

PROPUESTA METODOLÓGICA PARA LA DESAGREGACIÓN ESPACIO- TEMPORAL DE LOS INDICADORES DE DEMOGRAFÍA EMPRESARIAL.

Resumen:

En esta comunicación se presenta un diseño metodológico para obtener, para Cataluña (extensible a otras Comunidades Autónomas), indicadores sobre la demografía empresarial a nivel sectorial, desagregando tanto en función del tamaño empresarial, como según la condición jurídica de las empresas. 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. Para su implementación se utiliza Python, un lenguaje de programación de código abierto que permite el procesamiento de grandes volúmenes de información con unos requerimientos mínimos de memoria.

Clasificación JEL R12 · R30 · C8

Duque, Juan C.

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

Garcia, José R.

AQR Research Group-IREA.
Universidad de Barcelona, Avda Diagonal, 690, Barcelona, 08034, Spain
e-mail: jrgarcia@ub.edu

Suriñach, J.¹

AQR Research Group-IREA.
Universidad de Barcelona, Avda Diagonal, 690, Barcelona, 08034, Spain
e-mail: [jsurinach@ub.edu](mailto:jurinach@ub.edu)

¹ J. Suriñach agradece la financiación recibida de ECO2009-12678/ECON.

1.- INTRODUCCIÓN²

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 y 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 (Van Wissen, 2002). Su utilización como seguimiento de la actividad empresarial y económica es menor por la dificultad de disponibilidad de la misma. Anualmente, en España se publica dicha información, a nivel estatal, por parte del INE (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 (INE, 2008). No obstante, a nivel autonómico, el INE únicamente publica el stock de empresas existentes, sin entrar en la dinámica de las altas y bajas que se producen.

En este artículo, se presenta el diseño metodológico seguido para obtener, para la Comunidad Autónoma de Cataluña, información sobre la demografía empresarial, a nivel trimestral. De todos modos, 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.

El objetivo será explicar el procedimiento a seguir para poder obtener con la máxima fiabilidad posible, datos sobre diferentes conceptos de demografía empresarial a un nivel de

² Los autores quieren agradecer al Departament de Treball de la Generalitat de Catalunya la ayuda aportada para la realización de esta investigación, aunque la responsabilidad es única de los autores.

detalle sectorial, para las 4 provincias de Cataluña. Esta información se pretende obtener desagregando tanto en función del tamaño empresarial (en función del número de trabajadores de la empresa), como por condición jurídica (según titularidad de la empresa). Para alcanzar tales objetivos, se tendrá que avanzar en la resolución de problemas como: proponer una definición de altas, bajas y otros conceptos de demografía empresarial a nivel trimestral, conseguir ajustar la información trimestral estimada con los datos anuales que finalmente publique de manera oficial el INE, conciliar las estimaciones provinciales por condición jurídica y por tamaño empresarial, y, especialmente, utilizar un software adecuado que permita alcanzar todos estos objetivos, dada la elevada cantidad de datos a utilizar en tal proceso. Debe destacarse que, para la estimación trimestral-regional propuesta, se requiere trabajar simultáneamente con una base de datos que debe incluir, para realizar alguna de las operaciones, más de 7 millones de registros, cada uno de ellos con múltiples campos. Esto da idea de la gran cantidad de datos que han de tratarse simultáneamente. Es por ello, que la aproximación propuesta está basada en el software Python, en la línea de otros trabajos ya realizados para el Defence Logistic Agency, en EEUU, en los que ha colaborado alguno de los autores de este artículo, supone uno de los principales valores añadidos del mismo (Duque y Rey, 2008). Desde el punto de vista académico, además de este sistema de tratamiento de grandes cantidades de información, 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 algoritmo sugerido de conciliación de datos por condición jurídica versus tamaño empresarial.

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. Asimismo, tiene una clara relevancia para los policy makers (Van Dijk y Pellenburg, 2000a,b). Su disponibilidad permitiría, por ejemplo, (a) avanzar en el conocimiento de las zonas donde se concentran las bajas empresariales puede ser informativo para proceder a introducir medidas fiscales que las dinamicen; (b) permitir analizar la formación o desaparición

de nuevos clusters empresariales, relacionándolo con los momentos del ciclo económico; (c) conocer los efectos la instalación de una gran superficie comercial, sobre el resto de establecimientos comerciales de distintos tamaños o de establecimientos complementarios al nuevo implantado; entre otras.

