

# **PROCESAMIENTO Y DEPURACIÓN DE LA MUESTRA CONTINUA DE VIDAS LABORALES (MCVL) PARA EL ESTUDIO DEL MERCADO LABORAL GALLEGO**

**Noa Veiguela Fernández, Pilar Romero Martínez<sup>1</sup>**

**Instituto Galego de Estatística**

## **Resumen:**

La MCVL es un conjunto de microdatos anonimizados procedentes de diversos registros administrativos: de la Seguridad Social, del Padrón Municipal de Habitantes y de la Agencia Tributaria. Constituye una muestra representativa de todas las personas que, durante el año de referencia, han mantenido relación con la Seguridad Social, bien porque estuvieron afiliadas en situación de alta laboral, o bien porque percibieron algún tipo de prestación, subsidio por desempleo, o pensión contributiva. Su gran tamaño (más de un millón de individuos para el conjunto de España en cada edición) permite realizar estudios minuciosos sobre las características del mercado laboral en Galicia, tanto de tipo transversal como de corte longitudinal. En el IGE el manejo de la MCVL nos ha permitido difundir información a la que hasta ahora no era posible acceder a través de otras fuentes, como el número de días trabajados a lo largo del año en estudio, el tiempo de permanencia en la empresa, la duración de la vida laboral, las bases de cotización y retribuciones salariales, y cruzar toda esta nueva información con variables como el sexo, la edad, el sector de actividad, el grupo de cotización, etc. Ahora bien, y debido precisamente a la gran cantidad de información que ofrece, el procesamiento de la MCVL es una tarea ardua. Además, al haber sido extraída de registros administrativos, su contenido no está exento de los problemas que conlleva este tipo de fuentes; en particular, no ha sido sometida a un proceso de depuración, como en el caso de las encuestas; en ocasiones la información no aparece cumplimentada o debidamente actualizada y, al tratarse de una amalgama de registros procedentes de distintos organismos, puede resultar contradictoria.

El objetivo de esta ponencia es presentar los procedimientos informáticos que hemos desarrollado para manejar el ingente volumen de información que constituye la MCVL. Abordaremos las dificultades a las que nos hemos enfrentado en su tratamiento y soluciones adoptadas. En particular, nos centraremos en dos de ellas:

- La delimitación de los períodos en los que cada persona seleccionada ha mantenido relación con la Seguridad Social y el tipo de relación de que se trate (de trabajo, cobro de prestación por desempleo o de pensión contributiva).
- La identificación de las distintas relaciones contractuales que ha mantenido cada sujeto a lo largo de su vida laboral.

**Palabras clave:** MCVL, vida laboral, relación contractual

---

<sup>1</sup> e-mail: [noa.veiguela.fernandez@ige.eu](mailto:noa.veiguela.fernandez@ige.eu); [pilar.romero@ige.eu](mailto:pilar.romero@ige.eu)

## **1. Presentación de la MCVL**

La Muestra Continua de Vidas Laborales (MCVL) es un conjunto de microdatos anonimizados procedentes de diversos registros administrativos. Aporta información sobre algo más de un millón de personas en cada edición, y constituye una muestra representativa de todos los individuos que, durante el año de referencia, han mantenido relación con la Seguridad Social. La MCVL es el resultado de años de trabajo por parte de los responsables en estadística de este organismo, quienes, ante la avalancha de peticiones de información que recibían cada año, decidieron elaborar una muestra representativa de los ficheros en su haber<sup>2</sup>. Puesto que la información contenida en las bases de datos de la Seguridad Social, si bien amplia, es limitada, dicho organismo decidió aunar esfuerzos con el Instituto Nacional de Estadística (INE) y la Agencia Tributaria para completar sus carencias informativas.

El resultado es una base de datos que permite reconstruir la vida laboral de los afiliados a la Seguridad Social desde 1967 (año a partir del cual se conservan registros informatizados): los episodios en los que el individuo ha estado trabajando y las características laborales de estos episodios (duración, tipo de contrato, régimen, bases de cotización, etc.), así como los períodos de desempleo. También ofrece información sobre las pensiones de carácter contributivo percibidas por los integrantes de la muestra. El cruce de los ficheros de la Seguridad Social con el Padrón municipal de habitantes del INE tiene por finalidad completar los datos personales de los seleccionados, con información relativa a su domicilio de residencia, a las personas con las que conviven, etc. Por otra parte, la información aportada por la Agencia Tributaria procede del Modelo 190, declaración anual que tienen la obligación de presentar todas las entidades que pagan salarios, pensiones o prestaciones por desempleo sujetas a IRPF. Dicho modelo constituye un resumen de las retenciones e ingresos a cuenta aplicados en el ejercicio fiscal por los pagadores a sus trabajadores y a los perceptores de prestaciones.

La primera edición de la MCVL data del año 2004 (2006 con datos fiscales) y su carácter es continuo, en el sentido de que los individuos seleccionados en la primera edición permanecen en la muestra mientras no se extinga su vínculo con la Seguridad Social. Para sustituir a los que abandonan la muestra (por haberse extinguido su vínculo con dicho organismo), cada edición de la MCVL incorpora a nuevos integrantes, manteniendo un tamaño prácticamente estable (si cabe, en ligero aumento). Además, y esta es una peculiaridad que la distingue de otras

---

<sup>2</sup> Seguridad Social (2009 a); págs. 2-3.

fuentes estadísticas, cuando un antiguo integrante retoma su vínculo con la Seguridad Social, vuelve a formar parte automáticamente de la MCVL.

Esta peculiaridad y el hecho de que incluya información de las vidas laborales de los sujetos desde 1967, la convierten en una fuente idónea para la realización de estudios de tipo longitudinal, si bien la mayor parte de las investigaciones realizadas hasta el momento son de corte transversal. Su gran tamaño también permite abordar el estudio de colectivos relativamente pequeños a los que no es posible llegar por medio de otras fuentes. En el Instituto Galego de Estatística (IGE) la edición correspondiente al año 2011 nos ha permitido realizar un estudio minucioso sobre las características laborales de la población juvenil de Galicia (población de 16 a 29 años). Aunque disponíamos de cierta información sobre este colectivo<sup>3</sup>, la MCVL nos ha permitido ahondar en su conocimiento. Por poner un ejemplo, sabemos, gracias a la Encuesta de población activa (EPA) que en Galicia hay unos 111.900 asalariados de entre 16 y 29 años, prácticamente la mitad sujetos a un contrato temporal y la otra mitad a uno de duración indefinida. Pero lo que no sabíamos hasta la explotación realizada de la MCVL es que el 40,1% de ellos lleva menos de un año trabajando para la misma empresa, o que el salario anual medio de los que llevan menos de un año contratados por la misma empresa es significativamente inferior al que perciben los que llevan cinco o más años contratados (7.527€ para los primeros frente a 16.608€ para los segundos)<sup>4</sup>.

La MCVL nos ha permitido difundir información sobre el número de días trabajados por la población juvenil gallega a lo largo del año en estudio, el tiempo de permanencia en la empresa, la duración de la vida laboral, las bases de cotización y las retribuciones salariales, y cruzar toda esta nueva información con variables como el sexo, la edad, el sector de actividad, el grupo de cotización, etc. Ahora bien, y debido precisamente a la gran cantidad de información que ofrece, el procesamiento de la MCVL es una tarea ardua. El objetivo de esta ponencia es presentar algunas de las dificultades a las que nos hemos enfrentado en su tratamiento y soluciones adoptadas.

El artículo se estructura en cuatro apartados. En el primero se describen los procedimientos informáticos aplicados para delimitar los períodos en los que cada integrante ha mantenido relación con la Seguridad Social y el tipo de relación de que se trate. Se explicarán con detalle las distintas casuísticas que se pueden presentar y se introducirán los programas diseñados para automatizar y agilizar el

---

<sup>3</sup> En 2008 el INE y el IGE firmaron un convenio de colaboración para la duplicación de la muestra de la Encuesta de población activa (EPA) en el ámbito territorial gallego. La ampliación de dicha muestra ha permitido desde entonces estudiar con mayor grado de detalle determinados colectivos de personas y su relación con la actividad económica, entre ellos el de la población juvenil.

<sup>4</sup> En la siguiente dirección web se accede a los resultados completos del estudio sobre la población juvenil gallega realizado a partir de la MCVL:  
[http://www.ige.eu/web/mostrar\\_actividade\\_estadistica.jsp?idioma=gl&codigo=0204015](http://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0204015).

proceso. En el segundo apartado se define el concepto de relación contractual y se presentan las dificultades que acarrea la identificación de las distintas relaciones contractuales de cada integrante de la MCVL. En el tercer apartado se presentan los resultados más interesantes de la explotación de la edición de 2011 sobre el colectivo juvenil de Galicia. Finalmente, en el apartado de conclusiones resaltaremos, a modo de resumen, los puntos más importantes del artículo.

