

## **CAPÍTULO IV:**

# **DURACIÓN DEL DESEMPLEO Y DE LA OCUPACIÓN**

---

### **1. INTRODUCCIÓN**

En el capítulo segundo abordamos, desde un punto de vista meramente descriptivo, el proceso de búsqueda y duración del primer y último empleo de los técnicos y técnicos superiores que componen nuestra muestra.

En este presente capítulo nuestro objetivo es comprender cómo se produce el proceso de inserción de los titulados en el mercado de trabajo y cuál es la trayectoria ocupacional subsiguiente una vez insertos en el mismo, analizando qué factores influyen en la mayor o menor duración de cada etapa de dicha trayectoria ocupacional. Para ello nos basaremos en la metodología de modelos de duración, centrándonos, en esta ocasión, en tres etapas que consideramos fundamentales en la historia laboral de los titulados:

1. El período de desempleo previo al primer empleo
2. El primer período de ocupación
3. El último período de ocupación

En lo que respecta a la primera etapa, nuestro interés se centra en la situación en la que el titulado, que ya ha finalizado su período de formación técnico-profesional, toma la decisión de incorporarse al mercado de trabajo. Es a partir de este momento, cuando nos cuestionamos por los factores que pueden ser influyentes en la duración de su primer período de desempleo. Es decir, qué factores provocan que, una vez que el titulado ha decidido comenzar a buscar empleo, su proceso de búsqueda se prolongue durante más o menos tiempo.

No obstante, no finaliza en ese punto nuestro interés en cuanto al proceso de inserción de los titulados al mercado de trabajo, puesto que, de poco vale que los titulados

manifiesten un corto período inicial de desempleo si, una vez que han conseguido su primer empleo, la permanencia en el mismo resulta ser corta. Es por ello por lo que consideramos pertinente el análisis de una segunda etapa en el proceso de inserción laboral, es decir, el estudio de la duración del episodio de ocupación.

En lo que respecta a dicho estudio, el interés se centra de nuevo en la detección de los factores que pueden influir en su mayor o menor duración, tratando de distinguir si la influencia de los mismos se mantiene invariante a lo largo del tiempo. Para ello centramos nuestro análisis en dos episodios de ocupación diferentes: el primero de ellos será aquél que haga referencia al primer empleo conseguido por técnicos y técnicos superiores; mientras que el último, hará referencia al último puesto de trabajo de aquellos titulados que habían experimentado algún episodio de movilidad laboral hasta el momento de realización de la encuesta.

El fenómeno de la duración del desempleo ha sido objeto de interés de numerosas investigaciones, realizadas tanto dentro, como fuera de España. No obstante, en la mayoría de las ocasiones, los estudios, lejos de centrarse exclusivamente en el primer período de desempleo tras la finalización de una titulación determinada, se centran en la duración del período de desempleo en cualquier instante de la vida laboral (entre los más recientes cabe citar a Bover, Arellano y Bentolila, 1997; Bratberg y Nilsen, 1998; Stancanelli, 1998; Melkersson, 1999; Alba-Ramírez y Álvarez-Llorente, 2001; Petrongolo, 2001; Gámez-Amián y García-Pérez, 2002 y Kettunen, 2002).

Aún así, nos encontramos con algunas investigaciones recientes en España, que centran parte de su análisis en aquel período de desempleo que se produce nada más abandonar el sistema educativo. Este es el caso de la tesis doctoral realizada por Gil-Jurado (1998), quien analiza la duración del primer período de desempleo de los titulados universitarios de la Universidad de Las Palmas de Gran Canaria entre 1988 y 1996; la realizada por Salas-Velasco (2000), quien analiza lo propio para los titulados universitarios de la Universidad de Granada registrados en Colegios Profesionales; así como la realizada por Aguilar-Ramos (2001), quien analiza la duración de dicho período de desempleo para dos cohortes diferentes de jóvenes de la Encuesta Sociodemográfica de 1991.

De la misma manera, otro estudio reciente en España, es el que realizan Congregado-Ramírez y García-Pérez (2002) utilizando datos del módulo de transición de la educación inicial al mercado laboral, que incorpora la EPA en el segundo trimestre de 2000. Estos investigadores estiman modelos de duración para el período de transición entre la educación y el mercado laboral de los individuos entre 16 y 35 años, con la finalidad de indagar en los factores económicos e individuales que facilitan/dificultan dicha transición en Andalucía. De esta manera, entre otros factores, encuentran evidencias de la influencia del nivel educativo, tanto propio como de sus progenitores, así como de la situación profesional de los últimos, sobre la tasa de salida del

desempleo. De hecho, comprueban cómo, cuando sus progenitores están ocupados y su nivel de estudios es alto, así como cuando su propio nivel educativo es alto, la duración en el primer período de desempleo de los individuos es menor. Igualmente observan que los especializados en disciplinas de técnicas y biomédicas tienen una mayor probabilidad de salida del desempleo, que los especializados en arte y humanidades.

Por otra parte, Biggeri et al. (2001) realizan un análisis similar, utilizando para ello los datos de una encuesta de oportunidades laborales realizada por el Instituto Nacional de Estadística Italiano a los titulados universitarios graduados en 1992. A partir de la misma, analizan las transiciones desde el sistema educativo hacia el mercado laboral, centrándose en el tiempo necesario hasta obtener el primer empleo. De esta manera, encuentran evidencias de la importancia de la finalización de los estudios en el tiempo establecido, de la necesidad de obtención de notas altas en el caso de las mujeres (y no de los hombres), así como de poseer experiencia laboral previa, para reducir la duración de dicho período de desempleo.

En lo que respecta a la duración del período de ocupación, a pesar de ser éste un fenómeno de interés para los economistas, la pauta común en la mayoría de los estudios, consiste en centrarse en el análisis de la duración del período de ocupación, con independencia del momento de la vida laboral del individuo de que se trate.

Así, por ejemplo, nos encontramos con investigaciones que se refieren a la duración de un primer período de ocupación, como pueden ser las de Nilsen y Risa (1999), realizada para aquellos finlandeses que finalizaron sus estudios entre otoño de 1989 y primavera de 1990 y Bratberg y Nilsen (1998), quienes realizan su estudio para los finlandeses que completaron su educación entre enero de 1989 y septiembre de 1991. En cambio, por otro lado, nos encontramos con otra variedad de estudios, dentro de la literatura de modelos de duración, que, a pesar de analizar la duración del empleo, no se centran precisamente en el primer período de ocupación (Espinal-Berenguer, 1992; Toharia et al., 1998; Taylor, 1999; Alba-Ramírez y Álvarez-Llorente, 2001 y Gámez-Amián y García-Pérez, 2002).

En nuestro caso, como ya hemos comentado, al igual que ya hizo Ibáñez-Pascual (1999) para los jóvenes entre 19 y 29 años que residían en la zona central urbana de Asturias en 1992, centramos nuestro estudio acerca de la duración del episodio de ocupación, en dos momentos diferentes de la trayectoria ocupacional de los titulados que componen nuestra muestra: el primer y el último empleo.

Comenzamos nuestro estudio delineando, en el segundo epígrafe, un marco teórico en el que encuadrar las diferentes etapas de toda trayectoria ocupacional. En primer lugar, introducimos los principios fundamentales de aquellas teorías de competencia imperfecta que hablan de los motivos de existencia de un volumen de desempleo en el

mercado de trabajo. De igual forma, introduciremos aquellas otras teorías que hablan de los procesos de rotación en el mercado laboral, como resultado de la búsqueda de compatibilización, por parte de los individuos, entre la formación recibida y el puesto de trabajo.

A continuación, en el tercer apartado del presente capítulo, realizamos una revisión de la metodología de modelos de duración, que nos permita, posteriormente, optar por la herramienta más afín a nuestros objetivos. En él, revisaremos los conceptos fundamentales de la modelización no paramétrica, la semiparamétrica y, por último, de la paramétrica.

Finalmente, una vez establecido el marco teórico y metodológico, procedemos, en los apartados cuarto y quinto, a la estimación de modelos de duración, mediante métodos paramétricos y no paramétricos del primer período de desempleo, el primer período de ocupación y el último período de ocupación de los titulados que constituyen nuestra muestra.

## 2. MARCO TEÓRICO PARA LA ROTACIÓN LABORAL

Como hemos visto en los capítulos segundo y tercero, nuestros datos señalan que existe una determinada proporción de la muestra cuyo acoplamiento con el puesto de trabajo implica un desajuste educativo por exceso (sobre- educación y habilitación) o por defecto (sub- educación y habilitación) y, en algunos casos, incluso la falta de relación entre los estudios realizados y los puestos de trabajo ocupados. Esta situación, no obstante, no es exclusiva ni de la economía canaria, ni del mercado de trabajo de los técnicos y técnicos especialistas, sino que se da en muy diversos países y con diferente magnitud<sup>99</sup>. Igualmente hemos visto la explicación que aportan diversas teorías a la existencia del fenómeno de desajuste educativo. En este capítulo, sin embargo, veremos, entre otras cosas, cómo se relaciona dicho desajuste con la movilidad de los trabajadores dentro del mercado laboral.

Son diversas las teorías que pueden aportar luz sobre la cuestión desde diferentes perspectivas. Por una parte están las teorías de la competencia imperfecta relacionadas con la movilidad voluntaria de los trabajadores, como puede ser la teoría de '*insiders y outsiders*' (Lindbeck y Snower, 1986, 1988a y 1988b) o la de 'salarios de eficiencia' (Weiss, 1980 y 1990), y por otra parte las teorías que consideran la movilidad entre empresas como un mecanismo corrector del desajuste, como la teoría del '*job matching*' (Jovanovic, 1979a y 1979b y Sicherman, 1991) y la de la 'movilidad ocupacional' (Rosen, 1972 y Sicherman y Galor, 1990). En cualquier caso, todas ellas tratan de dar una explicación a la permanente existencia de desempleo y movilidad laboral dentro del mercado de trabajo.

### 2.1. TEORÍAS DE COMPETENCIA IMPERFECTA

En el marco de la literatura sobre desempleo la visión que aportan Lindbeck y Snower (1986, 1988a y 1988b) desde el punto de vista '*insider-outsider*' suele ser considerada como una alternativa a la visión que aporta Weiss (1980 y 1990) desde los 'salarios de eficiencia'. No obstante, más que como alternativas excluyentes, éstas se ven como complementarias. La visión de Lindbeck y Snower considera el desempleo como una consecuencia del poder negociador de los trabajadores ya ocupados ('*insiders*') para elevar la renta partiendo de los costes laborales derivados de la movilidad. Mientras que la visión de Weiss explica el desempleo como consecuencia de la información asimétrica existente en el mercado de trabajo, que da poder de mercado a las empresas a la hora de determinar los salarios.

En cualquier caso, ambas teorías consideran que la movilidad de los trabajadores no es voluntaria y que, en caso de existir, ésta es baja. Por tanto, bajo estas perspectivas, el

<sup>99</sup> Para más detalle, ver la Tabla III.1 del apartado 3.2.3. (Capítulo III).

desajuste educativo existente en el mercado de trabajo es probable que sea un fenómeno persistente, dado que la movilidad externa no va a funcionar como mecanismo corrector.

A continuación, con el fin de ahondar más en cada una de las perspectivas propuestas por estos autores, las abordamos con más detalle. Comenzamos, en primer lugar, hablando de la teoría de 'salarios de eficiencia', por ser ésta la más antigua y por haber sido, de hecho, la precursora de la teoría de '*insiders-outsiders*'.

### 2.1.1. Teoría de 'salarios de eficiencia'

En los modelos tradicionales de mercado de trabajo los salarios vienen determinados a través de una sencilla igualdad entre oferta y demanda. Sin embargo, estos modelos no eran capaces de explicar fenómenos como la existencia de 'colas de empleo' y los consiguientes despidos ocasionados por la rigidez de los salarios. La teoría de 'salarios de eficiencia' desarrollada por Weiss (1980, 1990), sin embargo, aporta una explicación a estos fenómenos. Para ello, parte de dos supuestos: "(1) *los salarios que cobran los trabajadores no son proporcionales a su productividad, pero (2) los salarios de aceptación de los trabajadores son función creciente de su productividad.*" Estos supuestos le ayudan a generar un modelo de mercado de trabajo en el que el salario ofrecido por las empresas afecta, no sólo al número de candidatos, sino también a las características productivas esperadas en los potenciales contratados.

A pesar de que en el mercado de trabajo se puede observar cómo algunas compañías ofrecen a los trabajadores más productivos mayores probabilidades de ser promocionados, o preferencia a la hora de elegir sus vacaciones, o incluso son objeto de una menor supervisión, Weiss considera que es más realista suponer una remuneración idéntica entre trabajadores de un mismo nivel jerárquico, antes que una remuneración proporcional a las habilidades del individuo. Entre otros motivos, el más obvio radica en la consideración de que, obtener información precisa de la productividad de sus trabajadores puede resultar a la empresa más caro que los propios beneficios que se deriven de dicha información. Igualmente, las empresas tampoco pagarían salarios acordes a una productividad esperada, puesto que pagar salarios diferentes a trabajadores con las mismas características observables puede originar resentimientos que reduzcan la propia productividad de los trabajadores, causen fricciones que induzcan el aumento de los abandonos de empleo o inciten a unos trabajadores a dificultar la labor de sus compañeros. Aparte de todo esto, es ya una práctica habitual en el mercado de trabajo, la existencia de convenios colectivos que implican el pago homogéneo de salarios entre empleados de una misma categoría.

La idea central del análisis realizado por Weiss aporta una nueva explicación al hecho de que los salarios no disminuyan cuando aparece un exceso de oferta de trabajadores dispuestos a trabajar cobrando menos. Para ello se centra en el efecto clasificador de

los salarios, de forma tal que la información asimétrica, junto con la correlación positiva entre salarios de aceptación y potencialidad laboral<sup>100</sup>, permiten a la empresa ofrecer un salario superior al *market-clearing wage* con el fin de seleccionar a su personal de entre un 'mejor' conjunto de candidatos.

Según explica Weiss (1980), si con un salario determinado la empresa consigue atraer suficientes candidatos como para satisfacer su demanda, ofreciendo un salario superior será capaz de atraer trabajadores más competentes. De esta manera, a la hora de seleccionar aleatoriamente un trabajador del conjunto de demandantes, aumenta la probabilidad de seleccionar uno con características productivas superiores.

La empresa, en el modelo de 'salarios de eficiencia', no está interesada en elegir el salario mínimo con el que satisface su demanda de trabajadores, sino en elegir el salario que le ayude a minimizar el coste por unidad de trabajo eficiente. Dado que los salarios determinan la calidad del trabajador a contratar y las empresas buscan la máxima calidad, el salario que logra minimizar los costes por unidad eficiente de trabajo para cada empresa, puede llegar a ocasionar un exceso de oferta de trabajadores en el mercado. Sin embargo, un trabajador que se encuentre en la 'cola de empleo', no podrá salir de esa situación ofreciendo su fuerza laboral a cambio de un menor salario, puesto que eso indicaría a la empresa cuál es el límite superior en su salario de aceptación y, consecuentemente (por el segundo supuesto), indicaría cuál es el límite superior de su capacidad laboral. Precisamente es debido a esto, por lo que Weiss considera que las 'colas de empleo' seguirían existiendo.

Igualmente, la teoría de 'salarios de eficiencia' da una explicación a la razón por la que las empresas despiden trabajadores como respuesta a un descenso en la demanda de su producción. Según argumenta Weiss (1990), las empresas que tuvieran que hacer recorte de personal, si optaran por mantener el mismo número de trabajadores, pero pagándoles un salario inferior, probablemente perderían a los trabajadores más hábiles, quienes encontrarían fácilmente empleo en el mercado de trabajo. Para evitar esta situación, las empresas prefieren optar por un despido no selectivo de trabajadores antes que reducir el salario de los mismos y provocar el abandono de los más productivos.

Aparte de poder explicar la existencia de 'colas de empleo' y despidos en el mercado de trabajo, la teoría de 'salarios de eficiencia' es capaz de dar una explicación a la existencia de diferentes tasas de desempleo y probabilidades de despido entre grupos. Basta con determinar el salario que minimiza los costes de contratación por unidad de trabajo eficiente considerando cada grupo de forma diferente. A partir de ahí, para un determinado nivel de demanda agregada, unos grupos se enfrentarán a un exceso de demanda por sus servicios, recibiendo un salario superior al que realmente minimiza sus

---

<sup>100</sup> Durante el desarrollo de la teoría de 'salarios de eficiencia' hablaremos indistintamente de potencialidad laboral, características productivas y capacidad laboral.

costes, mientras que otros, para el salario que minimizaría sus costes, se enfrentarán a la existencia de 'colas de empleo' o, incluso, a la nula contratación de su colectivo.

Como vemos, bajo esta teoría, las empresas son las que tienen la totalidad del poder de mercado, pues son quienes, en un contexto de información asimétrica, deciden el salario y el nivel de ocupación. Además, éstas pueden pagar salarios por encima de su nivel de competencia con objeto de reducir las indeseables consecuencias de los abandonos. Este comportamiento sigue siendo un comportamiento racional, pues, como demuestra Martin (1997) a través de un sencillo estudio que analiza el efecto de los 'salarios de eficiencia' sobre la productividad y la movilidad, el coste marginal de aumentar el salario, es más que compensado por el beneficio marginal de la reducción en la movilidad externa de sus trabajadores y por el logro de una mayor productividad.

Bajo los supuestos de esta teoría, el desempleo es contemplado como un problema de entendimiento, medido en términos de conflicto de intereses, entre las empresas y los trabajadores desempleados, pues éstas no están interesadas en contratar a los desempleados pagándoles un salario inferior, ya que, para ellas, los salarios son una señal de su productividad. Ésta es, como veremos a continuación, la gran diferencia entre la teoría de 'salarios de eficiencia' y la de '*outsiders-insiders*'.

### 2.1.2. Teoría '*insider-outsider*'

La teoría '*insider-outsider*' de Lindbeck y Snower (1986, 1988a y 1988b), al igual que la de 'salarios de eficiencia' de Weiss, trata de dar una explicación a la existencia de desempleo involuntario en las economías de países desarrollados. Concretamente, se centra en la explicación del por qué los desempleados no son capaces de encontrar un trabajo, a pesar de estar preparados para trabajar, cobrando incluso salarios inferiores a los que cobran los trabajadores ocupados.

Según argumentan, en el mercado de trabajo no se suele dar una puja, ni por parte de los trabajadores desempleados ni por parte de las empresas, para desempeñar un empleo cobrando salarios inferiores a los vigentes en el mercado. En caso de darse esta práctica y existir desempleados dispuestos a asumir estas condiciones, se observaría la desaparición del desempleo involuntario, a la vez que se observaría en el mercado de trabajo un descenso gradual en los salarios.

Las preguntas que se plantean para desarrollar su hipótesis son realmente las mismas que se plantea la teoría de 'salarios de eficiencia', es decir: a. ¿por qué no pueden / quieren los desempleados hacer uso de esta estrategia para salir de su situación de desempleo? b. ¿Por qué los trabajadores despedidos no pueden / quieren mantener su empleo ofreciendo trabajar por un sueldo inferior?

La explicación más obvia a estas cuestiones radica en la existencia de normas sociales que consideran que este tipo de comportamiento no es aceptable. Desde el punto de vista social unos trabajadores no deben ‘robar’ el empleo a otros, mientras que las empresas, a su vez, no deben permitir la existencia de este tipo de ‘robos’<sup>101</sup>. El que los primeros no deban tener este tipo de comportamiento radica en el hecho de que, en caso de hacerlo, probablemente las relaciones personales que tendrían con sus compañeros de trabajo serían bastante hostiles. Mientras que la motivación de las empresas para evitar el ‘robo’ de empleos radica en la sospecha de que, en caso de permitirlo, es posible que los trabajadores de mayor experiencia opten por no cooperar con los nuevos trabajadores, con lo que la productividad de los últimos no sería la deseable.