El diseño propuesto ha sido implantado en una etapa inicial por el Gobierno Regional de Cataluña³, estando previsto en el futuro ampliar y mejorar los informes que se derivan de la obtención de dicha Base de datos regional-trimestral de demografía empresarial.

La comunicación se estructura de la siguiente manera. En la siguiente sección se comenta brevemente cuales son los datos oficiales de estructura empresarial de Cataluña, que publican el DIRCE (Directorio Central 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. A continuación, se explican los retos técnicos del cálculo y desagregación espacio-temporal de los conceptos de demografía empresarial. La sección quinta está dedicada a explicar el desarrollo metodológico. Finalmente, en la última sección se presentan las principales conclusiones centradas en las oportunidades que el trabajo abre para futuras líneas de investigación y los beneficios derivados de la utilización del programa Python.

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

³ Véase <http://www20.gencat.cat/portal/site/observatoritreball/menuitem.39202212e4b9b7c298740d63b0c0e1a0/?vgnextoid=d43fd247538af110VgnVCM1000000b0c1e0aRCRD&vgnextchannel=d43fd247538af110VgnVCM1000000b0c1e0aRCRD&vgnextfmt=default>

información publicada por el DIRCE corresponde al número de empresas, unidades locales y movimientos, utilizando como fuentes principales: (1) 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), impuesto sobre la renta de las personas físicas (IRPF); (2) de origen social: cuentas de cotización de la Seguridad Social; (3) de origen jurídico: registro mercantil, empresas públicas de la intervención general del Estado; y (4) 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 i 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 la 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 Instituto de Estadística de Cataluña) 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 (3 dígitos de desagregación).

Cabe destacar que la información oficial publicada por el DIRCE tiene carácter anual y en ningún caso presenta datos de estructura y demografía empresarial con una periodicidad inferior al año. Nuestro objetivo es generar indicadores de demografía empresarial con una periodicidad trimestral, por lo que la base de datos utilizada debía 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.

Por 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 a continuación:

1. **Empresas del trimestre t:** número de empresas que desarrollan actividad en Cataluña en el trimestre t.
2. **Altas del trimestre t:** empresas que no estaban activas en el trimestre (t-1) pero que sí aparecen como activas en el trimestre t.

Las altas pueden ser de dos tipos:

- a) **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 8 trimestres anteriores.
 - b) **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 8 trimestres anteriores.
3. **Bajas del trimestre t:** empresas que no están activas en el trimestre t pero que sí aparecían como activas en el trimestre (t-1).

4. **Permanencias del trimestre t:** empresas que están activas en el trimestre t y que, además, también aparecían como activas en el trimestre (t-1).
5. **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.

$$TBE_t = \frac{\text{Altas}_t}{\text{Empresas}_{t-1}}$$

6. **Tasa Bruta de Salida del trimestre t:** número de bajas de empresas del trimestre, respecto al total existente en el trimestre anterior.

$$TBS_t = \frac{\text{Bajas}_t}{\text{Empresas}_{t-1}}$$

7. **Tasa Neta de Entrada:** diferencia entre tasa bruta de entrada y tasa bruta de salida.

$$TNE_t = \frac{\text{Altas}_t - \text{Bajas}_t}{\text{Empresas}_{t-1}}$$

8. **Rotación empresarial:** altas más bajas de empresas en un trimestre en relación al total de empresas activas en el mismo trimestre.

$$RE_t = \frac{\text{Altas}_t + \text{Bajas}_t}{\text{Empresas}_t}$$

9. **Función de Supervivencia:** en términos generales, una función de supervivencia presenta la probabilidad de que una empresa que ha iniciado su actividad en el periodo t continúe activa en los periodos siguientes. Por tanto, 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.
10. **Función de Riesgo:** una función de riesgo presenta la 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. Esto hace que, a partir del segundo periodo analizado, la función de riesgo ya no sea 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.