Pero antes de entrar en la explicación de las tareas de explotación de la MCVL, dedicaremos unas líneas a describir los ficheros que componen la muestra. La Seguridad Social la distribuye fragmentada en archivos de texto. Dentro de los distintos ficheros, cada registro se refiere a un solo individuo y, dependiendo del fichero, habrá un sólo o varios registros por individuo. Para identificar a cada integrante de la muestra se incluye en todos ellos una variable anonimizada, basada en el NIF o código sustitutivo asignado por la Seguridad Social en el caso de que un individuo carezca de NIF (en lo sucesivo nos referiremos a ella como la variable *id\_per*<sup>5</sup>). La información personal de los integrantes de la muestra y de las personas con las que conviven se encuentra alojada, respectivamente, en los ficheros PERSONAL y CONVIVIR<sup>6</sup>. Las bases de cotización de los trabajadores contratados por cuenta ajena se encuentran en los archivos COTIZA y las de los trabajadores por cuenta propia, en COTIZA13. La información procedente del Modelo 190 de la Agencia Tributaria está en el fichero FISCAL; la vida laboral de los integrantes en los archivos AFILIAD; y la información referida al cobro de pensiones contributivas en el fichero PRESTAC.

Centrémonos en los dos últimos archivos comentados, ya que son a los que nos vamos a referir con mayor frecuencia a lo largo del artículo. El fichero AFILIAD<sup>7</sup> permite reconstruir la vida laboral de los integrantes de la muestra<sup>8</sup> a partir de las relaciones administrativas mantenidas por estos con la Seguridad Social. En esta historia se incluyen los episodios en los que el individuo ha estado trabajando y cotizando al sistema, y que corresponden a la mayor parte de los registros incluidos en el fichero, pero también las situaciones de desempleo (siempre y cuando el

---

<sup>5</sup> Los ficheros distribuidos por la Seguridad Social no incluyen cabeceras, esto es, el nombre de las variables, dejando al usuario libertad para definir los que mejor se adapten a sus necesidades. Los nombres que se presentan en este artículo son los que utilizamos en el IGE para referirnos a las distintas variables de la MCVL.

<sup>6</sup> El nombre completo de estos ficheros, al igual que los restantes, sería MCVL[AñoEdición]PERSONAL\_CDF y MCVL[AñoEdición]CONVIVIR\_CDF, pero, por comodidad, en lo sucesivo lo acortaremos a PERSONAL y CONVIVIR. Las letras CDF indican que se trata de la edición con datos fiscales. Por ejemplo, si estuviésemos trabajando con la edición correspondiente a 2011, el nombre de estos dos ficheros sería MCVL2011PERSONAL\_CDF y MCVL2011CONVIVIR\_CDF.

<sup>7</sup> Mejor dicho, el conjunto de los tres ficheros AFILIAD, ya que la Seguridad Social lo distribuye dividido en tres particiones debido a su gran tamaño.

<sup>8</sup> El fichero incluye una parte pequeña de las relaciones laborales anteriores a 1966, pero, tanto estas como las inmediatamente posteriores a esa fecha, deben analizarse con precaución pues pueden contener errores. No debe considerarse que la primera relación que figura en la muestra es necesariamente la primera relación laboral del individuo, ya que la base de datos de la que se ha extraído la muestra contiene lagunas importantes para esos años.

individuo cobrase una prestación o subsidio durante ese período). El fichero incluye un registro por cada una de las relaciones laborales distintas que haya mantenido cada sujeto en cuestión. Aunque luego ahondaremos en esta cuestión, "cada relación es lo que transcurre entre un alta y una baja en la Seguridad Social", y las altas y bajas en dicho organismo se producen bien porque el individuo cambia de empresa, o bien porque queda en el paro. Además de información referida a las características de cada relación laboral (tipo de contrato, régimen de cotización, fechas de alta y de baja, etc.), en el fichero también encontramos dos variables que nos permiten identificar al empresario o empresa en las bases de datos de la Seguridad Social y de la Agencia Tributaria (las variables *CCCSecundaria* e *id\_pagador*, sobre las que se volverá más adelante).

El fichero PRESTAC, por su parte, ofrece información sobre las pensiones de carácter contributivo percibidas desde 1996 por los individuos seleccionados en la MCVL (que es el año en el que fue creada la base de datos de la que se extrae el fichero). Hay que recalcar que ofrece únicamente información sobre pensiones contributivas, quedando fuera de este ámbito las prestaciones por desempleo (de las que disponemos de información en AFILIAD), las pensiones de carácter no contributivo y algunas prestaciones de la Seguridad Social que no suponen la ruptura del vínculo laboral entre el trabajador y la empresa. Contiene un registro por pensión y año natural de cobro de la misma, con información referida al momento en que fue reconocida la pensión, su importe y las características que influyen en su cálculo.

## **2. Delimitación y caracterización de la relación con la Seguridad Social**

La población de referencia de la MCVL incluye a todas las personas que han mantenido algún tipo de relación con la Seguridad Social; estas pueden agruparse en una de las siguientes categorías:

- Trabajadores afiliados a la Seguridad Social en alta laboral.
- Beneficiarios de prestaciones o subsidios por desempleo.
- Personas que reciben cualquier tipo de pensión contributiva de la Seguridad Social: jubilación, incapacidad permanente y supervivencia (viudedad y orfandad).
- Aquellas personas que mantienen un Convenio Especial con la Seguridad Social, según el cual continúan cotizando al sistema a pesar de no encontrarse en situación de alta laboral, con el fin de salvaguardar el derecho a percibir una pensión contributiva en el futuro (no los vamos a considerar en nuestro análisis).

Cada integrante de la muestra puede pertenecer a varias categorías distintas en el año de referencia, simultánea o sucesivamente<sup>9</sup>. Una de las primeras tareas a las que debe enfrentarse todo aquel que quiera utilizar la MCVL es la delimitación y caracterización de los distintos colectivos.

Como se mencionó en los párrafos precedentes, la vida laboral de los sujetos de la muestra se encuentra en los ficheros AFILIAD; cada fila del fichero representa un episodio laboral. Para averiguar si se trata de una relación de trabajo (una afiliación en alta laboral) o bien, de un episodio de desempleo, debemos fijarnos en los valores de las siguientes cinco variables del fichero:

- Las fechas de alta y de baja de la afiliación (en adelante *FecAlta* y *FecBaja*): determinan el inicio y fin de la relación con la Seguridad Social, así como de la obligación de cotizar.
- La variable tipo de relación laboral (*TRLDeseempleo*): permite identificar a los beneficiarios de prestaciones por desempleo y, por ende, a los que no lo son, esto es, a los que se encuentran en situación de alta laboral.
- El grupo de cotización (*GrupoCotiza*): determina grupos de categorías profesionales para trabajadores por cuenta ajena, a fin de limitar las bases de cotización con los topes mínimos y máximos establecidos anualmente para cada categoría profesional.
- El régimen de cotización (*RegCotiza*): permite delimitar a los colectivos de trabajadores por cuenta ajena, por cuenta propia, y a los cotizantes a la Seguridad Social mediante un Convenio Especial.

Para identificar a los trabajadores que durante algún momento del año en estudio estuvieron afiliados a la Seguridad Social en alta laboral, consideraremos únicamente los registros de la tabla que cumplan las siguientes restricciones:

- La fecha de baja de la relación tiene que ser posterior al uno de enero del año en estudio. La variable *FecBaja* se codifica de la siguiente forma *AAAAMMDD*, donde los cuatro primeros dígitos hacen referencia al año, los dos siguientes al mes y los dos últimos al día. Por tanto, si, por ejemplo, se trata de un estudio sobre la edición de 2012, la variable *FecBaja* debe ser mayor o igual que *20120101*.
- *GrupoCotiza*≠13; este dígito identifica a las personas en situación de baja por enfermedad, maternidad, paternidad, etc. que han dejado de cotizar a la Seguridad Social durante dicho período. No se consideran trabajadores afiliados en alta laboral porque no está previsto que se reincorporen a la empresa al finalizar la baja.

---

<sup>9</sup> Seguridad Social (2009 a); pág. 4.

- *TRLDeseempleo*≠400,751,752,753,754,755,756: se trata de códigos utilizados en las bases de datos de las que se ha extraído la MCVL para identificar circunstancias laborales particulares que llevan a que la relación no sea considerada un alta laboral<sup>10</sup>.
- No se consideran los registros cuyo código de régimen de cotización acaba en 40; estos dos dígitos identifican a las personas que cotizan mediante un Convenio Especial; estas situaciones no implican empleo.