Esta teoría, a diferencia de la de ‘salarios de eficiencia’, considera que no sólo las empresas tienen poder negociador en el mercado laboral, sino que una parte de ese poder también está en manos de los trabajadores. La idea subyacente implica la existencia de costes laborales derivados de la movilidad, que son susceptibles de generar rentabilidad económica. Estos son los que conceden el poder de mercado a los trabajadores ocupados (*‘insiders’*) quienes, en el proceso de determinación de los salarios, manipulan y explotan dichos costes teniendo en cuenta ante todo sus propios intereses y dejando fuera del proceso de negociación salarial a los trabajadores desempleados, lo que finalmente provoca la aparición del desempleo involuntario.

La distinción que realizan Lindbeck y Snower entre *‘insiders’* y *‘outsiders’* tiene mayores implicaciones que la mera distinción entre ocupados y parados. Estos conceptos realmente sirven para diferenciar a grupos de trabajadores con diferentes oportunidades laborales, puesto que los *‘insiders’* se enfrentan a oportunidades mucho más favorables que los *‘outsiders’*, al ser quienes utilizan deliberadamente los costes de movilidad con el único objetivo de incrementar sus propios salarios. La postura que los *‘insiders’* defienden en la negociación está basada en el hecho de que a la empresa le resultaría demasiado costoso reemplazar a su plantilla actual (*‘insiders’*) por nuevos trabajadores (*‘outsiders’*). Estos últimos preferirían estar ocupados, cobrando los salarios de los *‘insiders’* y con las condiciones laborales de estos, antes que continuar en su situación de desempleo. No obstante son víctimas de un proceso de discriminación, puesto que, en el caso de conseguir el empleo ofertando trabajar por un salario inferior, reciben una menor cooperación, con la consiguiente disminución en su productividad, y sufren un mayor acoso por parte de los propios *‘insiders’*, con la consiguiente pérdida de funcionalidad del propio puesto de trabajo para la empresa. De hecho, los *‘outsiders’* estarían dispuestos a trabajar cobrando un salario más bajo que el de los *‘insiders’* de forma tal que la diferencia compensara a la empresa por sus menores habilidades de trabajo cooperativo. Pero, aún así, son incapaces de encontrar empleo. Entonces, dado que los propios *‘outsiders’* llegan a considerarse a sí mismos como trabajadores menos

---

<sup>101</sup> Utilizamos el término ‘robar’ manteniendo la terminología utilizada por Lindbeck y Snower (1986) para referirse a las formas aceptables de comportamiento social.

productivos y, desde su punto de vista, el puesto de trabajo disminuye su utilidad, resulta imposible fijar un salario que consiga a la vez inducir a las empresas a contratar *'outsiders'* e inducir a estos últimos a aceptarlo. Es en este sentido, en el que surge el desempleo involuntario.

Bajo condiciones idénticas de empleo, un *'outsider'* sería, en términos de salario de eficiencia, más barato para la empresa que un *'insider'*, debido a las diferencias de productividad, pluses de antigüedad, etc. Sin embargo, la razón por la que las empresas no sustituyen a los *'insiders'* por los *'outsiders'* es que esos trabajadores realmente no compiten en igualdad de condiciones. De hecho, bajo condiciones reales, en las cuales los *'insiders'* reciben cooperación de sus compañeros y no sufren ningún tipo de acoso mientras que los *'outsiders'* se enfrentan a las condiciones opuestas, estos últimos resultan más caros a la empresa en términos de salario de eficiencia. Por tanto, la contratación de nuevos empleados por menor salario es un hecho que no se da debido a que los propios *'insiders'* desincentivan a las empresas. Vemos, por tanto, cómo el resultado del conflicto de intereses entre *'insiders'* y *'outsiders'* finalmente conlleva a un nivel de salarios que ocasiona la aparición de desempleo involuntario. Además, estos últimos son incapaces de salir de esa situación ofreciendo cobrar salarios más bajos, puesto que los propios *'insiders'* se ocupan de que esto le resulte más costoso a la empresa<sup>102</sup>.

Son los propios Lindbeck y Snower (1988) quienes se plantean posibles objeciones a su teoría para, posteriormente, tratar de dar respuesta a las mismas. Así, las objeciones que se plantean son las siguientes:

1. Si los trabajadores ocupados son capaces de restringir la movilidad laboral en sus empresas, ¿cómo es posible que en la práctica las empresas sigan presentando grandes ratios de movilidad?

A lo que argumentan que, si su modelo contemplase las jubilaciones y los abandonos voluntarios de empleo, también recogería cierta movilidad laboral, aunque siempre limitada a un máximo.

2. ¿La existencia de desempleo voluntario radica en el hecho de que el número de empresas del mercado es fijo? ¿Desaparecería este desempleo en caso de dejar entrar más empresas en el mercado?

A este respecto argumentan que el mero hecho de que los *'insiders'* dejen fuera a los *'outsiders'* no es motivo suficiente para que nuevas empresas consideren rentable contratar a estos *'outsiders'*. Es más, dado que, en principio éstas no podrían

---

<sup>102</sup> En general, los costes de movilidad que argumentan los *'insiders'* en su proceso de negociación, pueden venir disfrazados de diferentes formas. Aunque aquí hemos hablado fundamentalmente de la capacidad de los *'insiders'* para decidir colaborar o no con los compañeros y crear o no un buen ambiente de trabajo (con el consiguiente efecto sobre la productividad y utilidad del empleo), estos costes pueden tomar la forma de costes de reclutamiento de nuevos empleados, formación específica de recién contratados o incluso despido de *'insiders'* (Lindbeck y Snower, 1986).

contratar a ‘insiders’, la productividad de sus trabajadores sería inferior y, además, tendrían que incurrir, entre otros, en costes de establecimiento.

3. Dado que los ‘insiders’ pueden evitar que los ‘outsiders’ consigan empleos, ¿pueden evitar también el aumento de empleo tras un período de recesión?

Finalmente, en lo que respecta a esta última objeción, tras desarrollar su teoría bajo tres escenarios diferentes, concluyen que es imposible que esta potencial objeción se mantenga.

## 2.2. TEORÍAS DE MOVILIDAD ENTRE EMPRESAS

Las teorías de competencia imperfecta tienen como objetivo fundamental a lo largo de toda su argumentación la explicación del porqué de la existencia en el mercado de trabajo de desempleo involuntario. Sin embargo, en el presente capítulo, nuestro interés no se centra únicamente en el desempleo, sino también, en el análisis de la situación de empleo. Es por esto, por lo que consideramos pertinente introducir la visión que aportan las teorías referentes a la movilidad entre empresas.

Como ya introdujimos en el capítulo tercero, las teorías del ‘*job matching*’ y de la ‘movilidad ocupacional’ consideran que la primera etapa del proceso de inserción laboral suele dar lugar a la existencia de desajuste educativo. Es precisamente debido a dicho desajuste, por lo que los trabajadores deciden comenzar un proceso de rotación en el mercado de trabajo, en la búsqueda del mejor ajuste posible dada su preparación.

Por tanto, la teoría del ‘*job matching*’ y la de movilidad ocupacional, al contrario que las teorías basadas en la competencia imperfecta, subrayan el carácter de movilidad entre empresas como un mecanismo corrector del desajuste educativo, contemplando en su argumentación aspectos tales como la antigüedad en la empresa, el emparejamiento o grado de ajuste y la movilidad laboral.

### 2.2.1. Teoría del ‘*job matching*’

La teoría del ‘*job matching*’, desarrollada originalmente por Jovanovic (1979a, 1979b) permite descubrir una cierta relación entre el desajuste educativo y la movilidad laboral. Según argumenta Jovanovic, si en el mercado de trabajo se observa la existencia de trabajadores con características que difieren de los requerimientos del puesto que ocupan, entonces nos encontramos ante una señal de mal ajuste. Éste, como ya hemos visto en el capítulo tercero, podrá tratarse de un mal ajuste por exceso o por defecto de conocimientos / habilidades. En el caso de tratarse de un desajuste por sobreeducación, es de esperar que la educación formal del individuo realmente le permitiera alcanzar un mejor puesto de trabajo que el que está ocupando, por lo que resulta más probable que los trabajadores sobreeducados opten por cambiar de empleo. En cambio, en el caso

contrario (subeducación), sería de esperar que la permanencia del trabajador en su puesto de trabajo fuera más prolongada, puesto que cambiar de empleo a uno adecuado a sus conocimientos le implicaría acudir a uno de menor categoría, con la consiguiente disminución en los salarios. Claro que, desde la perspectiva de la empresa, la subeducación también se puede considerar como un mal ajuste y, por tanto, a la empresa le podría llegar a interesar inducir al trabajador a abandonar su empleo.

Como vemos, el modelo de *'job matching'* enfatiza el carácter temporal de la inadecuación. Según argumenta, los individuos que presentan unas habilidades personales determinadas pueden resultar más adecuados para unos empleos que para otros. Entonces, asumiendo que los individuos buscan compatibilizar su formación con los requerimientos del empleo, de forma que la discrepancia sea la menor posible, la existencia de un desajuste educativo es un indicador de una mala compatibilización, que les lleva a decidir subsanar dicha situación. Es precisamente esta relación causa-efecto, lo que lleva a Jovanovic a catalogar a su modelo entre aquellos que predicen la existencia de movilidad laboral como consecuencia de la obtención de información sobre el grado de ajuste existente en el actual empleo. A este tipo de modelos los denomina modelos para los que el trabajo es un bien *'experimental'*, en el sentido de que, el único modo posible para determinar la calidad de un determinado ajuste consiste en llevarlo a cabo y experimentar el grado de ajuste o desajuste.

Los supuestos de los que parte Jovanovic para desarrollar su modelo son tres: 1. Cada trabajador presenta su propia función de productividad para diferentes trabajos; 2. Los empleadores pueden contratar a trabajadores basándose en sus características individuales; 3. Existe información imperfecta en ambas partes del mercado acerca de cuál sería la asignación óptima de trabajadores a empleos, por lo que se hace una asignación inicial y, a medida que se va obteniendo información, se van realizando posteriores reasignaciones<sup>103</sup>.

A partir de estos supuestos, el modelo desarrollado por Jovanovic (1979a) predice que los trabajadores permanecerán en empleos en los que su productividad se mantenga relativamente alta y abandonarán aquellos empleos para los que perciban que su productividad está siendo baja. Dado que la determinación de salarios se realiza en función de la productividad marginal esperada del individuo, el modelo muestra cómo el salario se va incrementando a medida que aumenta la antigüedad en la empresa. Además, al existir una correlación positiva entre antigüedad y experiencia laboral, se espera un aumento salarial a lo largo del ciclo vital. El modelo de *'job matching'* también predice que la probabilidad de separación del empleo decrece con los años de antigüedad en el mismo, debido a que la detección de los desajustes se realiza fundamentalmente al comienzo. Además, según argumenta, el mecanismo de

---

<sup>103</sup> Según comenta Jovanovic (1979b): "No existen *'buenos'* trabajadores y *'buenos'* empresarios, sino solamente *'buenos ajustes'*".

aprendizaje funciona de tal manera que, cuanto más tiempo lleve el individuo en su puesto de trabajo, mayor será su probabilidad de permanencia.

Jovanovic (1979b) continúa el razonamiento explicando que la probabilidad de separación del puesto, dada la antigüedad, es menor entre aquellos que presentan un buen ajuste, puesto que estos dedican menos tiempo a buscar un empleo alternativo y, además, es menos probable que acepten una oferta en caso de recibirla. Por otra parte, dado el ajuste o desajuste entre trabajador y puesto de trabajo, cuanto más específico sea el capital que el trabajador haya invertido en el emparejamiento, más probable será que continúe en el puesto, de donde se deduce que la antigüedad crece con la especificidad del capital, y en consecuencia la probabilidad de salida disminuye con dicha especificidad. Por tanto, como vemos, la teoría del '*job matching*' predice la existencia de un efecto negativo de la antigüedad, y uno positivo del desajuste entre trabajador y puesto de trabajo, sobre la movilidad externa.

No obstante, formulaciones posteriores a la teoría del '*job matching*', que integraban la misma con la del '*job training*', sugieren que las transiciones, en lugar de disminuir, aumentarían con la antigüedad. De hecho, Mortensen (1988), sostiene como válida esta afirmación, pues considera que la dispersión de los salarios futuros dentro de un mismo empleo decrece con la antigüedad y, por tanto, los aumentos salariales esperados también disminuyen, con lo que, para el trabajador, el valor de continuar ocupado en el mismo empleo disminuirá con la antigüedad. Como consecuencia, dada la preferencia de los trabajadores por una mayor dispersión en las ofertas salariales futuras, aunque esto incorpore un factor riesgo superior, estos tienden a cambiar de empresa con mayor intensidad a medida que se incrementan los años de permanencia en la misma. Además, esta tendencia, según comenta Mortensen, será mucho más pronunciada entre los trabajadores que se encuentren sobreeducados, ya que estos son los que tienen una mayor probabilidad de encontrar mejores ofertas salariales al cambiar de empleo.

Finalmente, la teoría de '*job matching*', también hace aportaciones acerca de la existencia de desempleo. Según explica Jovanovic (1979b), los desempleados serán, predominantemente, aquellos trabajadores jóvenes que han tenido poca suerte en la búsqueda de un empleo para el que presenten buen ajuste y que, por tanto, cobraron salarios bajos y permanecieron poco tiempo en la empresa. El tiempo dedicado a la búsqueda de empleo será mayor cuanto peor sea el grado de ajuste y cuanto menor inversión en capital específico presente el individuo.

### **2.2.2. Teoría de la 'movilidad ocupacional'**

La teoría de la 'movilidad ocupacional', desarrollada inicialmente por Rosen (1972) y ampliada posteriormente por Sicherman y Galor (1990), se basa en la hipótesis de que los individuos aprenden de su experiencia laboral, modelizando así el papel del mercado

de trabajo como transmisor de habilidades y conocimientos. De esta manera distingue entre un 'mercado de oportunidades de aprendizaje' y un mercado de ocupaciones.

Esta teoría parte de la base de que los individuos, con el fin de maximizar sus expectativas de rentas, eligen cómo distribuir su tiempo, decidiendo qué nivel de formación alcanzar y qué tipo de ocupación querrían obtener. No obstante, Rosen (1972) remarca que gran parte de la formación que presentan los individuos no proviene directamente de la educación formal, sino de la propia experiencia laboral. Según sus propias palabras: *"La ecuación no se produce solamente en los colegios, y el proceso de aprendizaje no cesa tras la graduación"*. Sea como sea, a partir de su nivel formativo, el individuo podrá obtener dos tipos de renta: a) una renta directa, como es la posibilidad de obtener mejores ingresos potenciales cuanto mayor capital específico aporte el individuo a su puesto de trabajo; y b) una renta indirecta, como es el aumento en la probabilidad de ser promocionado dentro de la empresa, o incluso fuera de ella, cuanto mayor nivel educativo demuestre poseer.

Dado que la regla de comportamiento de los trabajadores, bajo la teoría de la 'movilidad ocupacional', sigue el patrón de comportamiento previsto por la teoría del capital humano, los individuos decidirán si abandonar o continuar en un puesto de trabajo en función del valor actualizado de los ingresos futuros que esperan obtener mediante su comportamiento, analizando el valor actual de los ingresos futuros esperados en el caso de optar por una u otra decisión. De esta manera, la teoría sugiere que si en una ocupación los rendimientos del nivel de formación son bajos, entonces, la probabilidad de promoción ocupacional aumenta. Así, considera que los trabajadores pueden estar interesados en permanecer temporalmente en empleos que les doten de determinadas habilidades, si posteriormente pueden utilizar las mismas en otros empleos de mayor categoría. No obstante, las empresas no ofertan oportunidades de aprendizaje sin recibir nada a cambio, por lo que, en el mercado de trabajo, se observa una verdadera 'venta' de trabajos entre potenciales trabajadores. Estos demandan oportunidades de aprendizaje y están dispuestos a pagar por ellas, puesto que saben que sus habilidades y conocimientos, y por consiguiente sus ingresos futuros, se verán incrementados.

Por tanto, según este razonamiento, para un trabajador podría ser beneficioso permanecer cierto tiempo trabajando en un puesto de trabajo para el que realmente estuviese sobreeducado cobrando un salario inferior al adecuado a su formación. La diferencia entre la verdadera rentabilidad de sus habilidades y el salario que realmente está cobrando, constituye el precio sombra que el individuo está pagando por su proceso de aprendizaje. El fenómeno de la sobreeducación quedaría explicado así como un proceso de optimización. A través de dicho proceso, el individuo adquiere herramientas de capital humano, que utiliza posteriormente para acceder a ocupaciones de mayor nivel a través de la promoción interna o externa. Por tanto, según la teoría de la 'movilidad ocupacional', la sobreeducación no es más que un desajuste temporal,

puesto que los trabajadores con un mayor nivel educativo, en un futuro serán capaces de promocionar hacia empleos de mayor nivel.

En su desarrollo de la teoría de la 'movilidad ocupacional', Rosen (1972), pone sus miras en las posibilidades de promoción que tienen los individuos dentro de la propia empresa. Según argumenta, los trabajadores diseñan de antemano sus propios planes de carrera, tomando en consideración los diferentes puestos u ocupaciones y jerarquizándolos según el nivel de responsabilidad que esperan ir desempeñando a lo largo de su vida laboral. Sin embargo, Sicherman y Galor (1990) incorporan a la movilidad interna contemplada por Rosen el aspecto de movilidad externa. Así, proponen un modelo en el que un mismo trabajador se enfrenta ante dos caminos diferentes: 1. promocionar dentro de su misma empresa; 2. abandonar la misma y acceder a otra empresa diferente. De esta manera argumentan que, si un trabajador prevé que en la empresa actual no podrá realizar sus expectativas, a pesar de cumplir con los requisitos necesarios para ser promocionado, entonces, probablemente, y fundamentando su razonamiento en las bases del capital humano, optará por abandonarla. Según comentan Sicherman y Galor, la movilidad interna (o promoción) es algo que está sujeto a la decisión del empresario, mientras que la movilidad externa viene determinada única y exclusivamente por decisiones personales del propio trabajador, quien determina cuál es el mejor momento para abandonar su actual puesto de trabajo con el fin de maximizar sus expectativas de ingresos futuros.

La teoría de la 'movilidad ocupacional', por tanto, sugiere que, mientras que en determinadas ocupaciones el rendimiento de los años de educación se mide en términos de mayores salarios, en otras, dicho rendimiento se mide en mayores probabilidades de avanzar hacia empleos en los que se pagan mayores salarios. Por otro lado, también sugiere que, dada una ocupación de origen, los individuos con un mayor nivel educativo presentan una mayor probabilidad de promocionar hacia una ocupación de mayor nivel jerárquico. Igualmente postula que entre individuos que no han promocionado, aquellos con una mayor probabilidad de hacerlo, serán también quienes tengan una mayor probabilidad de abandonar la empresa. Finalmente, ésta teoría, al contrario de la teoría del '*job matching*', prevé una relación positiva entre la antigüedad en la empresa y la movilidad laboral. Este hecho lo argumenta basándose en la hipótesis de elección voluntaria de desajuste educativo por parte de los individuos, con el único objeto de adquirir formación y experiencia que les capacite para posteriormente poder acceder a otra ocupación mejor. De esta manera postula que los sobreeducados serán quienes presenten una mayor movilidad hacia otros trabajos, pues tras su estancia en el 'mercado de oportunidades de aprendizaje' irán en la búsqueda de empleos mejor remunerados o con mejores expectativas de promoción.

