

Problemas de calidad de las encuestas de hogares en presencia de flujos intensos de migraciones exteriores. Fuentes potenciales de error y su tratamiento

por
JORGE SARALEGUI GIL
Instituto Nacional de Estadística

RESUMEN

La fuerte intensidad de los flujos de migraciones exteriores que se vienen registrando recientemente en España plantea un riesgo de pérdida de calidad en las encuestas de hogares. En el documento se realiza un análisis de esta problemática y se aborda de manera particular el caso de los errores en el cálculo de los factores de expansión de los estimadores debidos a la existencia de flujos de migraciones exteriores 'ocultos' para las fuentes administrativas. Se describe cómo la explotación de resultados integrados de las encuestas de hogares disponibles en el sistema de estadísticas oficiales, proporciona una herramienta de gran potencial para el seguimiento y la evaluación de las proyecciones de población extranjera que intervienen en el cálculo de los pesos muestrales.

Palabras clave: Migraciones. Encuestas de hogares. Proyecciones de población. Paneles rotantes

Clasificación AMS:: 62P20; 91D20; 62D05

1. INTRODUCCIÓN

Los flujos recientes de inmigrantes hacia territorio español presentan una casuística asociada a diversos factores sistemáticos, coyunturales y erráticos que dificultan un adecuado registro administrativo, y por ende, estadístico, de este fenómeno. El comportamiento específico de estas subpoblaciones respecto a las características observadas para el resto de la población, plantea además la posibilidad de que se produzcan impactos indeseables en la calidad de algunas estimaciones obtenidas en las encuestas de hogares. Aparece en consecuencia un riesgo de incoherencia con otros indicadores estadísticos del mismo ámbito, estructurales o coyunturales, especialmente los relativos a variaciones de totales de una variable observada en una encuesta por muestreo y disponible también en otra fuente externa.

Aunque determinados diseños muestrales, como los que utilizan el esquema de panel rotante, estén concebidos para recoger en parte la dinámica de los cambios del universo objetivo, en los primeros momentos de instalación en el territorio de las nuevas subpoblaciones se producirán necesariamente distorsiones en la calidad del marco, en las probabilidades de selección, en determinados errores ajenos al muestreo y muy especialmente en los totales exógenos que intervienen en los factores de expansión de los estimadores. Ante estos fenómenos es indispensable, en primer lugar, estudiar y plantear las medidas para amortiguar los impactos de estas nuevas fuentes potenciales de error en las estimaciones, lo cual es el objeto de las dos primeras secciones. Posteriormente, en las secciones siguientes, se abordan de manera particular los problemas de calidad que plantean los flujos de migraciones exteriores 'ocultos' —al menos en un primer momento— para las fuentes administrativas, con la correspondiente consecuencia en las proyecciones de población que sirven de base de cálculo a los factores de expansión de las encuestas continuas de hogares.

2. LOS DIVERSOS IMPACTOS DE LAS MIGRACIONES EXTERIORES EN LA CALIDAD DE LAS ESTIMACIONES DE LAS ENCUESTAS DE HOGARES

2.1. Formulación de errores

Aunque el análisis que sigue es válido en un contexto general, es de especial aplicación en encuestas continuas de hogares, como la *EPA*. Sea una población objeto de observación muestral, un subconjunto de la cual está sometido a flujos de entrada-salida que pueden variar de intensidad bruscamente, como es el caso actualmente de la subpoblación de nacionalidad extracomunitaria (no UE) en

España. Esta subpoblación, tal como es posible observar en las incidencias que presentan las encuestas de hogares cuando se analizan en función de la nacionalidad, tiene una propensión a la falta de respuesta, o probabilidad de no respuesta total, diferente –sistemáticamente mayor- al resto de la población.

Además, según lo que muestran las evidencias empíricas, esta subpoblación aporta un error sistemático de medida en determinadas variables que no está presente en el resto de la población, por cuestiones relativas al idioma, diferentes esquemas de relaciones sociales, fuerte movilidad u otras causas muy relacionadas con su condición de inmigrante reciente.

Por añadidura, dada la tipología de esta subpoblación, existe otra componente potencial de error debido a que la mayor falta de respuesta respecto a otros colectivos, se presenta en combinación con diferencias significativas en la distribución de las principales variables objeto de observación. Por ejemplo, cuando se observa una mayor frecuencia dentro de la subpoblación de nuevos residentes de determinadas actividades económicas, o formas específicas en la composición del hogar u otras características socioeconómicas y demográficas.

Sea \bar{x}_1 la media o proporción poblacional de la subpoblación 1 (Ejemplo: proporción de extracomunitarios trabajando en la construcción) en un dominio cualquiera, cuyo índice omitimos (Ej: una provincia).

Sea \bar{x}_2 este valor para la subpoblación complementaria.

Sea \hat{P}_1 el estimador final, a partir de la muestra observada, del total de efectivos poblacionales 1.

Sea P_1 el valor poblacional de los efectivos de esta subpoblación en un momento dado, y P_2 el de su complementaria.

Sean \hat{P} y P , respectivamente, los totales estimados y poblacionales, respectivamente, de la unión de las subpoblaciones 1 y 2.

La diferencia entre el valor estimado y el poblacional como consecuencia de los diferentes efectos comentados anteriormente, en una media o proporción poblacional tendrá la forma

$$E = ((\bar{x}_1 + \epsilon_{\bar{x}_1}) \frac{\hat{P}_1}{\hat{P}} + (\bar{x}_2 + \epsilon_{\bar{x}_2}) \frac{\hat{P}_2}{\hat{P}}) - (\bar{x}_1 \frac{P_1}{P} + \bar{x}_2 \frac{P_2}{P}) \quad [1]$$

La componente de este error donde se recogen los efectos debidos a la población 1 se puede expresar, de manera aproximada como:

$$E_1 \equiv (\bar{x}_1 - \bar{x}_2) \left(\frac{\hat{P}_1}{\hat{P}} - \frac{P_1}{P} \right) + \epsilon_{\bar{x}_1} \frac{\hat{P}_1}{\hat{P}}; \quad [2]$$

donde $\epsilon_{\bar{x}_1}$ es el error total de la variable objetivo, \bar{x}_1 , correspondiente a esta subpoblación 'especial', debido a la componente sistemática de ocultación, por mala formulación de la pregunta, carácter 'sumergido' de la actividad, falta de respuesta y otros errores de observación y medida, así como a su propio error de muestreo. Suponemos que, en el caso de la subpoblación 2, estos errores ($\epsilon_{\bar{x}_2}$),

que introducirían el sumando $\epsilon_{\bar{x}_2} \frac{\hat{P}_2}{\hat{P}}$ en la expresión anterior, reciben el tratamiento habitual del proceso de encuesta y no serán objeto de análisis aquí.

2.2. Tratamiento de errores

En la expresión (2) podemos apreciar tres componentes del error cuyo control o anulación requiere acciones diversas y complementarias.

a) $\epsilon_{\bar{x}_1}$ disminuye por los siguientes procedimientos:

a.1) Con aumentos de tamaños muestrales de la subpoblación 1 mediante muestras suplementarias o selección controlada. Este tipo de muestreo, sin embargo, tiene diversos inconvenientes que es preciso tener en cuenta en el proceso de toma de decisiones. En primer lugar, en presencia de sesgos importantes debido a los errores ajenos al muestreo, la disminución del error medio cuadrático total por la vía del aumento del tamaño muestral no es fácilmente alcanzable, especialmente si los sesgos tienen su origen en la mayor predisposición a la falta de respuesta total de las personas u hogares de la subpoblación de inmigrantes recientes. Por otra parte, las encuestas continuas, con diseños complejos (como es el caso de la EPA), no se adaptan bien a diseños de geometría variable, dado que los mecanismos de actualización del marco muestral, o el esquema de rotación de las diversas submuestras, entre otros aspectos, pueden hacer que no resulte coste-eficiente añadir una dimensión adicional de complejidad a la encuesta continua. Hay que tener en cuenta también, que se producirán 'efectos de diseño' (aumento del error de muestreo) debidos a la variabilidad de los pesos que introduce la selección controlada de submuestras, lo cual puede tener consecuencias en la eficiencia de otras estimaciones relativas a la población en general (Kish, 1979).

a.2) Mejorando los instrumentos o mecánica de recogida y otras técnicas de trabajo de campo.