Por su parte, las relaciones referidas al cobro de una prestación o subsidio por desempleo deben verificar las siguientes condiciones:

- La fecha de baja de la relación debe ser posterior al uno de enero del año en estudio, como en el caso anterior.
- *TRLDeseempleo*=751,752,752,754,755,756: estos códigos se utilizan para consignar en las bases de datos que una persona se encuentra en situación de desempleo, percibiendo una de las prestaciones del Servicio Público de Empleo Estatal (SEPE).

Para la delimitación del colectivo de pensionistas, utilizaremos cinco variables del fichero PRESTAC:

- El identificador único de cada prestación (*id\_prestacion*): si un sujeto percibe más de una prestación, le corresponderá más de un *id\_prestacion*.
- La variable que informa sobre el año al que se refieren los datos que aparecen consignados en cada una de las filas de la tabla (*Año*).
- El tipo de pensión contributiva de que se trate (*TipoPension*): de jubilación, incapacidad, jubilación procedente de incapacidad, viudedad, orfandad o pensión a favor de familiares.
- La variable referida al régimen de cotización (*RegPension*) del trabajador durante su vida laboral, o bien la contingencia de tipo laboral (accidente de trabajo o enfermedad profesional) de la que deriva la prestación.

Debemos considerar únicamente los registros de la tabla que cumplan las siguientes características:

- Debemos filtrar por el año en estudio, para considerar sólo las pensiones en activo en dicho año (por ejemplo, si la edición considerada es la del 2012, *Año=2012*).
- *TipoPrestacion*≠23<sup>11</sup>: con este filtro eliminamos del estudio las ayudas previas a la jubilación ordinaria y anticipada, prestaciones que figuran en las bases de datos de las que se extrae la MCVL y, por tanto, en los ficheros de

<sup>10</sup> Para una descripción detallada del tipo de información que aporta cada variable o el significado de su codificación, consultar el diseño de registro de las tablas en Seguridad Social (2012).

<sup>11</sup> Para una descripción detallada del tipo de información que aporta cada variable o el significado de su codificación, consultar el diseño de registro de las tablas en Seguridad Social (2012).

esta, pero que no se consideran pensiones contributivas del sistema de la Seguridad Social.

- *RegPension#31,32,35,38,39,68*: estos códigos identifican prestaciones de regímenes que no constituyen pensiones básicas del sistema de la Seguridad Social, sino prestaciones complementarias a estas, como las ayudas de acción social, el complemento de acción social o las mutualidades de los funcionarios.

Además de todas las comentadas, utilizaremos también la variable que identifica a cada sujeto en la muestra (*id\_per*) para delimitar a los tres colectivos, pues es la que permite asociar las relaciones reflejadas en los ficheros a los distintos individuos.

Decíamos al comienzo de este apartado que cada integrante puede pertenecer a varias categorías distintas en el año de referencia, simultánea o sucesivamente. La tabla 1 ilustra este hecho mediante los registros de varios sujetos extraídos de la edición de 2012: el primero es un trabajador que llega a la edad de jubilación, se retira y comienza a percibir una pensión contributiva; el segundo, un trabajador que pierde su empleo durante unos meses y percibe una prestación por desempleo durante ese intervalo; el tercero es un trabajador contratado a jornada parcial en varias empresas; y el cuarto ilustra el caso de un individuo que deja de tener relación con la Seguridad Social.

Analicemos cada caso en detalle. El sujeto A estuvo afiliado en alta laboral desde septiembre de 1978 a febrero de 2012: trabajó por cuenta propia unos catorce años (su código de régimen de cotización durante ese período fue *0521*, que en las bases de datos del sistema identifica al Régimen especial de los trabajadores autónomos), y luego fue contratado por una empresa y comenzó a cotizar al Régimen general (*RegCotiza=0111*). En febrero de 2012 se retiró y comenzó a percibir una pensión contributiva de jubilación (el tipo de prestación es el *21*, que corresponde a las jubilaciones en general).

El sujeto B estuvo afiliado como trabajador al Régimen general desde agosto de 2004 a marzo de 2012. El día 22 perdió su empleo y comenzó a percibir una prestación por desempleo (*TRLDeseempleo=751*); el 21 de agosto dejó de recibir dicha prestación al ser contratado de nuevo por otra empresa.

El sujeto C ejemplifica el caso de un trabajador pluriempleado. La línea de tiempo de la figura 1 ilustra sus distintas relaciones laborales. En marzo de 2001 fue contratado con carácter indefinido a tiempo parcial por una empresa (*TipoContrato=200*). La MCVL nos permite saber qué porcentaje de la jornada habitual de un trabajador a tiempo completo en la empresa realiza este individuo: el 73,1% (*CoefTempParcial/100*). En agosto de ese mismo año fue contratado de



forma temporal, por circunstancias eventuales de la producción (*TipoContrato=502*), por otra empresa. Para esta trabajó un 10,2% de una jornada habitual de 8 horas hasta octubre de 2001. Simultaneó varios contratos más por obra y servicio con su trabajo principal durante 2001, 2006, 2011 y 2012. En el momento de extracción de la muestra el contrato principal, el que había iniciado en marzo de 2001, todavía seguía vigente (cuando una relación laboral, prestación o subsidio por desempleo está en activo cuando se extrae la muestra, en la variable *FecBaja* aparece el último día del año siguiente a la edición correspondiente<sup>12</sup>).

**Tabla 1.** Ejemplo de grabación de información referida a distintas relaciones administrativas con la Seguridad Social en la edición 2012 de la MCVL.

*Fichero de afiliados (AFILIAD)*

id_per	FecAlta	FecBaja	RegCotiza	TipoContrato	CoefTempParcial	TRLDesempleo
A	19780901	19920430	0521	000	0	000
A	19920428	20030115	0111	100	0	000
A	20030116	20120203	0111	100	0	000
B	20040701	20120322	0111	401	0	000
B	20120323	20120821	0111	000	0	751
B	20120822	20130201	0111	401	0	000
C	20010315	20131231	0111	200	731	000
C	20010707	20011006	0111	502	102	000
C	20011009	20011130	0111	501	103	000
C	20060417	20060428	0111	502	96	000
C	20110301	20111231	0111	200	128	000
C	20120416	20120611	0111	200	160	000
D	20111215	20120314	0111	502	600	000
D	20120318	20120717	0111	000	0	751
D	20121025	20130124	0111	502	900	000

*Fichero de pensiones (PRESTAC)*

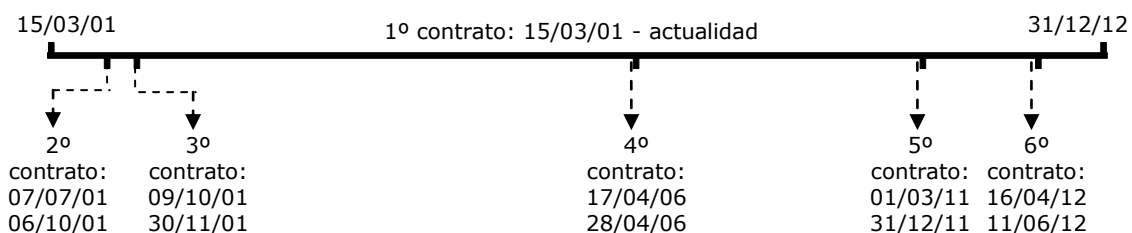
id_per	id_prestacion	Año	FecSituPrestacion	TipoPrestacion
A	0Y001L88000212L108000	2012	20120204	21

Fuente: elaboración propia a partir de los ficheros AFILIAD y PRESTAC correspondientes a la edición de 2012 de la MCVL

Nota: sólo se han reproducido las variables necesarias para el análisis realizado en este apartado. Se ha utilizado sólo una parte de los registros que incluye la tabla AFILIAD original para cada trabajador, con el objeto de facilitar el comentario. Se tratarán como si fuesen los únicos registros que contuviera la tabla original para los sujetos considerados. Además, hemos sustituido los valores originales de la variable *id\_per* (códigos numéricos de 15 dígitos) por letras para facilitar el comentario de la tabla.

<sup>12</sup> Por ejemplo, en la edición de 2012, en las relaciones en activo aparecería el valor 20131231 en la variable *FecBaja*.

