Proyecto de microdatos de vidas laborales

Carmen Marcos

Instituto de Estudios Fiscales

Almudena Durán

D.G. de Ordenación Económica de la Seguridad Social (MTASS)

El término análisis del mercado de trabajo suele referirse al estudio de la oferta y la demanda en un momento dado o a la evolución de aquéllas entre dos períodos. Sin embargo, también es importante conocer cómo evoluciona la situación laboral de las personas a lo largo del tiempo. El problema está en las carencias de información para realizar estos análisis. Aunque en teoría esto podría obtenerse mediante un panel de plazo largo, la fuente lógica son los registros de Seguridad Social. Debido a que la cuantía de las pensiones está ligada a lo cotizado a lo largo de la vida, la información se guarda mucho más tiempo que en cualquier otra entidad pública o privada.

Sin embargo, la organización de la Seguridad Social y la de la información que capta y almacena están diseñadas para la gestión, no para la explotación estadística ni para el análisis y la investigación. Ello significa que proporcionar a un investigador –y son muchos los que lo demandan– datos de forma que le resulten útiles requiere la dedicación de muchas horas de trabajo y de numerosas consultas simplemente para asegurar que cada uno entiende el sentido que el otro da a los términos; además de que la recuperación de las “vidas” de cientos de miles o de millones de personas es fuertemente consumidora de recursos de todo tipo.

Por ello se están desarrollando los trabajos para el diseño y obtención de una muestra de vidas laborales anónimas que responda a las necesidades habituales de los investigadores, a la vez que se definen y documentan los conceptos y variables utilizados.

Esta muestra podrá ser utilizada para diferentes estudios, a través de convenios con las correspondientes entidades investigadoras, lo que abrirá nuevas e importantes vías para el conocimiento del funcionamiento del mercado de trabajo y algunos de sus determinantes, para el análisis sobre la equidad o la rentabilidad del sistema público de pensiones y para el estudio del impacto de determinadas políticas públicas.

Población y muestras

La **población** de referencia está constituida por *todas aquellas personas que tienen una relación con la Seguridad Social*, bien como cotizante, bien como receptor de pensiones. El objetivo es que la población sea lo más exhaustiva posible

dentro del ámbito de la Seguridad Social. Por tanto no sólo recoge los ocupados (afiliados en alta laboral) sino también los desempleados (en la medida en que, por cobrar prestaciones, cotice por ellos el Servicio Público de Empleo Estatal), los que cotizan para futuras pensiones a través de Convenios Especiales y los pensionistas. Sólo se excluyen aquellos que únicamente son beneficiarios de asistencia sanitaria. Se estima que a finales de 2004 había 25 millones de personas en alguna de estas situaciones. Al referirse exclusivamente al ámbito de la Seguridad Social no están incluidos los que tienen sistemas de protección distintos como, por ejemplo, los funcionarios de Clases Pasivas.

La **muestra general** que se va a preparar en primer lugar cubrirá aproximadamente el 4 por ciento de todos los que a lo largo de 2004 tuvieron con una relación del tipo aludido con la Seguridad Social; esto significa algo más de un millón de personas.

Se ha optado por seleccionar un período y no una fecha concreta dada la importancia de los contratos de muy corta duración; de hecho se estima que en el número de personas que habiendo tenido alguna relación laboral a lo largo del año quedarían fuera de la muestra si esta se selecciona con referencia exclusivamente a una fecha sería superior a un millón de personas. Hay que tener en cuenta además que este tipo de contratos no se distribuyen aleatoriamente entre la población.

Para la extracción de la muestra se han seleccionado aleatoriamente algunos de los dígitos de determinadas posiciones de los números de identificación de las personas que componen la población. El método de selección garantiza que la muestra es representativa para cada uno de los años en los que se tome (permitiendo análisis transversales), al tiempo que se consigue continuidad en la muestra a lo largo del tiempo. Además, la ventaja de un sistema de este tipo es que, si se repite como está previsto la extracción en algún momento anterior y posterior, las personas seleccionadas serán comunes en una proporción muy elevada, con un enorme ahorro de coste de recuperación de vidas laborales, posibilitando además analizar las entradas y salidas del sistema entre dos períodos.

De todos los individuos seleccionados se recuperará, hasta donde los registros disponibles lo permitan, la historia de su relación con la seguridad social, bien como afiliado bien como receptor de prestaciones.

Está previsto completar esta muestra inicial con la extracción de muestras similares pero con otros periodos de referencia; es decir, personas con los mismos dígitos que estuvieron en relación con la Seguridad Social en otro momento, aunque ya no lo estén. También es previsible su actualización anual.

