

Propuesta metodológica para la desagregación espacio-temporal de los indicadores de demografía empresarial

Juan C. Duque *, José R. García **, Jordi Suriñach ***

RESUMEN: En este artículo se presenta un diseño metodológico para obtener, para el conjunto de las Comunidades Autónomas en España, indicadores sobre la demografía empresarial a nivel sectorial, desagregando tanto en función del tamaño empresarial, como según la condición jurídica de las empresas. Este diseño se aplica al caso de Cataluña. La información de base para su obtención son las cuentas de cotización (trimestrales) de la Seguridad Social. Por ello, se propone una adaptación trimestral de los conceptos habitualmente obtenidos a nivel anual, y una metodología que permita calcular la demografía empresarial a partir de los datos de registros de la Seguridad Social, que sea compatible con la información oficial.

Clasificación JEL: R12, R30, C8.

Palabras clave: demografía empresarial, análisis regional.

Methodological proposal for temporal break down of indicators of business demography

ABSTRACT: This article presents a methodology to calculate, for the whole of the Autonomous Communities in Spain, firm demography indicators at the sector level, disaggregating in terms of both business size and the legal status of companies. We provide empirical evidence using quarterly accounts of Social Security for Catalonia. As firm demography indicators are usually obtained annually, we propose new formulations that allow their calculation on quarterly bases, using information from the Social Security records, so that the results are consistent with the official information.

JEL Classification: R12, R30, C8.

Keywords: firm demography, regional analysis.

* Research in Spatial Economics (RISE-group), Department of Economics, EAFIT University, Carrera 49 7 Sur - 50, Medellín, Colombia. jduquec1@eafit.edu.co.

** AQR Research Group-IREA. Universitat de Barcelona, Avda. Diagonal, 690, Barcelona, 08034, España. jrgarcia@ub.edu.

*** AQR Research Group-IREA. Universitat de Barcelona, Avda. Diagonal, 690, Barcelona, 08034, España. jsurinach@ub.edu.

Recibido: 15 de septiembre de 2011 / Aceptado: 19 de abril de 2012.

1. Introducción¹

En el análisis de coyuntura y en el seguimiento de la actividad económica es habitual centrarse en el análisis de la evolución del mercado de trabajo, como por ejemplo la tasa de paro o el número de ocupados. Ello es debido a la facilidad de disposición (e inmediatez) de información sobre el número de afiliados a la Seguridad Social en la mayoría de países. Otras variables de interés son el PIB, el índice de producción industrial, y otras más indirectas como el consumo de electricidad, las ventas al detalle, el consumo de cemento, entre otras.

Un conjunto de información complementaria, muy relevante y que aporta muchos matices es la referente a la demografía empresarial, entendida como el conjunto de datos sobre altas y bajas empresariales, así como el *stock* de empresas existentes en cada momento (Hannan y Freeman, 1977; Hannan y Freeman, 1989). Su utilización como seguimiento de la actividad empresarial y económica es menor por la dificultad de disponibilidad de la misma (van Dijk y Pellenbarg, 1999). Anualmente, en España se publica dicha información, a nivel estatal, por parte del INE² (Instituto de Estadística Oficial), utilizando como base un conjunto muy heterogéneo de fuentes, como son el impuesto de actividades económicas, las retenciones sobre rentas del trabajo personal, el impuesto sobre el valor añadido, el impuesto de sociedades, las cuentas de cotización de la Seguridad Social y la información del registro mercantil, entre otras. Esta información, publicada por el Directorio Central de Empresas (DIRCE), ha sido utilizada en diversos estudios que analizan la creación, desaparición y supervivencia de empresas. Destaca el libro de Segarra *et al.* (2002), que es un estudio completo de la demografía empresarial de la industria en España en el periodo 1994-2000 en el que se realiza un análisis de la demografía empresarial para 24 sectores de la industria manufacturera. Asimismo, otros trabajos como el de Manjón *et al.* (2002) utilizan el Registro de Establecimientos Industriales y la Encuesta Industrial.

A nivel autonómico, en los últimos años el INE ha iniciado la publicación de la operación estadística denominada Demografía Armonizada de Empresas. En ella se proporciona información, de carácter anual, del *stock*, los nacimientos y las muertes de empresas. La información publicada corresponde a una desagregación de dos dígitos de la CNAE-2009 y realiza una diferenciación según el número de asalariados y la condición jurídica de las empresas. Asimismo, el INE, mediante el DIRCE, también publica para las Comunidades Autónomas el *stock* de empresas existentes, sin entrar en la dinámica de altas y bajas.

¹ Los autores desean agradecer al Departamento de Trabajo de la Generalitat de Cataluña su ayuda y colaboración, así como ser los verdaderos impulsores de la investigación. Asimismo, se desea hacer constar el agradecimiento al Ministerio de Educación y Ciencia, bajo el proyecto ECO2009-12678. Finalmente, agradecer los comentarios a los dos evaluadores anónimos que han efectuado observaciones a una versión inicial del artículo.

² La metodología a partir de la cual el INE obtiene indicadores de demografía empresarial se puede consultar en <http://www.ine.es/metodologia/t37/t3730204.pdf>.

En este artículo se presenta el diseño metodológico seguido para obtener, para la Comunidad Autónoma de Cataluña, la información sobre diferentes conceptos de la demografía empresarial a nivel trimestral. En este sentido, cabe destacar que se propone calcular no sólo el *stock* de empresas y los movimientos de altas y bajas, sino también otros conceptos que pueden ayudar a ampliar los análisis sobre la demografía empresarial, como pueden ser las altas por reactivación, las tasas brutas de entrada y salida, la tasa neta de entrada, la rotación empresarial, la función de supervivencia y la función de riesgo. Asimismo, la metodología propuesta, que se lleva a la práctica para el caso de Cataluña, es ampliable al resto de Comunidades Autónomas.

Un aspecto que también debe subrayarse es que se propone la obtención de los diferentes conceptos de demografía empresarial con una periodicidad trimestral. Éste es un aspecto novedoso del artículo, ya que habitualmente la información publicada y los análisis realizados en este ámbito toman como referencia datos anuales. Esto es así porque se considera que son más informativos y fácil de relacionar con otras variables. No obstante, entendemos que aunque la información anual es importante e imprescindible para la realización de muchos análisis, la obtención de la demografía empresarial trimestral puede ser también de gran utilidad y es complementaria a la información anual. Por ejemplo, el cálculo de la demografía empresarial trimestral puede ayudar a responder preguntas como: ¿existe estacionalidad en la demografía empresarial?, ¿en un sector determinado, la muerte de empresas se concentra en, por ejemplo, el segundo trimestre del año?, o, ¿es importante la magnitud de empresas cuya vida está constituida por altas y bajas debidas a la estacionalidad? Así, la información trimestral sobre demografía empresarial podría resultar clave para aumentar el conocimiento coyuntural de las empresas y para el diseño de políticas de apoyo a determinadas industrias. En resumen, la desagregación trimestral permitiría identificar patrones cíclicos intra-anales en el comportamiento de los diferentes indicadores de demografía empresarial.

El tipo de información primaria (base de datos) que se propone utilizar, esto es, el número de establecimientos empresariales registrados en la Seguridad Social, permite pensar en su extrapolación para otras regiones y países. Estudios previos han considerado el uso de los registros de la Seguridad Social como fuente de información para el cálculo de indicadores de demografía empresarial (Bartelsman, 2005; Duque *et al.*, 2011). Sin embargo, son pocos los detalles que se entregan sobre la parte procedimental de dichos cálculos.

El objetivo será explicar el procedimiento a seguir para poder obtener, con la máxima fiabilidad posible, datos sobre diferentes conceptos de demografía empresarial, como altas y bajas, a un nivel de detalle sectorial equivalente al de la información publicada por el DIRCE, para las cuatro provincias de Cataluña³. Esta información se pretende obtener desagregando tanto en función del tamaño empresarial (medido como el número de trabajadores de la empresa), como por condición jurídica (según titularidad de la propiedad de la empresa). Para alcanzar tales objetivos, se tendrá que

³ Cataluña es una Comunidad Autónoma dentro de España, que constituye un territorio de nivel NUTS II. La provincia es un nivel NUTS III.

avanzar en la resolución de problemas como: proponer una definición de altas, bajas y otros conceptos de demografía empresarial a nivel trimestral, conseguir cuadrar la información trimestral estimada con los datos anuales que finalmente publique de manera oficial el DIRCE y, finalmente, conciliar las estimaciones por condición jurídica y por tamaño empresarial.

Desde el punto de vista académico merece destacarse la propuesta de definición estacional de las altas y bajas empresariales; el proceso bietápico de estimación inicial de las estimaciones trimestrales (con respecto a las anuales oficiales), avalado por su evaluación *ex-post*, y el proceso posterior de revisión de los datos trimestrales cuando se conoce el dato anual; así como el modelo de optimización sugerido para la conciliación de datos por condición jurídica *versus* tamaño empresarial. El diseño propuesto ha sido implantado en una etapa inicial por el Gobierno Regional de Cataluña⁴.