**Figura 1.** Línea temporal donde se representan las relaciones laboral del sujeto C en la edición 2012 de la MCVL.



Fuente: elaboración propia a partir de los ficheros AFILIAD y PRESTAC correspondientes a la edición de 2012 de la MCVL

Finalmente, el sujeto D nos permite presentar un último supuesto que nos podemos encontrar al delimitar las relaciones de los individuos con la Seguridad Social: la falta de relación. El individuo en cuestión estuvo trabajando para una empresa a tiempo parcial entre diciembre de 2011 y marzo de 2012. En tal fecha perdió su empleo y percibió una prestación contributiva por desempleo hasta el 17 de agosto de 2012. En octubre de ese mismo año retomó la actividad laboral pero, ¿qué pasó entre el 18 de agosto y el 24 de octubre de 2012? No lo sabemos; durante ese período el individuo no ha tenido relación con la Seguridad Social; pero eso no debe llevarnos a concluir que dicha persona se haya mantenido en situación de inactividad. La MCVL no refleja las relaciones laborales de los colectivos que quedan al margen de la acción protectora de la Seguridad Social; entre ellos, los trabajadores de la economía sumergida, los desempleados que no perciben prestaciones de carácter contributivo, y las personas cubiertas por otros sistemas de previsión social (funcionarios de Clases Pasivas o algunos profesionales liberales). Por consiguiente, el que un sujeto desaparezca de la muestra durante un período de tiempo no implica que se encuentre en situación de inactividad. Puede haber formado parte durante el tiempo de no relación de uno de los colectivos no incluidos en la población de referencia de la MCVL<sup>13</sup>.

Como hemos visto en la tabla 1, la duración de los episodios con la Seguridad Social viene delimitado por las variables *FecAlta* y *FecBaja* de la tabla de afiliados (para las relaciones de trabajo y de desempleo) y por la variable *FecPension* de la tabla de prestaciones (para las pensiones contributivas). Cuando el número de sujetos analizado es pequeño, la caracterización y delimitación de las relaciones de los sujetos se puede realizar a mano, analizando uno a uno los registros que figuran en las tablas de la MCVL. Pero cuando los estudios abarcan colectivos de población grandes o bien, se quieren analizar las relaciones de los individuos para un conjunto de años (o incluso toda su vida laboral), es necesario recurrir a algún

<sup>13</sup> Lapuerta (2010); pág. 15.

software estadístico que automatice el proceso. En el IGE nos hemos servido de un gestor de bases de datos relacionales (SQL Server) y del programa de código abierto R para enfrentarnos a esta dificultad. Hemos desarrollado unas líneas de comandos, por medio de las cuales se crean cuatro matrices que identifican los días exactos del año en estudio en los que cada sujeto de la muestra se ha encontrado en cada una de las situaciones tipo: en alta laboral, cobrando una prestación o subsidio por desempleo, percibiendo una pensión contributiva o sin vínculo con la Seguridad Social.

La información en cada una de las matrices se organiza de la siguiente manera: las filas de la matriz son los meses del año y las columnas, los días de cada mes. El programa analiza cada registro de la tabla de afiliados y de prestaciones; si detecta que ha existido un vínculo entre un sujeto concreto y la Seguridad Social en el año en estudio, introduce en la matriz correspondiente una fila por cada mes en el que vínculo haya estado en activo (tomando en consideración las variables *FecAlta* y *FecBaja*). A continuación asigna valor 1 a las columnas correspondientes a los días comprendidos en el intervalo dado por *FecAlta* y *FecBaja*, o 0 en caso contrario. Si se trata de una relación laboral a tiempo parcial, en lugar de 1, el programa introduce el valor consignado en *CoefTempParcial*, dividido por 1.000; esto es, el porcentaje de la jornada laboral trabajada, en tanto por uno. Cuando ha terminado de analizar las tablas originales y construido las matrices por relaciones, las agrega por *id\_per* (la variable identificadora de cada individuo) y mes, sumando los valores consignados en las distintas filas para cada día del mes.

Para ayudar a clarificar el procedimiento, vamos a servirnos de los sujetos presentados en la tabla 1 y construyamos sus matrices de relaciones para el año 2012. El sujeto A estuvo afiliado en alta laboral durante todo el mes de enero y los tres primeros días del mes de febrero; la matriz de ocupados (ver tabla 2) debe contener dos registros para este individuo, uno con todas las columnas con valor 1 (el correspondiente al mes de enero); otro con valor 1 en las variables D1, D2, D3 (las variables que identifican los tres primeros días del mes de enero), y 0 en las restantes variables. A partir de febrero y hasta final de año, el sujeto percibió una pensión contributiva, por lo que en la matriz de pensiones deben figurar once filas, una por cada mes de percepción de la pensión. Para los sujetos B y D completaremos sus matrices de ocupación y paro sirviéndonos de la misma lógica que con el sujeto anterior. Los episodios de no relación son los que quedan fuera de los períodos delimitados por *FecAlta* y *FecBaja*. La matriz de no relación se obtiene una vez se han calculado las otras tres, por contraposición a estas: las celdas con valor 1 en las matrices de ocupados, parados y pensionistas tendrán valor 0 en la

matriz de no relación, y a la inversa. En el ejemplo de la tabla 2, dicha matriz contiene cuatro registros, uno por cada mes en el que el sujeto D no ha tenido relación con la Seguridad Social.

El supuesto C es un poco particular, ya que corresponde a un individuo contratado a tiempo parcial en varias empresas a lo largo del año. En este caso, en lugar del valor 1 para identificar los días en los que cada relación ha estado en activo, el programa introduce el valor del coeficiente de parcialidad (dividido por 1.000). El sujeto del ejemplo ha simultaneado dos contratos a tiempo parcial en 2012; el primero estuvo vigente todo el año (el coeficiente de parcialidad en este fue de 731); el segundo se inició el 16/4/2012 y finalizó el 11/06/2012 (con un coeficiente de parcialidad de 160). En un primer momento, el programa introduce doce filas en la matriz de ocupados para reflejar la primera relación laboral del sujeto, con valor 0,731 en sus columnas, y tres con valor 0,160 en las columnas correspondientes a los días comprendidos entre el 16/04/2010 y el 11/06/2012 para la segunda relación. Luego, se suman los valores de las filas referidas al mismo mes, resultando un coeficiente de parcialidad de 0,891 para los meses de abril, mayo y junio. Esto nos permite saber cuál ha sido el porcentaje total de la jornada laboral realizada por el trabajador C cada día del año.

Las matrices que acabamos de construir son muy útiles para delimitar las relaciones de los individuos con la Seguridad Social; en particular, los episodios de no relación. En realidad, para determinar la duración de las relaciones laborales, de desempleo y de percepción de pensiones contributivas, no es necesario recurrir a las matrices. Podría hacerse una aproximación comparando simplemente las fechas de alta y de baja de los registros de las tablas de afiliados y de prestaciones. Sin embargo, a la hora de abordar los episodios de no relación la cuestión no es sencilla. En las tablas originales no hay registros que informen de los momentos de no relación, sino que estos se obtienen por descarte, analizando todas las relaciones de los individuos y comparando las fechas de alta y de baja de un registro con las de los demás. Y tan importante es saber cuando un individuo tiene relación con la Seguridad Social como cuando no la tiene. Esto permite estimar, entre otras cosas, el volumen de personas que quedan fuera del sistema de previsión de la Seguridad Social.

Además, las matrices presentan la información de una forma clara e intuitiva. En primer lugar, separan los registros de las tablas originales según el tipo de relación que representen; de esta forma, para estudiar un colectivo concreto, basta con acudir a la matriz que corresponda a ese colectivo, sin tener que rebuscar por los ficheros originales la información. En segundo lugar, la dicotomía 1/0 permite que, con un simple vistazo, el usuario reconozca los períodos en los que cada sujeto

ha tenido relación con la Seguridad Social: el 1 indica relación, el 0 no relación. E, incluso, en el caso de las relaciones laborales a tiempo parcial, detectar qué porcentaje de una jornada laboral normal de 8 horas ha trabajado cada individuo.

Otra utilidad que presentan las matrices construidas es que permiten detectar incoherencias entre los registros de las bases de datos originales. Hasta ahora no se ha comentado nada al respecto, pero lo cierto es que la MCVL presenta los problemas típicos de las fuentes procedentes de registros administrativos: duplicidades en la información, valores mal grabados, variables sin actualizar, etc. En concreto, hemos detectado que la fecha de baja que figura en los registros referidos al cobro de prestaciones y subsidios por desempleo no siempre es correcta. Ello se debe a que la información de la MCVL se extrae de las bases de datos primarias de la Seguridad Social; los datos grabados en las mismas se someten a procesos de revisión y modificación, pero las modificaciones no siempre revierten en la MCVL. Un ejemplo de las inconsistencias de que hablamos sería el de un trabajador por cuenta propia que esté percibiendo una prestación o subsidio por desempleo. La normativa actual en materia laboral en España no permite la existencia de este tipo de situación; el sistema de la Seguridad Social tampoco contempla el cobro de una prestación o subsidio por desempleo por parte de un trabajador a jornada completa. Sin embargo, en las tablas originales de la MCVL aparecen registros simultáneos de trabajo por cuenta propia o por cuenta ajena a tiempo completo y de desempleo para un mismo individuo. Las matrices nos ayudan a detectar este tipo de inconsistencias y a depurarlas.