Además se preparará una muestra de extranjeros y otra muestra de fallecidos. El especial interés del fenómeno migratorio reciente y de los aspectos ligados a su integración laboral aconseja el diseño de una muestra específica. Por su parte, la selección de una muestra de fallecidos guarda relación con el interés de disponer de historias de relaciones con el sistema de la Seguridad Social no censuradas, es decir completas. A través de las pensiones de viudedad y orfandad, las consecuencias de una vida laboral no terminan con la muerte del cotizante.

Variables

Hay que hacer constar en primer lugar que, aunque muy ricos, los registros de la Seguridad Social tienen limitaciones que pueden ser importantes. Por poner un ejemplo, un trabajador no está obligado a comunicar su nivel de estudios, su estado civil ni sus circunstancias familiares excepto en los casos en que solicita una prestación basándose en estas. Tampoco es posible para este trabajo, dada la estructura de los ficheros, anotar para cada trabajador ciertas características del conjunto de la empresa en que trabaja, si esta dispone de varias cuentas de cotización, ya que la referencia es el centro de cotización y no la empresa. Igualmente ocurre que para algunas variables no relevantes para la gestión del sistema, aunque puedan tener gran interés analítico, no se conservan los datos históricos, por lo que algunas de las variables sólo podrán venir referidas a la fecha de extracción de la muestra.

Lo anterior no es óbice para señalar la enorme potencia analítica de las variables de las que en principio se piensa disponer de información.

Se espera reflejar los siguientes grupos de variables:

- 9 relacionadas con la persona física.
- 9 relacionadas con la empresa y la cuenta a través de la que cotiza; estas se recogerán tantas veces como relaciones laborales diferentes haya tenido el individuo, hasta donde la antigüedad de los registros lo permita.
- 11 relacionadas con el tipo y características de cada relación laboral; entre estas, las relativas a las cuantías de las bases de cotización se recogerán para cada año o fracción que dure esa relación
- 13 relacionadas con cada una de las pensiones recibidas. En este caso, las relativas a la cuantía de la prestación también se recogerán para cada año que dure esta.

Además de anonimizar los códigos identificadores de personas o empresas, será preciso llevar a cabo algunas transformaciones de variables originales, fundamentalmente por razones de confidencialidad. En este sentido será la información de localización geográfica la más afectada, en el caso de los municipios que no sean de gran tamaño; también la de nacionalidad en el caso de

países con un número reducido de trabajadores en España. Así mismo hay que llegar a un compromiso entre la calidad y cobertura de la información y su interés para el análisis.

Para cada variable se está preparando una ficha que recoja metainformación útil al usuario: definiciones, origen de la información, códigos, cobertura, actualización, etc. Ello no obsta para que la interpretación adecuada de algunas variables obligue al investigador a estudiar en detalle determinados aspectos legales y/o procedimentales.

Líneas de futuro

Ya se ha avanzado que se pretende extraer muestras referidas a períodos anteriores; y que además se seleccionará una muestra de extranjeros y otra de fallecidos. Además, la idea inicial es actualizar anualmente la muestra general y la de inmigrantes, alargando la historia de los seleccionados en el período anterior e incorporando a los nuevos sujetos de relaciones con la Seguridad Social. Sin embargo, dada la dedicación de recursos que esto requiere, la continuidad estará condicionada por el éxito entre los investigadores.

“ Lo que no está previsto todavía es cómo avanzar en la línea de fusión de ficheros, que abriría enormes posibilidades para el análisis ”

Lo que no está previsto todavía es cómo avanzar en la línea de fusión de ficheros, que abriría enormes posibilidades para el análisis. Así por ejemplo, la fusión con datos padronales permitiría incorporar a los microdatos de Seguridad Social variables tan relevantes como estado civil, número de hijos o personas residentes en el mismo domicilio y su evolución en el tiempo. La fusión con datos censales permitiría la recuperación de las características del hogar y otras variables sociodemográficas muy relevantes como, por ejemplo, el régimen de tenencia de la vivienda y, en su caso, la existencia de pagos pendientes. La fusión con datos tributarios abriría enormes posibilidades en muchos campos; entre otros, para el análisis de la vida laboral y sus determinantes ligados a renta, riqueza y patrimonio.

Esta línea de trabajo y la puesta a disposición para el análisis y la investigación de los ficheros de microdatos resultantes anonimizados está ampliamente desarrollada y desde hace muchos años en muchos países: Canadá, EEUU, Dinamarca, Reino Unido, Francia, etc., mientras que, en España es todavía hoy un territorio prácticamente inexplorado que confiamos en que pronto se desarrolle.