Éste es quizás el procedimiento que da mejores rendimientos para la corrección de sesgos debidos a los errores de marco, falta de respuesta total o parcial, res-

puestas incorrectas, errores de ocultación, etc., aunque requiere una inversión inicial en recursos para el trabajo de campo, no siempre disponibles.

Una opción de alto coste pero de gran rendimiento consiste en combinar ambos tipos de actuaciones, incluso con muestra independiente de la principal. En este último caso, la integración de la muestra complementaria de la subpoblación objetivo en la muestra principal se realizaría a posteriori, lo cual introduce cierta complejidad para los usuarios finales, que solo merecería la pena si las ganancias por las acciones del tipo a.2 son sustanciales.

b) En cuanto al primer término de (2), su análisis lo realizaremos teniendo en cuenta las proyecciones exógenas de población que son necesarias para la obtención de los factores de expansión que intervienen en el cálculo de totales estimados con los pesos muestrales. Obviamente, las proyecciones de población son uno de los productos estadísticos que más 'sufre' en presencia de flujos intensos y/o bruscos de migraciones exteriores, por razones obvias, por lo que lógicamente es necesario tener en cuenta la presencia de un error potencial en estas proyecciones, que no olvidemos que también son estimaciones predictivas. Para el análisis de las otras dos componentes de error mencionadas al comienzo de la sección 2.2, procedemos entonces de la siguiente manera.

Supongamos pues que disponemos de proyecciones de población y sus ratios entre la subpoblación objetivo y la total, $\left(\frac{\hat{P}_1}{\hat{P}}\right)$ con su propio error desconocido, entonces

$$\left(\frac{\hat{P}_1}{\hat{P}} - \frac{P_1}{P}\right) = \left[\left(\frac{\hat{P}_1}{\hat{P}} - \frac{\hat{P}_1}{\hat{P}}\right) + \left(\frac{\hat{P}_1}{\hat{P}} - \frac{P_1}{P}\right)\right] \quad [3]$$

El primer paréntesis del segundo miembro, en donde queda reflejada la segunda de las fuentes de error que estamos analizando, se anula con técnicas de postestratificación.

La postestratificación mediante calibrado de los pesos al nivel de hogar muestral es el procedimiento elegido por el INE para ajustar las poblaciones estimadas a estructuras externas por sexo y edad en las encuestas de hogares tanto continuas como esporádicas. A partir del primer trimestre de 2005 se ha anunciado oficialmente que está previsto incluir en el vector externo de variables auxiliares para el calibrado (para lo que el INE suele utilizar el software CALMAR producido por el Instituto Central de Estadística de Francia) las distribuciones de población por nacionalidad. Estas técnicas son muy útiles para reducir el error total en las estima-

ciones, pero hay que ser cuidadosos en la definición de los postestratos (en este caso serán agrupaciones de nacionalidades) con el fin de que proporcionen suficientes observaciones muestrales en cualquier periodo y, al mismo tiempo, sean muy heterogéneos entre sí respecto a las variables investigadas. No obstante, es necesario resaltar que los sesgos debidos a la falta de respuesta dentro de los postestratos son insensibles al calibrado, por lo que nunca se puede bajar la guardia en las acciones del tipo a2 mencionadas anteriormente.

Por último, la tercera fuente de error queda reflejada en el segundo paréntesis del segundo miembro de (3). Su efecto se controla, aunque en la práctica nunca será nulo, mejorando las estimaciones de estructuras poblacionales que proporcionen las proyecciones de población que vayan a intervenir en el cálculo de los factores de expansión. Estas mejoras se pueden lograr mediante nuevas fuentes externas o modelos de proyecciones complementarios adaptados a las nuevas condiciones introducidas por la intensificación de los flujos migratorios. Incluso es posible aprovechar el gran potencial de las operaciones estadísticas por muestreo corrientes del sistema para apoyar estas mejoras en la calidad de las cifras de efectivos de estas nuevas subpoblaciones, tal como desarrollamos en las secciones siguientes.

Con el fin de visualizar el impacto de las cifras de población exógenas que intervienen en los factores de expansión de los estimadores, es útil formular los mismos en formato 'estandarizado'. Por ejemplo, en las encuestas oficiales de hogares se utilizan habitualmente como estimador principal expresiones del tipo:

$$\hat{\theta} = \sum_{i=1}^p x_i w_i$$

Siendo:

p : tamaño muestral de un dominio, como estrato, provincia, comunidad autónoma, o de otro tipo, como los postestratos (ejemplo: un grupo de edad o categoría socioeconómica)

w_i : peso final de la unidad última (hogar, persona).

La práctica de los pesos estandarizados consiste en realizar en los pesos, w_i , un 'cambio de escala', como por ejemplo, dividirlos por su media en el dominio:

$$w'_i = \frac{w_i}{\bar{w}} \quad \text{con} \quad \bar{w} = \frac{\sum_{i=1}^p w_i}{p} = \frac{\hat{P}}{p}$$

El numerador es el estimador del total poblacional, \hat{P} , del dominio, siendo P desconocido, puesto que en el mundo real los marcos de selección son imperfectos o incompletos. Con ello nos queda la expresión del estimador principal:

$$\hat{\theta} = \sum x_i w_i' \cdot \frac{\hat{P}}{p}$$

w_i' : valores en torno a la unidad, que incorporan la probabilidad de selección, la corrección de falta de respuesta y el ajuste a poblaciones externas. Tienen las interesantes propiedades de que su suma en la muestra es el tamaño muestral, y las medias pueden ser estimadas con la media muestral.

En la práctica, se intenta mejorar el valor de \hat{P} sustituyéndolo por $\hat{\hat{P}}$, de una mayor calidad como, en particular, las proyecciones de población del INE obtenidas por un modelo no frecuentista, no basado en el diseño.

El valor (o "la distribución" en términos de modelos de superpoblación) de $\hat{\hat{P}}$ es indiferente para el cálculo de medias, ratios, tasas y proporciones, así como de los cambios de estas estimaciones entre dos periodos.

Sin embargo, los errores en $\hat{\hat{P}}$ tienen un gran impacto en la estimación de totales, diferencias y cambios de totales o diferencias.

Así pues, una vez que se hubieran tomado las medidas conducentes al control y reducción del error en medias, proporciones y totales en un periodo de encuesta, los cambios en medias y proporciones ya no quedan afectados significativamente por los flujos bruscos, aunque sean 'reales', de nuevas subpoblaciones. No así los cambios en totales, cuya estimación naturalmente recogerá estos cambios bruscos en el total exógeno que interviene en el factor de expansión. Efectivamente, en algunos periodos, como el que estamos viviendo actualmente en España, se pueden producir estas afluencias 'reales', bruscas, de inmigrantes. En este caso, para determinados estudios enfocados prioritariamente al cambio, estos cambios bruscos 'reales' de un total solo se podrán soslayar mediante suavizados en periodos más largos de las nuevas proyecciones.

3. LOS PROBLEMAS DE MEDICIÓN DE LOS FLUJOS MIGRATORIOS

Entre las fuentes de error analizadas anteriormente, destacan las debidas a los lógicos desfases en las propias proyecciones de población, en presencia de flujos bruscos aparentes o reales, pero difíciles de prever, de migrantes en los registros administrativos que vayan a ser utilizados como fuente de información para la

elaboración de las proyecciones, como es el caso del Padrón Continuo en España, en el cual tienen derecho y obligación de inscribirse los inmigrantes.