Finalmente, en este artículo hemos utilizado las matrices para determinar los períodos de relación con la Seguridad Social sólo para el año en estudio. También pueden emplearse para determinar y clasificar las relaciones de los sujetos para un conjunto de años o bien, a lo largo de toda su vida laboral.

**Tabla 2.** Matrices de relaciones de los sujetos presentados en la tabla 1.

*Matriz de ocupados*

id_per	mes	D1	D2	D3	D4	...	D11	D12	D13	D14	D15	D16	...	D21	D22	D23	D24	D25	...	D30	D31
A	01	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	1
A	02	1	1	1	0	...	0	0	0	0	0	0	...	0	0	0	0	0	...	-	-
B	01	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	1
B	02	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	-	-
B	03	1	1	1	1	...	1	1	1	1	1	1	...	1	1	0	0	0	...	0	0
B	08	0	0	0	0	...	0	0	0	0	0	0	...	0	1	1	1	1	...	1	1
B	09	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	-
B	10	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	1
B	11	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	-
B	12	1	1	1	1	...	1	1	1	1	1	1	...	1	1	1	1	1	...	1	1
C	01	0,731	0,731	0,731	0,731	...	0,731	0,731	0,731	0,731	0,731	0,731	...	0,731	0,731	0,731	0,731	0,731	...	0,731	0,731
C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
C	04	0,731	0,731	0,731	0,731	...	0,731	0,731	0,731	0,731	0,731	0,891	...	0,891	0,891	0,891	0,891	0,891	...	0,891	-
C	05	0,891	0,891	0,891	0,891	...	0,891	0,891	0,891	0,891	0,891	0,891	...	0,891	0,891	0,891	0,891	0,891	...	0,891	0,891
C	06	0,891	0,891	0,891	0,891	...	0,891	0,731	0,731	0,731	0,731	0,731	...	0,731	0,731	0,731	0,731	0,731	...	0,731	-
C	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
D	01	0,600	0,600	0,600	0,600	...	0,600	0,600	0,600	0,600	0,600	0,600	...	0,600	0,600	0,600	0,600	0,600	...	0,600	0,600
D	02	0,600	0,600	0,600	0,600	...	0,600	0,600	0,600	0,600	0,600	0,600	...	0,600	0,600	0,600	0,600	0,600	...	-	-
D	03	0,600	0,600	0,600	0,600	...	0,600	0,600	0,600	0,600	0	0	...	0	0	0	0	0	...	0	0
D	10	0	0	0	0	...	0	0	0	0	0	0	...	0	0	0	0	0,900	...	0,900	0,900
D	11	0,900	0,900	0,900	0,900	...	0,900	0,900	0,900	0,900	0,900	0,900	...	0,900	0,900	0,900	0,900	0,900	...	0,900	-
D	12	0,900	0,900	0,900	0,900	...	0,900	0,900	0,900	0,900	0,900	0,900	...	0,900	0,900	0,900	0,900	0,900	...	0,900	0,900

*Matriz de parados*

id_per	mes	D1	D2	D3	D4	D5	D6	D7	D8	...	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
B	03	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
B	04	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-
B	05	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
B	06	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-
B	07	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
B	08	1	1	1	1	1	1	1	1	...	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
D	03	0	0	0	0	0	0	0	0	...	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	04	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-
D	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
D	07	1	1	1	1	1	1	1	1	...	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*Matriz de pensionistas*

id_per	id_prestacion	mes	D1	D2	D3	D4	D5	D6	D7	D8	...	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31	
A	0Y001L88000212L108000	02	0	0	0	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	-	-
A	0Y001L88000212L108000	03	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1
A	0Y001L88000212L108000	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
A	0Y001L88000212L108000	12	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1

*Matriz de no relación*

id_per	mes	D1	D2	D3	D4	D5	D6	D7	D8	...	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31
D	07	0	0	0	0	0	0	0	0	...	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	08	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D	09	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-
D	10	1	1	1	1	1	1	1	1	...	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0

Fuente: elaboración propia a partir de los ficheros AFILIAD y PRESTAC correspondientes a la edición de 2012 de la MCVL.

Nota: (...) significa que se van a repetir los valores del mes que figura en la fila superior o los valores de las columnas anterior y posterior.

### 3. Identificación de las relaciones laborales

Como hemos visto en la introducción y en el apartado anterior, la MCVL es una fuente de información muy útil para realizar investigaciones sobre la relación de las personas con la actividad económica. En estas investigaciones la unidad de análisis es el individuo. Sin embargo, la muestra también puede utilizarse para estudiar las características de las relaciones laborales en España. En este segundo tipo de estudios el foco de atención deja de ser el individuo, para centrarse en el vínculo laboral en sí mismo. En Lapuerta (2010) se aclara la diferencia entre ambos tipos de estudios: "... no es lo mismo hablar de la distribución según sexo y edad de los afiliados dados de alta en la Seguridad Social [...] que de las altas en la Seguridad Social. El resultado de este último indicador siempre será más elevado, debido a que un mismo individuo puede presentar varias altas en un mismo año e incluso puede tener dos altas simultáneas en regímenes distintos"<sup>14</sup>. En este apartado nos vamos a centrar en el segundo tipo de análisis, en el de las situaciones. Para abordarlo es necesario definir el concepto de relación o vínculo laboral, y analizar cómo extraer información al respecto de los ficheros de la MCVL.

Según la bibliografía disponible en la web del organismo, "cada relación es lo que transcurre entre un alta y una baja en la Seguridad Social"<sup>15</sup> y, normalmente, dan origen y extinción al vínculo laboral entre un empleado y un empleador. En la MCVL cada fila del fichero de afiliados contiene un vínculo laboral distinto, pero existen excepciones. Todo cambio en algún aspecto fundamental del contrato queda reflejado en dicho fichero mediante la introducción de registros nuevos. Así, cuando la modalidad contractual se ve modificada sin que ello lleve consigo la extinción de la relación laboral entre empleado y empleador<sup>16</sup>, la información previa y posterior a la modificación se consigna en dos registros distintos. Igual sucede cuando cambia el grupo de cotización, el coeficiente de tiempo parcial, o cuando se produce alguna de las circunstancias que dan lugar a la suspensión momentánea de la relación, pero se mantiene la reserva del puesto de trabajo o el derecho a la reincorporación laboral (baja por servicio militar, incapacidad transitoria, excedencia para el cuidado de hijos, etc.)<sup>17</sup>.

Por tanto, debemos establecer una diferenciación clara entre relación laboral y relación contractual: la primera hace referencia a un vínculo ininterrumpido entre un trabajador y un empresario; la segunda identifica cada uno de los períodos que tienen lugar entre la firma de dos contratos sucesivos entre ambas partes. En el

---

<sup>14</sup> Lapuerta (2010); pág. 22.

<sup>15</sup> Seguridad Social (2009 b); pág. 4.

<sup>16</sup> Un ejemplo de ello sería un trabajador que pasa de estar contratado a tiempo parcial a jornada completa.

<sup>17</sup> Lapuerta (2010); pág. 23.



fichero de afiliados la información relativa a cada relación laboral del trabajador puede estar contenida en una sola fila, dos, tres, ..., tantas como cambios en las características del contrato se hubiesen producido. Esto complica enormemente la labor de explotación de la muestra, ya que obliga a tomar en consideración un buen número de registros y variables para delimitar cada uno de los episodios laborales del trabajador.

Además, con el tiempo, la Seguridad Social ha cambiado el procedimiento de recogida de la información. Los registros más recientes de la MCVL incluyen ocho variables que informan sobre los cambios acaecidos en la relación laboral: *TipoContrato* refleja el último contrato (el más reciente) y *TipoContratoModif1* y *TipoContratoModif2* la modalidad contractual inicial y la primera modificación; igual sucede con las variables *CoefTempParcial* (refleja el coeficiente de tiempo parcial del contrato más reciente) y sus homólogas *CoefTempParcialModif1* y *CoefTempParcialModif2* (coeficientes de parcialidad del primer contrato y de la primera modificación). Las columnas *FecContratoModif1* y *FecContratoModif2* informan sobre las fechas en las que se produjeron las primeras modificaciones de contrato, *GrupoCotizaModif1* informa sobre el cambio en el grupo de cotización y *FecGrupoCotizaModif1* la fecha en la que se produjo dicho cambio. De este modo, toda la información referida a una misma relación laboral, incluidas las modificaciones contractuales sucesivas<sup>18</sup>, puede condensarse en un solo registro.