### 3. MODELOS DE DURACIÓN

Los modelos de duración constituyen un instrumento muy adecuado para tratar de contrastar determinadas hipótesis cuando se dispone de información sobre el movimiento o permanencia de los individuos en diferentes situaciones. La variable de interés en el análisis de duración es el tiempo que transcurre entre el momento en que un fenómeno empieza y el momento en el que acaba o tiene lugar la medición.

El análisis de datos de duración se complica muchas veces por el mero hecho de que la medida, habitualmente, viene expresada de forma incompleta. Es decir, en el momento de la recolección de los datos, es posible que algunas de las observaciones no hubiesen finalizado su proceso, puesto que la medición se hace en un único momento para todos los individuos. Es debido a esto por lo que, para algunos de ellos, la medición puede tener lugar antes de finalizar el fenómeno, en cuyo caso estaríamos ante una situación de censura en los datos.

Al hablar de variables relacionadas con el factor tiempo nuestro interés puede estar centrado en la duración del fenómeno de estudio, pero también puede interesarnos conocer algo sobre la verosimilitud de que el suceso finalice en un instante próximo sabiendo que el mismo ha llegado hasta la actualidad. Para este tipo de estudios es para lo que se utiliza el ‘análisis estadístico de datos de duración’, aunque, dependiendo del ámbito de aplicación sea conocido por diferentes nombres<sup>104</sup>.

A continuación, vamos a dedicar el presente epígrafe al estudio de las herramientas y técnicas que han demostrado ser útiles en el análisis de datos de duración, poniendo especial énfasis en lo que respecta al ámbito de la econometría aplicada.

#### 3.1. DISTRIBUCIONES DE DURACIÓN

Los procedimientos utilizados en el análisis de duración, como veremos, son bastante útiles para organizar, resumir e interpretar datos. Así, a través de su utilización, podremos representar la información recabada en términos de probabilidades condicionales, lo que finalmente resulta teórica e intuitivamente más atractivo.

---

<sup>104</sup> La implantación de este tipo de técnicas en el ámbito de la ingeniería y de la bioestadística ha sido bastante habitual. Los ingenieros llevan décadas realizando análisis de calidad de sus productos, estando interesados en el análisis del tiempo que transcurre hasta que se llega a producir un fallo en la maquinaria objeto de estudio. Debido a esto, los modelos utilizados con tal fin son conocidos entre los ingenieros como ‘modelos de análisis de fiabilidad’ o ‘modelos de tiempos de fallo’. Por otro lado, los médicos también han estado interesados en los modelos de duración, aplicando los mismos, sobre todo, al análisis del tiempo de supervivencia de un paciente una vez que se le ha diagnosticado una enfermedad. Así, en éste ámbito, se habla de ‘modelos de análisis de supervivencia’. Sin embargo, en el ámbito de la economía y la sociología, la aplicabilidad de este tipo de modelos ha sido más tardía, recibiendo en este último caso el nombre de ‘análisis de historia de eventos’.

### 3.1.1. Función de densidad, de supervivencia y de riesgo

Como introducción a la nomenclatura de este tipo de modelos, comenzaremos viendo algunas de las funciones más útiles para el análisis de duración<sup>105</sup>.

Sea  $T$  una variable aleatoria no negativa continua, que representa el tiempo que transcurre desde el inicio de un evento hasta que éste finaliza. Su función de densidad se denota por  $f(t)$ , en donde  $t$  es una realización de dicha variable aleatoria.

Puesto que el tiempo es una variable continua, desde el punto de vista de la teoría de la probabilidad no tiene sentido calcular la probabilidad de que un suceso finalice en un punto temporal concreto. Sin embargo, sí será interesante dar una cierta medida del riesgo de finalización tras un período determinado, al igual que nos interesará obtener otras medidas, como pueden, entre otras, ser el tiempo medio de supervivencia o la variabilidad de esa media. Es decir, nos interesará el estudio de la distribución de probabilidad de la variable  $T$ .

Suele ser habitual describir  $T$  a partir de su función de distribución, puesto que ésta nos informa acerca de la probabilidad de que  $T$  no tome un valor superior a un período concreto  $t$ . La expresión analítica de dicha función de distribución es la siguiente:

$$F(t) = \int_0^t f(s) ds = \Pr(T \leq t) \quad [\text{Eq. IV.1}]$$

No obstante, más que en la función de distribución, habitualmente estaremos más interesados en la probabilidad de que la duración de  $T$  sea *como mínimo*  $t$ . Este dato se corresponde con el valor complementario de la función de distribución, y es la información que nos proporciona la función de supervivencia  $S(t)$ . Ésta, por tanto, es la que nos mide la probabilidad de que un suceso se prolongue durante  $t$  o más tiempo; probabilidad que decrece con el tiempo.

$$S(t) = 1 - F(t) = \Pr(T \geq t) \quad [\text{Eq. IV.2}]$$

Las expresiones que hemos visto hasta ahora son un acercamiento desde el punto de vista incondicional pero, en la práctica, suele ser más conveniente razonar a través de probabilidades condicionadas. De esta manera, si la probabilidad de que la duración  $T$  finalice en el corto intervalo de tiempo  $(t, t + \Delta t)$  se calcula de la siguiente manera:

$$l(t, \Delta t) = \Pr(t \leq T \leq t + \Delta t | T \geq t) \quad [\text{Eq. IV.3}]$$

<sup>105</sup> Ver, entre otros, Kiefer (1988) y Devine y Kiefer (1991)

dividiendo esta expresión entre el propio incremento temporal podemos determinar cuál es la probabilidad de que la duración de T finalice en el siguiente intervalo de tiempo, de longitud  $\Delta t$ , dado que ya hemos llegado hasta el momento t:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta S(t)} = \frac{f(t)}{S(t)} \quad [\text{Eq. IV.4}]$$

Esta expresión es lo que se conoce con el nombre de ‘función de riesgo’, ‘función de *hazard*’ o ‘tasa instantánea de fallo’ y coincide con la inversa de la ratio de Mill utilizada para corregir problemas de sesgo de selección en los modelos de regresión múltiple<sup>106</sup>. Ésta nos indica, de manera aproximada, cuál es la velocidad a la que se van completando los procesos tras haber llegado a una duración t, puesto que nos informa de la tasa instantánea a la cual finaliza un período de permanencia en un estado, transcurrido un período de tiempo t.

En los modelos de duración, a la función de densidad también se la conoce como tasa de fallo incondicional, en contraposición a la función de riesgo, a la que se conoce como tasa de fallo condicional, puesto que mide la tasa instantánea de fallos en el intervalo  $(t, t + \Delta t)$  condicionada a que el evento T ha durado hasta el instante t.

A la hora de analizar la duración de un fenómeno concreto podemos pensar que, cuanto más haya durado el mismo, mayor será la probabilidad de que termine. Sin embargo, resulta igualmente plausible razonar en el sentido inverso, deduciendo que cuanto más tiempo haya durado el evento en cuestión, mayor dificultad debe existir para que finalice, por lo que es menos probable que acabe en un período corto de tiempo. Obviamente, todo dependerá del tipo de fenómeno estemos analizando y de cuánto tiempo haya transcurrido desde el comienzo del mismo. De hecho, no es sólo la duración en si misma lo que interesa, sino también, la probabilidad de que ésta vaya a terminar en el próximo período teniendo en cuenta el tiempo transcurrido. Por tanto, será precisamente la forma de la función de riesgo la que nos proporcione información acerca de cuál es el proceso estocástico subyacente.

Si al calcular la derivada de la función de riesgo con respecto al tiempo,

$$\frac{\delta h(t)}{\delta t} > 0 \quad [\text{Eq. IV.5}]$$

obtenemos un coeficiente positivo, entonces se dice que el proceso muestra ‘dependencia de duración positiva’, lo que implica que la probabilidad de finalización de un proceso aumenta con el tiempo.

Si al calcular dicha derivada obtenemos un coeficiente negativo,

<sup>106</sup> Consultar el apartado 4.1.2 del Capítulo III para más detalle.

$$\frac{\delta h(t)}{\delta t} < 0 \quad [\text{Eq. IV.6}]$$

entonces se dice que el proceso muestra 'dependencia de duración negativa', implicando que la probabilidad de finalización del proceso disminuye con el tiempo.

Mientras que, si dicha derivada toma el valor cero,

$$\frac{\delta h(t)}{\delta t} = 0 \quad [\text{Eq. IV.7}]$$

nos encontramos ante un proceso que 'no tiene memoria' y que, por tanto, viene definido por la función exponencial.

Sin embargo, tratar de caracterizar las distribuciones como monótonas crecientes o decrecientes, puede llevar a realizar afirmaciones bastante poco realistas, puesto que los datos reales difícilmente muestran un comportamiento tan estable.

Las cuatro funciones enumeradas hasta ahora, es decir, la función de densidad, la función de distribución, la de supervivencia y la de riesgo, están claramente relacionadas. De hecho, a partir de [Eq. IV.1] y [Eq. IV.2] podemos obtener:

$$f(t) = \frac{\delta F(t)}{\delta t} = \frac{\delta [1 - S(t)]}{\delta t} = -S'(t) \quad [\text{Eq. IV.8}]$$

Además, teniendo en cuenta la expresión [Eq. IV.4], podemos deducir:

$$h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = -\frac{\delta \ln S(t)}{\delta t} \quad [\text{Eq. IV.9}]$$

Y finalmente, a partir de la expresión [Eq. IV.4], también podemos obtener una última expresión, que relaciona a la función de densidad con la de riesgo y la de supervivencia:

$$f(t) = h(t) \cdot S(t) \quad [\text{Eq. IV.10}]$$

Otra función que resulta útil en el análisis de modelos de duración es la función integrada de riesgo o función de tasa de fallo integrada:

$$H(t) = \int_0^t h(s) ds \quad [\text{Eq. IV.11}]$$

Ésta, según comenta Lancaster (1992), se puede ver como un residuo generalizado, por lo que resulta especialmente útil para verificar la especificación de modelos de duración. A partir de la expresión [Eq. IV.9], la función integrada de riesgo puede tomar la expresión siguiente:

$$H(t) = -\ln S(t) \quad [\text{Eq. IV.12}]$$

Por tanto,

$$S(t) = e^{-H(t)} = e^{-\int_0^t h(s) ds} \quad [\text{Eq. IV.13}]$$

y, por ende, de [Eq. IV.10] obtenemos una nueva expresión de la función de densidad, medida solamente en términos de la función de riesgo:

$$f(t) = h(t) \cdot S(t) = h(t) \cdot e^{-\int_0^t h(s) ds} \quad [\text{Eq. IV.14}]$$

### 3.1.2. Mecanismos de censura

Habitualmente, en el análisis de datos de duración, nos podemos encontrar con dos tipos de observaciones. Un primer tipo se refiere a aquellas observaciones cuya duración del evento objeto de estudio ha finalizado en el momento de recabar la información. Por otro lado, el segundo tipo, se refiere a las observaciones cuya duración no se observa completamente, puesto que al recabar la información el evento aún no había finalizado. Como ya anticipamos al comienzo del epígrafe, éste es un problema bastante habitual en el manejo de datos de duración. A este segundo tipo de observaciones es a lo que se denota por observaciones censuradas, y habrá que tener en cuenta esta casuística en el proceso de estimación del modelo propuesto.

Sea  $T^*$  la duración real del evento objeto de estudio en caso de no existir censura, y sea  $C$  el momento de censura. La variable aleatoria que realmente somos capaces de observar es tan sólo  $T = \min \{T^*, C\}$ . Si las observaciones censuradas realmente presentan la función de supervivencia  $G(C)$ , con su correspondiente función de densidad  $g(C)$ , y si el momento de censura es totalmente independiente de la duración real del fenómeno objeto de estudio, entonces la función de densidad de  $T$  será la siguiente:

$$k(T) = G(T) \cdot f(T) + S(T) \cdot g(T) \quad [\text{Eq. IV.15}]$$

El primer sumando de esta expresión recoge la probabilidad de que el evento haya finalizado en el momento  $T$ . Es decir, recoge la probabilidad de que el evento finalice justo en el momento  $T$  o de que el punto de censura  $C$  sea posterior a  $T$ . Mientras que el segundo sumando recoge la probabilidad de que  $T$  sea un punto de censura.

Así como la distribución del tiempo que transcurre desde que se inicia el evento hasta que éste finaliza habitualmente es desconocida, la distribución de los momentos de censura está parcialmente bajo el control del investigador que diseña la muestra. De esta manera, tal y como comenta Neumann (1997), nos podemos encontrar con la *censura de tipo I*, en donde la muestra es observada por un período  $Z$ , de forma que todas las duraciones que no hayan concluido en  $Z$  estarán censuradas. Otros mecanismos de censura incluyen la *censura de tipo II*, donde se mantiene el proceso de muestreo hasta llegar a observar el  $i$ -ésimo período de duración más pequeño; o la *censura progresiva de tipo II*, donde una fracción dada de la muestra puede aparecer censurada en más de un momento.

El problema que surge, en cualquier caso, está relacionado con la cantidad de información que, a pesar de la censura, sigue incorporando la muestra. De hecho, habrá que plantearse si la muestra de la que se dispone proporciona información suficiente acerca del proceso estocástico observado, a pesar de la censura.

### 3.1.3. Estimación no paramétrica

La estimación no paramétrica se basa en la representación gráfica de la función de supervivencia, siendo ésta es la manera más sencilla de describir los datos de duración.

Existen dos métodos para la estimación de las curvas de supervivencia mediante la metodología no-paramétrica, el actuarial y el del producto de Kaplan y Meier. Ambos métodos estiman la supervivencia para un determinado período de tiempo en función de todas las supervivencias anteriores. Pero, la principal diferencia entre ambos, radica en que el método actuarial, a diferencia del de Kaplan y Meier, define previamente intervalos de tiempo.

El método actuarial es el adecuado cuando el número de observaciones es grande (Rivas-López y López-Fidalgo, 2000). Éste parte de la división de la muestra en  $k$  intervalos. De esta manera, el tiempo de observación quedará dividido arbitrariamente a priori en:  $[0, t_1), [t_1, t_2), [t_2, t_3), \dots, [t_{k-1}, t_k]$ .

La probabilidad de sobrevivir a  $[t_{j-1}, t_j)$  sabiendo que se 'estaba vivo' al comienzo será

$$p_j = \Pr(T \geq t_j | T \geq t_{j-1}) \quad [\text{Eq. IV.16}]$$

La base de este método consiste en estimar, para cada intervalo, estas probabilidades de sobrevivir  $p_j$ .

Obviando los aspectos relacionados con la censura en los datos, a partir de una muestra de tamaño  $N$  perteneciente a una población homogénea, la función de supervivencia empírica tiene la siguiente especificación:

$$\hat{S}(t) = \frac{P(T \geq t)}{N} \quad [\text{Eq. IV.17}]$$

Esta función tendrá una forma escalonada decreciente, con aparición de peldaños en los momentos en los que se produce la finalización del evento objeto de estudio para alguna de las observaciones.

En caso de disponer de muestras con observaciones censuradas, se suele incorporar a la base de datos una nueva variable dicotómica  $\delta$ , que nos indique si la observación está o no censurada. De esta manera,  $\delta$  tomará el valor 1 si la observación está censurada y el valor 0 en caso contrario.

Al trabajar con datos censurados, en la estimación no paramétrica se asume que si  $T=t$  y  $\delta=1$ , la censura aparecerá justo inmediatamente tras  $T$ . Teniendo en cuenta esta consideración, supongamos que  $t_1 < t_2 < \dots < t_k$  son las diferentes duraciones observadas en la muestra de tamaño  $N$ . Llamemos  $d_j$  al número de episodios completos de duración  $t_j$  (con  $j=1,2,\dots, k$ ) y sea  $m_k$  el número de observaciones con duración superior a  $t_k$ , que es la mayor de las duraciones no censuradas. Entonces, el conjunto de riesgo, es decir el conjunto de observaciones que podrían ser susceptibles de finalizar su duración en el momento  $t_i$ , viene definido por:

$$n_i = \sum_{j \geq i} (m_j + d_j) \quad [\text{Eq. IV.18}]$$

Dado que  $h(t)$  es la probabilidad de que el evento finalice en un intervalo cualquiera  $(t, t + \Delta t)$  tomando en consideración que éste ya ha durado hasta el momento  $t$ , un estimador de  $h(t)$  sería:

$$\hat{h}(t_i) = \frac{d_i}{n_i} \quad [\text{Eq. IV.19}]$$

La estimación actuarial de la supervivencia  $\hat{S}(t_i)$  no es más que el producto de las probabilidades de sobrevivir al intervalo  $[t_{j-1}, t_j)$  y a todos los intervalos anteriores:

$$\hat{S}(t_i) = p_1 \cdot p_2 \cdot \dots \cdot p_i = \prod_{j=1}^i (1 - \hat{h}_j) = \prod_{j=1}^i \frac{(n_j - d_j)}{n_j} \quad [\text{Eq. IV.20}]$$

La estimación de la varianza asintótica, se realiza mediante la siguiente expresión:

$$\text{Var}(\hat{S}(t_i)) = \hat{S}^2(t_i) \sum_{j \leq i} \frac{d_j}{n_j(n_j - d_j)} \quad [\text{Eq. IV.21}]$$

Las expresiones [Eq. IV.20] y [Eq. IV.21] coinciden con los estimadores máximo verosímiles de la función de supervivencia derivados por Kaplan y Meier (1958), puesto

que el estimador de la curva de supervivencia propuesto por ellos se basa en el mismo principio que el actuarial, es decir, calcular la supervivencia como producto de probabilidades condicionadas, pero llevando la partición del tiempo de estudio en intervalos al caso extremo de considerar que cada intervalo contenga sólo la observación correspondiente a un individuo, sea esta finalización real o censura.

La función de supervivencia Kaplan-Meier, al igual que la estimación actuarial, sirve para estimar la proporción de supervivientes en cada uno de los tiempos en los que se ha producido un suceso, únicamente a través de la propia información que aportan los datos y sin necesidad de utilizar variables independientes explicativas de dicha proporción.

#### 3.1.4. Estimación paramétrica

La base teórica a considerar a la hora de especificar una función de riesgo, parte de los modelos de elección óptima para los agentes<sup>107</sup> cuyas transiciones van a ser estudiadas. Un modelo así puede ser especificado con gran detalle, dando lugar a una función de riesgo determinada, prácticamente en su totalidad, por la teoría económica, con la excepción de algunos parámetros desconocidos.

A continuación presentaremos brevemente algunas de las familias de distribuciones cuya forma funcional no ha sido dictada precisamente por consideraciones económicas teóricas, pero que, sin embargo, han sido bastante utilizadas en diferentes investigaciones econométricas. Estas familias tienen la ventaja de permitir considerar los siguientes aspectos:

- a. Las distribuciones de duración difieren entre personas porque, por ejemplo, se enfrentan a diferentes precios, tienen diferentes niveles de riqueza e ingresos o tienen diferentes niveles y tipos de capital humano.
- b. Las anteriores fuentes de variabilidad pueden ser representadas por un vector de covariables  $X$  para cada persona, donde  $X$  puede tener componentes que, de acuerdo con la teoría económica, deberían haber sido observadas pero no lo fueron.
- c. El vector de covariables  $X$  puede tener elementos que varían con el tiempo, como puede ser el estado civil, el número de hijos, el nivel educativo...<sup>108</sup>
- d. Los datos de duración suelen presentar censura pues, aunque la medición se realice en un momento dado, habitualmente el proceso continúa su desarrollo.
- e. El interés económico se centra en la función de riesgo, que puede ser función tanto de  $t$  como de  $X$ .<sup>109</sup>

<sup>107</sup> Estos agentes podrán ser personas, empresas,....