Podemos clasificar estas fuentes potenciales de error en las cifras población migrante (flujos) o de los correspondientes efectivos poblacionales afectados por estos flujos en un momento dado (stocks), en una fuente registral en los siguientes tipos:

- Altas retardadas ('aflorados') debido al conocido fenómeno de que el alta administrativa de un inmigrante está fuertemente condicionada por causas no medibles, como la coyuntura reglamentaria administrativa, la conveniencia subjetiva u otras, independientemente de la fecha de llegada al territorio.

- Bajas ficticias, cuando una persona opta por la salida del registro aunque no haya movimiento migratorio, debido a causas similares al caso anterior.

- Bajas ocultas (ej.: los extranjeros retornados al país de origen, que no declaran su baja), junto con el más clásico de los 'duplicados'.

- Errores administrativos, de técnicas de actualización de registros, no específicos de este colectivo.

Aquí nos referiremos a las tres primeras y para una fuente estadística de importancia primordial para la elaboración de las proyecciones de población como es el Padrón Continuo administrado por el INE, desde donde se coordinan y supervisan los procesos de actualización realizados por los ayuntamientos. Se trata de tres tipos de error de cobertura cuya corrección puede beneficiarse de operaciones complementarias por muestreo, mientras que la última entra en el ámbito de mantenimiento de registros con un tratamiento más general.

Tampoco tratamos aquí sobre los errores de especificación del modelo en las proyecciones a medio o largo plazo, particularmente presentes en estos colectivos (extranjeros o nacidos en el extranjero, cuya diferente casuística por el momento es indiferente tratar aquí). El impacto en las diversas componentes del modelo de proyección (como las de movimiento natural por defunción, nacimiento etc., o las entradas futuras, tan dependientes del ciclo económico y factores erráticos), incluso en periodos cortos, presenta dificultades obvias de predicción por los modelos.

4. APLICACIÓN DE LAS ENCUESTAS POR MUESTREO PARA ESTIMAR LOS FLUJOS MIGRATORIOS

4.1. Estimación basada en el cambio en EPA, de las altas netas de inmigrantes

La hipótesis subyacente consiste en admitir que existe un subconjunto, muy mayoritario, que puede ser definido con gran precisión tanto por variables de encuesta como por la fuente externa (padrón continuo, proyecciones de población, etc.) y cuyas cifras de efectivos son de alta fiabilidad.

En lo que sigue, designamos por conveniencia a esta subpoblación como 'subpoblación de referencia' (SR). Implícitamente, se puede tratar por ejemplo de 'españoles', 'españoles nacidos en España', 'nacidos en España', etc. En cualquier caso, la delimitación de SR a los efectos que nos ocupan debe responder a dos criterios básicos:

– Tener un comportamiento significativamente diferente respecto a determinadas variables investigadas en las encuestas de hogares (Ejemplo: totales de ocupados por ramas de actividad) en relación con otras subpoblaciones.

– Su definición mediante características recogidas por la encuesta debe ser en la práctica exactamente igual que la que permite definir la subpoblación en la fuente externa.

En contrapartida, la subpoblación cuyos efectivos en un momento dado tratamos de estimar mejor, que aquí llamaremos subpoblación objetivo (SO), podrá estar constituida por 'extranjeros residentes', 'extracomunitarios', 'extracomunitarios nacidos en el extranjero', u otras análogos. La elección de la delimitación de la SO responderá a los mismos criterios básicos que la de la SR. En las pruebas que se describen en este trabajo, la SR está constituida por 'residentes en viviendas principales, de nacionalidad española', y la SO por 'residentes en viviendas principales, de nacionalidad extranjera'. No se ha podido diferenciar la población UE de la no-UE, al no disponer de proyecciones a ese nivel en el momento de desarrollar el estudio, que no obstante resulta igualmente ilustrativo.

Describimos aquí un procedimiento basado en la explotación de una fuente pre-existente: las diversas encuestas de hogares del INE de tipo transversal, combinadas con la EPA. Esta última, como es sabido, es una fuente independiente del Padrón Continuo, por lo que sus observaciones muestrales no están afectadas por la casuística administrativa que condiciona los movimientos en este registro.

Se trata de aprovechar especialmente dos de las características que confieren a la EPA un perfil especial en las encuestas por muestreo, y que han demostrado su robustez durante las ya cuatro décadas de existencia de la encuesta:

a) Su diseño bietápico, con actualización del listado de viviendas dentro de las unidades de primera etapa (secciones censales, subdivisiones geográficas dentro de un municipio conteniendo unas 500 viviendas principales en media) mediante un barrido exhaustivo de las mismas cada seis trimestres (una sexta parte al trimestre) es decir, una actualización en mundo real, independiente del registro administrativo nominal.

b) Su carácter rotante en segunda etapa, con seis submuestras o turnos de rotación, que permite una renovación total de la muestra de viviendas (unas 18 unidades aproximadamente dentro de cada sección) cada seis trimestres, pero manteniendo fija la muestra de primera etapa, salvo actualización esporádica por cambios en la probabilidad de selección a ese nivel, o por agotamiento de viviendas dentro de la sección. Esto permite captar la dinámica 'neta' de las personas que entran o salen del territorio, proporcionando una estimación del cambio en las estructuras poblacionales SR y SO por procedimientos, conviene recalcar, que son independientes de las incidencias registrales. Hay que recordar que las unidades de primera etapa de la EPA, alrededor de 3.500 secciones, suponen aproximadamente un 10 por ciento del territorio nacional, por lo que la cobertura de la población migrante 'neta', en caso de muestreo exhaustivo de viviendas dentro de la sección en segunda etapa, sería muy eficiente. De hecho, los estudios de evaluación de la primera etapa de la EPA, en los que se comparan datos de la estimación a ese nivel con el censo más reciente, ponen de manifiesto la muy buena cobertura de la población extranjera registrada que se alcanzaría si fueran investigadas todas las viviendas de la muestra de secciones, lo cual demuestra que los defectos de cobertura de población extranjera en la EPA son debidos a las incidencias de no respuesta en segunda etapa.

El procedimiento se basa en la explotación de la submuestra (turno de rotación) que en cada trimestre entra en primera entrevista y consiste en lo siguiente:

Sea un periodo, de n trimestres consecutivos EPA, próximo al trimestre t de referencia de la estimación basada en el cambio que deseamos obtener, con $n \geq 8$, para cubrir al menos dos turnos de rotación con los que estimar el cambio en las estructuras poblacionales en la primera entrevista.

Sea x_a el colectivo de la subpoblación objetivo (SO). Ej: extranjeros en un grupo de edad, a , que deberá ser amplio, para disponer de muestra eficiente.

Sea y el subconjunto de referencia SR, cuya evolución puede ser proyectada con gran fiabilidad. Ej: población española mayor de 15 años.

El procedimiento consiste en estimar, en primer lugar, el cambio en el ratio x/y , a partir de la muestra de viviendas en primera entrevista de EPA, en m trimestres, respecto a los m anteriores en los que se renovó por última vez el mismo turno de rotación, e imputar este cambio a una población base, cuyos niveles y estructuras (de SR y SO) se suponen de gran fiabilidad en un momento dado. La serie de estimaciones EPA permite detectar que, aunque los niveles de este ratio están subestimados, su cambio parece muy ajustado a la situación actual, con incorporación sostenida de extranjeros a la muestra efectiva.

Con el mencionado criterio de mínimo se incluiría en la estimación la muestra de dos turnos en primera entrevista, es decir, un tercio de la muestra total, unas 1.200 secciones, que en el nivel nacional suponen una muestra muy eficiente, aunque no sería satisfactoria en algunas comunidades autónomas. Esta muestra correspondería a los trimestres (t) y $(t-1)$, respecto a $(t-7)$ y $(t-8)$. Para estimar el cambio neto en el ratio x/y basado en toda la muestra EPA de primera etapa serán necesarios doce trimestres consecutivos de observaciones muestrales, cada uno de los cuales contribuirá con la submuestra correspondiente al turno de rotación que en cada trimestre en particular se encuentre en primera entrevista, por renovación de la submuestra de viviendas de segunda etapa.