Por lo tanto, en la MCVL coexisten dos procedimientos distintos de grabar la historia laboral: en filas distintas o por medio de las variables presentadas en el párrafo anterior. En la tabla 3 ejemplificamos lo expuesto<sup>19</sup>. En la misma se presenta, en primer lugar, el caso de un sujeto en el que la modificación de uno de los parámetros contractuales da origen a la grabación de varios registros; en segundo lugar, un trabajador en que las modificaciones contractuales se condensan en un único registro; y, en tercer lugar, dos ejemplos de interrupciones laborales que se analizarán más adelante. Para identificar a cada trabajador utilizamos la variable *id\_per*; para identificar a la empresa, la variable *CCCSecundaria*<sup>20</sup>, esto es, el código de la cuenta de cotización al que figuran asociados los datos identificativos del empresario y al que se vinculan sus trabajadores por cuenta ajena.

---

<sup>18</sup> Hasta un máximo de tres modificaciones en realidad. En resumen, se dispone de información de las características del contrato más reciente, del primer contrato y de la primera modificación en el mismo, pero no de las modificaciones sucedidas entre el segundo contrato y el último.

<sup>19</sup> El ejemplo lo hemos tomado prestado, cambiando los registros originales por unos referidos a la edición de 2012, de Lapuerta (2010); pág. 25.

<sup>20</sup> Aunque también podríamos haber utilizado la variable *id\_pagador*, como comentamos en la introducción.

**Tabla 3.** Ejemplo de los distintos procedimientos de grabación de las relaciones laborales en la edición 2012 de la MCVL.

id_per	CCC Secundaria	FecAlta	FecBaja	Tipo Contrato	CoefTemp Parcial	Causa Baja	FecContrato Modif1	TipoContrato Modif1	CoefTemp Parcial Modif1	FecContrato Modif2	Tipo Contrato Modif2	CoefTemp Parcial Modif2
E	Empresa 1	20100202	20100221	502	500	54	-	-	-	-	-	-
E	Empresa 1	20100222	20100323	402	0	54	-	-	-	-	-	-
E	Empresa 1	20100324	20100422	502	500	54	-	-	-	-	-	-
E	Empresa 1	20100511	20100525	410	0	54	-	-	-	-	-	-
E	Empresa 1	20100802	20101018	510	750	54	-	-	-	-	-	-
F	Empresa 2	19920801	20131231	100	0	00	20120118	100	0	20120831	100	750
G	Empresa 3	19980101	19981031	100	0	96	-	-	-	-	-	-
G	Empresa 3	19981101	20020211	109	0	68	-	-	-	-	-	-
G	Empresa 3	20021009	20030704	100	0	68	-	-	-	-	-	-
G	Empresa 3	20040704	20070801	100	0	99	-	-	-	-	-	-
G	Empresa 3	20070802	20080411	100	0	54	-	-	-	-	-	-
H	Empresa 4	20020916	20030305	401	0	54	-	-	-	-	-	-
H	Empresa 4	20030310	20040914	401	0	54	-	-	-	-	-	-
H	Empresa 4	20041015	20041106	150	0	51	-	-	-	-	-	-

Fuente: elaboración propia a partir de los ficheros AFILIAD correspondientes a la edición de 2012 de la MCVL

Nota: sólo se han reproducido las variables necesarias para el análisis realizado en este apartado. Se ha utilizado sólo una parte de los registros que incluye la tabla AFILIAD original para cada trabajador y empresario, con el objeto de facilitar el comentario. Se tratarán como si fuesen los únicos registros que contuviera la tabla original para los sujetos considerado. Además, hemos sustituido los valores originales de la variable id\_per (códigos numéricos de 15 dígitos) por letras para facilitar el comentario de la tabla; tampoco se han reproducido los valores originales de la variable CCCSecundaria por ser muy largos.

El sujeto E mantiene un vínculo laboral ininterrumpido con la empresa 1 desde el 2 de febrero de 2010 hasta el 22 de abril de ese mismo año. Hasta esta fecha el vínculo laboral se formalizó por medio de un contrato eventual por circunstancias de la producción a tiempo parcial. El 22 de febrero, trabajador y empresa firmaron un nuevo contrato a tiempo completo, que estuvo vigente durante un mes. Pasado ese tiempo, las condiciones de la relación laboral experimentaron una nueva modificación, y ambas partes volvieron a vincularse mediante un contrato a tiempo parcial hasta el 22 de abril de 2010. El trabajador prestó de nuevo sus servicios a la empresa 1 hasta en otras dos ocasiones, como se puede ver en la tabla 4. Sin embargo, estos no forman parte de la primera relación laboral, ya que no se producen en momentos consecutivos del tiempo. Es decir, aunque los cinco registros estén referidos al mismo trabajador y empresa, se consideran tres relaciones laborales distintas. El sujeto F, por su parte, mantiene un vínculo laboral con la empresa 2 desde el 1 de agosto de 1992. Durante este período se han producido dos modificaciones en el tipo de contrato que han sido recogidas en un único registro: hasta el 17 de enero de 2012 el vínculo laboral adoptó la forma de un contrato indefinido a tiempo completo; durante los ocho meses siguientes el contrato siguió teniendo carácter indefinido, pero pasó a la modalidad de tiempo parcial, con un coeficiente de parcialidad del 75%. El 31 de agosto de 2012 el trabajador volvió a trabajar para la empresa a tiempo completo.

La existencia de los dos procedimientos de grabación puede llegar a confundir al estadístico en la tarea de explotación de la muestra. Para solucionar este inconveniente, algunos autores optan por la unificación en un solo registro de las relaciones laborales recogidas en varias filas, sirviéndose de las variables introducidas a partir de 2005 (que aparecerán en blanco en estos casos). Pero este procedimiento sólo se puede implementar cuando el número de modificaciones contractuales es igual o inferior a dos porque, de hacerlo para un número mayor, obligaría a prescindir de parte de la información o bien, a incluir variables adicionales en la tabla. Otros autores, en cambio, abogan por el procedimiento contrario, que consistiría en dividir las relaciones laborales consignadas en un solo registro en varios. En este caso es necesario asignar algún identificador a las filas que procedan de la división de otra para poder mantener identificado cada vínculo laboral.

En el IGE nos hemos decantado por el segundo procedimiento: primero dividimos las relaciones laborales registradas en una sola fila en varias para, después, crear un identificador único de cada relación laboral. De nuevo, hemos utilizado el software libre R para automatizar el proceso. A partir de una subdivisión del fichero de afiliados con información exclusivamente de las relaciones en alta

laboral, separamos los registros con modificaciones en el tipo de contrato, el coeficiente de parcialidad o grupo de cotización; esto es, los registros en los que los campos *TipoContratoModif1*, *TipoContratoModif2*, *CoefTempParcialModif1*, *CoefTempParcialModif2* o *GrupoCotizaModif1* contienen información. Las variables fecha de alta y de baja en las nuevas filas creadas se actualizan a partir de los valores de *FecContratoModif1*, *FecContratoModif2* y *FecGrupoCotizaModif1*; reflejarán los períodos transcurridos entre cada una de las modificaciones en las características de la relación laboral. Siguiendo con los sujetos introducidos en la tabla 3, vamos a detallar cuál sería el resultado de aplicar la división a los registros de esos individuos (ver la tabla 4).

**Tabla 4.** Ejemplo de división en varios registros de las relaciones laborales grabadas en una única fila en la edición 2012 de la MCVL.