<sup>108</sup> Aunque nosotros no abordemos específicamente esta situación, habría que tener presente la posibilidad de su existencia.

<sup>109</sup> En el caso de considerar conjuntamente la influencia del tiempo y de las covariables, nos encontraríamos ante un caso de estimación semiparamétrica.

Varios de estos aspectos justifican la construcción de familias paramétricas de distribuciones de duración, permitiendo de esta manera realizar inferencia por métodos de máxima verosimilitud, en contraposición a los métodos de análisis de datos de duración en función de X por mínimos cuadrados, que podrán ser incluso no lineales. Concretamente, la posible dependencia del tiempo de los elementos de X, la presencia frecuente de censura en los datos y el interés teórico en la función de riesgo aconsejan la utilización de una metodología máximo-verosímil, puesto que el análisis de regresión no permite manejar con tanta facilidad regresores que varían con el tiempo o censura en los datos.

En la Tabla IV.1 recogemos de manera resumida algunas de las familias de distribuciones que han resultado ser herramientas econométricas bastante utilizadas en los trabajos empíricos y que desarrollaremos con mayor detalle más adelante.

**Tabla IV.1: Modelos paramétricos de funciones de distribución**

Modelo	Función de densidad $f(t)$	Función de supervivencia $S(t)$	Función de riesgo $h(t)$	Función integrada de riesgo $H(t)$
Exponencial	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	$\lambda$	$\lambda t$
Weibull	$\alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha}$	$e^{-(\lambda t)^\alpha}$	$\alpha \lambda^\alpha t^{\alpha-1}$	$\lambda^\alpha t^\alpha$
Gamma generalizada	$\frac{\alpha \lambda^\alpha t^{\alpha m-1} e^{-(\lambda t)^\alpha}}{\Gamma(m)}$	$1 - \int_1^t f(s) ds$	$\frac{f(t)}{1 - \int_1^t f(s) ds}$	$\int_1^t h(s) ds$
Log-logística	$\frac{\lambda \rho (\lambda t)^{\rho-1}}{1 + (\lambda t)^\rho}$	$(1 + \lambda t^\rho)^{-1}$	$\frac{\lambda \rho (\lambda t)^{\rho-1}}{1 + (\lambda t)^\rho}$	$\ln(1 + (\lambda t)^\rho)$
Gompertz	$\lambda e^{\gamma t} e^{\frac{\lambda}{\gamma}(1-e^{\gamma t})}$	$e^{\frac{\lambda}{\gamma}(1-e^{\gamma t})}$	$\lambda e^{\gamma t}$	$\frac{\lambda}{\gamma}(e^{\gamma t} - 1)$
Log-normal	$\frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}$	$1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)$	$\frac{f(t)}{S(t)}$	$-\ln S(t)$

Como vemos a partir de las expresiones de funciones de riesgo recogidas en la Tabla IV.1 la dependencia de duración definida mediante las ecuaciones [Eq. IV.5] a [Eq. IV.7] no se mantiene constante a lo largo de toda la curva. Es por esto por lo que debería venir definida en un punto  $t^*$ , puesto que ésta puede cambiar en función del momento del tiempo que estemos considerando.

Así, algunas funciones de riesgo presentan tasas de riesgo monótonas, como es el caso de la distribución Weibull, pero éstas pueden convertirse en tasas no monótonas simplemente añadiendo potencias superiores de t en su especificación. Elegir una forma funcional inapropiada para la función de riesgo puede llegar a hacer irreconocibles los

datos. Por ejemplo, ajustar un modelo exponencial a datos que presentan una tasa de fallo creciente producirá un patrón de desajuste similar al que se produce cuando se ignora la tendencia en el tratamiento de series temporales. De la misma manera, unos datos que presenten una tasa de fallo creciente a lo largo de prácticamente todo el período considerado, excepto al final, en donde presenta un decrecimiento, estarían mal representados si le ajustáramos una función de riesgo monótona como puede ser la distribución Weibull. Por tanto, una vez realizada la estimación paramétrica, es necesario utilizar mecanismos de diagnóstico que nos permitan determinar si el ajuste es o no el más adecuado.

Lo primero que debemos hacer, por tanto, es elegir una familia de distribuciones que se ajuste lo mejor posible a los datos recogidos en nuestra muestra. Para ello podemos representar gráficamente dichos datos y tomar una familia de distribuciones cuya densidad tenga una forma parecida a la de estos.

Una vez que sabemos con qué familia vamos a trabajar, hemos de estimar los parámetros que hagan que la densidad se ajuste lo más posible a los datos. Esta estimación se hará utilizando la función de verosimilitud  $L$ , que nos da la probabilidad de que una variable tome un determinado valor para distintos valores del parámetro. A continuación explicaremos de manera genérica cómo se desarrolla el proceso de estimación paramétrica.

Supongamos que se especifica una familia de distribuciones de duración  $F$  en función de un vector de parámetros finito  $\theta$ , y que el mecanismo de censura  $G$  también se especifica en función de un vector de parámetros finito  $\nu$ . Los datos están compuestos por  $N$  observaciones de la variable  $T$  con su correspondiente valor de  $\delta$ , que indica si la observación en cuestión está o no censurada. Dependiendo del valor que tome  $\delta$ , la contribución a la función de verosimilitud de la observación  $i$ -ésima será:

$$\begin{aligned} L_i(\theta, \nu | t_i, \delta) &= G(t_i, \nu) \cdot f(t_i, \theta) & \text{si } \delta_i = 1 (\text{censura}) \\ L_i(\theta, \nu | t_i, \delta) &= S(t_i, \theta) \cdot g(t_i, \nu) & \delta_i = 0 (\text{no censura}) \end{aligned} \quad [\text{Eq. IV.22}]$$

Si el mecanismo de censura es independiente de la variable  $T$ ,  $\nu$  no proporciona información sobre  $\theta$ , y la función de verosimilitud construida a partir de [Eq. IV.22] se puede factorizar en términos relacionados con el mecanismo de censura  $G$  y en términos relacionados con la distribución de fallo<sup>110</sup>  $F$ . Bajo estas circunstancias  $\theta$  puede ser estimada sin tener en cuenta cuál es la forma precisa que adopta la censura. El logaritmo de la función de verosimilitud para la muestra, llevado al límite, se podrá especificar entonces como muestra la ecuación [Eq. IV.23]

<sup>110</sup> Al hablar de 'distribución de fallo' estamos adoptando la terminología utilizada en el ámbito de la ingeniería, en donde el término 'fallo' hace referencia al momento en el que se completa una duración.

$$\ln L(\theta | \delta) = \sum_{i=1}^N \delta_i \ln h(t_i, \theta) - \sum_{i=1}^N H(t_i, \theta), \quad [\text{Eq. IV.23}]$$

en donde se utiliza la relación entre la función de densidad y la de supervivencia expresada en [Eq. IV.4]. La representación de esta función de verosimilitud enfatiza el acercamiento a través de la función de riesgo y, de hecho, en los trabajos empíricos, la mayoría de las caracterizaciones de las distribuciones de datos de duración se hacen en términos de funciones de riesgo. Por supuesto, en los casos en los que no existe una única solución que nos aproxime a la forma de la función de riesgo ni a la función integrada de riesgo, la expresión recogida en [Eq. IV.23] no tiene utilidad alguna (Neumann, 1997).

Una vez definida la función de verosimilitud en función del vector de parámetros  $\theta$ , habrá que determinar un valor estimado de  $\theta$  que maximice el valor de  $L$ , es decir, que maximice la probabilidad de obtener esos valores muestrales en concreto al realizar  $n$  veces una toma de datos. El método de máxima verosimilitud proporciona una estimación para cada uno de los parámetros recogidos en el vector  $\theta$  a través de la resolución de las ecuaciones de verosimilitud:

$$\frac{\delta \ln L}{\delta \theta_1} = 0, \quad \frac{\delta \ln L}{\delta \theta_2} = 0, \quad \dots, \quad \frac{\delta \ln L}{\delta \theta_k} = 0 \quad [\text{Eq. IV.24}]$$

Una vez obtenidas las estimaciones  $\hat{\theta}$  habrá que comprobar si el ajuste de la distribución con esos parámetros a los datos reales es o no satisfactoria, para lo que se suele recurrir a la realización de pruebas  $\chi^2$ .

Como vemos, a la hora de ajustar modelos paramétricos, la mayoría de los estudios empíricos han enfatizado la metodología máximo verosímil. Sin embargo, Neumann (1997) comenta que éste no es el único método de aproximación, demostrando que, cuando se dispone de datos no censurados, se puede tratar de ajustar por mínimos cuadrados el logaritmo de los datos de duración, para obtener los parámetros de interés una vez transformadas las estimaciones MCO. Sin embargo, advierte que estas estimaciones podrían ser ineficientes, en comparación con las de máxima verosimilitud, en el caso de que el modelo ajustado por este último método sea el correcto. Además, comenta que, en este caso, mediante la comparación de la varianza de los residuos con su valor esperado, podemos obtener 'pistas' acerca del grado de adecuación del modelo<sup>111</sup>.

<sup>111</sup> Dado que la característica típica de los modelos de duración es la existencia de censura en los datos y que la utilización de modelos de regresión en datos con censura es relativamente reciente en economía aplicada, existen pocos ejemplos disponibles que utilicen la metodología que comenta Neumann.

### 3.1.5. Concentración de las distribuciones de duración

Suele ser habitual comparar las diferentes familias de distribuciones de duración en términos de sus funciones de riesgo. Sin embargo, como recuerda Alegría-Hernández (1999), hay otras formas de comparar distribuciones, como puede ser el estudio del grado de concentración de las mismas.

Al estudiar distribuciones de ingresos o riquezas suele ser habitual conocer el grado de desigualdad o concentración de dicha distribución. Sin embargo, algo similar puede hacerse con las distribuciones de duración. Supongamos que el evento que estamos estudiando sea la duración del período de desempleo de un colectivo en concreto. La muestra extraída de la población nos informa acerca de la duración de este período, en el cual los individuos pudieron salir de esa situación de acuerdo con una distribución de duración con su correspondiente función de supervivencia. Sería razonable preguntarse cómo se distribuye la carga total experimentada de desempleo entre esas personas. ¿Se trata de un colectivo en el que unas pocas personas han sufrido un período largo de desempleo o, por el contrario, es un colectivo en el que todos han sufrido un período corto de desempleo?

Al igual que la curva de Lorenz y el coeficiente de Gini sirven para estudiar el grado de concentración de los ingresos, algo semejante se puede hacer con la concentración de las duraciones de desempleo.

Sea  $f(t)$  la función de densidad correspondiente a la función de supervivencia  $S(t)$  y a la de distribución  $F(t)$ . Además, sea  $\mu = E(t)$ . La proporción del tiempo total de desempleo sufrido por aquellas personas cuyas duraciones no fueron mayores a  $t$ , viene dada por

$$G(t) = \frac{\int_0^t s \cdot f(s) ds}{\mu} \quad [\text{Eq. IV.25}]$$

La curva de Lorenz resulta de graficar los puntos  $(F(t), G(t))$ , mientras que el coeficiente de Gini se define como dos veces el área entre esta curva y la línea  $G=F$ , como queda recogido en la ecuación

$$g = 2 \int_0^1 F dG - 1 \quad [\text{Eq. IV.26}]$$

Como  $\int_0^1 F dG = \int_0^\infty F(s) \cdot s \cdot f(s) ds / \mu$ , integrando por partes se puede demostrar que el índice de Gini se puede describir como:

$$g = 1 - \frac{\int_0^{\infty} S^2(s) ds}{\mu} = 1 - \frac{\int_0^{\infty} S^2(s) ds}{\int_0^{\infty} S(s) ds} \quad [\text{Eq. IV.27}]$$

En donde hemos tenido en cuenta que la media de la duración coincide con la integral de la función de supervivencia.

De manera similar, la curva de Lorenz se puede expresar como:

$$G(F) = \frac{\int_0^t S(s) ds - t \cdot S(t)}{\int_0^{\infty} S(s) ds} \quad [\text{Eq. IV.28}]$$

Estas expresiones de la curva de Lorenz y el índice de Gini en términos de la función de supervivencia y, por tanto, en términos de la función integrada de riesgo, es útil para aquellos datos de duración que vengan expresados a partir de sus funciones de riesgo.

En el caso concreto de la distribución exponencial con parámetro  $\lambda$ , la función de supervivencia toma la expresión  $S(t) = e^{-\lambda t}$  y, como  $\int_0^{\infty} S^k(s) ds = \frac{1}{k\lambda}$ , resulta que  $g = \frac{1}{2}$  independientemente del valor de  $\lambda$ . La curva de Lorenz, entonces, vendrá dada por

$$G = F + (1-F) \ln(1-F) \quad [\text{Eq. IV.29}]$$

Un valor de  $\frac{1}{2}$  para el coeficiente de Gini es bastante grande, en comparación con los valores que se obtienen en las distribuciones de ingreso (Lancaster, 1992). Estos últimos, en las economías desarrolladas, suelen oscilar en torno al 0.35.

Para ratificar esta afirmación, Alegría-Hernández (1999) recoge algunos puntos interesantes de la curva de Lorenz obtenida bajo la hipótesis de distribución exponencial. A partir de estos valores hemos generado la Tabla IV.2.

**Tabla IV.2: Curva de Lorenz para la distribución exponencial**

F	G	1-F	1-G
0,05	0,001	0,95	0,999
0,10	0,005	0,90	0,995
0,50	0,15	0,50	0,85
0,90	0,67	0,10	0,33
0,95	0,80	0,05	0,20

Así podemos observar, por ejemplo, cómo una tercera parte de todo el desempleo lo experimenta el 10% de las personas, mientras que el 85% de la duración total de desempleo es experimentada por tan sólo el 50% de la población.

A partir del análisis de la curva de Lorenz para diversas funciones de distribución, Lancaster (1992) llega a la conclusión de que para cualquier función de riesgo decreciente, es decir, para las funciones de riesgo que muestren dependencia negativa de duración, el índice de Gini tomará valores superiores a  $\frac{1}{2}$ .

### **3.2. MÉTODOS DE ESTIMACIÓN**

Los datos económicos rara vez son homogéneos. Difícilmente tenemos observaciones que puedan ser consideradas como obtenidas a partir de diferentes selecciones de una misma distribución de probabilidad. Por esto, normalmente tenemos que admitir la existencia de diferencias sistemáticas, medibles o no, entre los agentes que estemos estudiando. Es la aceptación de esta realidad lo que llevó a los investigadores a la utilización de los llamados modelos paramétricos.

Los modelos no paramétricos de duración, a diferencia de los paramétricos, no consideran la existencia de factores externos que pudieran afectar a la distribución de duraciones. Estos factores son los que se conocen con el nombre de covariables ( $X$ ). Sin embargo, al modelizar datos de duración mediante técnicas paramétricas, sí se incorpora en la especificación la información aportada por estas covariables, con el fin de contemplar su influencia sobre la función de riesgo. El caso más sencillo de introducción de covariables es el de aquellas que no cambian con el tiempo, como pueden ser la raza y el género de una persona. De manera más general, las características más relevantes de un agente en el momento de comenzar el evento objeto de estudio, es decir, sus características biográficas en ese momento del tiempo, constituyen un conjunto de regresores independientes del tiempo que, efectivamente, pueden afectar a la duración total de dicho evento. Otras características, ciertamente, pueden cambiar durante el período de tiempo considerado, pero normalmente éstas cambian a un ritmo lo suficientemente pausado, en comparación con las duraciones del evento objeto de estudio, como para ser tratadas, por motivos técnicos, como si se mantuvieran constantes (Lancaster, 1992). Ejemplos de este último caso, en un estudio de la duración del desempleo, podrían ser la edad del individuo y el estado del ciclo económico.

En presencia de regresores independientes del tiempo recogidos en el vector  $X$ , la función de riesgo en el momento  $t$  que habíamos definido mediante la expresión [Eq. IV.4] estará condicionada por los valores de  $X$ :

$$h(t/X) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq T < t + \Delta t | T \geq t, X)}{\Delta t} \quad [\text{Eq. IV.30}]$$

De este modo,  $h(t/X)dt$  nos informa de qué fracción, del total de supervivientes en  $t$ , finaliza su período de permanencia en el estado objeto de estudio, en un corto intervalo de tiempo  $(t, t + \Delta t)$  para una población homogénea en lo que respecta a  $X$ .

La función de riesgo, dada  $X$ , estará relacionada con la función de densidad y la función de supervivencia de  $T$ , como ya comentamos en el apartado 3.1.1, a través de la siguiente expresión

$$h(t/X) = \frac{f(t/X)}{S(t/X)} \quad [\text{Eq. IV.31}]$$

con

$$S(t/X) = e^{-\int_0^t h(s/X)ds} \quad [\text{Eq. IV.32}]$$

El hecho de introducir covariables en los modelos de duración complica la interpretación de sus resultados, puesto que, excepto en el caso del modelo lineal, no podemos hablar de la existencia de un punto de partida a partir del cual las covariables consigan aumentar únicamente el valor medio, manteniendo el resto de los momentos invariables. Debido a esto, en los modelos de duración, los coeficientes de esas covariables no tienen una interpretación tan sencilla como puede ser la de las derivadas parciales.

En el análisis de duración de datos, la mayoría de los estudios siguen una misma pauta de comportamiento, y es que la inmensa mayoría de ellos centran sus esfuerzos en la estimación de la función de riesgo, para, a partir de la misma, deducir las implicaciones sobre la función de distribución.

Dado que los acercamientos econométricos se han centrado fundamentalmente en la especificación paramétrica o semiparamétrica de dicha función de riesgo en presencia de covariables, dedicaremos este apartado al estudio de este tipo de especificaciones. No obstante, hemos considerado apropiado introducir un apartado previo, en el que veamos cómo se realiza también la estimación no paramétrica de los modelos de duración, puesto que ésta permite realizar una aproximación acerca de la familia de distribución más acorde con la realidad que manifiesta la muestra.

### 3.2.1. Métodos no paramétricos

El tiempo de supervivencia, como ya hemos apuntado más arriba, es una variable continua y, por tanto, la representación de medidas relacionadas con ella ha de

demostrar dicha continuidad. No obstante, una oportuna discretización de esta variable, mediante la agrupación de los datos en intervalos, nos puede permitir elaborar una tabla de gran utilidad práctica, que es conocida en la literatura como 'tabla de vida'.

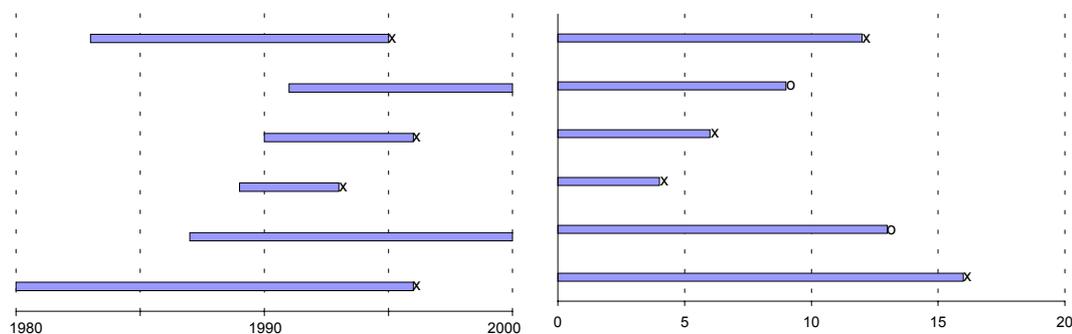
Para crear dicha tabla necesitamos subdividir el intervalo temporal de observación desde el punto inicial en intervalos menores. Se contará el número de observaciones cuya duración, en lo que respecta al evento objeto de estudio, aún sigue vigente en cada intervalo, con el fin de poder determinar la probabilidad genérica de que la duración de una observación se prolongue hasta un momento determinado.