Veamos la expresión del estimador para el caso de utilizar la muestra completa en primera entrevista.

Sea Δr_a el cambio en el ratio de la población extracomunitaria del grupo a , x_a (perteneciente al colectivo objetivo, SO) respecto a la población y del colectivo de referencia SR. Se puede estimar por diferencia mediante:

$$\Delta \hat{r}_a = \left[\frac{\sum_{j=7}^{j=12} \hat{x}_{aj} \cdot C_x}{\sum_{j=7}^{j=12} \hat{y}_j} - \frac{\sum_{j=1}^{j=6} \hat{x}_{aj} \cdot C'_x}{\sum_{j=1}^{j=6} \hat{y}_j} \right] \quad [4]$$

Se ha omitido el índice de territorio de estudio (Ej.: comunidad autónoma).

\hat{x}_{aj} : estimador del colectivo SO, grupo a , en el trimestre j , basado exclusivamente en la submuestra del turno de rotación de primera entrevista de la EPA

j : trimestre de primera entrevista. Varía, en el caso de la fórmula, entre 7 y 12 en el minuendo y entre 1 y 6 en el sustraendo, considerando una muestra completa de EPA en dos periodos de seis trimestres consecutivos de primeras entrevistas, con $j=1$ en el trimestre más lejano. No obstante, estos periodos se pueden acotar de otro modo, por ejemplo, dentro de un periodo de 10 trimestres consecutivos de

primeras entrevistas EPA, de 7 a 10 y de 1 a 4 respectivamente, o el ya citado anteriormente de dos turnos de rotación solamente.

\hat{y}_j : estimador del colectivo SR de referencia (Ej.: españoles mayores de 15 años), obtenido a partir de la misma muestra.

Para la obtención de estas estimaciones, obviamente, los pesos no podrán estar calibrados a postestratos definidos de igual forma que los colectivos objetivo o de referencia, por lo cual, a partir de 2005, cuando en el INE se comience a realizar el calibrado de pesos según nacionalidad, será necesario estimar este cambio con los pesos antes de calibrado a nacionalidad.

c_x : propensión diferencial de falta de respuesta del colectivo x_a respecto al colectivo y . Esta propensión diferencial se refiere a la diferencia en probabilidad de respuesta (sistemáticamente menor, según observaciones empíricas de las que se da un ejemplo más adelante) del colectivo SO respecto a SR y es lo que perturba las estructuras muestrales respecto de las poblacionales. En la fórmula se supone variable entre los dos periodos consecutivos de seis trimestres. También se puede calcular 'separado', es decir, variando entre estratos, grupos de edad, o territorios. En el apartado siguiente se entra en algún detalle sobre este coeficiente.

Aunque existen diversas alternativas para el cálculo de estimaciones SO basadas en el cambio EPA, elegimos la siguiente:

$$\hat{X}'_a(T) = (1/6) \left(\frac{\hat{X}'_a(T-6)}{\hat{Y}(T-6)} + \Delta r_a \right) \cdot \hat{Y}(T) - \hat{X}'_a(T-6) + \hat{X}'_a(T-1); \quad [5]$$

Donde \hat{X}' se refiere a totales corregidos de 'altas netas' del colectivo SO en estudio e \hat{Y} a las proyecciones originales del colectivo SR, ambos del trimestre T-6 T-1 o T. Se trata de imputar al último trimestre una llegada de inmigrantes 'neta' estimada como media móvil del cambio en los últimos seis trimestres de EPA.

El efecto neto estimado sobre la proyección original, \hat{X}_a , sería:

$$(AN) = \frac{\hat{X}'_a(T) - \hat{X}_a(T)}{\hat{X}_a(T)} (\%) \quad [6]$$

4.2. El problema de la falta de respuesta de la subpoblación de nuevos residentes

Idealmente, la estimación del cambio en el ratio mencionado se deberá hacer a partir de la muestra teórica en primera entrevista, para lo que es necesario conocer la nacionalidad de los ilocalizados, ausentes y negativos, además de, obviamente, las unidades de la muestra efectiva (que contesta). Si fuera posible aplicar la fórmula a la muestra teórica y se conociera la nacionalidad de todas las personas que la integran, entonces, $c_x = 1$ en la fórmula [4]. Para ello sería necesario incluir en los cuestionarios de la muestra efectiva y en los de control de falta de respuesta, durante el trabajo de campo, las características que definen a los colectivos SR y SO, especialmente la nacionalidad y país de nacimiento, año de llegada, u otras, (según la definición por la que optemos) de los residentes en las viviendas afectadas por la falta de respuesta aunque, obviamente, algunas de estas variables van a ser de calidad muy pobre por su dificultad y costes implícitos.

En el caso más general, la fórmula [4] deberá ser calculada con la muestra efectiva, por lo que será necesario disponer de una fuente externa que proporcione una estimación de la propensión diferencial a la falta de respuesta entre ambos colectivos. Hay que tener en cuenta que, al tratarse de un ratio, es suficiente estimar la corrección de falta de respuesta diferencial entre la población en estudio, x , y la de referencia, y , designada como c_x en (4).

La corrección de falta de respuesta diferencial se puede estimar a su vez de manera exógena a la propia EPA, a partir del análisis de incidencias en otras fuentes disponibles en el INE. Para ello, dada una encuesta de hogares lo más amplia posible (como las encuestas oficiales representativas por comunidades autónomas del INE) que utilice como marco el Padrón Continuo, será posible conocer la nacionalidad (u otras características que definen SO), de los residentes en las viviendas de la muestra teórica a partir de información contenida en el marco. El coeficiente de corrección se puede aproximar mediante el estimador

$$\hat{c}_x = \frac{\hat{V}_{xt} / \hat{V}_{xt}}{\hat{V}_{yt} / \hat{V}_{yr}} \quad [7]$$

Es decir, el cociente entre los inversos de la probabilidad de pertenecer a la muestra efectiva, r , de las viviendas de la muestra teórica, t , con algún residente del colectivo objetivo, x , respecto a este mismo indicador para la población de

viviendas con ocupantes del colectivo y de referencia. Una media para un estrato o colectivo, durante un periodo determinado, para todas las encuestas de hogares aproximadamente simultáneas que utilicen como marco el Padrón Continuo del INE (por lo que el propio marco da una estimación de las características de la muestra teórica necesarias para estimar la propensión diferencial de la falta de respuesta según la nacionalidad) es con toda probabilidad una aproximación muy robusta para este tipo de coeficientes.

En el cuadro que sigue se presentan los datos de base para la estimación de la propensión diferencial a la falta de respuesta según la nacionalidad de la fórmula [7] elaborados a partir de los datos publicados por el INE sobre las incidencias de dos encuestas 'transversales' de las que utilizan el Padrón Continuo como marco, a saber, la *Encuesta de Tecnologías de la Información y de la Comunicación en los Hogares (TIC-H 03)* y la *Encuesta de Empleo del Tiempo (EET 2002-2003)*. El INE sólo ha explotado los datos de incidencias en el nivel de viviendas, por lo que no ha sido posible estimar la falta de respuesta de personas, por lo que hemos definido el coeficiente 'c' para viviendas. Cabe suponer que los coeficientes en el nivel de personas serán más elevados que los calculados aquí, dado el mayor tamaño medio de los hogares con extranjeros.

Según la población teórica que se considere, definida al pie del **cuadro 1**, la estimación del coeficiente varía. Un coeficiente de 1,22 en la TIC-H, por ejemplo, para las incidencias del **tipo 2**, significa que un ratio estimado en la encuesta referido a características de la población extranjera respecto a la española, estaría subestimado en un 22 por ciento aproximadamente en media nacional.