Estas dos funciones, la de supervivencia y la de riesgo, se acostumbran 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. En este sentido, se procede de la siguiente manera:

- Para ambas funciones -supervivencia y riesgo- 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 8 trimestres anteriores.
- 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:
 - a) Funciones de riesgo i supervivencia calculadas de un trimestre respecto a los trimestres siguientes (a 1, 2, 3,trimestres vista). Esta opción no elimina la estacionalidad.
 - b) Funciones de riesgo y supervivencia calculadas de un trimestre respecto a los mismos trimestres de los años siguientes. Esta opción elimina, en parte, la

estacionalidad.

4.- RETOS TÉCNICOS EN LA TRIMESTRALIZACIÓN DE LOS INDICADORES DE DEMOGRAFÍA EMPRESARIAL

Uno de los principales retos a la hora de afrontar una desagregación espacio-temporal de los indicadores de demografía empresarial consiste en el manejo eficiente del volumen de información requerida para la generación de cada uno de los indicadores demográficos. Para ilustrar la magnitud de la exigencia en los cálculos presentamos algunas cifras:

- Un trimestre promedio involucra 276,438 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, cada registro con múltiples variables.
- La puesta a punto de los datos requiere que cada registro sea evaluado por aproximadamente 60 condicionales tipo “SI..ENTONCES”. Si se incluyen los demás cálculos de estructura y demografía empresarial, el número de condicionales evaluados asciende a 216.

Para el manejo eficiente de tal volumen de información y operaciones se utilizó Python, un lenguaje de programación de código abierto y gratuito creado a finales de la década de los ochenta por Guido Van Rossum. Entre las principales características de Python, es su (1) manejo automático de memoria que permite el procesamiento de grandes volúmenes de información con unos requerimientos mínimos de memoria; (2) una variada tipología de datos entre las cuales destacan los conjuntos (set) y los diccionarios (dict), los cuales permiten operar y navegar en un gran volumen de datos de una forma rápida y eficiente; (3) al tratarse de un lenguaje de alto nivel, la escritura de código en Python es compacta y fácil de interpretar; (4) Python permite integrar librerías como lp_solve para la solución de problemas de optimización lineal y Numerical Python (NumPy) que posibilita el uso de complejas operaciones matemáticas (Oliphant 2006); y (5) Python puede ser ejecutado en múltiples sistemas operativos.

5.- METODOLOGÍA

En este apartado presentaremos los pasos para la generación de los informes de demografía empresarial.

5.1. Lectura y depuración de la base de datos

Esta etapa está diseñada como un módulo dentro del sistema de demografía empresarial. Dicho módulo está encargado de ingresar los datos al sistema y ponerlos a punto para su utilización en el cálculo de los diferentes índices de demografía empresarial. Para cada trimestre se realizan tres actividades 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; (c) unificación de registros RGSS y RETA.

Las tablas 1 y 2 resumen los criterios de clasificación utilizados en las variables de tabulación: estratos de asalariados y condición jurídica, que están clasificadas en cinco y cuatro estratos respectivamente. Con respecto a la estratificación por actividad económica, se utilizó la variable CCAE (Clasificación Catalana de Actividades Económicas, adaptación de la CNAE), 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 IPF (tipo de documento + documento) 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.

- 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.

Los dos casos mencionados anteriormente no son muy comunes y estos ajustes se realizan en un porcentaje muy reducido de registros.

Tabla 1. Criterios utilizados para la clasificación de registros por estrato de asalariado

Variable disponible	Estratos definidos
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 2. Criterios utilizados para la clasificación de registros por condición jurídica

Tipo de documento	Estratos definidos
DNI, Pasaporte o NIE	Personas físicas
CIF comenzado con “A”	Sociedad Anónima
CIF comenzado con “B”	Sociedad Limitada
Demás códigos CIF	Otras formas jurídicas

5.2. 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 reportan en dos 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 al módulo de lectura y depuración, la generación de los reportes de estructura no conlleva mayores complicaciones técnicas. No obstante, 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 programas. 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 conjuntos resultado, se extrae la información de las variables de clasificación iterando únicamente sobre los registros del conjunto resultado. Este procedimiento reduce significativamente el número de iteraciones requeridas bajo una programación de tipo procedimental.