Las hipótesis que han de cumplirse para construir una tabla de vida son las siguientes:

- Que las condiciones experimentales de supervivencia no cambien a lo largo del estudio
- Que una observación que se comience a estudiar ahora responda de la misma manera que si se hubiera introducido en el estudio con anterioridad
- Que las observaciones censuradas no difieran de las que no lo son, puesto que, de no ser así, las muertes/fallos se producirían de manera no aleatoria influenciadas por variables que no se están teniendo en cuenta

La representación de los datos observados se puede hacer como la que aparece en la Ilustración IV.1. Ésta se podrá hacer en tiempo real, es decir, con momentos iniciales y finales reales, como se muestra en el gráfico de la izquierda; o poniendo a cero todos los momentos iniciales, como se muestra en el gráfico de la derecha. En ambos casos una 'x' indicará que el suceso ha finalizado, mientras que una 'o' indicará que el dato está censurado. Mediante esta última gráfica podemos determinar la amplitud del intervalo más adecuada para discretizar la variable, puesto que podremos intuir la distribución de las observaciones dentro de cada uno de los intervalos.

**Ilustración IV.1: Representación gráfica de tiempos de vida**



En los intervalos definidos podemos encontrarnos con 4 tipos de observaciones diferentes que deben venir perfectamente identificadas mediante la tabla de vida. Si

miramos, por ejemplo, el segundo intervalo de la Ilustración IV.1 derecha, podemos comprobar cómo nos encontramos con:

- a. Observaciones que terminan dentro del propio intervalo
- b. Observaciones terminales en un momento posterior al intervalo
- c. Observaciones censuradas en el intervalo
- d. Observaciones censuradas en un momento posterior al intervalo

En cualquier caso, la tabla de vida recogerá la siguiente información en sus columnas, aparte de informar al final de la misma del valor que toma la mediana del tiempo de supervivencia:

1. Los extremos superior e inferior del intervalo
2. Los supervivientes hasta el inicio del intervalo en cuestión
3. El número de observaciones censuradas en el intervalo
4. El número de observaciones que finalizan su duración en el intervalo considerado
5. La proporción de sucesos terminales 'q': Es una estimación de la probabilidad de que una observación que entra en el intervalo  $i$ -ésimo finalice dentro de él.
6. La proporción de supervivientes 'p': Es una estimación de la probabilidad de supervivencia de una observación que entra en el intervalo  $i$ -ésimo.
7. La proporción acumulada de supervivientes: Es una estimación de la probabilidad de supervivencia al final del intervalo  $i$ -ésimo. Se suele denotar por supervivencia acumulada. Mediante esta estimación se podrá representar posteriormente la función de supervivencia empírica.
8. La densidad de probabilidad: Es una estimación de la probabilidad de finalización del evento por unidad de tiempo. Mediante esta medida se intenta expresar una probabilidad instantánea, propia de la función de densidad probabilística, a través de una distribución discreta. Dicha estimación se obtiene mediante la fracción entre la disminución de probabilidad acumulada de supervivientes del intervalo y la amplitud del mismo.
9. La razón de riesgo: Es una estimación de la probabilidad de finalización del evento por unidad de tiempo, suponiendo que la observación ya ha durado hasta el inicio del intervalo. Se trata de una estimación de la función de riesgo en el instante  $t_i$ . Como ya hemos comentado, esta función ofrece en cada momento la idea del riesgo que tiene una observación de finalizar su duración 'en breve'. Este valor se obtiene como cociente entre la densidad de probabilidad y la supervivencia.

Los datos recogidos en la tabla de vida, que nos proporcionan estimaciones de las funciones de supervivencia, densidad y riesgo, nos permiten dibujar una primera aproximación a dichas funciones. La columna 'proporción acumulada de supervivientes' representa una estimación de la función de supervivencia empírica. Ésta nos proporciona valores al final del intervalo, por lo que, al graficarlos, la función tomará

siempre el valor 1 en el primer intervalo, a partir del cual, tomará de forma escalonada valores cada vez menores.

La representación de la función de densidad se hace a partir de los datos de la columna de 'densidad de probabilidad'. Ésta nos da una idea de las zonas de mayor riesgo y viene dada por los valores de la tabla al inicio de cada intervalo.

Finalmente, la función de riesgo empírica se obtiene graficando los datos recogidos en la columna 'razón de riesgo'. Ésta también viene dada por los valores de la tabla en el inicio de cada intervalo, y proporciona la idea del riesgo de finalización en cada momento del tiempo para los que han llegado hasta ahí.

### 3.2.2. Métodos semiparamétricos

Los datos de duración pueden ser descritos de dos maneras equivalentes: especificando la función de densidad de una distribución paramétrica o especificando la función de riesgo. Esta última forma será la que nos permita acercarnos a cuestiones tales como el efecto que tienen determinadas variables exógenas sobre la duración del evento objeto de estudio.

La función de riesgo, como ya hemos comentado, nos proporciona la probabilidad de que un suceso finalice en el instante inmediato, dado que ya ha llegado hasta el momento  $t$ . Esta duración dependerá de la naturaleza del propio fenómeno, pero también es posible que existan variables exógenas  $X$ , o covariables, que influyan en dicha duración. Por tanto, si quisiéramos estudiar la duración de una variable  $T$  teniendo en cuenta el efecto de dichas covariables, lo lógico sería integrarlas a través de la imposición de una estructura tipo 'modelo de regresión' en la función de riesgo. De esta manera, la función de riesgo quedaría especificada de manera multiplicativa en función, tanto de las covariables, como del tiempo:

$$h(t/X) = h_0(t) \cdot g(X) \quad [\text{Eq. IV.33}]$$

La función  $h_0(t)$  recoge cómo cambia la función de riesgo dependiendo únicamente del tiempo de supervivencia, mientras que la función  $g(X)$  recoge cómo cambia la función de riesgo atendiendo a las covariables. Estas dos funciones han de ser elegidas de manera que  $h(t/X) > 0$ , puesto que ésta mide una probabilidad que, por definición, ha de ser siempre positiva.

En el caso en que  $g(X) = 1$ , entonces  $h_0(t)$  será la función de riesgo que habíamos denotado de manera genérica. Por esto es por lo que  $h_0(t)$  se suele denotar como

‘función de riesgo base’<sup>112</sup>. Bajo el modelo [Eq. IV.33], la ratio de funciones de riesgo para dos individuos con vectores de covariables  $X_1$  y  $X_2$  será:

$$\frac{h(t|X_1)}{h(t|X_2)} = \frac{h_0(t)g(X_1)}{h_0(t)g(X_2)} = \frac{g(X_1)}{g(X_2)} \quad [\text{Eq. IV.34}]$$

Vemos, por tanto, que la tasa de riesgo depende sólo de la función  $g(X)$ . Debido a esto, si la función de tasas de riesgo se puede interpretar fácilmente, la forma funcional que adopte la ‘función de riesgo base’ será de escasa importancia.

### 3.2.2.1. El modelo de Cox

Cox (1972) fue el primero en proponer una especificación como la recogida en [Eq. IV.33], pero sugiriendo como especificación de  $g(X)$  una función exponencial, de manera que  $g(X) = e^{X\beta}$ . Una función de riesgo que depende de covariables,  $h(t|X)$ , en principio, podría permitir una gran interacción entre las duraciones  $t$  y los regresores  $X$ . Por tanto, para poder obtener un modelo más general, éste tendría que restringir el número posible de interacciones. Fue así como Cox (1972) introdujo su especificación de la función de riesgo, restringiendo la interacción entre  $t$  y  $X$  a una mera interacción multiplicativa entre la ‘función de riesgo base’  $h_0(t)$  y las covariables recogidas mediante una ‘función vínculo’  $e^{X\beta}$ . De esta manera, la especificación de la función de riesgo de Cox adopta la siguiente forma

$$h(t|X) = h_0(t) \cdot g(X) = h_0(t) \cdot e^{X\beta} \quad [\text{Eq. IV.35}]$$

siendo

$h(t|X)$  la función de riesgo, dependiente del tiempo, teniendo en cuenta la información suministrada por el conjunto de  $p$  variables  $X$ ;

$h_0(t)$  la función de riesgo sin considerar el efecto de las variables  $X$ ; y

$g(X)$  una función de las  $p$  variables independientes.

Por tanto, en el modelo de Cox, la probabilidad de que un suceso que ha durado hasta el instante  $t$ , finalice en el intervalo  $(t, t + \Delta t)$ , realmente se obtiene como el producto de dos funciones, una que depende solamente de  $t$ ,  $h_0(t)$ , y otra que solamente depende de  $X$ ,  $g(X)$ ; centrándose el problema en la estimación de los parámetros  $\beta_1, \beta_2, \dots, \beta_p$ .

<sup>112</sup> Traducción literal del término anglosajón ‘baseline hazard function’.

La tasa de riesgo en este caso viene expresada por:

$$\frac{h(t | X_1)}{h(t | X_2)} = e^{\beta(X_1 - X_2)}. \quad [\text{Eq. IV.36}]$$

Este modelo se suele conocer en la literatura como ‘modelo de Cox’, ‘modelo de riesgo proporcional de Cox’ o, sencillamente, ‘modelo de riesgo proporcional’. Parte del atractivo de este tipo de modelos radica precisamente en la posibilidad de interpretación de la [Eq. IV.36] como una tasa de riesgo relativo.

La expresión del modelo de Cox de [Eq. IV.35] es la más habitual de las posibles funciones de riesgo recogidas en [Eq. IV.33]. El término de ‘riesgo proporcional’ se refiere precisamente al hecho de que en [Eq. IV.35] las funciones de riesgo se relacionan de forma multiplicativa, por lo que su ratio se mantiene constante a lo largo del período de supervivencia.

Las funciones de riesgo [Eq. IV.33] y [Eq. IV.35] se denominan funciones semiparamétricas por el hecho de no describir de manera explícita la ‘función de riesgo base’  $h_0(t)$ .

Como ya comentamos en el apartado 3.1, los datos de duración pueden venir caracterizados a partir de su función de supervivencia, por lo que resulta conveniente conocer la especificación de dicha función en el modelo de Cox.

Teniendo en cuenta la expresión [Eq. IV.13], en donde la función de supervivencia venía expresada en términos de la función de riesgo, la función de supervivencia de un modelo semiparamétrico tomará la siguiente especificación:

$$S(t | X) = e^{-H(t|X)} \quad [\text{Eq. IV.37}]$$

donde  $H(t | X)$  es la función integrada de riesgo en el momento  $t$  para un individuo con vector de covariables  $X$ . Asumiendo que el tiempo de supervivencia es una variable continua, la función integrada de riesgo se puede expresar como

$$H(t | X) = \int_0^t h(s | X) ds = g(X) \int_0^t h_0(s) ds = g(X) \cdot H_0(t) \quad [\text{Eq. IV.38}]$$

Sustituyendo la expresión [Eq. IV.38] en [Eq. IV.37] se obtiene la función de supervivencia para un modelo genérico semiparamétrico, que valdrá

$$S(t | X) = e^{-g(X)H_0(t)} = \left( e^{-H_0(t)} \right)^{g(X)} = \left( S_0(t) \right)^{g(X)}, \quad [\text{Eq. IV.39}]$$

siendo  $S_0(t)$  la ‘función de supervivencia base’<sup>113</sup>.

Bajo el modelo de Cox, la función de supervivencia tendrá la especificación

$$S(t | X) = (S_0(t))^{e^{X\beta}} \quad [\text{Eq. IV.40}]$$

La forma específica de la expresión de la función de supervivencia de [Eq. IV.40] se debe únicamente a la relación multiplicativa entre la ‘función de riesgo base’ y la función exponencial que describe el efecto de las covariables. En cualquier caso, el valor de la ‘función de supervivencia base’ oscilará entre cero y uno, como sucede en el caso de cualquier otra función de supervivencia.

Cox demostró que el parámetro  $\beta$  de la función de riesgo podía ser estimado sin necesidad de especificar qué forma adoptaba  $h_0(t)$ . Para ello, lo que hay que hacer es ordenar las diferentes duraciones muestrales de manera creciente ( $t_1 < t_2 < \dots < t_k$ ). A partir de ahí, la probabilidad condicional de que la observación 1 finalice su duración en  $t_1$ , dado que el conjunto de riesgo que podía haber finalizado estaba compuesto por N observaciones, se obtiene (ignorando cualquier problema de censura) mediante la siguiente expresión:

$$\frac{h(t_1 | X)}{\sum_{i=1}^N h(t_i | X_i)} = \frac{e^{X\beta}}{\sum_{i=1}^N e^{X_i\beta}}. \quad [\text{Eq. IV.41}]$$

De manera similar, la probabilidad condicional de la j-ésima duración más corta se obtiene mediante el cálculo de la tasa de riesgo para el individuo que finaliza su duración en el período  $t_j$  entre el sumatorio de los riesgos para los individuos cuyas duraciones aún no habían finalizado justo antes del momento  $t_j$ . En caso de darse censura entre las duraciones  $t_j$  y  $t_{j+1}$ , ésta aparecería en el sumatorio del denominador de [Eq. IV.41] para las observaciones 1 hasta j, pero no para ninguna otra, y jamás entraría en el numerador. De esta manera, incorporamos la información de que existía un proceso que aún no había finalizado y que, por tanto, era susceptible de hacerlo, hasta un momento determinado, a partir del cual, no tenemos más información, por lo que desconocemos su verdadera duración.

En cualquier caso, como se observa en la ecuación [Eq. IV.41] la ‘función de riesgo base’ se simplifica al calcular la probabilidad condicional, por lo que no es necesario conocerla para estimar los valores de  $\beta$ .

<sup>113</sup> Traducción literal de ‘baseline survivorship function’.

Si exactamente un individuo finaliza su duración en cada momento del tiempo y no existen observaciones censuradas, la función del logaritmo de verosimilitud parcial será (Cox, 1972):

$$\ln L = \sum_{i=1} \left( X\beta - \ln \sum_{j \in R_i} e^{X_j\beta} \right) \quad [\text{Eq. IV.42}]$$

en donde  $R_i$  es el conjunto de observaciones en riesgo de finalización justo antes del momento  $t_i$ , es decir, el conjunto de individuos con duraciones superiores o iguales a  $t_i$ .

Si, en cambio,  $m_i \geq 1$  individuos finalizan su duración en el mismo momento  $t_i$ , entonces, el logaritmo de verosimilitud parcial será la suma de las verosimilitudes individuales siguiente

$$\ln L = \sum_{i=1} \left( \beta' \sum_{j \in t_i} X_j - m_i \ln \sum_{j \in R_i} e^{X_j\beta} \right) \quad [\text{Eq. IV.43}]$$

En caso de existir observaciones censuradas, como comentamos previamente, éstas entrarán a formar parte del conjunto de riesgo  $R_i$ , pero no contribuirán a formar parte del numerador de la función de verosimilitud parcial.

La ventaja de la especificación semiparamétrica de Cox se debe a la facilidad de interpretación de sus coeficientes. De hecho, Kiefer (1988) anota que a la especificación de riesgo proporcional se le puede dar una interpretación similar a la de los modelos lineales, puesto que satisface la siguiente condición

$$-\ln(H_0(t)) = G(t) = X\beta + v \quad [\text{Eq. IV.44}]$$

donde  $G(\cdot)$  es una función monótona y  $v$  es una variable aleatoria que se distribuye como un 'valor unitario extremo', es decir  $F(v) = 1 - e^{-e^v}$  con  $-\infty < v < \infty$ . Esta característica de los modelos de riesgo proporcional es, precisamente, lo que facilita el desarrollo de contrastes de validación.

### 3.2.3. Métodos paramétricos de duración

En el análisis de datos de duración, uno de los objetivos principales consiste en elegir una distribución de probabilidad apropiada a los datos con la que trabajar. Una vez elegida la familia de distribuciones, es preciso conocer el valor de los parámetros que proporcionan el ajuste particular más adecuado a los datos dentro de esa familia. Por tanto, una vez más, los datos nos conducen hacia el modelo. Pero ¿cuál es el modelo más adecuado?, ¿cómo tomar la decisión de elegir entre uno u otro? La representación gráfica probabilística que hemos visto en la estimación no paramétrica, es una técnica

adecuada para seleccionar una familia razonable, a partir de la cual entraremos a utilizar los métodos de estimación paramétricos.

Cuando se pretende hacer inferencia por métodos de máxima verosimilitud, la posible dependencia del tiempo de los regresores, la presencia de censura en los datos y el interés en la función de riesgo, pueden justificar la construcción de modelos paramétricos.

En este tipo de modelos, al igual que en los modelos semiparamétricos, el objetivo fundamental se centra en la modelización de la función de riesgo. No obstante, a diferencia de estos últimos, en esta ocasión la función de riesgo tiene que estar perfectamente especificada. Las especificaciones más utilizadas para esta función son, como recogíamos en la Tabla IV.1, las de la distribución Exponencial, la distribución Weibull, la Gamma generalizada, la log-normal y la log-logística.

Todas ellas tienen una característica común, y es que sirven para modelizar variables aleatorias no negativas y, puesto que estamos modelizando la variable  $T$ , que mide el tiempo, éste es un requisito indispensable. No obstante, el comportamiento que describe cada una de ellas es totalmente diferente al de las otras. Así, la función de riesgo para una distribución Exponencial será constante; la de una distribución Weibull será, en función de su parámetro  $p$ , monótona creciente o decreciente; la de las distribuciones log-normal y log-logística serán primero crecientes y luego decrecientes; mientras que la función de riesgo de una distribución Gamma generalizada permite una gran variedad de formas posibles dependiendo del valor que tomen sus parámetros.

En este apartado vamos a suponer que las covariables son constantes a lo largo del tiempo en que los individuos son observados. Asumimos este supuesto, únicamente con la finalidad de simplificar la exposición de los modelos paramétricos, puesto que lo habitual es que esta hipótesis no se cumpla en el manejo de datos reales.

A pesar de haber visto la ventaja de los modelos semiparamétricos en cuanto a la escasez de hipótesis de partida necesarias para especificar la función de riesgo, es posible que, en ocasiones, a partir de investigaciones previas, la distribución del tiempo de supervivencia de una determinada variable, tenga una forma paramétrica conocida, que justifique el uso de modelos paramétricos, con el objetivo de realizar determinados análisis.

Estos modelos, como comentan Hosmer y Lemeshow (1999), también presentan determinadas ventajas, como pueden ser:

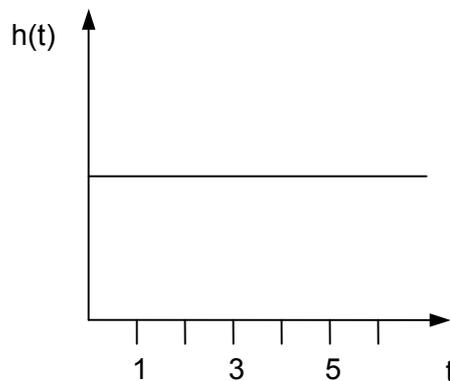
1. Se pueden utilizar funciones de máxima verosimilitud completas para estimar los parámetros.