Cuadro 1

PROPENSIÓN DIFERENCIAL A LA FALTA DE RESPUESTA SEGÚN LA NACIONALIDAD ©

Encuesta TIC-H 2003

	<i>Muestra efectiva</i>	<i>Muestra teórica 1 (a)</i>	<i>Muestra teórica 2 (b)</i>
Viviendas con al menos 1 extranjero	339	775	585
Viviendas con solo españoles	10.101	16.343	14.265
Propensión diferencial a la falta de respuesta		1,41	1,22

Encuesta EET 2002-2003

	<i>Muestra efectiva</i>	<i>Muestra teórica 1 (a)</i>	<i>Muestra teórica 2 (b)</i>
Viviendas con al menos 1 extranjero	528	1.132	829
Viviendas con solo españoles	14.255	22.312	19.715
Propensión diferencial a la falta de respuesta		1,37	1,13

(a) Incluye: muestra efectiva + vacías + ilocalizables + ausentes + negativas + incapacidad para contestar

(b) Incluye: muestra efectiva + ausentes + negativas + incapacidad para contestar

(c) Según fórmula (7)

4.3. Solución con modelo multiplicativo

Ante los problemas teóricos que puede plantear la estimación de la propensión diferencial a la falta de respuesta entre los colectivos SO y SR, es posible abordar una solución con modelo multiplicativo, que consistiría en partir de la hipótesis de que la propensión diferencial a la falta de respuesta permanece prácticamente constante entre dos periodos de comparación, con lo cual la estimación del cambio en el ratio de la fórmula [4] puede obviar la utilización del corrector de falta de respuesta diferencial del colectivo SO respecto al SR. El estimador con este modelo quedaría de la forma:

$$\Delta \hat{r}'_a = \left[\frac{\sum_{j=7}^{j=12} \hat{x}_{aj}}{\sum_{j=7}^{j=12} \hat{y}_j} : \frac{\sum_{j=1}^{j=6} \hat{x}_{aj}}{\sum_{j=1}^{j=6} \hat{y}_j} \right] \quad [8]$$

La proyección estimada basada en el cambio observado en EPA se puede calcular con la expresión:

$$\hat{X}'_a(T) = (1/6) * \left(\frac{\hat{X}'_a(T-6)}{\hat{Y}(T-6)} * \Delta r_a \right) * \hat{Y}(T) - \hat{X}'_a(T-6) + \hat{X}'_a(T-1) \quad [9]$$

4.4. Periodo base

Para el arranque de la serie corregida es necesario partir de un periodo base, b , en el que se considere que el ratio objetivo tiene una gran fiabilidad (como por ejemplo, en un periodo censal, o de actualización especial de registros, etc.). Con lo cual podemos igualar la proyección o dato inicial base con la corregida por el modelo:

$$\hat{X}'_a(T_b) = \hat{X}_a(T_b)$$

y así calcular el ratio para el periodo base, b . Pero este ratio debe inicializarse para un periodo completo de recorrido de todo el panel de primera etapa de la EPA, es decir seis trimestres centrados en b . De no existir datos fiables para todo ese periodo, la estimación inicial del ratio para los trimestres centrados en b se haría utilizando el estimador, r , del cambio en el periodo de encuesta disponible en el momento del 'cambio de base' más próximo a b , como veremos a continuación.

5. RESULTADOS

En el experimento que se ilustra aquí, las proyecciones obtenidas con estimadores asistidos por modelos se han obtenido utilizando un ciclo completo de turnos de rotación en primera entrevista de la EPA, con objeto de que la muestra base que se utiliza para el cálculo de los cambios en el ratio de SO/SR sea suficientemente amplia incluso a nivel de comunidades autónomas, cubriendo de esta manera toda la muestra de primera etapa de esta encuesta continua, y con el modelo de la fórmula [9]. En la serie de tablas 1 del anexo se presenta una serie de datos desde el primer trimestre de 2000 hasta el segundo de 2004 de ámbito nacional y para Cataluña, Madrid y Navarra respectivamente, con el fin de reflejar el fenómeno en tres comunidades autónomas de tipología diversa. En las columnas (I), (II) y (III) de la serie de tablas 1, aparece el desglose por nacionalidad de las estimaciones de la EPA publicadas por el INE. Las columnas (V) y, (VI) de la serie de tablas 2 presentan la 'repercusión', para cada trimestre, de la estimación basada en la submuestra

en primera entrevista (aproximadamente un sexto de la muestra total, unas 10.000 viviendas en el nivel nacional) en los totales de los diferentes colectivos.

En las columnas (IX) y (X) aparecen los totales 'exógenos' de las proyecciones de población empleados por el INE en los factores de expansión de los estimadores sin desagregar UE/no UE, pues no se utilizan las proyecciones en ese nivel. Como se aprecia comparando la columna (IV) de la serie 1 con la columna (X) de la serie 2, la subestimación de la muestra efectiva de la EPA (efecto de la falta de respuesta diferencial del colectivo de extranjeros) prácticamente supera el 50 por ciento en todo el periodo.

En esta explotación, se considera como colectivo objetivo (SO) el de 'extranjeros', y como colectivo de referencia (SR) el de 'españoles'. El estudio se realizó tomando como trimestre 'base' el cuarto de 2001, o trimestre censal, en el cual se iguala la estimación del ratio x/y con el dato censal. Por supuesto, esta elección también es experimental, de manera que en otra especificación del modelo se puede adoptar como periodo base un trimestre en el que el ratio mencionado se considere de muy alta calidad, y a ser posible dentro de un periodo de estabilidad en los flujos migratorios.

En primer lugar se calculó el cambio de ratio en el 'periodo de arranque', centrado en el trimestre base (2001/IV) y basado solamente en la muestra correspondiente a tres turnos de rotación para centrar lo más posible la estimación del cambio en el trimestre censal tomado como base. Para ello con los datos estimados por la EPA, columnas (V) y (VI) desde 2000/III hasta 2001/I, a.i., se calculan los denominadores de la fórmula [8], tanto de españoles como de extranjeros (datos x e y según la fórmula, respectivamente). Asimismo, se calculan los numeradores de la fórmula con los datos del primer turno de rotación de la EPA desde 2002/I hasta 2002/III a.i.

El valor obtenido 1,523088, col. (VIII), da idea del fuerte ritmo de entrada de extranjeros en España y representa la variación media en el ratio SR/SO en un periodo móvil de seis trimestres que contienen el trimestre censal (2001/IV).

Es necesario insistir aquí de nuevo en que este dato es endógeno a EPA, no influido ni por el Padrón Continuo ni por ningún otro dato censal o registral, ni por las proyecciones oficiales de población, con lo que se demuestra la capacidad de la EPA, como panel rotante de viviendas, manteniendo fija la primera etapa de unidades geográficas (secciones) para detectar el cambio neto de las nuevas subpoblaciones de residentes.

A partir de este dato, se ha estimado la cifra de residentes SO (extranjeros) en un periodo base que va desde 2001/I hasta 2002/III, necesario para estimar las proyecciones en trimestres sucesivos. Con objeto de representar la serie completa

de los trimestres sobre los que se disponía de una serie homogénea (después de los recientes cambios metodológicos en la encuesta) se ha prolongado la serie hacia atrás hasta el 2000/I, enlazando con las proyecciones oficiales de la columna (X) para todo el año 2000.