Número registro	id_per	CCC Secundaria	FecAlta	FecBaja	Tipo Contrato	CoefTemp Parcial	Causa Baja	id contrato
1º	E	Empresa 1	20100202	20100221	502	500	54	1
2º	E	Empresa 1	20100222	20100323	402	0	54	2
3º	E	Empresa 1	20100324	20100422	502	500	54	3
4º	E	Empresa 1	20100511	20100525	410	0	54	4
5º	E	Empresa 1	20100802	20101018	510	750	54	5
6º	F	Empresa 2	19920801	20120117	100	0	-	1
7º	F	Empresa 2	20120118	20120830	100	750	-	1
8º	F	Empresa 2	20120831	20131231	100	0	00	1
9º	G	Empresa 3	19980101	19981031	100	0	96	1
10º	G	Empresa 3	19981101	20020211	109	0	68	2
11º	G	Empresa 3	20021009	20030704	100	0	68	3
12º	G	Empresa 3	20040704	20070801	100	0	99	3
13º	G	Empresa 3	20070802	20080411	100	0	54	3
14º	H	Empresa 4	20020916	20030305	401	0	54	1
15º	H	Empresa 4	20030310	20040914	401	0	54	1
16º	H	Empresa 4	20041015	20041106	150	0	51	2

*Fuente: elaboración propia a partir de los ficheros AFILIAD correspondientes a la edición de 2012 de la MCVL*

*Nota: sólo se han reproducido las variables necesarias para el análisis realizado en este apartado. Se ha utilizado sólo una parte de los registros que incluye la tabla AFILIAD original para cada trabajador y empresario, con el objeto de facilitar el comentario. Se tratarán como si fuesen los únicos registros que contuviera la tabla original para los sujetos considerados. Además, hemos sustituido los valores originales de la variable id\_per (códigos numéricos de 15 dígitos) por letras para facilitar el comentario de la tabla; tampoco se han reproducido los valores originales de la variable CCCSecundaria por ser muy largos.*

Solamente habría que reformar el registro referido al sujeto F, ya que es el único que contiene información en las variables mencionadas. Habría que modificar la fila original y añadir dos nuevas. La primera refleja la primera relación laboral del trabajador con la empresa 2 (en la variable *FecBaja* hay que introducir la fecha inmediatamente anterior al valor consignado en *FecContratoModif1*; en las variables *TipoContrato* y *CoefTempParcial* deben figurar los valores de *TipoContratoModif1* y *CoefTempParcialModif1*). La segunda fila contiene la información referida al segundo contrato (en la variable *FecAlta* se introduce el valor que figura en *FecContratoModif1*, y en *FecBaja* la fecha inmediatamente anterior a la firma del contrato actual, esto es, a *FecContratoModif2*; en *TipoContrato* y *CoefTempParcial* debe figurar la información contenida en *TipoContratoModif2* y *CoefTempParcialModif2*, respectivamente). La tercera fila debe contener la información referida al contrato más reciente (en las variables *FecBaja*, *TipoContrato* y *CoefTempParcial* se introduce la información que figura en el registro original para estas variables, y en *FecAlta* se introduce el valor de *FecContratoModif2*).

A continuación, introducimos en la tabla dividida un campo, *id\_contrato*, que identificará unívocamente cada una de las relaciones contractuales sucesivas. Hemos supuesto, además, que para que la relación laboral entre trabajador y empresario se considere interrumpida, deben mediar al menos 7 días de distancia entre la finalización de un contrato y el inicio del siguiente. Ahora bien, si la interrupción es superior a 7 días pero se debe a una de las siguientes causas: baja por servicio militar (*CausaBaja=61*), por huelga legal o cierre patronal (*CausaBaja=64*), por agotamiento de incapacidad transitoria (*CausaBaja=65*), por excedencia para cuidado de hijos (*CausaBaja=68*), por excedencia para cuidado de familiares (*CausaBaja=73*), se considera que la relación laboral se halla en suspenso, pero no rota, pues se mantiene la reserva del puesto a favor del trabajador<sup>21</sup>. En estos supuestos, las filas correspondientes a la situación previa a la interrupción e inmediatamente posterior a ella, deben contener idéntico valor en la variable *id\_contrato*, siempre y cuando no se hayan modificado las condiciones del contrato.

El resultado de aplicar el programa a los registros de la tabla 4 se presenta en la propia tabla, en la variable *id\_contrato*. El sujeto E ha mantenido cinco relaciones contractuales distintas con la empresa 1; el sujeto F sólo una con la empresa 2; el trabajador G hasta tres relaciones contractuales distintas con la empresa 3; y el último individuo, el sujeto H, ha mantenido dos relaciones contractuales distintas con la empresa 4.

---

<sup>21</sup> Lapuerta (2010); pág. 23.

¿Cuál es la utilidad del procedimiento descrito? Al introducir este apartado comentamos la distinción entre utilizar la muestra para análisis centrados en el individuo o análisis en los que la unidad de observación sea la relación. En este segundo conjunto, la utilidad de la tabla que acabamos de construir es inmediata: disponer de una variable que permita identificar cada contrato entre trabajador y empresario es vital para estudiar las relaciones. Por el momento, la información difundida en la web del IGE se refiere a individuos; el análisis de las relaciones es un proyecto que hemos iniciado con la edición 2012 de la muestra y del que esperamos poder difundir resultados a final de año.

#### 4. Resultados.

En este apartado presentamos algunos de los resultados más interesantes que hemos obtenido para la población joven de Galicia mediante el tratamiento de la muestra con los procedimientos presentados en el artículo. La tabla siguiente presenta una clasificación del colectivo estudiado según su relación con la Seguridad Social a lo largo del año 2011.

**Tabla 5.** Población de 16 a 29 años según la relación con la Seguridad Social en el año de referencia. Galicia. Año 2011.

	<b>Personas</b>	<b>Porcentajes</b>
Total	217.000	100,0
Trabajando	154.050	71,0
Todo el año	75.700	34,9
Parte del año	78.350	36,1
Cobrando una prestación o subsidio por desempleo	7.400	3,4
Todo el año	1.075	0,5
Parte del año	6.325	2,9
Cobrando una pensión contributiva de la Seguridad Social	5.400	2,5
Trabajando y cobrando una prestación o subsidio por desempleo	48.500	22,4
Sin períodos de no relación con la Seguridad Social	10.750	5,0
Con períodos de no relación con la Seguridad Social	37.750	17,4
Trabajando y cobrando una pensión contributiva de la Seguridad Social	1.425	0,7
Resto de situaciones	*	*

*Fuente: elaboración propia a partir de la Muestra Continua de Vidas Laborales facilitada por la Seguridad Social*

*Nota: los períodos de no relación con la Seguridad Social son aquellos días en los que la persona no ha estado afiliada en alta laboral, ni percibiendo una prestación o subsidio por desempleo, ni una pensión contributiva de la Seguridad Social. (\*) Pocas observaciones muestrales, por lo que el dato puede estar afectado por un elevado error de muestreo.*

Un total de 217.000 jóvenes gallegos de entre 16 y 29 años tuvieron relación con la Seguridad Social durante 2011, el 71% de los cuales sólo mantuvo relaciones de alta laboral. Un total de 48.500 personas trabajaron y cobraron una prestación o subsidio por desempleo, lo que supone el 22,4% de la población de referencia. Otros 7.400 jóvenes únicamente mantuvieron relaciones con la Seguridad Social por el cobro de prestaciones o subsidios por desempleo; esto es, no estuvieron afiliados en alta laboral en ningún momento del año. En ambos casos, los individuos han podido encadenar estas con períodos en los que no tuvieron nexos con la Seguridad Social, por ejemplo, porque se les acabó el derecho al cobro de la prestación o subsidio. Finalmente, 5.400 jóvenes percibieron una pensión contributiva del sistema en 2011 (la mayor parte de orfandad), sin tener otro tipo de relación, y algo más de 1.400 cobraron una pensión contributiva y estuvieron afiliados en alta laboral.

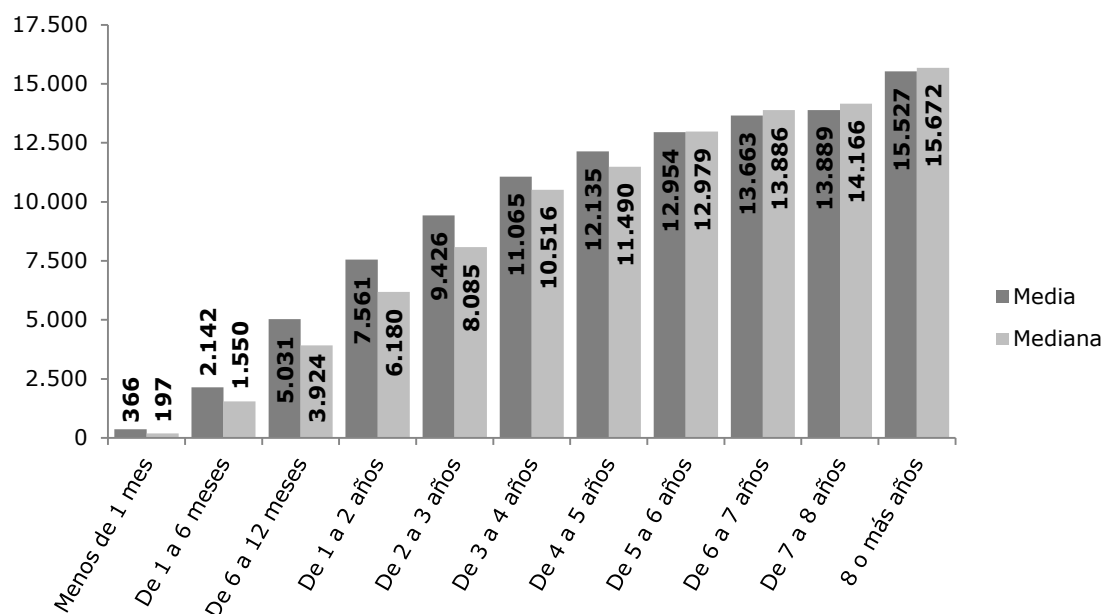
Para el subconjunto de los trabajadores afiliados por cuenta ajena a la Seguridad Social en algún momento de 2011 (el 91,3% del total de ocupados), se ha calculado el número total de días en alta laboral durante el año, el número medio de días de permanencia en la empresa y estimaciones de las bases de cotización mensuales medias. Hemos observado que el 13,3% de los gallegos menores de 30 años ha trabajado menos de 6 meses a lo largo de toda su vida laboral. Si nos ceñimos a los menores de 25 años, el porcentaje sube hasta el 26,7%, mientras que para el grupo de 25 a 29 se sitúa por debajo del 5%. Para obtener estos resultados, calculamos la matriz de ocupación descrita en el apartado 2 para toda la vida laboral del colectivo en estudio.