Sin duda, la disponibilidad de la información sobre demografía empresarial mencionada introduce un conocimiento sobre la situación empresarial de un territorio que va más allá de su análisis coyuntural (van Dijk y Pellenbarg, 1999; van Wissen, 2002). Su disponibilidad permitiría a instituciones públicas y privadas analizar la formación o desaparición de nuevos *clusters* empresariales, relacionándolo con los momentos del ciclo económico. También podría servir, por ejemplo, para conocer los efectos de la instalación de una gran superficie comercial, sobre el resto de establecimientos comerciales de distintos tamaños o de establecimientos complementarios al nuevo implantado; o, en línea con los programas existentes en EEUU, la delimitación espacial de zonas de comercio históricamente subutilizada (HUB Zones, por sus siglas en inglés) a través de las cuales se busca incentivar la creación de pequeñas empresas que generen plazas de trabajo destinadas, en su mayoría, al fomento del empleo para las personas que habitan dichas zonas⁵.

El artículo se estructura de la siguiente manera. En la siguiente sección se comenta brevemente cuáles son los datos oficiales de estructura empresarial de Cataluña, que publican el DIRCE (Directorio Central de Empresas), la Demografía Armonizada de Empresas y el Idescat (Instituto de Estadística de Cataluña), y se presenta la base de datos utilizada para la realización del trabajo. En la sección tercera se definen los indicadores de estructura y demografía empresarial que se utilizarán, haciendo especial mención al hecho de trabajar con información trimestral en lugar de anual, que es la situación habitual. La sección cuarta está dedicada a explicar el desarrollo metodológico que ha permitido obtener indicadores de estructura y demografía empresarial a nivel trimestral, compatibles con la información oficial. Finalmente, en la última sección, se presentan las principales conclusiones y se comentan las oportunidades que el trabajo abre para futuras líneas de investigación.

⁴ Véase <http://www20.gencat.cat/portal/site/observatoritreball/menuitem.39202212e4b9b7c298740d63b0c0e1a0/?vgnextoid=d43fd247538af110VgnVCM1000000b0c1e0aRCRD&vgnnextchannel=d43fd247538af110VgnVCM1000000b0c1e0aRCRD&vgnnextfmt=default>.

⁵ Para mayor información sobre este programa, véase «Small Business Reauthorization Act of 1997», *Public Law*, 2 de diciembre de 1997, 105-135..

2. Base de datos

El Instituto Nacional de Estadística (INE) recopila y sistematiza en el Directorio Central de Empresas (DIRCE) información básica (identificación, localización en el territorio y clasificación por dimensión y sector de actividad económica) de las empresas y establecimientos ubicados en el territorio español, a partir de fuentes administrativas y ficheros estadísticos, complementados con información procedente de las propias operaciones del INE. La información publicada por el DIRCE corresponde al número de empresas, unidades locales y lo que denomina movimientos DIRCE.

Las principales fuentes del DIRCE son: i) de origen tributario: impuesto de actividades económicas (IAE), retenciones sobre rentas del trabajo personal (RTP), impuesto sobre el valor añadido (IVA), impuesto de sociedades (IS) e impuesto sobre la renta de las personas físicas (IRPF); ii) de origen social: cuentas de cotización de la Seguridad Social; iii) de origen jurídico: registro mercantil, empresas públicas de la intervención general del Estado, y iv) estadísticas: operaciones estadísticas corrientes. La información que proporcionan estas fuentes es de tipo anual y la referencia concreta temporal de los datos es el 1 de enero del año correspondiente, ya que la práctica totalidad de la información proviene de declaraciones o inscripciones activas en esta fecha. No obstante, estas fuentes tienen diversas limitaciones que hacen que el DIRCE cubra todas las actividades económicas excepto la producción agraria y pesquera, los servicios administrativos de la administración central, autonómica y local (incluida la Seguridad Social), las actividades de las comunidades de propietarios, las actividades de los hogares y los organismos extraterritoriales.

Asimismo, en el DIRCE se registran dos tipos de unidades: empresa y local. La empresa se define como una organización sometida a una autoridad rectora que puede ser, según los casos, una persona física, una persona jurídica o una combinación de ambas y que está constituida con miras a ejercer, en uno o diversos lugares, una o diversas actividades de producción de bienes y servicios. Por otra parte, el local o establecimiento corresponde a una empresa o a una parte de ésta, situada en una ubicación geográfica concreta y desde el que se ejercen actividades económicas por cuenta de la empresa.

La información que proporciona el DIRCE para las Comunidades Autónomas corresponde al número de locales y empresas por estrato de asalariados. Por su parte, el Idescat da, para Cataluña, información por establecimientos (concepto análogo al local del DIRCE) y empresas que tienen algún establecimiento en Cataluña, así como empresas con sede social en Cataluña. Esta información se proporciona por estrato de asalariados y en el caso de empresas, además, por condición jurídica. Pero en ningún caso se proporciona información sobre demografía empresarial para Cataluña. Adicionalmente, y únicamente para el conjunto español, el DIRCE proporciona información de movimientos, que corresponden a altas, permanencias y bajas de empresas.

Respecto a la agrupación sectorial, el DIRCE publica información de las empresas por Comunidad Autónoma por grupos de la CNAE (a tres dígitos de desagregación).

Utilizando como punto de partida la información del DIRCE, el INE proporciona, mediante la Demografía Armonizada de Empresas, información relativa a la población de empresas implantadas en el territorio español, tomando en consideración el dinamismo empresarial. Para ello utiliza una metodología que ha sido acordada en el seno de la Unión Europea para que todos los Estados miembros aporten información sobre demografía empresarial con periodicidad anual. Por tanto, dicha metodología se desarrolla para calcular indicadores a nivel de Estado. Esto hace que la utilización de una metodología análoga en las Comunidades Autónomas dé lugar a indicadores que no serían consistentes con el global estatal. En este sentido, un cambio de emplazamiento de una empresa de una comunidad a otra podría suponer un alta de empresa en una Comunidad Autónoma y una baja en otra, mientras que a nivel estatal no se produciría ningún movimiento. No obstante, en la Demografía Armonizada de Empresas se realiza una desagregación de la información estatal y se proporciona, para las Comunidades Autónomas, información sobre tres conceptos: *stock*, nacimientos y muertes de empresas⁶.

A pesar de que los datos del INE incluyen parte de la información que se pretende obtener con la metodología presentada en este artículo, existen diferencias importantes. Por una parte, tanto la información oficial publicada por el DIRCE como la de la Demografía Armonizada de Empresas, y también la del Idescat, tienen carácter anual y sólo la de la Demografía Armonizada de Empresas presenta datos de demografía empresarial. Por otra parte, la información contenida en el DIRCE toma como periodo de referencia el 1 de enero del año correspondiente. De esta manera, si una empresa se encuentra activa el 1 de enero se incluye en la contabilización del número de empresas del año, mientras que si el 1 de enero no está activa no se considera. En la Demografía Armonizada se considera como el *stock* de empresas de un año a aquellas que han permanecido activas durante todo el año o parte del mismo. De esta manera, se contabilizan todas las empresas que en un momento u otro han estado activas a lo largo del año, sin tener en cuenta si en una fecha concreta estaban activas o no. Esta diferente apreciación hace que la información del número de empresas que hay a partir del DIRCE no corresponda con el *stock* de empresas según la Demografía Armonizada de Empresas, tal y como se puede comprobar en la siguiente tabla:

Tabla 1. Número de empresas. Año 2009

<i>DIRCE</i>		<i>Demografía Armonizada de Empresas</i>	
España	Cataluña	España	Cataluña
3.355.830	619.624	3.631.232	672.327

Fuente: DIRCE y Demografía Armonizada de Empresas.

En este entorno, nuestro objetivo es generar indicadores de demografía empresarial con una periodicidad trimestral, por lo que la base de datos utilizada debía

⁶ La información de demografía armonizada de empresas publicada hasta hoy para el conjunto español corresponde a los años 2008 y 2009. Sin embargo, para las Comunidades Autónomas únicamente se tiene disponible la del año 2009.

adaptarse a esta circunstancia. Los registros de la Seguridad Social cumplen con este requisito y, adicionalmente, es posible disponer de los datos con un desfase de tiempo reducido respecto al periodo de referencia. Por tanto, la información de base a partir de la que se obtienen los indicadores de estructura y demografía empresarial para Cataluña son las cuentas de cotización de la Seguridad Social, tanto del régimen general como del de autónomos (RGSS y RETA). Los ficheros que se utilizan tienen una periodicidad trimestral y han sido facilitados por el Departamento de Trabajo de la Generalitat de Cataluña.