Para la estimación de los trimestres del periodo arranque se ha empleado la fórmula:

$$\hat{X}'(t_n) = \hat{Y}(t_n) * (X_b / Y_b) * \Delta r_a^{(n/6)} \quad [10]$$

donde:

$\hat{X}'(t_n)$ es la estimación de la población SO en t (trimestre del periodo arranque)

(X_b / Y_b) es el ratio del trimestre base (censal 2001/IV)

$\hat{Y}(t)$ es la proyección de población SR de que disponemos en las proyecciones, y

Δr_a es el estimador del cambio en el ratio SO/SR estimado con el turno de rotación de primera entrevista EPA según la fórmula [8] y por el procedimiento indicado anteriormente para el periodo de seis trimestres centrados en el trimestre censal.

n varía de (-3) en 2001/I hasta (+3) en 2002/III. En el momento censal se parte de la identidad, con n=0 en la fórmula [10].

En cuanto a las estimaciones de la serie inmediatamente anterior al periodo de referencia, que constituyen el objetivo final de esta explotación, en primer lugar es necesario calcular las estimaciones del cambio en el ratio según la fórmula [8], que dan como resultado los datos de la columna (VII) de la serie 2. Como se aprecia, la entrada de inmigrantes se siguió produciendo a un ritmo muy alto en el periodo postcensal, según las estimaciones de EPA. El dato de la columna (VIII), a partir de 2002/III se obtiene ahora ya como estimación con la fórmula [9], tratándose pues de una predicción obtenida exclusivamente a partir de EPA. La serie presenta un perfil ligeramente diferente al de la columna (X), de la proyección 'oficial', aunque es obvio que ambas fuentes, a pesar de su independencia estadística, y éste es el hecho más destacado, recogen de manera análoga la entrada neta reciente de extranjeros en España. Para utilizar con mayor exactitud la EPA como instrumento evaluador de las proyecciones de población a corto plazo, es necesario realizar los cálculos con datos de la encuesta algo más desplazados hacia delante, para lo que hay que esperar unos pocos trimestres a fin de que el periodo de estimación quede mejor centrado en el trimestre de referencia de las cifras de población. En efecto, en el cálculo del cambio del ratio x / y , al utilizarse medias móviles en las fórmulas

(4) y (8) (al fin y al cabo, estamos tratando con 'predicciones') se producirá un retardo en la referencia del ratio respecto a la referencia del efectivo de población que estamos estimando. La reducción en el número de trimestres de primera entrevista a utilizar (con un mínimo de dos, para garantizar representatividad de la muestra de primera etapa) ayuda a 'acercar' la referencia del ratio x/y respecto del periodo de actualización, tal como se ha hecho para el periodo base, pero aumenta el error del muestreo, especialmente para las estimaciones por comunidades autónomas. En los gráficos 1 y 2, se representan las estimaciones basadas en el cambio EPA (columna (VIII) de la serie de tablas 2) frente a las proyecciones exógenas a la encuesta, columna (X), donde se aprecia la calidad del ajuste de la estimación basada en el cambio de la EPA respecto de la proyección en el nivel nacional. En las comunidades autónomas el perfil ofrece mayores diferencias, con evolución menos 'suavizada' en el caso de la estimación EPA. Hay que tener en cuenta el efecto del error de muestreo de las estimaciones de encuesta, así como que la observación directa del trabajo de campo, recoge de manera natural una mayor variabilidad en mundo real, sin olvidar el predominio de la componente de tendencia a largo plazo, más suave, que suelen incorporar los modelos implícitos en la elaboración de las proyecciones oficiales de población.

6. CONCLUSIONES

Una explotación de resultados integrados de las principales encuestas de hogares disponibles en el sistema de estadísticas oficiales, en un periodo determinado, proporciona una herramienta de gran potencial para el seguimiento y evaluación de las cifras relativas a los flujos de migraciones exteriores captados en las fuentes administrativas y, por lo tanto, las estructuras según la nacionalidad (y/o país de origen) de la población residente en el ámbito territorial en un momento dado.

En particular la disponibilidad de una macroencuesta como la EPA, de calidad contrastada, permite estimar el cambio neto en las estructuras según la nacionalidad de la población residente en territorio español, de manera continua. En efecto, el diseño bietápico de la encuesta mantiene fijas en el tiempo (salvo pequeños cambios por actualización de probabilidades de selección) las unidades de primera etapa, constituidas por pequeñas áreas geográficas (secciones) lo cual es esencial para la medida del cambio al tratarse de una muestra de gran tamaño que cubre un 10 por ciento del territorio nacional. La actualización, según un esquema rotante, de las listas de viviendas principales dentro de los límites de las secciones, asegura que la muestra teórica de segunda etapa, antes de falta de respuesta, tenga una gran calidad de cobertura. A su vez, el mismo esquema rotante de segunda etapa,

permite estimar el cambio en la tipología de los habitantes en las viviendas principales que van incorporándose por primera vez a la muestra.

Otras encuestas transversales que utilizan el registro permanente de la población residente en España (Padrón Continuo) como marco de selección, pueden proporcionar las estimaciones de bajas ocultas o detectar el comportamiento específico de estas subpoblaciones frente a la falta de respuesta, dado que la muestra teórica siempre dispone de la característica nacionalidad (o país de origen) junto con otras características demográficas, presentes en el marco de selección.

Las técnicas de posestratificación son muy útiles para corregir el sesgo en la estimación de las estructuras muestrales por nacionalidad, debido a la mayor falta de respuesta de los colectivos de inmigración reciente, según se viene observando en diferentes operaciones estadísticas. No obstante, estos procedimientos exigen una buena calidad de las estructuras externas de calibrado, por lo que es fundamental una evaluación permanente de las fuentes externas que se utilicen a este fin.

Para estimaciones detalladas de características socioeconómicas de los nuevos residentes, será necesario aportar muestra complementaria en las encuestas de hogares del sistema público de encuestas, dados sus escasos tamaños muestrales que no permiten cubrir con eficiencia estas subpoblaciones. Estos procedimientos permiten mejorar no solo la estimación del cambio en totales, sino también otros parámetros como medias o proporciones de determinadas características de estos colectivos especiales. No obstante, no hay que olvidar que estos estudios son de coste muy elevado, si se pretende alcanzar cotas de calidad aceptables en cobertura, tasas de respuesta, y errores de observación y de medida. Por ello es muy recomendable formular proyectos mixtos, que combinen la fuente administrativa con la observación directa. Con este objetivo, es conveniente diseñar protocolos de colaboración con las autoridades implicadas en el registro de las migraciones exteriores, de manera que sea posible introducir cuestionarios obligatorios para observación con fines estadísticos –contemplando incluso esquemas muestrales específicos- en algunos de los procedimientos administrativos que registran el fenómeno migratorio.

7. AGRADECIMIENTOS

Debo agradecer a los funcionarios del INE, D^a Juana Porras Puga, Subdirectora General Adjunta de Diseño de Muestras, y D. Juan Jiménez Llorente, experto en evaluación de resultados de encuestas de hogares, su extraordinaria colaboración en la resolución de consultas y en el suministro de datos a medida sin los que no hubiera sido posible el desarrollo de los análisis que se presentan en este trabajo.