Al cruzar esta matriz con la información procedente de la Agencia Tributaria, obtuvimos también el salario medio y mediano en el año 2011 según el tiempo trabajado, la permanencia en la empresa, la retribución en el primer empleo, etc. Algunos de los resultados más interesantes que obtuvimos son que la población juvenil de Galicia gana, de media, unos 9.463 euros al año en concepto de retribuciones salariales; confirmamos la creencia popular de que los hombres ganan más que las mujeres, incluso en el grupo de edad analizado, o la de que el salario se incrementa conforme a la edad del trabajador. Además, el importe anual de las percepciones salariales parece estar estrechamente relacionado con la duración de la vida laboral, como se desprende del gráfico 1: cuanto mayor es la experiencia laboral del trabajador, mayor es su retribución. También comprobamos que existe una relación positiva entre el tiempo de permanencia en la empresa y el salario pagado, algo que ya apuntamos en el primer apartado del artículo. En la página

web del IGE encontrarán los resultados completos del estudio realizado sobre la población juvenil de Galicia<sup>22</sup>.

**Gráfico 1.** Media/mediana del salario anual de las personas de 16 a 29 años que en 2011 estuvieron afiliadas en alta laboral por cuenta ajena según el tiempo trabajado a lo largo de su vida laboral. Galicia

Unidad: euros



Fuente: elaboración propia a partir de la Muestra Continua de Vidas Laborales (MCVL) facilitada por la Seguridad Social

## 5. Conclusiones.

A lo largo del artículo hemos ido desgranando las principales características de la MCVL de la Seguridad Social: su contenido, su utilidad, algunas de las dificultades que plantea su explotación, etc. Una de sus mayores virtudes frente a otras fuentes de información es su tamaño: en cada edición, la muestra se compone de una cuarta parte de la población de referencia. Este tamaño es muy superior al de otras encuestas relacionadas sobre el mercado laboral, como la Encuesta de población activa o la Encuesta de coste laboral, lo que permite abordar el estudio de colectivos o situaciones que otras estadísticas no permiten.

Otra ventaja de la MCVL es su carácter continuo: su contenido no está referido a un momento concreto del tiempo sino que abarca toda la vida laboral de los sujetos incluidos. De esta forma, resuelve el problema que presentan otras

<sup>22</sup> Consultar: [http://www.ige.eu/web/mostrar\\_actividade\\_estadistica.jsp?idioma=gl&codigo=0204015](http://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0204015)



fuentes estadísticas estáticas, como la explotación del fichero de afiliaciones de la Seguridad Social, referido a una fecha concreta. Mediante el análisis de esta fuente podemos saber el número de afiliaciones al sistema registrado en dos momentos distintos del tiempo, pero no lo que ha sucedido entre ambos; en cambio, la MCVL sí permite este tipo de análisis. Más aún, puesto que sus integrantes permanecen en la muestra mientras no se extinga su vínculo con la Seguridad Social, resulta una fuente idónea para la realización de estudios longitudinales.

Pero quizás, su mayor ventaja sea, a nuestro entender, su contenido. Un contenido relativamente novedoso, puesto que parte de la información que ofrece ya está disponible desde hace años en otras fuentes, como en la mencionada explotación estadística que realiza la Seguridad Social de sus bases de datos. Sin embargo, el hecho de que en la MCVL la información de este organismo se haya cruzado con la del Padrón municipal de habitantes del INE y el Modelo 190 de la Agencia Tributaria, añade valor a la información extraída de la Seguridad Social. Esta peculiaridad, unido a las otras dos grandes ventajas (su tamaño y su carácter continuo), convierten a la MCVL en una fuente con grandes potencialidades para el estudio del mercado laboral en España.

Sin embargo, no podemos realizar un estudio sobre la MCVL comentando las ventajas que presenta su utilización, sin hablar de sus inconvenientes. Su contenido procede de registros administrativos, que no han sido diseñados para ser explotados estadísticamente, sino para la gestión administrativa. Esto conlleva que existan registros incompletos, variables sin actualizar, duplicidades e incluso incoherencias. No ha sido posible abordar en este trabajo este tipo de cuestiones, pero sí nos gustaría resaltar que el contenido de la MCVL no ha sido sometido a un proceso de depuración, hecho que ha de tenerse en cuenta en su explotación. Otro aspecto que deberán tener presente los usuarios de la MCVL es que su población objetivo no cubre a todo el colectivo de ocupados. Quedan fuera los trabajadores cubiertos por sistemas de previsión social distintos a la Seguridad Social (como los funcionarios de clases pasivas) y los trabajadores ilegales.

No obstante, la principal dificultad que acarrea la explotación de la muestra tiene que ver, aunque pueda parecer contradictorio, con su tamaño. La gran cantidad de tablas, variables y registros que contiene requiere que los usuarios dispongan de equipos y programas que permitan manejarlos. Para tratar de facilitar la tarea a los que se embarquen en su explotación, hemos presentado dos procedimientos que permiten, entre otras cosas...

- Clasificar a los integrantes dentro de los tres grandes colectivos que componen la muestra (trabajadores afiliados en alta laboral, perceptores de

prestaciones o subsidios por desempleo y perceptores de pensiones contributivas, como las de jubilación o viudedad).

- Delimitar cuál ha sido su relación con la Seguridad Social a lo largo de un período de tiempo, si han pertenecido al mismo colectivo o han experimentado cambios en algún momento del año; e, incluso, si se han encontrado en varias situaciones distintas a la vez, como por ejemplo, trabajando a tiempo parcial y cobrando una prestación por desempleo.
- Diferenciar las distintas relaciones contractuales establecidas entre un trabajador y un empresario, la duración de los episodios y sus características.

Las limitaciones de espacio y tiempo no nos han permitido ahondar en el detalle de cada uno de los procedimientos presentados. Nos hemos limitado a introducirlos y comentar sucintamente en qué consisten, poniendo algunos ejemplos de los resultados que arrojan. Sin embargo, algunas veces los ejemplos prácticos pueden ser mucho más clarificadores que la mejor de las explicaciones teóricas. Esperamos que este sea el caso y que los procedimientos desarrollados en el IGE puedan servir de ayuda a otros usuarios de la MCVL.

## **BIBLIOGRAFÍA**

LAPUERTA, Irene (2010), "Claves para el trabajo con la Muestra Continua de Vidas Laborales", en *DemoSoc Working Paper, Paper Number 2010-37*, Universitat Pompeu Fabra, Barcelona, en [http://www.upf.edu/demosoc/\\_pdf/DEMOSOC37.pdf](http://www.upf.edu/demosoc/_pdf/DEMOSOC37.pdf)

Seguridad Social (2009 a), *Descripción general de la muestra*, en [http://www.seg-social.es/Internet\\_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm](http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm)

Seguridad Social (2009 b), *Cómo se organiza la información de la muestra*, en [http://www.seg-social.es/Internet\\_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm](http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra Continua de Vidas Laborales/CONTUTIL2009/index.htm)

Seguridad Social (2012), *Diseño de registro de la MCVL sin datos fiscales*, en [http://www.seg-social.es/Internet\\_1/Estadistica/Est/Muestra Continua de Vidas Laborales/Descripción de ficheros y variables/MCVLSINDATOS/index.htm](http://www.seg-social.es/Internet_1/Estadistica/Est/Muestra Continua de Vidas Laborales/Descripción de ficheros y variables/MCVLSINDATOS/index.htm)