Asimismo, la información oficial de referencia que utilizamos es la que publica el DIRCE y no la de la Demografía Empresarial de Empresas. La decisión de utilizar la información del DIRCE se basa fundamentalmente en que al presentar una metodología para obtener la demografía empresarial trimestral compatible con la información oficial del INE, se debe imponer que el número de empresas trimestrales de un trimestre (IV trimestre del año) sea equivalente al número de empresas «oficiales» en un periodo concreto. Esta equivalencia tiene sentido realizarla con la información del DIRCE, que corresponde al 1 de enero, mientras que no es adecuado compatibilizar la información trimestral con la de la Demografía Armonizada de Empresas, ya que ésta no se refiere a las empresas existentes en una fecha concreta, sino a las que en algún momento a lo largo del año han estado activas.

En cuanto a la distinción entre empresas con sede social en el territorio y empresas con actividad en el territorio, la información de la Seguridad Social corresponde a registros (o establecimientos) con actividad en el territorio. El análisis del presente trabajo se hace a partir de esta información y, por tanto, el objetivo es obtener instrumentos para calcular indicadores de estructura y demografía empresarial de las empresas que tienen actividad en Cataluña.

Respecto a las fechas de la información, los datos del DIRCE, y también del Idescat, son anuales y se refieren al 1 de enero de ese año (que es equivalente al 31 de diciembre del año anterior). Esta información se publica en el tercer o cuarto trimestre del año. Debido a que en el presente trabajo se utiliza información de las cotizaciones a la Seguridad Social como base de datos para hacer los análisis correspondientes, se considera que los registros de la Seguridad Social de los cuartos trimestres son análogos a lo que el DIRCE y el Idescat asignan como dato para el año siguiente. Por tanto, por ejemplo, en el tratamiento del fichero de la Seguridad Social del año 2008, los datos de 31 de diciembre tendrían su analogía con el dato anual del DIRCE y del Idescat del año 2009.

3. Definiciones

La estructura empresarial de una área económica/geográfica se obtiene a partir del recuento del número de empresas que desarrollan su actividad en aquella área o bien que tienen su sede social. En este sentido, la información de base que se utiliza en este estudio, que proviene de la Seguridad Social, corresponde a las empresas que tienen actividad en Cataluña. Además, su periodicidad es trimestral y, por tanto,

se utiliza el trimestre como periodo de referencia para todos los conceptos. Como usualmente el periodo de referencia de los indicadores de estructura y demografía empresarial es anual, ha sido necesario hacer una adaptación de las definiciones de demografía empresarial a una periodicidad trimestral. Los criterios establecidos han dado lugar a las definiciones de estructura y demografía empresarial que se muestran en la tabla 2:

Tabla 2. Definiciones de estructura y demografía empresarial

<i>Concepto</i>	<i>Descripción</i>
Empresas del trimestre t	Número de empresas que desarrollan actividad en Cataluña en el trimestre t .
Altas del trimestre t	Empresas que no estaban activas en el trimestre $(t-1)$ pero que sí aparecen como activas en el trimestre t .
Altas puras del trimestre t	Empresas que no estaban activas en el trimestre $(t-1)$ pero que sí aparecen como activas en el trimestre t , y que, además, no estuvieron activas en ninguno de los ocho trimestres anteriores.
Altas por reactivación del trimestre t	Empresas que no estaban activas en el trimestre $(t-1)$ pero sí aparecen como activas en el trimestre t , y que, además, estuvieron activas en alguno de los ocho trimestres anteriores.
Bajas del trimestre t	Empresas que no están activas en el trimestre t pero que sí aparecían como activas en el trimestre $(t-1)$.
Permanencias del trimestre t	Empresas que están activas en el trimestre t y que, además, también aparecían como activas en el trimestre $(t-1)$.
Tasa Bruta de Entrada del trimestre t	Número de altas de empresas en el trimestre t , respecto al total de empresas existentes el trimestre anterior. $TBE_t = \frac{\text{Altas}_t}{\text{Empresas}_{t-1}}$
Tasa Bruta de Salida del trimestre t	Número de bajas de empresas del trimestre, respecto al total existente en el trimestre anterior. $TBS_t = \frac{\text{Bajas}_t}{\text{Empresas}_{t-1}}$
Tasa Neta de Entrada	Diferencia entre tasa de entrada y tasa de salida. $TNE_t = \frac{\text{Altas}_t - \text{Bajas}_t}{\text{Empresas}_{t-1}}$
Rotación empresarial	Altas más bajas de empresas en un trimestre en relación al total de empresas activas en el mismo trimestre. $RE_t = \frac{\text{Altas}_t + \text{Bajas}_t}{\text{Empresas}_t}$

Tabla 2. (continuación)

<i>Concepto</i>	<i>Descripción</i>
Función de Supervivencia	Probabilidad de que una empresa que ha iniciado su actividad en el periodo t continúe activa en los periodos siguientes. Para su cálculo, se ha de hacer un seguimiento individualizado de las empresas que han iniciado su actividad en un periodo —que se toma como base— y el número de éstas que continúan activas en los periodos siguientes. En este sentido, se pueden definir diversas funciones de supervivencia según los periodos base considerados y el número de periodos incluidos.
Función de Riesgo	Probabilidad de que una empresa que ha comenzado su actividad en un periodo t cierre en el periodo siguiente. Para su cálculo, también se debe realizar un seguimiento individualizado de las empresas que han iniciado su actividad en un periodo t y el número de éstas que suponen una baja en el periodo siguiente, $t+1$. De esta manera, para el primer periodo analizado, las funciones de riesgo y supervivencia son complementarias. A continuación, se hace el seguimiento individualizado de las empresas que continúan activas (de las que habían iniciado la actividad en el periodo t) y se comprueba qué parte se da de baja en el periodo $t+2$. Esto supone que en el segundo periodo cambia la base del cálculo de la función de riesgo, y ya no es el número de empresas que se dieron de alta en el periodo t sino el número de empresas dadas de alta en el periodo t que en el periodo $t+1$ continuaban activas. Así, a partir del segundo periodo analizado, la función de riesgo ya no es complementaria a la función de supervivencia. De esta manera, se pueden definir diversas funciones de riesgo según el periodo origen considerado y el número de periodos adelante a analizar.

Con respecto a las dos últimas funciones, la de supervivencia y la de riesgo, se acostumbra a calcular utilizando información anual. El hecho de utilizar información trimestral supone un problema metodológico de adaptación de las definiciones para datos anuales a una periodicidad trimestral. Así, para ambas funciones se consideran como altas únicamente aquellas empresas consideradas como altas puras de un trimestre. Se ha definido una alta pura como aquella empresa que no estaba activa en el trimestre $(t-1)$ pero que sí aparecía como activa en el trimestre t , y que, además, no estaba activa en ninguno de los ocho trimestres anteriores. También se debe tener en cuenta que, cuando se considera información trimestral, el cálculo de las funciones de supervivencia y riesgo está influido por la estacionalidad. Por este motivo se calculan dos tipos de funciones: i) funciones de riesgo y supervivencia calculadas de un trimestre respecto a los trimestres siguientes (a 1, 2, 3, ... trimestres vista), y ii) funciones de riesgo y supervivencia calculadas de un trimestre respecto a los mismos trimestres de los años siguientes. Esta segunda función, a diferencia de la primera, permite eliminar, en parte, la estacionalidad.

4. Metodología

En esta sección se presenta el diseño metodológico que permite estimar (y se ha aplicado para la Comunidad Autónoma de Cataluña) los indicadores de demo-

grafía empresarial, y que incluye: puesta a punto de los datos de entrada, cálculo de los indicadores de demografía a partir de los registros de la Seguridad Social, transformación de los resultados de registros de la Seguridad Social a número de empresas y armonización de los resultados con los datos oficiales publicados por el DIRCE.

4.1. Lectura y depuración de la base de datos

En esta etapa se leen y se ponen a punto los registros que serán utilizados en el cálculo de los diferentes índices de demografía empresarial. Para cada trimestre se realizan tres operaciones básicas: *a)* lectura de archivos RGSS y RETA; *b)* clasificación de cada registro de acuerdo a las tres dimensiones: estrato de asalariados, condición jurídica y actividad económica, y *c)* unificación de registros RGSS y RETA.

Las tablas 3, 4 y 5 resumen los criterios de clasificación utilizados en cada una de las dimensiones analizadas. Las variables dimensión y condición jurídica se clasifican en cinco y cuatro estratos respectivamente. Con respecto a la estratificación por actividad económica, se utilizó la variable CCAE (Clasificación Catalana de Actividades Económicas), disponible en las bases de datos RGSS y RETA, para agrupar las actividades de tal forma que coincidieran con la clasificación sectorial utilizada por el Instituto de Estadística de Cataluña.