Tabla 1
 ESTIMACIONES DE POBLACIÓN DE 16 Y MÁS AÑOS EN LA EPA SEGÚN
 NACIONALIDAD. Considerando todos los turnos de rotación

<i>Año/ trimestre</i>	<i>Españoles (I)</i>	<i>UE (II)</i>	<i>No UE (III)</i>	<i>Total extranjeros (IV)</i>
1.1 TOTAL NACIONAL				
2000/1	32.802.223	125.068	234.720	359.788
2000/2	32.881.232	138.747	250.112	388.859
2000/3	32.977.799	128.371	272.004	400.375
2000/4	33.046.139	139.043	301.077	440.120
2001/1	33.091.265	144.640	341.470	486.110
2001/2	33.132.134	149.327	370.081	519.408
2001/3	33.187.834	149.122	388.755	537.877
2001/4	33.205.315	153.330	441.230	594.560
2002/1	33.215.356	159.011	494.565	653.576
2002/2	33.271.413	154.536	506.970	661.506
2002/3	33.287.208	160.703	548.999	709.703
2002/4	33.268.472	161.092	631.340	792.432
2003/1	33.228.572	174.796	717.077	891.873
2003/2	33.260.310	163.201	751.994	915.195
2003/3	33.255.499	158.710	816.382	975.092
2003/4	33.250.552	170.831	864.278	1.035.109
2004/1	33.255.461	183.743	897.006	1.080.749
2004/2	33.284.911	183.660	913.681	1.097.341
1.2 CATALUÑA				
2000/1	5.118.483	12.360	56.920	69.280
2000/2	5.134.021	13.778	54.412	68.190
2000/3	5.149.238	14.253	53.161	67.414
2000/4	5.160.567	15.096	55.435	70.530
2001/1	5.167.111	17.747	57.974	75.721
2001/2	5.170.459	17.317	64.080	81.397
2001/3	5.171.639	19.204	70.035	89.238
2001/4	5.160.175	21.073	88.654	109.728
2002/1	5.165.814	26.157	86.198	112.354
2002/2	5.175.878	29.472	80.329	109.801
2002/3	5.183.861	32.633	76.694	109.327
2002/4	5.178.071	30.500	92.125	122.625
2003/1	5.159.860	32.724	114.880	147.604
2003/2	5.144.553	27.043	141.881	168.923
2003/3	5.140.776	24.448	154.276	178.724
2003/4	5.142.021	23.566	159.927	183.493
2004/1	5.127.226	28.750	174.990	203.740
2004/2	5.124.950	31.938	178.961	210.899

Tabla 1 (Conclusión)

ESTIMACIONES DE POBLACIÓN DE 16 Y MÁS AÑOS EN LA EPA SEGÚN
NACIONALIDAD. Considerando todos los turnos de rotación

<i>Año/ trimestre</i>	<i>Españoles (I)</i>	<i>UE (II)</i>	<i>No UE (III)</i>	<i>Total extranjeros (IV)</i>
1.3 MADRID				
2000/1	4.191.018	10.547	84.772	95.319
2000/2	4.199.781	20.203	85.613	105.817
2000/3	4.219.044	19.135	86.676	105.811
2000/4	4.225.795	26.077	92.241	118.318
2001/1	4.221.186	24.729	114.283	139.012
2001/2	4.237.170	24.713	111.217	135.930
2001/3	4.238.516	29.336	118.157	147.493
2001/4	4.250.790	28.279	119.842	148.120
2002/1	4.242.703	29.913	138.399	168.312
2002/2	4.254.502	24.913	142.906	167.819
2002/3	4.250.402	23.427	159.796	183.223
2002/4	4.250.818	25.071	169.043	194.114
2003/1	4.241.417	24.849	189.227	214.076
2003/2	4.252.728	22.992	189.595	212.587
2003/3	4.241.900	21.841	211.390	233.231
2003/4	4.231.668	30.165	223.120	253.285
2004/1	4.249.447	34.622	209.881	244.503
2004/2	4.261.611	32.703	207.815	240.518
1.4 NAVARRA				
2000/1	448.371	977	776	1.753
2000/2	448.308	1.347	1.303	2.650
2000/3	448.749	1.471	1.571	3.042
2000/4	447.075	1.778	3.772	5.550
2001/1	446.749	2.010	4.591	6.600
2001/2	447.783	1.405	4.775	6.180
2001/3	446.437	1.308	6.832	8.140
2001/4	446.373	1.755	7.062	8.817
2002/1	445.391	2.472	7.886	10.358
2002/2	444.646	2.402	9.212	11.613
2002/3	443.772	2.327	10.670	12.996
2002/4	442.299	2.576	12.403	14.978
2003/1	443.967	2.966	10.809	13.776
2003/2	442.649	3.187	12.327	15.514
2003/3	442.698	2.669	13.218	15.888
2003/4	439.729	1.957	17.321	19.277
2004/1	442.028	1.515	15.877	17.392
2004/2	442.957	2.159	14.709	16.868

Tabla 2
 ESTIMACIONES DEL CAMBIO DE POBLACIÓN EXTRANJERA DE 16 Y MÁS AÑOS EN LA EPA
 Considerando solo el turno de rotación en 1ª Entrevista

Año/ trimestre	Repercusión del 1er. Turno de Rotación en el estimador del total		Incremento ratio (SO/SR) (VII)	Estimación EPA basada en el cambio 1ª entrevista Extranjeros (VIII)	Proyección exógena de población de 16 y más años (base 2001)	
	Españoles (V)	Extranjeros (VI)			Españoles (IX)	Extranjeros (X)
2.1 TOTAL NACIONAL						
2000/1	5.214.794	58.606		750.377	32.943.158	748.008
2000/2	5.260.275	64.570		826.120	32.977.161	823.512
2000/3	5.192.891	62.040		901.865	33.011.142	899.018
2000/4	5.389.563	91.468		977.608	33.045.139	974.522
2001/1	5.383.945	81.432		1.057.946	33.074.548	1.054.606
2001/2	5.589.995	108.967		1.135.648	33.099.392	1.139.266
2001/3	5.546.540	95.748		1.219.056	33.124.229	1.223.929
2001/4	5.548.325	110.599		1.308.589	33.149.065	1.308.589
2002/1	5.536.149	118.097		1.404.257	33.163.503	1.403.661
2002/2	5.514.094	119.230		1.505.487	33.146.436	1.530.210
2002/3	5.366.136	130.593	1,523088	1.614.016	33.129.391	1.656.758
2002/4	5.541.740	176.738	1,558170	1.719.779	33.112.348	1.783.307
2003/1	5.559.972	180.970	1,623173	1.846.401	33.125.030	1.895.668
2003/2	5.582.325	169.244	1,604222	1.978.170	33.148.054	2.013.132
2003/3	5.334.193	174.058	1,572809	2.112.316	33.171.085	2.130.597
2003/4	5.620.413	189.020	1,615206	2.267.263	33.194.109	2.248.061
2004/1	5.604.041	190.294	1,574468	2.422.878	33.213.998	2.358.709
2004/2	5.545.763	188.232	1,445340	2.552.007	33.230.751	2.462.543
2.2 CATALUÑA						
2000/1	806.208	9.753		190.538	5.118.185	183.180
2000/2	840.978	11.614		199.410	5.126.568	191.709
2000/3	800.751	13.462		208.281	5.134.946	200.238
2000/4	898.592	16.811		217.153	5.143.325	208.767
2001/1	904.639	13.760		227.226	5.150.547	218.451
2001/2	935.657	13.747		234.880	5.156.613	229.291
2001/3	886.095	17.426		242.792	5.162.679	240.130
2001/4	890.974	23.436		250.970	5.168.749	250.970
2002/1	831.677	18.575		258.980	5.165.974	270.654
2002/2	894.346	19.707		266.787	5.154.352	299.181
2002/3	920.924	15.937		274.829	5.142.729	327.709
2002/4	949.410	30.602	1,532830	295.390	5.131.111	356.236
2003/1	853.819	40.981	1,694976	323.091	5.130.900	378.963
2003/2	901.870	46.845	1,738683	353.478	5.132.472	405.502
2003/3	808.993	34.755	1,826182	388.652	5.134.053	432.041
2003/4	869.760	32.931	1,908296	428.730	5.135.625	458.580
2004/1	906.174	38.414	2,090315	478.559	5.136.655	483.622
2004/2	913.547	40.915	1,910931	523.516	5.137.138	507.167

Tabla 2 (Conclusión)
ESTIMACIONES DEL CAMBIO DE POBLACIÓN EXTRANJERA DE 16 Y MÁS AÑOS EN LA EPA
 Considerando solo el turno de rotación en 1ª Entrevista