2. Los coeficientes de algunos modelos estarán relacionados con los coeficientes del modelo de riesgo proporcional de Cox, con la consiguiente simplicidad en su interpretación.
3. Los valores ajustados del modelo pueden proporcionar estimaciones de la función de riesgo.
4. Se pueden computar los residuos como diferencia de los valores observados y los que predice el modelo, puesto que la estimación paramétrica, realmente puede tener la apariencia de un modelo de regresión lineal.

### 3.2.3.1. La distribución exponencial

En este apartado vamos a describir las propiedades de la distribución exponencial de parámetro  $\lambda$ , que es una de las distribuciones más importantes en la modelización y análisis de datos de duración, puesto que, como demuestra Lancaster (1992), sea cual sea la forma de la función de riesgo de una distribución, ésta podrá ser transformada en otra de riesgo constante, es decir, de forma exponencial, a través de un cambio en la escala de tiempo. Este hecho es particularmente importante, puesto que dicha transformación será de utilidad en el análisis de bondad del ajuste de los modelos de duración.

Ilustración IV.2: Función de riesgo exponencial



Cuando la función de riesgo es independiente de la duración transcurrida decimos que nos encontramos ante un proceso de Poisson. En este caso, dicha función de riesgo presenta una gráfica como la reflejada en la Ilustración IV.2 y la integral de la función de riesgo que aparece en la expresión [Eq. IV.13] se transforma en

$$\int_0^t h(s) ds = \lambda t, \quad [\text{Eq. IV.45}]$$

por lo que la función de supervivencia toma la expresión

$$S(t) = e^{-\lambda t}, \quad \lambda > 0, \quad t \geq 0, \quad [\text{Eq. IV.46}]$$

y la función de densidad

$$f(t) = \lambda e^{-\lambda t} \quad [\text{Eq. IV.47}]$$

Esta es la función de densidad exponencial definida para todos los valores positivos de  $\lambda$ . En ese caso podríamos decir que la variable de duración  $T$  se distribuye como una exponencial de parámetro  $\lambda$ .

A partir de la función generatriz de momentos acumulada de  $\log T$  para una variable  $T$  con distribución exponencial, como demuestra Lancaster (1992), podemos obtener una ecuación de regresión con la siguiente especificación:

$$\log T = -\log \lambda + U, \quad [\text{Eq. IV.48}]$$

donde el término de perturbación aleatoria  $U$  tiene media y varianza conocidas e independientes del valor que tome  $\lambda$ , pudiendo catalogar, por tanto, a la variable  $\log T$  como homocedástica<sup>114</sup>. La función de densidad de  $U = \log(\lambda T) = \log Z$  se puede obtener a partir de la función de densidad exponencial unitaria de  $Z$  cambiando el valor de  $Z$  por el de  $U$  y utilizando  $du = dz/z$ , obteniendo

$$f(u) = e^u e^{-e^u} \quad -\infty < u < \infty. \quad [\text{Eq. IV.49}]$$

Una manera sencilla de introducir el vector de regresores  $X$ , que recoge la influencia de las covariables, en la función de riesgo independiente del tiempo, es decir, en la función de riesgo exponencial, sería la siguiente

$$h(t | X) = e^{-X\beta} \quad [\text{Eq. IV.50}]$$

donde  $X$  incluirá una variable unitaria para permitir la existencia de constante en la función logarítmica del riesgo. La especificación recogida en la ecuación [Eq. IV.50] asegura la no negatividad de la función de riesgo sin necesidad de imponer restricciones a  $\beta$ . Además, este modelo permite que los datos puedan haber sido obtenidos de distribuciones exponenciales con medias diferentes, obtenidas a partir de la expresión  $E(T | X) = \lambda^{-1} = e^{X\beta}$ , en función de la composición del vector de regresores.

Con esta especificación, la ecuación [Eq. IV.48] se transforma en

$$\log T = X\beta + U \quad [\text{Eq. IV.51}]$$

donde  $U$  tiene, tal y como se expresa en la nota a pie de página número 114, media conocida,  $\psi(1)$ , que vendrá recogida en el término constante de  $X\beta$ , a la vez que varianza conocida,  $\psi'(1)$ , independiente de la media. Por tanto, con este modelo,  $\log T$

<sup>114</sup> La media de la perturbación aleatoria,  $\psi(1)$ , y su varianza,  $\psi'(1)$ , se obtienen a partir de las siguientes expresiones:  $E(\log T) = -\log \lambda + \psi(1)$  y  $Var(\log T) = \psi'(1)$

vendría especificado de manera lineal mediante una regresión homocedástica en función de  $X$ .

Este sería un modelo muy restrictivo, puesto que la función de densidad no incorpora parámetros desconocidos. Por tanto, considerar este tipo de modelos para el tratamiento de datos económicos resulta poco atractivo. Sin embargo, como veremos a continuación, sí que sirve como punto de partida a partir del cual poder realizar generalizaciones que nos encaminen hacia otros tipos de modelos más apropiados.

### 3.2.3.2. Modelos de duración Weibull

Aunque hayamos comentado que la manera natural de especificar un modelo de duración es a través de la especificación de su función de riesgo, existe una tipología de modelos que se obtiene, sencillamente, a través de generalizaciones de ecuaciones de regresión. Entre estos se encuentra la familia de modelos de duración Weibull, que es una de las más utilizadas hasta ahora en la mayoría de los estudios empíricos.

Partiendo de la especificación exponencial recogida en la ecuación [Eq. IV.48], ésta se puede describir como

$$\log(\lambda T) = U, \quad [\text{Eq. IV.52}]$$

donde  $\lambda = \lambda(X)$ , por ejemplo  $\lambda = e^{X\beta}$ , como se supuso para la ecuación [Eq. IV.51]. Dado que  $\lambda T$  es la función de riesgo integrada,  $e^U$  se distribuirá como una exponencial unitaria. Considerando que  $X$  no varía con el tiempo, la ecuación [Eq. IV.52] nos proporciona, como ya hemos visto, un modelo de regresión para  $\log T$  con una perturbación aleatoria cuya distribución no tiene ningún parámetro desconocido.

Ahora consideremos una generalización de la distribución de la perturbación aleatoria  $U$ , de manera que obtengamos una formulación más flexible que incorpore parámetros desconocidos.

Una generalización podría consistir en introducir una constante de proporcionalidad en el propio término de error en [Eq. IV.52] escribiendo

$$\log(\lambda T) = \frac{U}{\alpha}, \quad \alpha > 0 \quad [\text{Eq. IV.53}]$$

donde, al igual que antes,  $e^U \sim \varepsilon(1)$ . Esto permite que la varianza del error de  $\log T$  tome cualquier valor positivo, puesto que  $Var(\log T)$  es  $\alpha^{-2}$  multiplicado por la varianza del logaritmo de una variable exponencial unitaria y, por tanto, según la nota a pie de página 114, será igual a  $\alpha^{-2}\psi'(1)$ , cuyo valor depende de  $\alpha$ . Éste modelo se reduce al caso de la distribución exponencial en el caso en que  $\alpha = 1$ . La distribución de  $T$  se distingue fácilmente pues, dado que  $\lambda T = e^{U/\alpha}$ , entonces  $(\lambda T)^\alpha = e^U \sim \varepsilon(1)$ . Por

consiguiente  $(\lambda T)^\alpha$  es la función integrada de riesgo y, a partir de la ecuación [Eq. IV.13],

$$S(t) = e^{-(\lambda t)^\alpha} \quad [\text{Eq. IV.54}]$$

y

$$f(t) = \alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha} \quad [\text{Eq. IV.55}]$$

con la función de riesgo,  $f(t)/S(t)$ , dada por

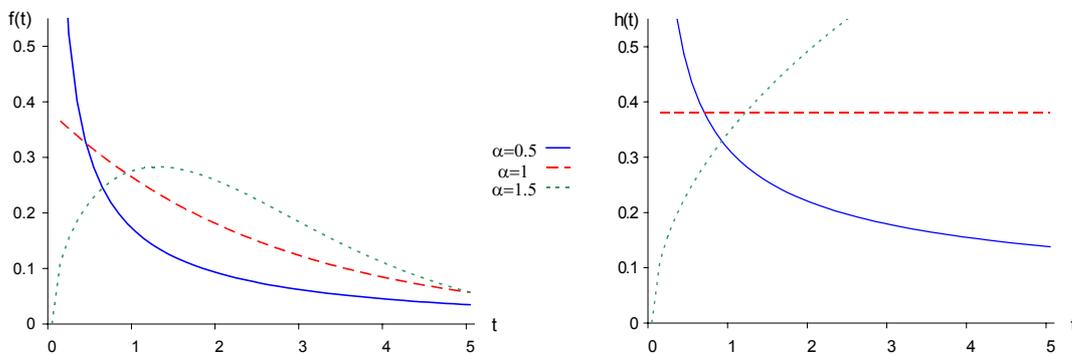
$$h(t) = \alpha \lambda^\alpha t^{\alpha-1} \quad [\text{Eq. IV.56}]$$

La función de riesgo [Eq. IV.56] depende del tiempo, excepto en el caso en que  $\alpha = 1$  que, como ya hemos comentado, coincide con el caso exponencial. A partir de esta especificación podemos comprobar cómo el riesgo presenta un comportamiento monótono creciente o decreciente en función del valor que tome  $\alpha$ . Si éste es superior (inferior) a 1 presentará un comportamiento creciente (decreciente).

Es de resaltar el hecho de que, si introducimos covariables independientes del tiempo en la función de riesgo, por ejemplo a través de la expresión  $\lambda = e^{-X\beta}$ , el efecto proporcional de cada elemento de X en el riesgo es el mismo para cada momento del tiempo<sup>115</sup>.

Mediante la Ilustración IV.3 recogemos la representación gráfica de algunas funciones de riesgo y de densidad tipo Weibull.

**Ilustración IV.3: Funciones de densidad y de riesgo Weibull**



<sup>115</sup> Esto implicaría, por ejemplo, que, si tener una determinada edad, disminuye la probabilidad de salir de una situación de desempleo al día siguiente en un porcentaje determinado, también disminuirá la probabilidad de salida al cabo de un tiempo superior, en exactamente la misma cuantía.

Por tanto, vemos cómo, a partir de la generalización del modelo exponencial mediante la introducción de un escalar en la distribución de su término de error hemos obtenido una nueva familia de distribuciones, la Weibull. Sin embargo, éste modelo sigue siendo bastante restrictivo pues, como se desprende de la Ilustración IV.3, la función de riesgo es necesariamente monótona, y por tanto no se podrá utilizar para el caso en el que la función de riesgo empírica presente, por ejemplo, crecimiento al comienzo y decrecimiento con posterioridad. Sin embargo, el modelo Weibull es de aplicabilidad bastante sencilla, por lo que ha sido utilizado en gran número de investigaciones empíricas.

### 3.2.3.3. Modelos de duración Gamma generalizada

Una segunda extensión de la especificación exponencial nos lleva a determinar otro tipo de familias de modelos de duración, conocidas con el nombre de distribuciones Gamma generalizada. En este caso, la generalización consiste en mantener el parámetro  $\alpha$ , pero asumir que  $e^U$  no se distribuye como una exponencial unitaria, sino como una distribución Gamma estándar de parámetro  $m$ , denotada por  $\zeta(m)$ . Es decir, si  $Y = e^U$ , entonces

$$f(y) = \frac{y^{m-1} e^{-y}}{\Gamma(m)}, \quad y \geq 0, \quad m > 0, \quad [\text{Eq. IV.57}]$$

expresión que se reduce al modelo que acabamos de explicar en el caso en que  $m = 1$ . O dicho de otra manera, la distribución  $\zeta(1)$  coincide exactamente con la distribución  $\varepsilon(1)$ . Considerando que  $\log(\lambda T) = U/\alpha$ , entonces tendremos que

$$(\lambda T)^\alpha = e^U = Y. \quad [\text{Eq. IV.58}]$$

Para obtener la función de densidad de  $T$ , sustituiremos la variable de [Eq. IV.57] utilizando  $\alpha \lambda^\alpha t^{\alpha-1} dt = dy$  para obtener

$$f(t) = \frac{\alpha \lambda^{\alpha m} t^{\alpha m - 1} e^{-(\lambda t)^\alpha}}{\Gamma(m)}. \quad [\text{Eq. IV.59}]$$

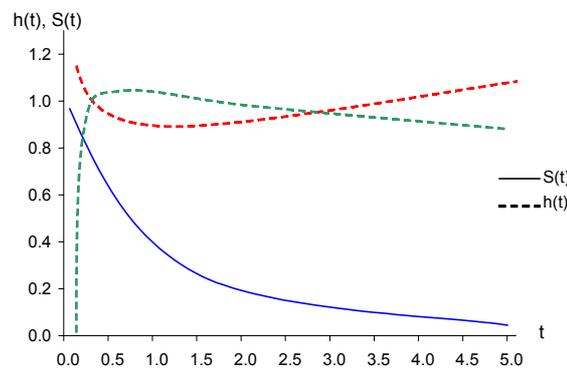
Ésta será, por tanto, la función de densidad de la familia Gamma generalizada. En el caso concreto de  $m = 1$ , quedará reducida a la distribución Weibull; si le imponemos la restricción de que  $\alpha = 1$ , entonces obtendríamos la distribución Gamma biparamétrica; mientras que, si imponemos una restricción que implique  $m = \alpha = 1$ , entonces estaríamos ante el caso exponencial.

La función de supervivencia, a partir de la sustitución de la variable  $t$  por  $t^\alpha$ , se convierte en una integral incompleta de Gamma, por lo que no se puede escribir mediante una especificación cerrada. Es por esto, por lo que tampoco se puede expresar con una única fórmula la función de riesgo. Esta última no tiene por qué ser

necesariamente monótona. De hecho, para  $\alpha m > 1$  y  $\alpha < 1$  tiene forma de U invertida, variando desde cero en  $t=0$  hasta cero cuando  $t \rightarrow \infty$ ; mientras que, en el caso de  $\alpha m < 1$  y  $\alpha > 1$  tiene forma de U, variando desde  $\infty$  cuando  $t \rightarrow 0$  hasta  $\infty$  cuando  $t \rightarrow \infty$ . Fuera de esas regiones, los valores de la función de riesgo varían monótonamente entre cero e  $\infty$ , aumentando si  $\alpha > 1$  y disminuyendo en caso contrario.

En la Ilustración IV.4 recogemos algunos ejemplos de distribuciones Gamma generalizadas.

**Ilustración IV.4: Funciones de riesgo y de supervivencia Gamma generalizadas**



Vemos cómo, a partir de una función de supervivencia monótona decreciente, dependiendo de los valores que tomen  $\alpha$  y  $m$ , la función de riesgo va a tener una forma u otra.

**3.2.3.4. Modelos de duración acelerada**

Los modelos recogidos en los dos últimos epígrafes son claramente casos especiales de una familia general de modelos conocida por el nombre de modelos de duración acelerada. En este tipo de modelos, la duración del evento objeto de estudio para una observación bajo la influencia de un vector de regresores independiente del tiempo X, se puede escribir como

$$T = \frac{T_0}{\lambda(x'\beta)}. \tag{Eq. IV.60}$$

En donde  $T_0$  representa a una variable aleatoria con una distribución no afectada ni por X ni por  $\beta$ . En el caso de la distribución Weibull, por ejemplo,  $T_0$  se distribuía como una potencia positiva de una variable exponencial unitaria, mientras que en el caso de la distribución Gamma generalizada se distribuía como una potencia positiva de una variable Gamma estándar.

El nombre de esta familia de distribuciones proviene del hecho de que la duración de una observación con su vector de covariables  $X$  sea acelerada (decelerada), en comparación con  $T_0$ , por el hecho de que  $\lambda$  sea superior (inferior) a la unidad.

Tomando logaritmos en la expresión [Eq. IV.60] obtenemos

$$\log T = -\log \lambda(x' \beta) + \log T_0. \quad [\text{Eq. IV.61}]$$

De aquí deducimos que los modelos de duración acelerada serán aquellos en los que el  $\log T$  satisfaga un modelo de regresión como el recogido en [Eq. IV.61] con un término de error cuya distribución sea la de  $\log T_0$ . Esta regresión será lineal en parámetros únicamente en el caso en que  $\lambda(\cdot)$  sea la función exponencial, en cualquier otro caso estaremos ante un caso no-lineal. En cualquier caso, Lancaster (1992) afirma que éste será siempre un modelo homocedástico.

### 3.2.3.5. Modelo de riesgo proporcional

A pesar de haber hablado de este tipo de modelos en el apartado dedicado a los modelos semiparamétricos, hemos considerado conveniente introducirlos nuevamente aquí, puesto que, como veremos a continuación, estos no son más que la especificación genérica de algunos de los modelos paramétricos.

Hasta ahora hemos comentado repetidamente que la función integrada de riesgo es una variable exponencial unitaria, lo que implicaba que, dado  $X$ ,

$$Z = \int_0^T h(s | X) ds \sim \varepsilon(1) \quad [\text{Eq. IV.62}]$$

o, lo que es lo mismo

$$\log Z(T | X) = U, \quad [\text{Eq. IV.63}]$$

donde  $e^U \sim \varepsilon(1)$  dado  $X$ . A partir de aquí hemos generado todas las anteriores familias de modelos de duración a partir de una restricción que imponía que  $Z(T | X)$  tomara la forma  $\lambda(X)T$ . Sin embargo, manteniendo la hipótesis de que  $e^U \sim \varepsilon(1)$ , pero considerando formas funcionales más genéricas para  $Z(T | X)$ , se obtiene otro tipo de familias de modelos de duración que también han tenido una gran aceptación entre los econométricos. La aceptación de este tipo de modelos se debe a que permiten incorporar recomendaciones de la teoría económica a la hora de elegir la forma funcional de la función de riesgo. A través de la utilización de esta metodología surgen los modelos de riesgo proporcional.

Estos, son modelos en los que, si las covariables son independientes del tiempo, como hemos visto, la función de riesgo toma la forma

$$h(t | X) = h_0(t)g(X), \quad [\text{Eq. IV.64}]$$

siendo  $h_0$  y  $g$  la misma función para todos los individuos.

Esta función de riesgo se conoce como función de riesgo proporcional porque el riesgo para dos individuos con vector de regresores  $X_1$  y  $X_2$  mantendrá la misma proporción  $g(X_1)/g(X_2)$  en todo momento  $t$ . Por ejemplo, la función exponencial con  $h(t | X) = e^{\beta'x}$  se trata de un modelo de riesgo proporcional en donde se ha supuesto que  $h_0 \equiv 1$ , siendo  $h_0$  la ‘función de riesgo base’.

Asumir un modelo de riesgo proporcional simplifica mucho la realización de inferencia en modelos de duración porque, como comenta Cox (1972), es posible estimar los parámetros desconocidos de  $g(X)$  sin necesidad de especificar la forma funcional de  $h_0$ , proporcionando así un estimador semiparamétrico de esas constantes. Sin embargo, habitualmente, los economistas han proporcionado una especificación completamente paramétrica del modelo, aunque han adoptado la hipótesis multiplicativa de [Eq. IV.64].

### 3.2.3.6. Modelo de riesgo log-lineal

Este tipo de modelos, en la misma línea que el modelo de riesgo proporcional, serán aquellos cuya función de riesgo se pueda escribir como

$$\log h(t | X) = \sum_{j=1}^K \gamma_j g_j(X(t), t), \quad [\text{Eq. IV.65}]$$

donde los valores de  $g_j$  son funciones conocidas, y en los que el concepto de linealidad hace referencia a linealidad en los parámetros  $\gamma_j$  y no en  $X$  ni en  $t$ .