Por último, las bases de datos RGSS y RETA son unidas en una única base de datos. Este proceso se realiza en dos etapas:

- Conciliación de registros con un mismo Identificador de Persona Física (IPF) dentro de cada base de datos RGSS y RETA: en este caso, se procede a la unificación de estos registros en un único registro donde la variable «número de trabajadores» es agregada, y las variables «estrato de asalariados» y «condición jurídica» son tomadas del registro con el mayor número de trabajadores. En caso de existir empate en el número de trabajadores, se tomarán los valores del registro que corresponda a la provincia con mayor importancia económica de acuerdo con el siguiente orden: Barcelona, Tarragona, Girona y Lleida. Por último, en caso de que persista la igualdad entre registros, se tomarán los valores del registro con mayor importancia en la actividad económica siguiendo el siguiente orden: Id CCAE-1993: 21, 23, 24, 16, 20, 35, 31, 18, 27, 17, 29, 19, 33, 15, 9, 32, 30, 3, 28, 26, 5, 2, 34, 14, 4, 10, 22, 12, 8, 7, 6, 25, 11, 13, 1; Id CCAE-2009: 23, 36, 34, 22, 27, 20, 35, 24, 38, 18, 28, 29, 2, 31, 9, 37, 32, 21, 6, 25, 33, 12, 3, 5, 26, 11, 16, 7, 10, 30, 19, 8, 13, 14, 17, 4, 15, 1.

Las anteriores condiciones fueron suficientes para evitar cualquier conflicto en el proceso de puesta a punto de los datos.

- Eliminación de duplicados entre las bases de datos RGSS y RETA: una vez que cada una de las bases de datos, RGSS y RETA, ha sido puesta a punto, se

procede a la unificación de dichas bases en una única base de datos. En este último paso es posible encontrar autónomos con asalariados, lo cual implica que estas personas se encuentran registradas en ambas bases de datos. Para su identificación comparamos el campo IPF de la base de datos RETA con el campo IPF de la base de datos RGSS calculado como la concatenación de la variable «tipo de empresario» con la variable NIF cuyo tipo de documento sea DNI, Pasaporte o NIE. En este caso se mantiene la información de la base de datos RETA pero modificando las variables «número de trabajadores» y «estrato de dimensión» a partir de los valores registrados en la base de datos RGSS. Una vez se realiza dicho ajuste, se procede a eliminar el registro de la base de datos RGSS.

Tabla 3. Criterios utilizados para la clasificación de registros por estrato de asalariado

<i>Variable disponible</i>	<i>Estratos definidos</i>
Número de trabajadores	Sin asalariados
	Entre 1 y 9 asalariados
	Entre 10 y 49 asalariados
	Entre 50 y 199 asalariados
	Más de 200 asalariados

Tabla 4. Criterios utilizados para la clasificación de registros por condición jurídica

<i>Tipo de documento</i>	<i>Estratos definidos</i>
DNI, Pasaporte o NIE	Personas físicas
CIF comenzado con «A»	Sociedad Anónima
CIF comenzado con «B»	Sociedad Limitada
Demás códigos CIF	Otras formas jurídicas

4.2. Cálculo de indicadores de demografía a partir de los registros de la Seguridad Social

En este módulo se realizan tres tipos de cálculos: *a)* tablas de estructura empresarial, *b)* indicadores de demografía empresarial, y *c)* indicadores de supervivencia y riesgo. Al igual que el DIRCE, los resultados de estos cálculos se presentan en dos

Tabla 5. Criterios utilizados para la clasificación de registros por condición jurídica

<i>Id</i>	<i>CCAE – 1993</i>	<i>Actividad económica</i>	<i>Id</i>	<i>CCAE - 2009</i>	<i>Actividad económica</i>
1	10, 11, 12, 13, 14, 23, 40, 41	Industrias extractivas, petróleo y energía	1	05, 06, 07, 08, 09	Industrias extractivas
2	15, 16	Alimentación, bebidas y tabaco	2	10, 11, 12	Alimentación, bebidas y tabaco
3	17, 18, 19	Textil, confección, cuero y calzado	3	13, 14, 15	Textil, confección, cuero y calzado
4	20	Industrias de la madera y corcho	4	16	Ind. de madera y corcho, exc. muebles; cestería y espartería
5	21, 22	Papel, edición, artes gráficas y reprografía	5	17, 18	Papel, artes gráficas y soportes grabados
6	24	Industrias químicas	6	19, 20, 21	Refino de petróleo, ind. química y farmacéutica
7	25	Caucho y materias plásticas	7	22	Fabricación productos caucho y materias plásticas
8	26	Otros productos minerales no metálicos	8	23	Fabricación otros productos minerales no metálicos
9	27, 28	Metalurgia y productos metálicos	9	24, 25	Metalurgia y productos metálicos
10	29	Maquinaria y equipos mecánicos	10	26, 27	Fabricación de productos informáticos, electrónicos, ópticos y eléctricos
11	30, 33	Máquinas de oficina e instrumentos	11	28	Fabricación maquinaria y equipos ncop
12	31, 32	Equipos eléctricos y electrónicos	12	29, 30	Fabricación materiales de transporte
13	34, 35	Fabricación materiales de transporte	13	31, 32	Muebles y otras industrias manufactureras
14	36, 37	Industrias manufactureras diversas	14	33	Reparación e instalación de maquinaria y equipos
15	451, 455	Preparación obras, alquiler equipos construcción	15	35	Energía eléctrica, gas, vapor y aire acondicionado
16	452	Construcción inmuebles y obras ingeniería civil	16	36, 37, 38, 39	Agua, saneamiento y gestión de residuos
17	453	Instalaciones de edificios y obras	17	411	Promoción inmobiliaria
18	454	Acabado de edificios y obras	18	412	Construcción de edificios
19	50	Venta y reparación de vehículos de motor	19	42	Ingeniería civil

<i>Id</i>	<i>CCAE - 1993</i>	<i>Actividad económica</i>	<i>Id</i>	<i>CCAE - 2009</i>	<i>Actividad económica</i>
20	51	Comercio al por mayor, exc. vehículos motor	20	43	Actividades de construcción especializada
21	52	Comercio al por menor, exc. vehículos motor; reparaciones	21	45	Venta y reparación vehículos de motor y motocicletas
22	551, 552	Hoteles, camping y otros alojamientos	22	46	Comercio al por mayor e intermediarios, exc. vehículos
23	553, 554, 555	Restauración	23	47	Comercio al por menor, exc. vehículos motor
24	60, 61, 62, 63	Transportes y almacenaje	24	49	Transporte terrestre y por tubería
25	64	Correos y telecomunicaciones	25	50, 51, 52, 53	Transp. marítimo, aéreo, almacenamiento y correos
26	65, 66, 67	Mediación financiera	26	55	Servicios de alojamiento
27	70	Actividades inmobiliarias	27	56	Servicios de comidas y bebidas
28	71, 72, 73	Act. de alquiler, informáticas e I+D	28	58, 59, 60, 61, 62, 63	Información y comunicaciones
29	741	Act. jurídicas, consultoría y gestión empresarial	29	64, 65, 66	Actividades financieras y seguros
30	742, 743	Servicios técnicos	30	68	Actividades inmobiliarias
31	744, 745, 746, 747, 748	Otros servicios empresariales	31	69, 70	Act. jurídicas, consultoría y gestión empresarial
32	80	Educación	32	71, 72	Servicios técnicos e I+D
33	85	Act. sanitarias y veterinarias, servicios sociales	33	73, 74, 75	Publicidad y otras activ. profesionales y técnicas
34	90, 91	Actividades de saneamiento público y asociativas	34	77, 78, 79, 80, 81, 82	Actividades administrativas y servicios auxiliares
35	92, 93	Actividades recreativas y de servicios personales	35	85	Educación
			36	86, 87, 88	Actividades sanitarias y de servicios sociales
			37	90, 91, 92, 93	Act. artísticas, recreativas y de entretenimiento
			38	94, 95, 96	Otros servicios

tipos de tablas de doble entrada que cruzan el sector de actividad económica con el estrato de asalariados, y el sector de actividad económica con el tipo de condición jurídica.

Gracias a la etapa de puesta a punto y unificación de registros, la generación de los resultados de estructura no conlleva mayores complicaciones técnicas. Pero los cálculos de los índices de demografía y las funciones de riesgo y supervivencia implican el tratamiento simultáneo de grandes volúmenes de información que rápidamente sobrepasan los límites de muchos programas⁷. Para abordar esta problemática, se ha utilizado álgebra de conjuntos para simplificar los cálculos. Así, los códigos IPF en cada trimestre se asignan a un conjunto determinado. En esta operación no se tiene en cuenta ninguna otra variable de los registros, únicamente su código IPF. Con estos conjuntos, la determinación de las empresas que nacen, mueren o sobreviven se reduce a una serie de operaciones básicas de unión, resta e intersección de conjuntos. Una vez se tienen los resultados de las operaciones de conjuntos, se extrae la información de las variables de clasificación iterando únicamente sobre los registros de dichos resultados. Este procedimiento reduce significativamente el número de iteraciones requeridas bajo una programación de tipo procedimental.