Año/ trimestre	Repercusión del 1er. Turno de Rotación en el estimador del total		Incremento ratio (SO/SR) (VII)	Estimación EPA basada en el cambio 1ª entrevista Extranjeros (VIII)	Proyección exógena de población de 16 y más años (base 2001)	
	Españoles (V)	Extranjeros (VI)			Españoles (IX)	Extranjeros (X)
2.3 MADRID						
2000/1	689.032	14.106		123.854	4.333.074	104.709
2000/2	734.858	18.910		157.052	4.327.052	132.776
2000/3	699.292	9.301		190.252	4.321.024	160.844
2000/4	720.883	31.245		223.451	4.315.003	188.911
2001/1	631.910	31.308		255.684	4.309.795	216.162
2001/2	712.566	23.508		268.310	4.305.407	242.595
2001/3	650.307	25.016		281.558	4.301.016	269.028
2001/4	756.364	24.546		295.461	4.296.628	295.461
2002/1	686.671	24.840		310.002	4.291.578	322.556
2002/2	697.804	27.883		325.208	4.285.866	350.311
2002/3	601.076	40.684		341.159	4.280.154	378.067
2002/4	678.334	38.441	1,454065	360.996	4.274.443	405.822
2003/1	691.000	48.534	1,485067	383.343	4.275.385	429.958
2003/2	756.784	38.318	1,530967	409.160	4.277.806	455.643
2003/3	685.408	41.779	1,485745	434.054	4.280.231	481.327
2003/4	759.065	48.077	1,614389	467.289	4.282.652	507.012
2004/1	711.239	40.881	1,474335	494.344	4.284.481	531.174
2004/2	701.830	35.966	1,321468	513.895	4.285.721	553.814
2.4 NAVARRA						
2000/1	80.222	776		14.849	441.877	16.347
2000/2	77.116	395		15.418	442.630	16.973
2000/3	77.162	678		15.986	443.383	17.598
2000/4	68.976	2.864		16.554	444.137	18.224
2001/1	73.634	1.851		17.073	444.942	18.795
2001/2	78.416	533		18.099	445.805	19.310
2001/3	84.070	2.021		19.187	446.668	19.825
2001/4	78.113	1.269		20.340	447.530	20.340
2002/1	76.152	2.050		21.538	447.879	21.369
2002/2	76.151	2.698		22.779	447.715	22.912
2002/3	73.244	3.017		24.093	447.549	24.455
2002/4	69.445	3.875	2,096153	27.422	447.384	25.998
2003/1	77.936	1.637	1,775843	29.928	448.658	26.218
2003/2	73.823	933	1,588790	31.940	448.819	27.660
2003/3	79.053	1.863	1,352973	33.219	448.980	29.102
2003/4	67.587	5.013	1,657983	35.737	449.141	30.544
2004/1	72.621	1.804	1,381334	37.289	449.256	31.904
2004/2	71.580	1.944	0,912822	36.909	449.333	33.182

Figura 1

Comparación proyecciones de población extranjera y estimaciones basadas en el cambio de EPA

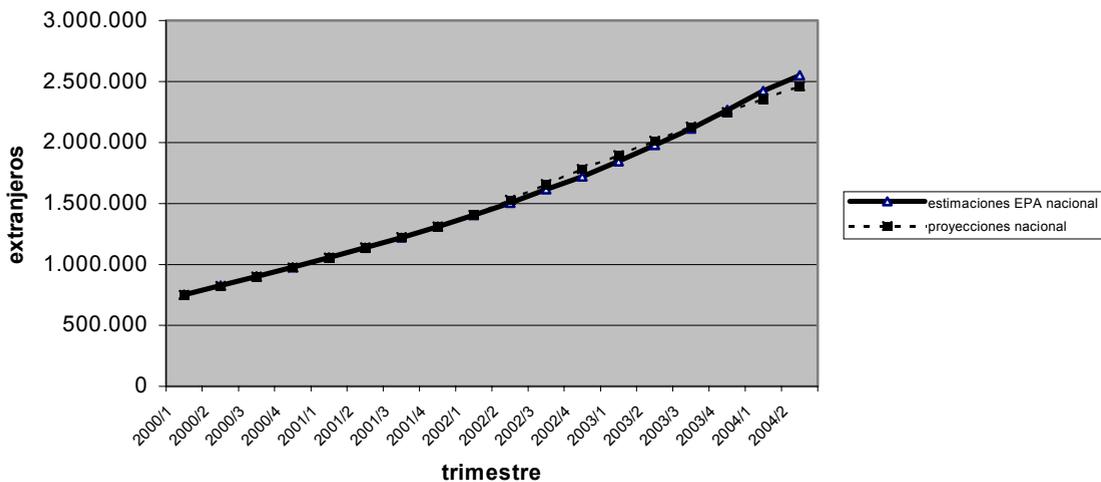
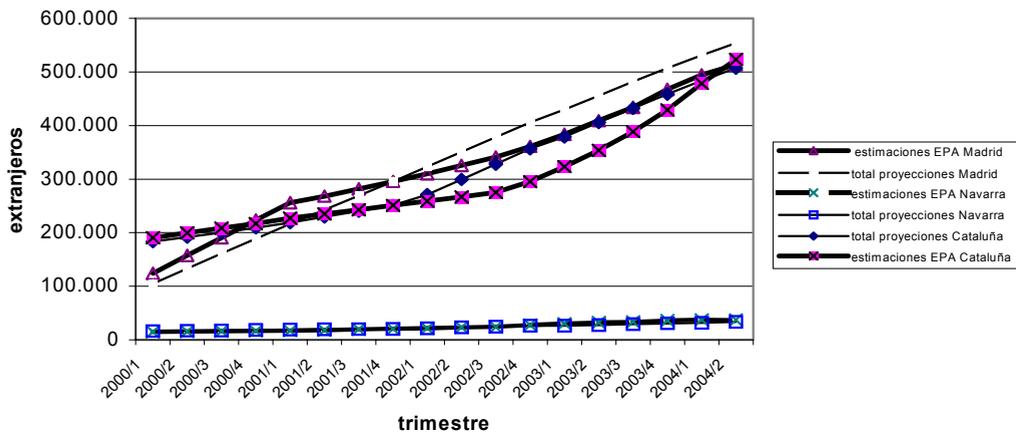


Figura 2

Comparación proyecciones de población extranjera y estimaciones basadas en el cambio de EPA



REFERENCIAS

COCHRAN, W.G. (1977) «Sampling Techniques». Wiley.

DEVILLE, J.C. and SANDARL, C.E. (1992) «Calibration Estimators in Survey Sampling». JASA.

DUNCAN, G. et al. (1989). «Panel Surveys». Wiley

INE (Trimestral) . «Encuesta de Poblacion Activa. (EPA) ».

INE (anual) . «Encuesta de Poblacion Activa. (EPA). Evaluación de Resultados» .

INE (2003) . «Encuesta de Empleo del Tiempo. Evaluación de Resultados».

INE (2001) . «Informe Técnico de la EPA».

KISH, L. (2003). «Selected Papers». Wiley.

SARNDAL, C.A. et al. (1991) «Model Assisted Survey Sampling». Springer.

QUALITY PROBLEMS IN HOUSEHOLDS SURVEYS IN THE PRESENCE OF INTENSIVE EXTERNAL MIGRATION FLOWS. ERRORS TYPOLOGY AND CORRECTION TECHNIQUES

ABSTRACT

The strong intensity of recent immigration flows , implies a risk of loss of quality in households surveys. The document analyses the typology of these negative impacts and approaches in particular a procedure to counteract the potential errors incurred in the calculation of survey weights in the presence of migrations flows ' hidden' for the administrative sources. It describes how the analysis of integrated results of household surveys currently available within the system of official statistics, provides an efficient tool to monitor the quality of population forecasts which are actually used to calculate final survey weights.

Key Words: Migration flows. Households Surveys. Population Forecasts. Rotating Panels.

AMS Classification: 62P20; 91D20; 62D05