Los modelos de riesgo log-lineal ofrecen la ventaja de que, como se puede demostrar, su función de verosimilitud, para muestras aleatorias obtenidas de distribuciones de  $T$ , que incluso pueden estar sesgadas por la derecha, es cóncava. Y este hecho resulta beneficioso a la hora de realizar estimaciones máximo verosímiles.<sup>116</sup>

---

<sup>116</sup> El modelo de riesgo proporcional  $h(t) = g(x)\alpha t^{\alpha-1}$ , coincide con el modelo Weibull que derivamos en el apartado 3.2.3.2 a través de la generalización del término de error. Siguiendo la notación de ese apartado,  $g$  tomará la expresión  $g = \lambda^\alpha$ . No obstante, la elección de forma funcional más habitual suele ser  $g(x) = e^{\beta'x}$ , en cuyo caso, dicho modelo Weibull se transforma en un modelo log-lineal.

### 3.2.3.7. Modelo de riesgo log-logístico

Como ya vimos al explicar el modelo Weibull, éste tenía un comportamiento monótono. Una función de riesgo igual de sencilla, pero que permite un comportamiento no monótono es la siguiente

$$h(t) = \frac{g(X)\alpha t^{\alpha-1}}{1 + g(X)t^\alpha} \quad [\text{Eq. IV.66}]$$

con la función de supervivencia siguiente

$$S(t) = (1 + g(X)t^\alpha)^{-1}, \quad [\text{Eq. IV.67}]$$

y la función de densidad siguiente

$$f(t) = \frac{g(X)\alpha t^{\alpha-1}}{(1 + g(X)t^\alpha)^2}. \quad [\text{Eq. IV.68}]$$

Hacer el cambio de variable  $y = \log t$  en [Eq. IV.68] proporciona

$$f(y) = \frac{\alpha e^{\alpha(y-\mu)}}{(1 + e^{\alpha(y-\mu)})^2} \quad -\infty < y < \infty, \quad [\text{Eq. IV.69}]$$

para  $\mu = \alpha^{-1} \log g(X)$ .

Esta función es la que se conoce como función de densidad logística de media  $\mu$  y varianza  $\pi^2/3\alpha^2$ . Por tanto,  $\log T$  presentará una distribución logística cuando su función de riesgo adopte la forma [Eq. IV.66] y tendrá una regresión lineal homocedástica sobre  $X$  cuando  $g(X)$  tome la forma  $e^{x^\beta}$ .

La función de riesgo de este tipo de modelos se comporta de la siguiente manera:

- Cuando  $\alpha < 1$  decrece monótonamente desde  $\infty$ , en el origen, hasta cero, a medida que  $t \rightarrow \infty$
- Cuando  $\alpha = 1$  decrece monótonamente desde  $g$ , en el origen, hasta cero, a medida que  $t \rightarrow \infty$
- Cuando  $\alpha > 1$  aumenta desde cero, en el origen, hasta un máximo en  $t = \left(\alpha - 1/k_1\right)^{1/\alpha}$  para, posteriormente, irse aproximando a cero a medida que  $t \rightarrow \infty$

Por tanto, presenta una forma generalizada de U invertida siempre que  $\alpha$  sea mayor que la unidad.

### 3.2.3.8. Modelo de riesgo Gompertz

Si escribimos el modelo de Weibull recogido en la nota a pie de página 116 a través de la expresión siguiente

$$h(t) = g(X)\alpha e^{(\alpha-1)\log t}, \quad [\text{Eq. IV.70}]$$

y aplicando la generalización que se deriva de la transformación de Box-Cox

$$t^{(\lambda)} = \frac{t^\lambda - 1}{\lambda}$$

$$\lim_{\lambda \rightarrow \infty} t^{(\lambda)} = \lim_{\lambda \rightarrow \infty} \frac{t^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow \infty} t^\lambda \log t = \log t \quad [\text{Eq. IV.71}]$$

podemos expresar la tasa de riesgo mediante expresión

$$h(t) = g(X)\alpha e^{\gamma t^{(\lambda)}}, \quad \gamma = \alpha - 1, \quad \lambda \geq 0. \quad [\text{Eq. IV.72}]$$

Partiendo de esta nueva especificación de la función de riesgo, si optamos por un valor de  $\lambda = 0$ , obtenemos de nuevo la expresión de la función de riesgo Weibull; mientras que si optamos por un valor de  $\lambda = 1$ , entonces obtenemos la distribución de riesgo Gompertz, cuya expresión viene recogida en la ecuación [Eq. IV.73]

$$h(t) = g(X)\alpha e^{\gamma t - \gamma}. \quad [\text{Eq. IV.73}]$$

Ésta función de riesgo es capaz de recoger la información necesaria de una distribución de datos de duración que puede estar incompleta en el caso de  $\gamma < 0$ , o lo que es lo mismo, en el de  $\alpha < 1$ , puesto que

$$\int_0^t h(s) ds \alpha e^{\gamma t} \quad [\text{Eq. IV.74}]$$

converge en el caso de que  $\gamma < 0$ .

En este caso, la función de riesgo decrece muy rápidamente, de manera que todos los individuos presentan una probabilidad de abandono del estado objeto de estudio cercana a 1, mientras que una pequeña fracción de la cohorte de individuos entrantes en ese estado, permanecerán en él de manera permanente.

A pesar de que la expresión [Eq. IV.72] sea una generalización de la función Weibull, también tiene la restricción de que sólo permite variaciones monótonas, puesto que  $dt^{(\lambda)}/dt = t^{\lambda-1}$  es de signo constante.

### 3.2.3.9. Modelo de duración log-normal

Por último, un modelo de duración que también suele ser bastante habitual en los trabajos empíricos, pero cuya función de riesgo no se obtiene, ni a través de la generalización sucesiva de la forma de la distribución del error, ni a partir de una generalización de la propia función de riesgo, es, entre otros, el modelo de duración log-normal.

En este modelo se supone que  $\log T$  se distribuye como una normal de media  $\mu$  y varianza  $\sigma^2$ . De esta manera, como en el caso de los tres primeros modelos de duración paramétricos, el modelo log-normal podrá proporcionar una estimación lineal homocedástica de  $\log T$ . Si  $\phi$  es la función de densidad de la distribución Normal y  $\Phi$  su correspondiente función de distribución, con su función complementaria  $\bar{\Phi}$ , la función de densidad para  $T$  en el modelo log-normal será

$$f(t) = \frac{\phi(x)}{\sigma t}, \text{ donde } x = \frac{\log t - \mu}{\sigma}. \quad [\text{Eq. IV.75}]$$

La función de supervivencia vendrá dada por

$$S(t) = \bar{\Phi}(x) \quad [\text{Eq. IV.76}]$$

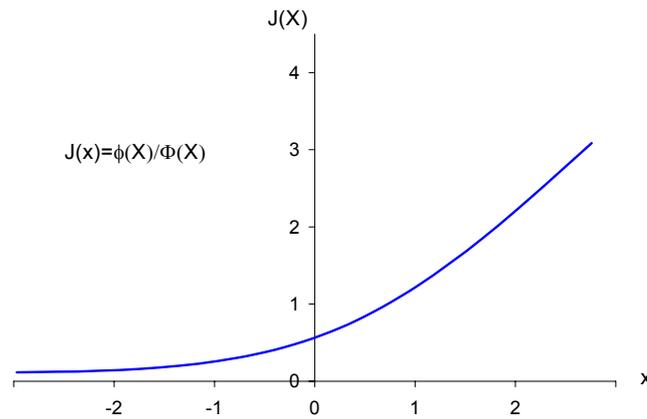
y la función de riesgo por

$$h(t) = \frac{\phi(x)}{\sigma t \bar{\Phi}(x)} = \frac{j(x)}{\sigma t}. \quad [\text{Eq. IV.77}]$$

Antes de entrar en las propiedades de la función de riesgo, vamos a analizar la función  $j(x) = \phi(x)/\bar{\Phi}(x)$ . Lancaster (1992) denota a esta función como función de riesgo normal estándar, aún a pesar de que la variable  $X$  no tome solamente valores positivos, motivo por el cual no debía catalogarla de función de riesgo. En cualquier caso, su forma viene recogida mediante la Ilustración IV.5.

Se puede observar cómo  $j(x)$  presenta un comportamiento monótono creciente desde cero a medida que  $X$  aumenta desde  $-\infty$  y que, a medida que  $x \rightarrow \infty$ ,  $j(x)$  se acerca a  $X$ .

**Ilustración IV.5: Función de riesgo normal estándar**



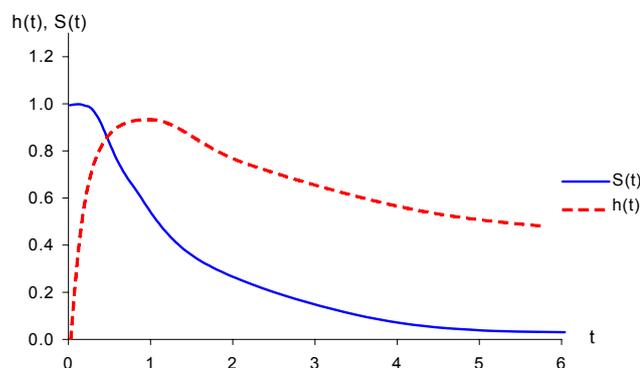
Volviendo ahora a la función de riesgo log-normal, podemos ver cómo, dado que  $j(x) \sim x$  para valores grandes de  $X$ , entonces  $h(t) \sim (\log t - \mu) / t\sigma^2$  para valores grandes de  $t$ , de manera que  $h(t)$  se aproxima a cero a medida que  $t \rightarrow \infty$ . Es más, dado que la función de densidad log-normal vale cero en el origen, entonces  $h(0) = 0$ . Por tanto, el riesgo comienza aumentando desde cero para, al final, volver a disminuir hasta cero. De hecho, tiene un único máximo en el punto  $x = (\log t - \mu) / \sigma$ , que satisface la ecuación

$$j(x) = \sigma + x, \quad [\text{Eq. IV.78}]$$

ecuación que puede resolverse a través de la Ilustración IV.5 para cualquier valor de  $\sigma$ .

La Ilustración IV.6 muestra la forma de una función de riesgo log-normal para  $\mu = 0$  y  $\sigma = 1$  con  $E(t) = 1.65$  y  $Var(T) = 4.67$ .

**Ilustración IV.6: Función de supervivencia y de riesgo log-normal**



#### 4. ESTIMACIÓN DE LA DURACIÓN DEL PERÍODO DE DESEMPLEO

El análisis de la duración del período de desempleo ha sido, como ya anticipábamos, uno de los principales objetos de estudio dentro del análisis de datos de duración bajo la perspectiva socioeconómica. Este fenómeno ha sido ampliamente analizado de forma empírica en diferentes países, como reflejan, entre otros, los estudios de Kiefer et al. (1985), Sider (1985), Han y Hausman (1990), Eckstein y Wolpin (1995) y Choi y Shin (2002) para la economía estadounidense, el estudio de Petrongolo (2001) para Gran Bretaña, el de Bratberg y Nilsen (1998) para la economía noruega, el de Melkersson (1999) para la economía sueca, el estudio de Kettunen (2002) para Finlandia, el de Theeuwes (1989) para los Países Bajos y el estudio de Biggeri et al. (2001) para Italia.

Igualmente, éste ha sido un objeto de preocupación entre investigadores españoles, por lo que son numerosos los estudios que se centran en el análisis de la duración del desempleo en la economía española. Así encontramos, entre otros, los de Ahn y Ugidos-Olazábal (1995), Lázaro-Pardo et al. (1995), Bover et al. (1997), Cañada-Vicinay et al. (1998), Toharia-Cortés et al. (1998), Ibáñez-Pascual (1999), Salas-Velasco (2000), Aguilar-Ramos (2001), Alba-Ramírez y Álvarez-Llorente (2001), Arranz y Muro (2001), Lassibille et al. (2001), Congregado-Ramírez y García-Pérez (2002) y Gámez-Amián y García-Pérez (2002). Sin embargo, con la excepción del análisis de Gil-Jurado (1998), existe una notable ausencia de estudios rigurosos en este tema para la economía canaria.

Este apartado se encargará de rellenar una parte de dicha laguna, mediante el estudio de la longitud del episodio del primer período de desempleo de los titulados de Formación Profesional Específica de cualquiera de los centros educativos de la isla de Gran Canaria, a través de la estimación de modelos de duración, con los datos procedentes de nuestra encuesta.

Ya en el capítulo segundo del presente trabajo, al analizar el proceso de búsqueda de empleo de los titulados, hicimos referencia, entre otros aspectos, a la duración de dicho proceso.

En lo que respecta a la búsqueda del primer empleo, y a partir de la información obtenida de las preguntas P.4<sup>117</sup>, P.11<sup>118</sup> y P.14<sup>119</sup>, dividimos el período que transcurre desde la finalización del Ciclo Formativo hasta la consecución del primer empleo -o hasta el momento de realización de la encuesta, para los que aún no hubiesen conseguido su primer empleo-, entre lo que llamamos primer período de inactividad y primer período de desempleo propiamente dicho.

---

<sup>117</sup> Pregunta la fecha de finalización del CF

<sup>118</sup> Pregunta la duración del período de búsqueda

<sup>119</sup> Pregunta la fecha de comienzo en el primer empleo

La duración de este primer período de desempleo, el previo a la consecución del primer trabajo tras la finalización del Ciclo Formativo, la obteníamos directamente de la respuesta a la pregunta P.11, mientras que la duración del período de inactividad se obtuvo a través de la siguiente diferencia:

$$T. \text{ de inactividad} = (\text{Fecha comienzo empleo} - \text{Fecha fin CF}) - \text{Busqueda} \quad [\text{Eq. IV.79}]$$

Dado que la muestra consta de titulados que ya disfrutaban de un empleo con anterioridad a la realización del Ciclo Formativo (*'titulados del grupo 1'*), titulados que consiguieron el empleo durante la realización del Ciclo (*'titulados del grupo 2'*), titulados que consiguieron el empleo tras finalizar el Ciclo (*'titulados del grupo 3'*) y titulados que no han conseguido un primer empleo tras la finalización del Ciclo (*'titulados del grupo 4'*), hemos considerado conveniente dirigir nuestro análisis al colectivo formado por los *'titulados del grupo 3'* y aquellos *'titulados del grupo 4'* que no estaban trabajando pero sí se encontraban buscando empleo, puesto que son los únicos para los que podemos hablar de un verdadero período de búsqueda de empleo *tras la finalización del Ciclo Formativo*.

Como vimos en el capítulo segundo del presente trabajo, el período medio que los *'titulados del grupo 3'* tardan en encontrar empleo, una vez superado su primer período de inactividad, es de alrededor de dos meses. Dicha duración es superior en el caso de las mujeres, quienes tardan cerca de tres meses, que en el de los hombres, quienes tardan en torno al mes y medio en conseguir empleo. De la misma manera, la duración del primer período de desempleo de los técnicos es dos semanas mayor a la de los técnicos superiores. Igualmente, detectamos grandes diferencias en la duración del período de desempleo en función de la Familia Profesional a la que pertenezca el Ciclo Formativo cursado, siendo los titulados en 'Textil, Confección y Piel', junto con los de 'Química', quienes presenta una mayor duración (en torno a 4 meses), encontrándonos, por el lado contrario, con los titulados de 'Mantenimiento y Servicios a la Producción', quienes no tardan dos semanas en conseguir su primer empleo. Las diferencias en la duración de dicho periodo debidas a la titularidad del centro de estudios, en cambio, resultaron ser prácticamente inexistentes.

Del total de titulados recogidos en nuestra muestra, en el momento de responder a la encuesta, el 8.2% aún no había podido incorporarse al mercado de trabajo como ocupado<sup>120</sup>. No obstante, de estos, tan sólo el 34% se encontraba en situación de desempleo, puesto que el 66% restante se declaró como inactivo. Por tanto, al incorporar a este 34% en el estudio de la duración de desempleo, realmente estamos incorporando un total de 78 titulados, cuya duración del primer período de desempleo presenta censura<sup>121</sup>. Pese a dicha censura, la duración media de su periodo de

---

<sup>120</sup> Estos son los *'Titulados del grupo 4'*.

<sup>121</sup> Estos 78 titulados representan una censura del 3.8% sobre la muestra seleccionada

desempleo a día de la encuesta, oscilaba en torno a los 6 meses y medio. De estos 78 titulados, el 75% presentaba una duración inferior al año y, tan sólo el 5% presentaba duraciones de desempleo superiores al año y medio, por lo que prácticamente no nos encontramos ante ningún caso de paro de larga duración.

El análisis de la duración del primer período de desempleo lo realizaremos en dos etapas. Comenzaremos realizando una estimación no paramétrica, para luego proceder a la estimación paramétrica. Así, inicialmente, analizaremos exclusivamente los períodos de duración, sin tomar en consideración ninguna otra variable que pueda influir sobre el mismo. Realizaremos los cálculos de tablas de vida y funciones de supervivencia para el colectivo de la muestra y, a continuación, volveremos a realizarlos diferenciando la misma en diversos subcolectivos, con la finalidad de detectar la posible existencia de diferencias entre grupos.

Una vez finalizado el análisis no paramétrico, podremos aventurarnos a realizar supuestos acerca de la posible distribución de las duraciones de salida de la situación de desempleo, a la vez que podemos intuir qué variables pueden afectar a dicha duración, por lo que procederemos a la realización del análisis paramétrico mediante la estimación de un modelo de distribución de supervivencia para el conjunto de la muestra seleccionada.