4.3. Paso de demografía según Seguridad Social a demografía según DIRCE

Presentación y evaluación de metodologías

No existe un procedimiento que permita, a partir de los datos de registros de la Seguridad Social, obtener datos de empresas. Uno de los objetivos de este artículo consiste en establecer dicho procedimiento para lograr una buena aproximación del número de empresas que permita construir los informes de estructura y demografía empresarial a nivel trimestral para Cataluña.

La información oficial de las empresas de Cataluña, que publica el Idescat a partir de la información del DIRCE, corresponde al número de empresas que tienen su sede social en Cataluña y al de empresas que desarrollan actividades en Cataluña. Pero ni el DIRCE ni el Idescat muestran información sobre la demografía empresarial.

En esta sección se presentan y analizan diversas metodologías para aproximar la estructura y demografía empresarial en Cataluña, tomando como base de datos los resultados de estructura y demografía que se han obtenido previamente y que están referidos al tratamiento de los registros de la Seguridad Social. Concretamente se

⁷ Un trimestre promedio involucra 276.439 registros RGSS y 583.587 registros RETA. Para el cálculo de, por ejemplo, las altas puras, es necesario cargar en memoria un total de nueve trimestres, lo cual implica que en algunos casos se estarían realizando operaciones con aproximadamente 7,7 millones de registros. Para el manejo eficiente de tal volumen de información se ha utilizado Python, un lenguaje de programación de código abierto y gratuito creado a finales de la década de los ochenta por Guido Van Rossum y cuyo manejo automático de memoria permite el procesamiento de grandes volúmenes de información con unos requerimientos mínimos de memoria.

proponen y analizan las estrategias definidas como: *Bottom-up*, *Top-down*, *Top-down para grandes sectores* y *Bottom-up de la media de los pesos*.

La metodología *Bottom-up* para aproximar el número de empresas consiste en obtener un peso del número de empresas oficiales a nivel sectorial según el Idescat (DIRCE) sobre el número de registros de la Seguridad Social. Una vez obtenido este peso, se aplica esta proporción al número de registros de la Seguridad Social del periodo siguiente para obtener una aproximación al número de empresas de cada sector y variable de tabulación (estrato de asalariados y condición jurídica). Posteriormente se agregan los resultados sectoriales para obtener los totales.

Así, a nivel de cada uno de los sectores de actividad considerados, se calcula el peso que tienen las empresas sobre los registros de la Seguridad Social en un periodo determinado:

$$\frac{\text{Empresas}_{it}}{\text{Registros SS}_{it}} = \text{Peso}_{it}$$

donde i corresponde al sector de actividad y t corresponde al periodo.

Se obtiene un peso para cada uno de los sectores, que se aplica a continuación a los registros del periodo siguiente:

$$\text{Peso}_{it} \times \text{Registros SS}_{i(t+1)} = \text{Aprox. Empresas}_{i(t+1)}.$$

Una vez que se obtiene una aproximación al número de empresas a nivel sectorial, haciendo las agregaciones correspondientes, se obtienen los totales.

La metodología *Top-down* para aproximar el número de empresas consiste en obtener un peso del número de empresas oficiales totales, según el Idescat, sobre el número de registros de la Seguridad Social. Una vez obtenido este peso, se aplica esta proporción al número de registros del periodo siguiente para obtener una aproximación al número de empresas. A continuación, este volumen total de empresas se distribuye sectorialmente y según las variables de tabulación (estrato de asalariados y condición jurídica) de manera proporcional según el peso de los sectores. De esta manera, a nivel del conjunto de empresas se calcula el peso que tienen las empresas sobre los registros de la Seguridad Social en un periodo determinado:

$$\frac{\text{Empresas}_t}{\text{Registros SS}_t} = \text{Peso}_t,$$

donde t corresponde al periodo.

Se obtiene un peso que relaciona el número de empresas con el total de registros. Este peso se aplica a continuación al número de registros total del periodo siguiente:

$$\text{Peso}_t \times \text{Registros SS}_{(t+1)} = \text{Aprox. Empresas}_{(t+1)}.$$

El resultado obtenido corresponde a la aproximación del número de empresas total. A continuación este total de empresas se distribuye sectorialmente según el peso de cada sector en el total. Así, se obtiene:

$$\text{Aprox. Empresas}_{(t+1)} \times \left(\frac{\text{Empresas}_{it}}{\text{Empresas}_t} \right) = \text{Aprox. Empresas}_{i(t+1)}$$

La metodología *Top-down para grandes sectores* consiste en calcular un peso del número de empresas oficiales de los sectores de la industria, de la construcción y de los servicios sobre el número de registros de la Seguridad Social. Una vez obtenido este peso, se aplica dicha proporción al número de registros del periodo siguiente para tener una aproximación al número de empresas. A continuación, este volumen total de empresas de los tres grandes sectores se distribuye para los niveles inferiores de sectores y según las variables de tabulación (estrato de asalariados y condición jurídica) de manera proporcional según el peso de dichos sectores. El total agregado de empresas se obtiene como agregación de estos tres grandes sectores. Esta metodología es muy parecida a la *top-down*. La única diferencia es que se utilizan tres totales diferentes (uno para cada gran sector) y no el total general.

La metodología propuesta denominada *Bottom-up de la media de los pesos* para aproximar el número de empresas es, en una primera etapa, igual que la metodología *Bottom-up*. Es decir, en primer lugar se obtiene un peso del número de empresas oficiales a nivel sectorial sobre el número de registros de la Seguridad Social. Este peso se obtiene para todos los periodos disponibles. Una vez calculados estos pesos, en una segunda etapa, se calcula la media de esta proporción para todos los periodos, obteniendo como resultado una media de los pesos de los sectores para cada variable de tabulación (estrato de asalariados y condición jurídica). Esta media de los pesos se aplica al número de registros de todos los periodos para obtener una aproximación al número de empresas de cada sector y variable de tabulación. Posteriormente se agregan los resultados sectoriales para obtener los totales.

Así, a nivel del sector de actividad considerado, se calcula el peso que tienen las empresas sobre los registros de la Seguridad Social en un periodo determinado:

$$\frac{\text{Empresas}_{it}}{\text{Registros SS}_{it}} = \text{Peso}_{it},$$

donde i corresponde al sector de actividad y t corresponde al periodo.

Se obtiene un peso de cada sector y , a continuación, se hace la media para todos los periodos disponibles:

$$\text{Peso Medio}_i = \frac{\sum_{t=1}^T \text{Peso}_{it}}{T}.$$

Y esta proporción se aplica a los registros de todos los periodos:

$$\text{Peso Medio}_i \times \text{Registros SS}_{it} = \text{Aprox. Empresas}_{it}$$

Una vez obtenida una aproximación al número de empresas a nivel sectorial, haciendo las agregaciones correspondientes, se calculan los totales.

Cuando se aplica cualquiera de las metodologías presentadas, los resultados no corresponden a números enteros, sino que hay decimales. Como no puede haber un número decimal de empresas, el resultado se redondea al número entero más próximo.

Estas metodologías han sido evaluadas mediante su aplicación para periodos para los que previamente ya se conocía el número de empresas y el cálculo de los porcentajes de error cometidos. En las tablas 6 y 7 se muestran resultados de esta evaluación para el total de empresas según las dos tabulaciones consideradas.

Tabla 6. Porcentaje de desviación entre el número de empresas y la estimación realizada. Estrato de asalariados

	<i>Sin asalariados</i>	<i>De 1 a 9 asalariados</i>	<i>De 10 a 49 asalariados</i>	<i>De 50 a 199 asalariados</i>	<i>Más de 200 asalariados</i>	<i>Total</i>
<i>Bottom-up</i>						
2007	-1,0%	-2,1%	-0,7%	-0,7%	-0,3%	-1,4%
2008	2,0%	-1,6%	-0,3%	-0,2%	-1,4%	0,3%
<i>Top-down</i>						
2007	-2,3%	0,4%	2,7%	0,4%	0,4%	-0,9%
2008	-0,3%	0,0%	1,8%	-1,3%	-4,2%	-0,1%
<i>Top-down grandes sectores</i>						
2007	-1,1%	-1,9%	-0,7%	-0,9%	-0,2%	-0,9%
2008	-0,3%	-0,1%	1,7%	-1,9%	-5,0%	-0,1%
<i>Bottom-up media de pesos</i>						
2008	1,5%	-2,6%	-0,7%	-0,5%	-1,8%	-0,3%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y datos del Idescat.