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

Presentación y evaluación de metodologías

No existe un criterio que permita, a partir de los datos de registros de la Seguridad Social, obtener datos de empresas. Uno de los objetivos de este trabajo consiste en construir los procesos necesarios en un paquete estadístico de tratamiento de datos, para obtener indicadores relevantes 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 social en Cataluña o

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 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 S.S. Una vez obtenido este peso, se aplica esta proporción al número de registros de la S.S. 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:

$$\frac{\text{Empresas}_{i t}}{\text{RegistrosSS}_{i t}} = \text{Peso}_{i t} \text{ donde } i \text{ corresponde al sector de actividad y } t \text{ corresponde}$$

al periodo.

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

$$\text{Peso}_{i t} * \text{RegistrosSS}_{i(t+1)} = \text{Aprox Empresas}_{i(t+1)}.$$

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 S.S. 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:

$$\frac{\text{Empresas}_t}{\text{Registros SS}_t} = \text{Peso}_t \text{ donde } t \text{ corresponde al periodo.}$$

Y 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:

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

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:

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

La metodología *Top-down para grandes sectores* para aproximar el número de empresas consiste en obtener 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 S.S. Una vez obtenido este peso, se aplica dicha proporción al número de registros del periodo siguiente para obtener 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 S.S. Este peso se obtiene para todos los periodos disponibles. Una vez obtenidos estos pesos, en una segunda etapa, se hace 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 agrega 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:

$\frac{Empresas_{it}}{Registros\ SS_{it}} = Peso_{it}$, 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: $Peso\ Medio_i = \frac{\sum_{t=1}^T Peso_{it}}{T}$. Y esta proporción se aplica a los registros

de todos los periodos: $Peso\ Medio_i * Registros\ SS_{it} = Aprox\ Empresas_{it}$

Una vez obtenida una aproximación al número de empresas a nivel sectorial, haciendo las agregaciones correspondientes, se obtienen 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 períodos para los que previamente ya se conocía el número de empresas y el cálculo de los porcentajes de error

cometidos. En las tablas 3 y 4 se muestran resultados de esta evaluación para el total de empresas según las dos tabulaciones consideradas.

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 de mejores resultados a todos los niveles requeridos, es decir, para los diferentes tramos de asalariados (y condición jurídica) y para la diferente desagregación sectorial. Así, si se considera la desagregación sectorial y los diferentes tramos de asalariados y condición jurídica, es complejo analizar cual 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.

Tabla 3. Porcentaje de desviación entre el número de empresas y la estimación realizada.

<i>Estrato de asalariados</i>						
	<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>
Bottom-up						
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%
Top-down						
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%
Top-down grandes sectores						
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%
Bottom-up media de pesos						
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.

Tabla 4. 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>
Bottom-up					
2007	1,2%	-2,1%	-4,5%	-7,5%	-1,6%
2008	4,4%	-0,3%	-3,3%	-11,1%	0,0%
Top-down					
2007	1,5%	6,6%	-4,8%	-4,7%	-0,9%
2008	2,9%	5,2%	-3,8%	-5,3%	-0,1%
Top-down grandes sectores					
2007	2,2%	-1,3%	-5,1%	-7,4%	-0,9%
2008	3,3%	-0,8%	-3,1%	-10,2%	-0,1%
Bottom-up media de pesos					
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 a :

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

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

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

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

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 coincida 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 período 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 calculen 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 la estimaciones provisionales para el cuarto trimestre del año 2009⁴ 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.

A continuación se presenta, en la tabla 5, un resumen de la obtención de estimaciones provisionales (EI^{pr}) y definitivas (EI^{def}) que se pueden obtener desde finales del año 2009 hasta finales del año 2010.

⁴ 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 i el Idescat corresponde a información referida al 1 de enero de ese año.

Tabla 5. Ejemplos de los periodos en los que se obtienen las estimaciones provisionales y definitivas.

