of Political Economy*, 83 (5), 1.065-1.072.
- SANTILLANA, I. (1981). "Los determinantes económicos de las migraciones internas en España, 1960-1973", *Cuadernos de Economía*, 9 (25), 381-407.
- STARK, O., HELMENSTEIN, C. y PRSKAWETZ, A. (1997). "A brain gain with a brain drain", *Economics Letters*, 55 (2), 227-234.
- STARK, O., HELMENSTEIN, C. y PRSKAWETZ, A. (1998). "Human capital depletion, human capital formation, and migration: a blessing or a "curse"?", *Economics Letters*, 60 (3), 363-367.
- STARK, O. y WANG, Y. (2002). "Inducing human capital formation: migration as a substitute for subsidies", *Journal of Public Economics*, 86 (1), 29-46.