Como puede comprobarse en las tablas, si el objetivo fuese únicamente estimar el total de empresas, la metodología que funciona mejor es la *Top-down*, ya que con ella se obtienen los porcentajes de error más pequeños. Pero se pretende utilizar la metodología que dé mejores resultados a todos los niveles requeridos, es decir, para los diferentes tramos de asalariados (y condición jurídica) y para la diferente desagregación sectorial⁸. Así, si se considera la desagregación sectorial y los diferentes tramos de asalariados y condición jurídica, es complejo analizar cuál es la mejor metodología. Lo que sí es cierto es que tanto la aproximación *Bottom-up* como la *Top-down* proporcionan resultados aceptables ya que los porcentajes de error son relativamente bajos.

⁸ En el anexo se muestran los porcentajes de error para los tres grandes sectores: industria, construcción y servicios, para las dos variables de tabulación.

Tabla 7. Porcentaje de desviación entre el número de empresas y la estimación realizada. Condición jurídica

	<i>Personas Físicas</i>	<i>Sociedades Anónimas</i>	<i>Sociedades Limitadas</i>	<i>Otras</i>	<i>Total</i>
<i>Bottom-up</i>					
2007	1,2%	-2,1%	-4,5%	-7,5%	-1,6%
2008	4,4%	-0,3%	-3,3%	-11,1%	0,0%
<i>Top-down</i>					
2007	1,5%	6,6%	-4,8%	-4,7%	-0,9%
2008	2,9%	5,2%	-3,8%	-5,3%	-0,1%
<i>Top-down grandes sectores</i>					
2007	2,2%	-1,3%	-5,1%	-7,4%	-0,9%
2008	3,3%	-0,8%	-3,1%	-10,2%	-0,1%
<i>Bottom-up media de pesos</i>					
2008	5,1%	-1,4%	-5,5%	-14,2%	-0,7%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y datos del Idescat.

Finalmente, se ha tomado la decisión de utilizar la metodología *Bottom-up* por dos motivos: 1) por sus buenos resultados prácticos, y 2) por sus características. Como el objetivo es obtener estimaciones para todos los sectores considerados y tabulaciones, la metodología *Bottom-up* tiene la característica que utiliza toda la información más reciente, es decir, la de los registros de la Seguridad Social en el periodo para el que se está haciendo la estimación para todos los sectores y tabulaciones consideradas. Con la metodología *Top-down* se estaría utilizando la información más reciente, la de los registros de la Seguridad Social, únicamente para los totales, mientras que para hacer el reparto sectorial y según las diferentes tabulaciones se estaría utilizando la distribución presentada por el Idescat (DIRCE), pero correspondiente al año anterior. De esta manera, si los resultados son buenos, como es el caso, es mejor considerar toda la riqueza informativa que contienen los registros de la Seguridad Social en el periodo más actual.

Asimismo, debe valorarse muy positivamente el hecho de que los errores cometidos mediante la metodología *Bottom-Up* son inferiores al 5% en el total de Cataluña y el total sectorial, y también para la mayoría de estratos de asalariados y condición jurídica.

Conciliación de resultados según variables de tabulación

Es evidente que la suma de empresas que hay según el número de asalariados debe ser igual a la suma de empresas según la condición jurídica, y que esta igualdad debe cumplirse para cualquier sector. Sin embargo, cuando se aplica la metodología finalmente escogida para estimar el número de empresas, metodología *Bottom-up*, no hay una completa coincidencia entre los resultados obtenidos para la tabulación de estrato de asalariados y los correspondientes a la de forma jurídica. No obstante, es necesario que para todos los sectores considerados, el número de empresas totales sea igual con independencia de la tabulación considerada. Para solucionar esta divergencia se procede a realizar una conciliación entre ambos resultados mediante la siguiente optimización:

Parámetros:

- a : Índice de actividad económica.
- d : Índice de estrato de dimensión.
- c : Índice de tipo de condición jurídica.
- $D_{a,d}$: Número de empresas en la actividad económica a del estrato de dimensión d .
- $C_{a,c}$: Número de empresas en la actividad económica a del tipo de condición jurídica c .

Variables de decisión:

- $FD_{a,d}$: Factor de ajuste aplicado al número de empresas en la actividad económica a del estrato de dimensión d .
- $FC_{a,c}$: Factor de ajuste aplicado al número de empresas en la actividad económica a del tipo de condición jurídica c .

Para cada actividad económica α :

$$\text{Minimizar } Z_a = \sum_d |1 - FD_{a,d}| + \sum_c |1 - FC_{a,c}| + \quad (1)$$

$$\sum_d D_{a,d} \times FD_{a,d} - \sum_c C_{a,c} \times FC_{a,c} = 0 \quad (2)$$

$$FD_{a,d} \geq 0 \quad \forall d \quad (3)$$

$$FC_{a,c} \geq 0 \quad \forall c \quad (4)$$

La función objetivo (1) busca minimizar los cambios realizados en los valores originales de número de empresas, $D_{a,d}$ y $C_{a,c}$. Dicho efecto se logra minimizando la desviación absoluta de cada factor con respecto a la unidad. El conjunto de restricciones (2) asegura que para cada actividad económica a , la suma de empresas por estrato de dimensión d coincide con la suma de empresas por condición jurídica c . El conjunto de restricciones (3) y (4) exige que el valor de los factores sea positivo para

asegurar la integridad de la solución. En caso de que se desee un mayor grado de restricción sobre los posibles valores de las variables de decisión, es posible reemplazar los conjuntos de restricciones (3) y (4) por restricciones del tipo $I \leq FD_{a,d} \leq S \forall d$ y $I \leq FC_{a,c} \leq S \forall c$, donde I y S son los límites inferior y superior deseado. Dichos límites no se deben fijar muy cerca a la unidad, pues se podrían generar infactibilidades. A partir de experimentos previos, sugerimos $I = 0,8$ y $S = 1,2$.

Armonización de los resultados con nuevos datos oficiales: informes provisionales y definitivos

La información de la Seguridad Social y la que publica el Idescat difieren en varios aspectos, dos de los cuales son de especial relevancia para la aplicación de nuestra metodología. En primer lugar, la información de la Seguridad Social tiene una periodicidad trimestral, mientras que la del Idescat es anual. Y en segundo lugar, los datos de la Seguridad Social se tienen disponibles con una diferencia de tiempo con el periodo de referencia muy breve, mientras que el tiempo transcurrido entre la información publicada por el Idescat y el periodo de referencia es considerablemente más elevado.

Estas cuestiones provocan que se proponga calcular dos tipos de informes/resultados:

- Informes provisionales: son aquellos que se obtienen cuando el Idescat todavía no ha publicado los datos de referencia anuales correspondientes. Estos informes tienen validez hasta que el Idescat publica el dato anual de la estructura empresarial.
- Informes definitivos: son aquellos que se obtienen cuando el Idescat ya ha publicado los datos de referencia anuales correspondientes. Estos informes definitivos sustituyen a los informes provisionales.

Así por ejemplo, a finales del año 2009 se pudo obtener $\alpha_{2009} = \frac{EI_{2009}}{ES_{2008.IV}}$, donde EI_{2009} corresponde al número de empresas según el Idescat (dato oficial) y $ES_{2008.IV}$ es el número de empresas (registros) calculadas según la Seguridad Social. Una vez obtenido el valor de α_{2009} se obtienen las estimaciones provisionales para el cuarto trimestre del año 2009⁹ y para los tres primeros de 2010. Sin embargo, a finales del año 2010 se puede obtener $\alpha_{2010} = \frac{EI_{2010}}{ES_{2009.IV}}$ y con él se calculan las estimaciones definitivas del cuarto trimestre del año 2009 y de los tres primeros de 2010. Cabe destacar que operando de esta manera se garantiza que la estimación definitiva del cuarto trimestre coincida exactamente con el valor publicado oficialmente.

⁹ La estimación del número de empresas de un cuarto trimestre tiene su equivalencia con el dato anual del año siguiente, ya que la información oficial recogida por el DIRCE y el Idescat corresponde con información referida al 1 de enero de ese año.

En la tabla 8 se resumen las estimaciones provisionales (EI^{pr}) y definitivas (EI^{def}) que se pueden obtener desde finales del año 2009 hasta finales del año 2010.