Estimación provisional	Estimación definitiva
$EI_{2009.IV}^{pr} = \alpha_{2009} * ES_{2009.IV} = EI_{2010}^{pr}$ $EI_{2010.I}^{pr} = \alpha_{2009} * ES_{2010.I}$ $EI_{2010.II}^{pr} = \alpha_{2009} * ES_{2010.II}$ $EI_{2010.III}^{pr} = \alpha_{2009} * ES_{2010.III}$ donde $\alpha_{2009} = \frac{EI_{2009}}{ES_{2008.IV}}$ que se puede obtener a finales del año 2009	$EI_{2009.IV}^{def} = \alpha_{2010} * ES_{2009.IV} = EI_{2010}^{def} = EI_{2010}$ $EI_{2010.I}^{def} = \alpha_{2010} * ES_{2010.I}$ $EI_{2010.II}^{def} = \alpha_{2010} * ES_{2010.II}$ $EI_{2010.III}^{def} = \alpha_{2010} * ES_{2010.III}$ donde $\alpha_{2010} = \frac{EI_{2010}}{ES_{2009.IV}}$ que se puede obtener a finales del año 2010
$EI_{2010.IV}^{pr} = \alpha_{2010} * ES_{2010.IV} = EI_{2011}^{pr}$ $EI_{2011.I}^{pr} = \alpha_{2010} * ES_{2011.I}$ donde $\alpha_{2010} = \frac{EI_{2010}}{ES_{2009.IV}}$ 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 trabajo, 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 3 ó 4 registros. Estas diferencias deben atribuirse a pequeñas inconsistencias en la base de datos, y son consideradas como descuadres muy 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 lo 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, 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 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 S.S. era el 4º trimestre de 2008 de 716.451. Pero si se hace el cálculo a partir del número de empresas del 3er trimestre de 2008 más las altas y menos las bajas que hubo el 4º 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 (N/A)).

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.

$$\text{Así } TBE_t = \frac{Altas_t}{Empresas_{t-1}}, TBS_t = \frac{Bajas_t}{Empresas_{t-1}} \text{ y } TNE_t = \frac{Altas_t - Bajas_t}{Empresas_{t-1}}.$$

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 2008-9 i se generan las tablas de estructura empresarial (entre otras). Tres meses después (2008-12) 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 2008-9, ya que han pasado tres meses en los cuales algunas empresas han cambiado su clasificación. Lo que se hace es generar en 2008-12 una versión actualizada de la tabla de estructura de 2008-9: 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 N/A) 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.

6.- 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 ha surgido en un periodo continúe o cese en su actividad en periodos posteriores? etc. es crucial para aumentar el conocimiento que se tiene de la actividad económica de una zona tanto a nivel general como a nivel sectorial. Este mayor conocimiento permite detectar dinámicas muy diversas que se pueden estar produciendo: concentración de empresas de determinados sectores, desertización económica en algún sector concreto, formación de clusters, etc. Pero además, este mayor conocimiento facilita el diseño de actuaciones y planes estratégicos que permitan orientar/potenciar la economía en determinados sectores.

En este contexto, el trabajo 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.

REFERENCIAS

- Duque, J. C. y Rey, S. J. (2008) A network based approach towards industry clustering. En Blien, U. y Maier, G., editores, *The economics of regional clusters: Networks, technology and policy* (New Horizons in Regional Science Series). Edward Elgar Publishing Ltd (1 Jun 2008), ISBN-10: 1847205151, ISBN-13: 978-1847205155.
- INE (Instituto Nacional de Estadística) (2008). Demografía armonizada de empresas: Nota metodológica. <http://www.ine.es/metodologia/t37/t3730204.pdf>.
- Oliphant, T. (2006) *Guide to NumPy*. Trelgol Publishing, USA.
- 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.
- Van Dijk, J. y Pellenbarg, P. H. (2000a). Spatial perspectives on firm demography. *Papers in Regional Science*, 79(2), 107–110.
- Van Dijk, J. y Pellenbarg, P.H. (eds) (2000b). *Demography of Firms. Spatial Dynamics of Firm Behaviour*, Nederlandse Geografische Studies 262, KNAG, Utrecht/Groningen.