Tabla 8. Ejemplos de los periodos en los que se obtienen las estimaciones provisionales y definitivas

<i>Estimación provisional</i>	<i>Estimación definitiva</i>
$EI_{2009.IV}^{pr} = \alpha_{2009} \times ES_{2009.IV} = EI_{2010}^{pr}$ $EI_{2010.I}^{pr} = \alpha_{2009} \times ES_{2010.I}$ $EI_{2010.II}^{pr} = \alpha_{2009} \times ES_{2010.II}$ $EI_{2010.III}^{pr} = \alpha_{2009} \times ES_{2010.III}$ donde $\alpha_{2009} = \frac{EI_{2009.IV}}{ES_{2008.IV}}$ que se puede obtener a finales del año 2009.	$EI_{2009.IV}^{def} = \alpha_{2010} \times ES_{2009.IV} = EI_{2010}^{def} = EI_{2010}$ $EI_{2010.I}^{def} = \alpha_{2010} \times ES_{2010.I}$ $EI_{2010.II}^{def} = \alpha_{2010} \times ES_{2010.II}$ $EI_{2010.III}^{def} = \alpha_{2010} \times ES_{2010.III}$ donde $\alpha_{2010} = \frac{EI_{2010.IV}}{ES_{2009.IV}}$ que se puede obtener a finales del año 2010.
$EI_{2010.IV}^{pr} = \alpha_{2010} \times ES_{2010.IV} = EI_{2011}^{pr}$ $EI_{2011.I}^{pr} = \alpha_{2010} \times ES_{2011.I}$ donde $\alpha_{2010} = \frac{EI_{2010.IV}}{ES_{2009.IV}}$ que se puede obtener a finales del año 2010.	Se podrá calcular cuando el Idescat publique los datos del número de empresas del año 2011.

Incumplimiento de algunas igualdades

Respecto a los resultados de demografía empresarial calculados en el presente artículo, se deben realizar algunas consideraciones:

1. La igualdad $Altas_t + Permanencias_t = Total_t$, se cumple para muchos sectores. Pero en algún caso concreto hay pequeñas diferencias que afectan como mucho a tres o cuatro registros. Estas diferencias deben atribuirse a pequeñas inconsistencias en la base de datos, y son consideradas como descuadras menores.
2. La igualdad $Total_{t-1} + Altas_t - Bajas_t = Total_t$, para el total de sectores (teniendo en cuenta los no clasificados) se cumple. Pero si en lugar de considerar el total de registros se consideran las agrupaciones por sectores, estrato de asalariados y condición jurídica, esta igualdad no acostumbra a cumplirse. Este hecho se debe atribuir a las reclasificaciones que se producen en un periodo respecto al anterior. De esta manera, cuando hay una alta, alta por reactivación o permanencia, la clasificación en el sector, estrato de asalariados y condición jurídica depende de lo que aparezca en el registro en el

periodo t con independencia de lo que constaba en el periodo $t-1$. Por tanto, es posible que se hayan producido reclasificaciones, en el sentido que haya una variación en el estrato de asalariados, condición jurídica o actividad de un periodo respecto al periodo anterior. Siguiendo las prácticas del DIRCE, esto no supone una baja en una categoría y una alta en otra categoría, sino que es una permanencia, que es clasificada según la información del registro t . Este hecho hace que la igualdad mencionada anteriormente no se cumpla y, además, hace imposible establecer algún procedimiento para que se cumpla. En este sentido, se debe remarcar que en los datos que el DIRCE publica para el conjunto del Estado español esta igualdad tampoco se cumple. A título de ejemplo, según los datos del DIRCE, en el año 2009 había en España un total de 3.355.830 empresas. Pero si se obtiene la cifra a partir de las empresas que había en el año 2008 más las altas y menos las bajas que hubo en el año 2009, el resultado es de 3.358.082. La diferencia entre ambas supone una desviación del 0,07%. Para Cataluña, según la información de la Seguridad Social, el número de empresas en la Seguridad Social era en el cuarto trimestre de 2008 de 716.451. Pero si se hace el cálculo a partir del número de empresas del tercer trimestre de 2008 más las altas y menos las bajas que hubo en el cuarto trimestre, el resultado es de 716.647. La diferencia entre ambas supone, en este caso, una desviación del 0,03% del total (sin tener en cuenta las empresas no clasificadas).

3. El cálculo de las tasas brutas de entrada (TBE), salida (TBS) y las tasas netas de entrada (TNE) implican la utilización de información de dos periodos diferentes.

$$\text{Así } TBE_t = \frac{\text{Altas}_t}{\text{Empresas}_{t-1}}, TBS_t = \frac{\text{Bajas}_t}{\text{Empresas}_{t-1}} \text{ y } TNE_t = \frac{\text{Altas}_t - \text{Bajas}_t}{\text{Empresas}_{t-1}}$$

Este hecho supone que para su cálculo se utilizan las altas y/o las bajas de un periodo y el número de empresas del periodo anterior. Pero, en el momento en que se calculan las TBE, TBS, TNE, algunas (no muchas) empresas del denominador ($t-1$) han cambiado de condición jurídica, o de estrato de asalariados o incluso de actividad económica (ha pasado un trimestre) en el periodo t . Por este motivo, si se calculan estas tasas utilizando en el denominador los valores de las tablas de estructura (que se generan un trimestre antes) los resultados no son iguales que si se hace el cálculo a partir de todos los resultados obtenidos en el trimestre actual. Nosotros realizamos el cálculo a partir de la información más reciente posible. Por ejemplo, supongamos que estamos en septiembre de 2008 y se generan las tablas de estructura empresarial (entre otras). Tres meses después (diciembre de 2008) se generan otras tablas, entre ellas las TBE, TBS, TNE. Para su cálculo es necesario el listado de identificadores de las empresas que están en el periodo $t-1$. A partir de este identificador, se clasifica cada empresa por su sector, tramo de asalariados y condición jurídica, utilizando la información más actualizada que hay en ese momento. Obsérvese que tener en cuenta esto es diferente a tomar los resultados de estructura que se generan en septiembre de 2008, ya que han pasado tres meses en los cuales algunas empresas han cambia-

do su clasificación. Lo que se hace es generar en diciembre de 2008 una versión actualizada de la tabla de estructura de septiembre de 2008: el total de empresas es el mismo, pero su distribución en la tabla cambia un poco. No obstante, para el total de empresas (sin distinguir condición jurídica, estrato de asalariados ni actividad económica, pero teniendo en cuenta las empresas no clasificadas) el resultado no se ve modificado. De esta manera, las diferencias que se dan en los diversos sectores y condición jurídica y tramos de asalariados en las TBE, TBS y TNE son análogas a las que se producen al hacer $Total_{t-1} + Altas_t - Bajas_t = Total_t$. Es decir, son producidas por las reclasificaciones.

5. Conclusiones

El conjunto de información referida a la estructura empresarial de un área geográfica y de su evolución a lo largo del tiempo es un elemento básico para realizar un seguimiento de la actividad económica que vaya más allá de la mera descripción de una situación. Poder dar respuesta a preguntas como: ¿cuántas empresas realizan sus actividades en un sector en concreto?, ¿cuáles son sus características en cuanto a tamaño o condición jurídica?, ¿cuántas empresas han surgido en un periodo determinado?, ¿cuántas han desaparecido?, ¿cuál es la probabilidad de que una empresa que ha surgido en un periodo continúe o cese en su actividad en periodos posteriores?, etc., es crucial no sólo para aumentar el conocimiento que se tiene de la actividad económica de una zona sino para diseñar actuaciones y planes estratégicos que permitan orientar/potenciar la economía en determinados sectores.

En este contexto, el artículo desarrollado se ha dirigido a diseñar una metodología para obtener indicadores de estructura y demografía empresarial a nivel sectorial para la Comunidad Autónoma de Cataluña (y también para sus cuatro provincias) a partir de los registros de la afiliación a la Seguridad Social.

La metodología propuesta permite su extensión al resto de Comunidades Autónomas españolas y viene a cubrir un vacío en la disponibilidad de información de indicadores de demografía empresarial a nivel de Comunidades Autónomas, ya que la fuente oficial de información, el DIRCE, únicamente proporciona datos de movimientos para el conjunto español. Asimismo, cabe destacar que el DIRCE proporciona información con una periodicidad anual, mientras que nuestro trabajo, realizado a partir de información de la Seguridad Social, permite su obtención a nivel trimestral, hecho que facilita un seguimiento más coyuntural.

La disponibilidad de esta información regional-temporal, como se ha explicado en el artículo, permitirá un mejor conocimiento de la demografía empresarial. Dicho conocimiento, a su vez, facilitará la toma de decisiones en ámbitos como el industrial o comercial, en términos de política industrial, fiscal, laboral, entre otras.

6. Referencias

- Bartelsman, E.; Scarpetta, S., y Schivardi, F. (2005): «Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries», *Industrial and Corporate Change*, 14(3), 365-391.
- Duque, J. C.; García, J. R., y Suriñach, J. (2011): *Propuesta metodológica para la desagregación espacio-temporal de los indicadores de demografía empresarial*, XIV Encuentro de Economía Aplicada, Huelva.
- Hannan, M. T., y Freeman, J. (1977): «The population ecology of organizations», *American Journal of Sociology*, 82(5): 929-964.
- (1989): *Organizational Ecology*. Cambridge, MA: Harvard University Press.
- INE (Instituto Nacional de Estadística) (2008): *Demografía armonizada de empresas: Nota metodológica*, <http://www.ine.es/metodologia/t37/t3730204.pdf>.
- Langtangen, H. (2006): *Python Scripting for Computational Science*, Springer.
- Manjón, M.; Segarra, A.; Martín, M., y Arauzo, J. M. (2002): «Demografía industrial y convergencia regional en España», *Papeles de Economía Española*, 93, 65-78.
- Segarra, A.; Arauzo, J. M.; Gras, N.; Manjón, M.; Mañe, F.; Teruel, M., y Theilen, B. (2002): *La creación y la supervivencia de las empresas industriales*, Madrid. Civitas.
- Van Dijk, J., y Pellenbarg, P. H. (1999): *The demography of firms: progress and problems in empirical research*, en van Dijk y Pellenbarg (eds.) (1999), 325-337A.
- (2000): «Spatial perspectives on firm demography», *Papers in Regional Science*, 79(2), 107-110.
- Van Wissen, L. J. G. (2002): «Demography of the Firm: A Useful Metaphor?», *European Journal of Population/Revue Européenne de Démographie*, 18(3), 263-279.

Anexo 1

Se presentan, a continuación, los resultados de la evaluación de las diferentes metodologías para estimar el número de empresas para el total y para los tres grandes sectores según las dos tabulaciones consideradas, estrato de asalariados y condición jurídica.

Por estrato de asalariados

Tabla A.1. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Bottom-up*. Estrato de asalariados

		<i>Sin asalariados</i>	<i>De 1 a 9 asalariados</i>	<i>De 10 a 49 asalariados</i>	<i>De 50 a 199 asalariados</i>	<i>Más de 200 asalariados</i>	<i>Total</i>
2007	Industria	-4,4%	-0,6%	1,2%	-0,1%	1,2%	-1,5%
	Construcción	-0,7%	-0,9%	-1,4%	-3,8%	29,4%	-0,8%
	Servicios	-0,8%	-2,5%	-1,3%	-0,5%	-2,5%	-1,5%
	Total	-1,0%	-2,1%	-0,7%	-0,7%	-0,3%	-1,4%
2008	Industria	-1,4%	1,8%	0,8%	4,3%	4,5%	0,7%
	Construcción	10,1%	-0,5%	-0,1%	-4,8%	-21,5%	4,8%
	Servicios	0,8%	-2,2%	-0,8%	-1,4%	-2,6%	-0,5%
	Total	2,0%	-1,6%	-0,3%	-0,2%	-1,4%	0,3%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.2. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Top-down*. Estrato de asalariados

		<i>Sin asalariados</i>	<i>De 1 a 9 asalariados</i>	<i>De 10 a 49 asalariados</i>	<i>De 50 a 199 asalariados</i>	<i>Más de 200 asalariados</i>	<i>Total</i>
2007	Industria	0,6%	7,1%	9,0%	5,8%	7,6%	5,3%
	Construcción	-5,9%	-2,6%	0,2%	-4,0%	-3,7%	-4,1%
	Servicios	-1,9%	0,1%	1,0%	-1,5%	-2,4%	-0,9%
	Total	-2,3%	0,4%	2,7%	0,4%	0,4%	-0,9%
2008	Industria	0,5%	5,3%	4,1%	4,0%	1,4%	3,5%
	Construcción	1,7%	-1,4%	3,5%	0,0%	-6,6%	0,5%
	Servicios	-0,7%	-0,3%	0,4%	-3,7%	-6,2%	-0,6%
	Total	-0,3%	0,0%	1,8%	-1,3%	-4,2%	-0,1%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.3. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Top-down* para grandes sectores. Estrato de asalariados

		<i>Sin asalariados</i>	<i>De 1 a 9 asalariados</i>	<i>De 10 a 49 asalariados</i>	<i>De 50 a 199 asalariados</i>	<i>Más de 200 asalariados</i>	<i>Total</i>
2007	Industria	-5,6%	0,4%	2,2%	-0,8%	0,9%	-1,3%
	Construcción	11,3%	0,6%	-1,6%	-4,1%	19,5%	8,3%
	Servicios	-3,0%	-2,7%	-1,5%	-0,4%	-1,8%	-2,6%
	Total	-1,1%	-1,9%	-0,7%	-0,9%	-0,2%	-0,9%
2008	Industria	-2,0%	2,7%	1,6%	1,4%	-1,1%	0,9%
	Construcción	5,2%	2,1%	7,1%	3,5%	-3,4%	4,0%
	Servicios	-1,2%	-0,8%	0,0%	-4,1%	-6,6%	-1,0%
	Total	-0,3%	-0,1%	1,7%	-1,9%	-5,0%	-0,1%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.4. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Bottom-up* de la media de los pesos. Estrato de asalariados

		<i>Sin asalariados</i>	<i>De 1 a 9 asalariados</i>	<i>De 10 a 49 asalariados</i>	<i>De 50 a 199 asalariados</i>	<i>Más de 200 asalariados</i>	<i>Total</i>
2008	Industria	-3,6%	1,5%	1,5%	4,2%	4,8%	0,0%
	Construcción	9,7%	-1,0%	-0,8%	-6,7%	-9,4%	4,3%
	Servicios	0,4%	-3,4%	-1,5%	-1,6%	-3,9%	-1,2%
	Total	1,5%	-2,6%	-0,7%	-0,5%	-1,8%	-0,3%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Por condición jurídica**Tabla A.5.** Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Bottom-up*. Condición jurídica

		<i>Personas físicas</i>	<i>Sociedades Anónimas</i>	<i>Sociedades Limitadas</i>	<i>Otras</i>	<i>Total</i>
2007	Industria	0,9%	-0,9%	-2,3%	-4,8%	-1,2%
	Construcción	1,1%	-3,1%	-2,7%	-7,3%	-1,1%
	Servicios	1,3%	-2,3%	-5,1%	-7,8%	-1,8%
	Total	1,2%	-2,1%	-4,5%	-7,5%	-1,6%
2008	Industria	3,1%	0,1%	1,0%	-2,0%	1,3%
	Construcción	11,6%	-1,2%	-0,9%	-7,7%	4,9%
	Servicios	3,1%	-0,4%	-4,4%	-12,6%	-1,1%
	Total	4,4%	-0,3%	-3,3%	-11,1%	0,0%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.6. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Top-down*. Condición jurídica

		<i>Personas físicas</i>	<i>Sociedades Anónimas</i>	<i>Sociedades Limitadas</i>	<i>Otras</i>	<i>Total</i>
2007	Industria	7,3%	9,5%	2,6%	3,8%	5,3%
	Construcción	-3,7%	4,1%	-4,4%	-6,7%	-4,1%
	Servicios	2,1%	5,7%	-5,9%	-5,1%	-0,9%
	Total	1,5%	6,6%	-4,8%	-4,7%	-0,9%
2008	Industria	6,2%	7,4%	0,7%	1,1%	3,5%
	Construcción	5,3%	3,8%	-4,7%	-6,6%	0,5%
	Servicios	2,2%	4,5%	-4,3%	-5,5%	-0,6%
	Total	2,9%	5,2%	-3,8%	-5,3%	-0,1%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.7. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Top-Down* para grandes sectores.

Condición jurídica

		<i>Personas físicas</i>	<i>Sociedades Anónimas</i>	<i>Sociedades Limitadas</i>	<i>Otras</i>	<i>Total</i>
2007	Industria	0,7%	2,7%	-3,8%	-2,6%	-1,3%
	Construcción	12,4%	-3,3%	-3,0%	-7,7%	8,3%
	Servicios	0,4%	-2,5%	-5,7%	-7,7%	-2,6%
	Total	2,2%	-1,3%	-5,1%	-7,4%	-0,9%
2008	Industria	3,4%	-0,1%	0,9%	-1,7%	0,9%
	Construcción	9,3%	-2,0%	-1,2%	-8,4%	4,0%
	Servicios	2,1%	-0,9%	-4,0%	-11,3%	-1,0%
	Total	3,3%	-0,8%	-3,1%	-10,2%	-0,1%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.

Tabla A.8. Porcentaje de desviación entre el número de empresas y la estimación realizada mediante la metodología *Bottom-up* de la media dos pesos.

Condición jurídica

		<i>Personas físicas</i>	<i>Sociedades Anónimas</i>	<i>Sociedades Limitadas</i>	<i>Otras</i>	<i>Total</i>
2008	Industria	3,6%	-0,4%	-0,1%	-4,3%	0,7%
	Construcción	12,2%	-2,8%	-2,2%	-11,1%	4,3%
	Servicios	3,8%	-1,6%	-6,9%	-15,6%	-1,8%
	Total	5,1%	-1,4%	-5,5%	-14,2%	-0,7%

Fuente: Elaboración propia a partir de los registros de la Seguridad Social y los datos del Idescat.