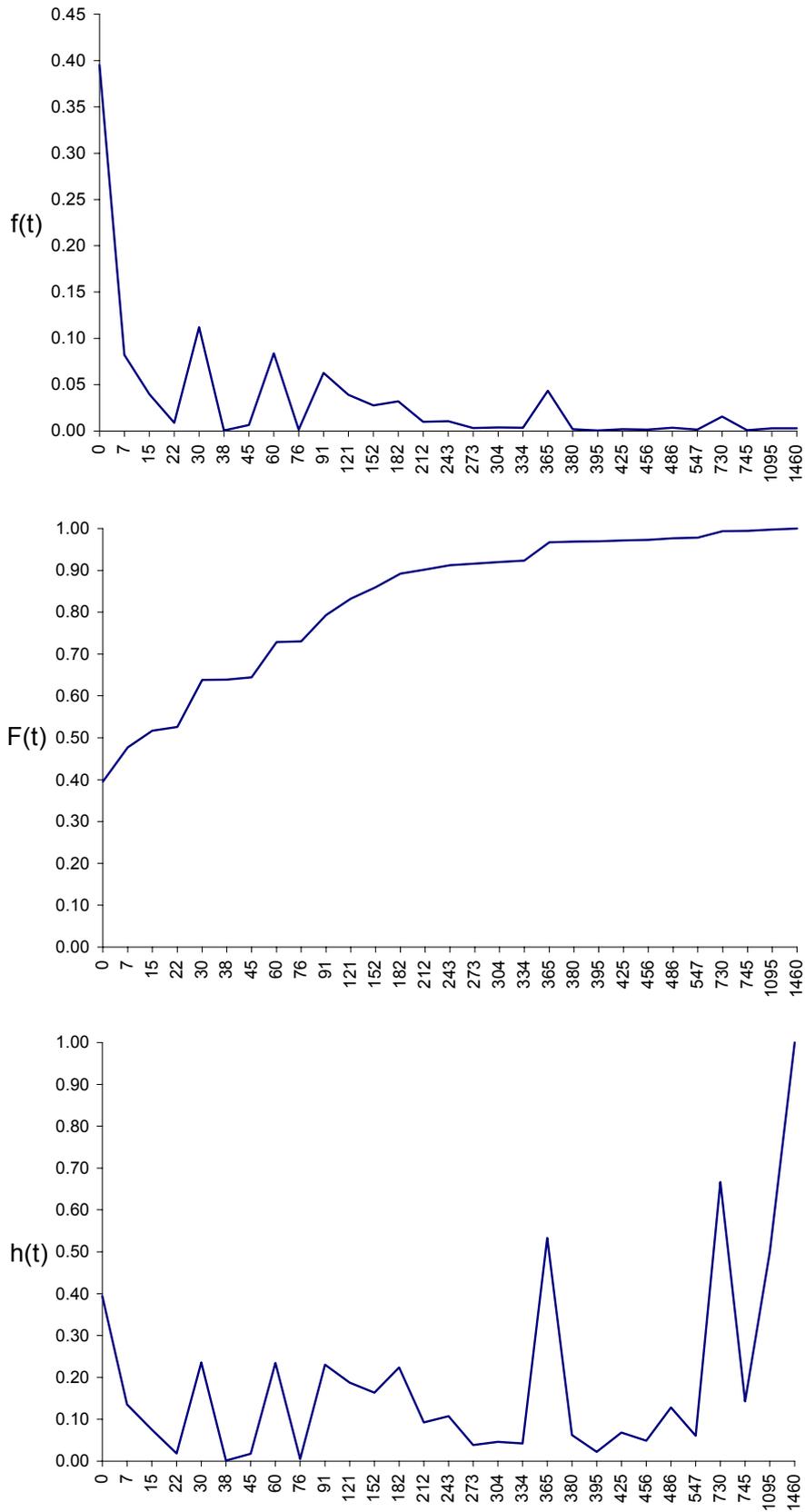
#### **4.1. ESTIMACIÓN NO PARAMÉTRICA**

##### **4.1.1. Los estimadores de Kaplan y Meier**

Comenzamos el análisis de los datos de duración del período de desempleo previo a la consecución del primer empleo, mediante la estimación de las funciones empíricas de densidad, distribución y riesgo, a través de la metodología propuesta por Kaplan y Meier (1958). La representación gráfica de dichas estimaciones viene recogida en la Ilustración IV.7.

Como ya hemos comentado previamente, la metodología propuesta por estos dos autores sugiere que el cálculo de las funciones se realice de forma tal que, al agrupar las duraciones por intervalos, estos recojan la información de un único individuo, sea ésta finalización real o censura. Por tanto, las categorías marcadas en el eje de abscisas de la Ilustración IV.7, responden a los diferentes períodos de desempleo, medidos en días, manifestados por los titulados de nuestra muestra.

Ilustración IV.7: Funciones empíricas de densidad, distribución y riesgo de la duración del desempleo previo al primer empleo (Kaplan-Meier)



La primera gráfica de la Ilustración IV.7, representa la función de densidad empírica, es decir, la distribución porcentual de la duración del primer período de desempleo de la muestra de técnicos y técnicos superiores. Como vemos, cerca del 40% de los titulados dijeron haber encontrado empleo nada más comenzar a buscarlo, por lo que presentan una duración de desempleo de 0 días. Igualmente se observa una cierta concentración en las duraciones de desempleo equivalentes a 1 semana (8%), 1 mes (11%), 2 meses (8%) y 3 meses (6%), distribuyéndose el resto de las duraciones, fundamentalmente entre los períodos de 4 y 6 meses (4% y 3% respectivamente) y de 1 año (4%). De esta distribución se deduce que los titulados presentan una cierta tendencia a ‘redondear’, a la hora de responder a la pregunta referida a la duración de su período de búsqueda de empleo. Aquellos que tardaron menos de un mes, optaban por ‘redondear’ en semanas, mientras que los que tardaron más de un mes, ‘redondeaban’, fundamentalmente, en meses. Este patrón de comportamiento nos facilitará, posteriormente, la agrupación de los datos en intervalos, con la finalidad de realizar la estimación a través del cálculo actuarial.

La segunda gráfica de la Ilustración IV.7 recoge la función de distribución empírica, es decir, la proporción acumulada de los distintos períodos de duración. A partir de la misma comprobamos que, cerca del 52%, presenta una duración del primer período de desempleo inferior a las dos semanas; en torno al 64% de los titulados, presenta una duración inferior al mes; el 73% permanece menos de 2 meses desempleado antes de lograr su primer trabajo; cerca del 90% ha conseguido su primer empleo transcurridos 6 meses de búsqueda; y, finalmente, el 96.6% lo ha conseguido una vez transcurrido el primer año de búsqueda.

Por último, la tercera gráfica de la Ilustración IV.7, muestra la función de riesgo empírica, es decir, la probabilidad condicional de que el individuo que ha estado parado hasta el instante  $t$  deje de estarlo en el período inmediatamente posterior. Vemos que los períodos que presentan una mayor probabilidad condicional de abandono de la situación de desempleo son, precisamente, el instante cero (39.5%), el año (53.3%), los dos años (63.3%) y los 4 años (100%)<sup>122</sup>. Aunque la probabilidad condicional de abandono también es relativamente alta para los que han llegado a una duración de desempleo de 1, 2, 3 y 6 meses (en torno al 23%).

#### 4.1.2. Estimación por el método actuarial

En la mayoría de los casos es preferible realizar la estimación de la función de riesgo una vez agrupados los datos en intervalos, puesto que de esta manera se puede detectar con mayor claridad qué tipo de relación de dependencia mantiene dicha función con el tiempo. Es por ello que optamos por la realización de las estimaciones

<sup>122</sup> Ésta es, precisamente, la máxima duración muestral, por lo que es de esperar que la probabilidad de abandono del desempleo a los 4 años sea la máxima.

nuevamente, pero ésta vez, a través del cálculo actuarial, obteniendo de esta manera la tabla de vida que refleja la Tabla IV.3.

**Tabla IV.3: Tabla de vida de la duración del desempleo previo al primer empleo**

Intervalo	Total	Finalizan	Censura	q	p	Función de Supervivencia	f	Función de Distribución	Función de Riesgo	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
0	7	2054	811	4	0,395	0,605	0,605	0,056	0,395	0,070
7	15	1239	168	2	0,136	0,864	0,523	0,010	0,477	0,018
15	30	1069	99	3	0,093	0,907	0,474	0,003	0,526	0,007
30	60	967	242	5	0,251	0,749	0,355	0,004	0,645	0,010
60	90	720	172	10	0,241	0,759	0,270	0,003	0,730	0,009
90	120	538	124	9	0,232	0,768	0,207	0,002	0,793	0,009
120	150	405	76	4	0,189	0,811	0,168	0,001	0,832	0,007
150	180	325	53	4	0,164	0,836	0,141	0,001	0,860	0,006
180	210	268	60	13	0,229	0,771	0,108	0,001	0,892	0,009
210	240	195	18	0	0,092	0,908	0,098	0,000	0,902	0,003
240	270	177	19	1	0,108	0,892	0,088	0,000	0,912	0,004
270	300	157	6	0	0,038	0,962	0,084	0,000	0,916	0,001
300	330	151	7	2	0,047	0,953	0,080	0,000	0,920	0,002
330	360	142	6	1	0,042	0,958	0,077	0,000	0,923	0,001
360	540	135	86	16	0,677	0,323	0,025	0,000	0,975	0,006
540	720	33	2	1	0,062	0,938	0,023	0,000	0,977	0,000
720	900	30	21	3	0,737	0,263	0,006	0,000	0,994	0,007
1080	.	6	6	0	1,000	0,000	0,000	.	1,000	.

La Tabla IV.3 nos proporciona información sobre la duración del primer período de desempleo de los 2054 *'titulados de los grupos 3 y 4'*, que constituyen nuestra muestra. Se trata de una distribución agrupada en la que la primera columna nos informa del período de tiempo considerado. Los intervalos, que al igual que antes vienen medidos en días, los hemos definido de manera que aparezcan cerrados por el lado izquierdo, y abiertos por el lado derecho. Hemos considerado conveniente agrupar la información comenzando por dos intervalos de una semana y uno de quince días; a continuación, pasamos a definir los intervalos de mes en mes, hasta llegar al año, a partir del cual, establecimos intervalos de 6 meses.

La segunda columna de la Tabla IV.3 muestra el número de personas que aparecen como aún desempleadas al comienzo del intervalo en cuestión. La tercera columna muestra el número de personas que finalizan su situación de desempleo durante el intervalo considerado. Por otro lado, la cuarta columna nos informa del número de observaciones que se pierden en el estudio en ese mismo intervalo de tiempo. Tal y como aparece recogida la información, para estas últimas personas sólo sabemos que la duración de su primer período de desempleo fue, como mínimo, del valor que aparece en el extremo inferior del intervalo, mientras que, para las que vienen recogidas en la columna tercera sí conocemos la duración completa de desempleo, con la precisión de la amplitud del intervalo. Estas cuatro primeras columnas constituyen un resumen de los datos, a través de los cuales se estiman las demás columnas de la tabla de vida.

La quinta columna de la Tabla IV.3 recoge así la proporción de sucesos terminales 'q', es decir, la probabilidad de que un individuo que entre en el intervalo encuentre empleo dentro de él. La sexta columna recoge la probabilidad complementaria a esta última 'p', por lo que muestra la proporción de supervivientes. A continuación se recoge la estimación de la función de supervivencia; la densidad de probabilidad, o estimación de la probabilidad de encontrar empleo por unidad de tiempo 'f'; la estimación de la función de distribución; y, finalmente, la estimación de la función de riesgo.

La estimación de la función de distribución a partir del cálculo actuarial, nos proporciona prácticamente los mismos valores que la estimación a través de la metodología de Kaplan y Meier, aunque con ligeras diferencias debidas a la corrección que se realiza en el cálculo actuarial ante la presencia de censura en los datos. Por tanto, las proporciones representadas en la segunda gráfica de la Ilustración IV.7 se corresponden básicamente con los valores recogidos en la penúltima columna de la Tabla IV.3.

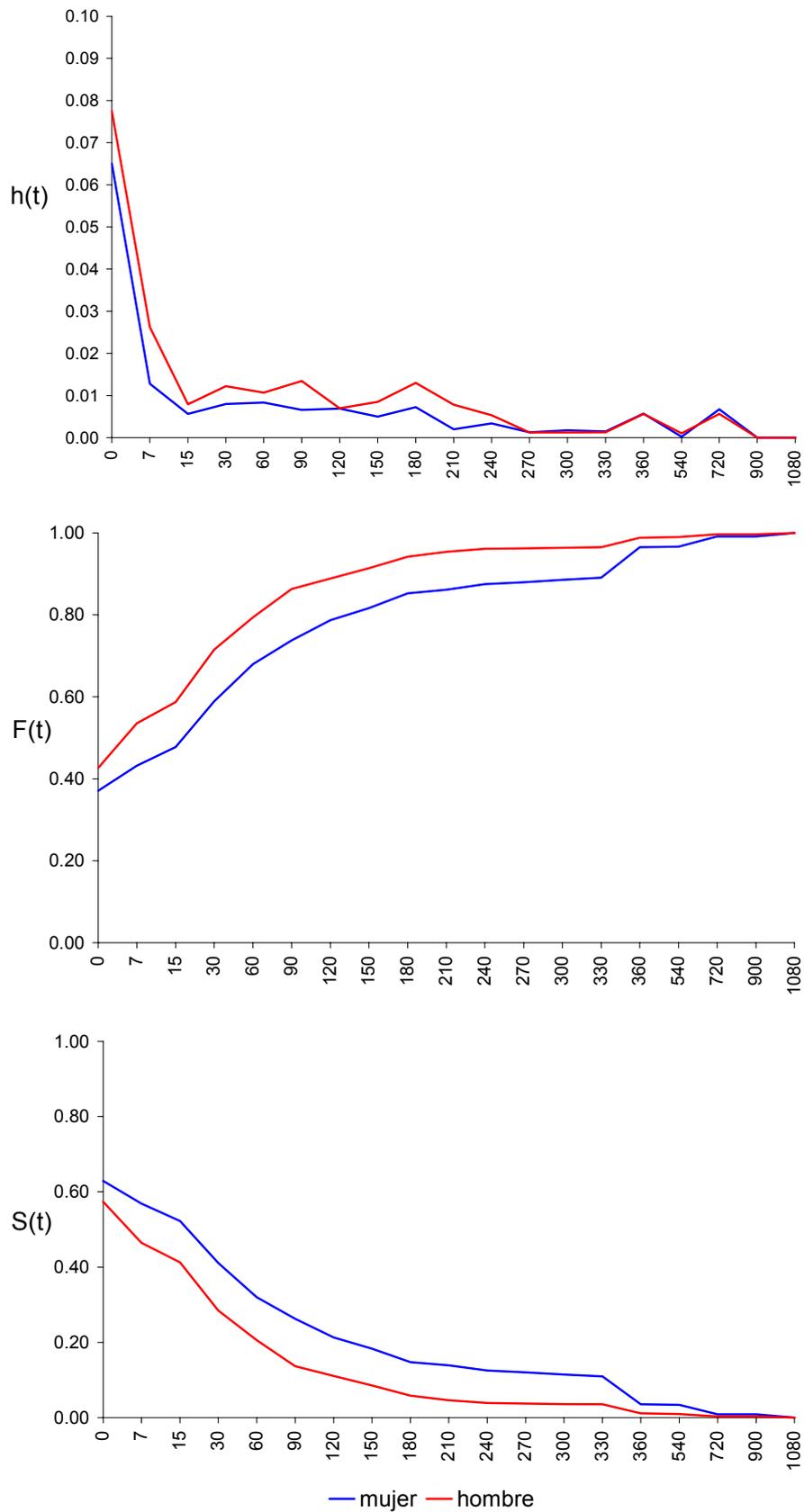
Sin embargo, a la hora de estimar la función de riesgo mediante el cálculo actuarial, los valores obtenidos son muy diferentes a los obtenidos a través de la estimación de Kaplan-Meier, debido al hecho de que la información, en esta ocasión, aparece agrupada en intervalos, con la consiguiente distribución del riesgo a lo largo del mismo. Es por ello por lo que el riesgo estimado de esta manera toma valores mucho menores. No obstante, esta metodología nos permite descubrir de forma más nítida la relación de dependencia existente entre la función de riesgo y el tiempo. En este caso, dicha dependencia es en el sentido negativo, puesto que la probabilidad condicionada de que un episodio de desempleo finalice resulta ser menor cuanto mayor ha sido la longitud de dicho episodio<sup>123</sup>. Según comentan Eckstein y Van Den Berg (2002), éste es el comportamiento habitual de la función de salida del desempleo, cuando se realiza la estimación paramétrica de máxima verosimilitud propuesta por Kaplan y Meier.

La representación gráfica de las funciones de riesgo, distribución y supervivencia obtenidas mediante el cálculo actuarial las recogemos desde la Ilustración IV.8 hasta la Ilustración IV.11. En lugar de realizar una representación conjunta para toda la muestra, hemos optado por graficar dichas funciones segregando la muestra en función del género (Ilustración IV.8), del nivel del Ciclo Formativo (Ilustración IV.9), de la titularidad del centro de estudios (Ilustración IV.10) y de la Familia Profesional a la que pertenece el Ciclo Formativo cursado (Ilustración IV.11), con la finalidad de detectar si existen diferencias entre estos colectivos en lo que respecta a la duración de ese primer período de desempleo.

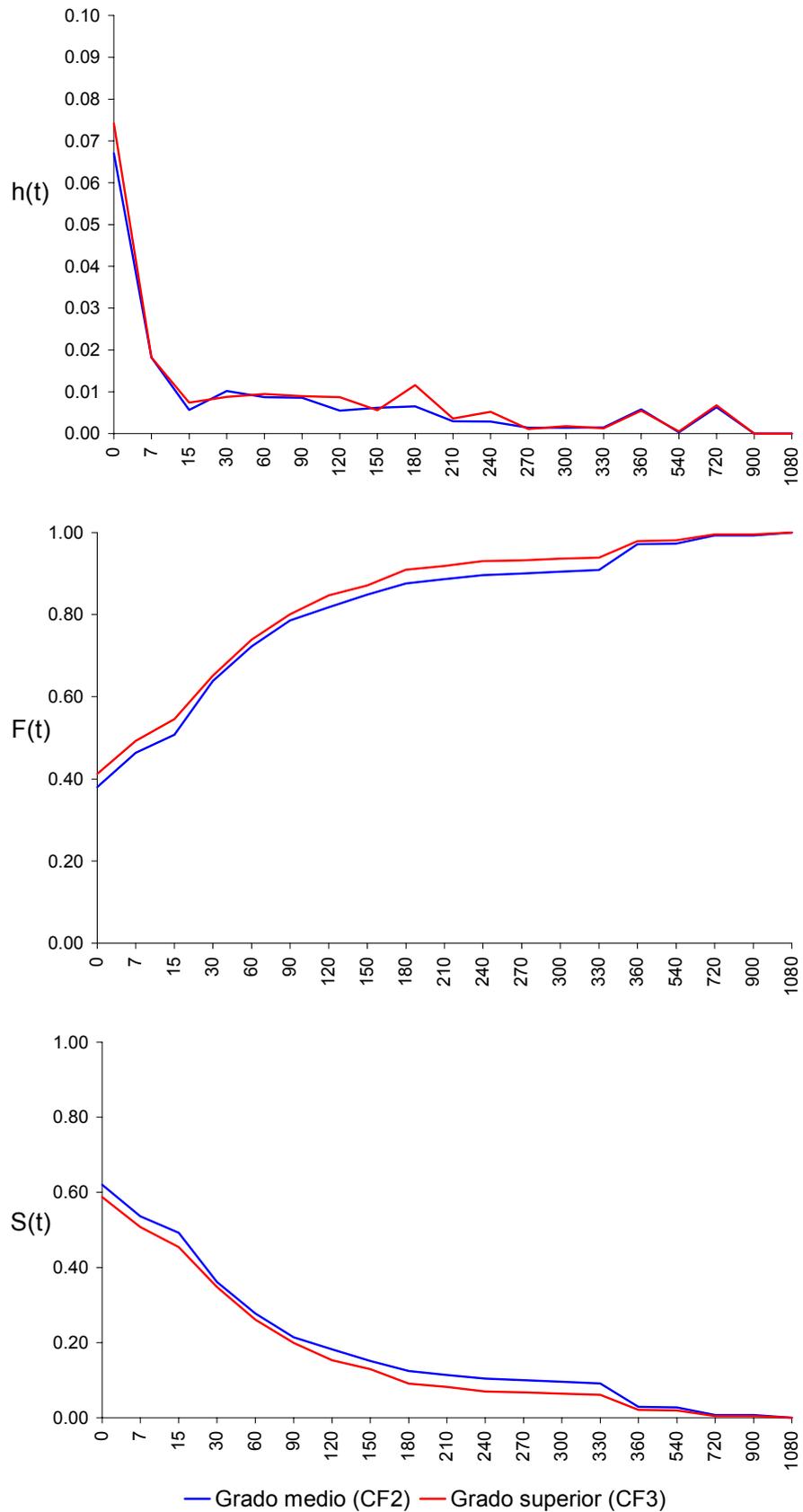
---

<sup>123</sup> Esta información nos será útil a la hora de sugerir una posible forma funcional en la estimación paramétrica.

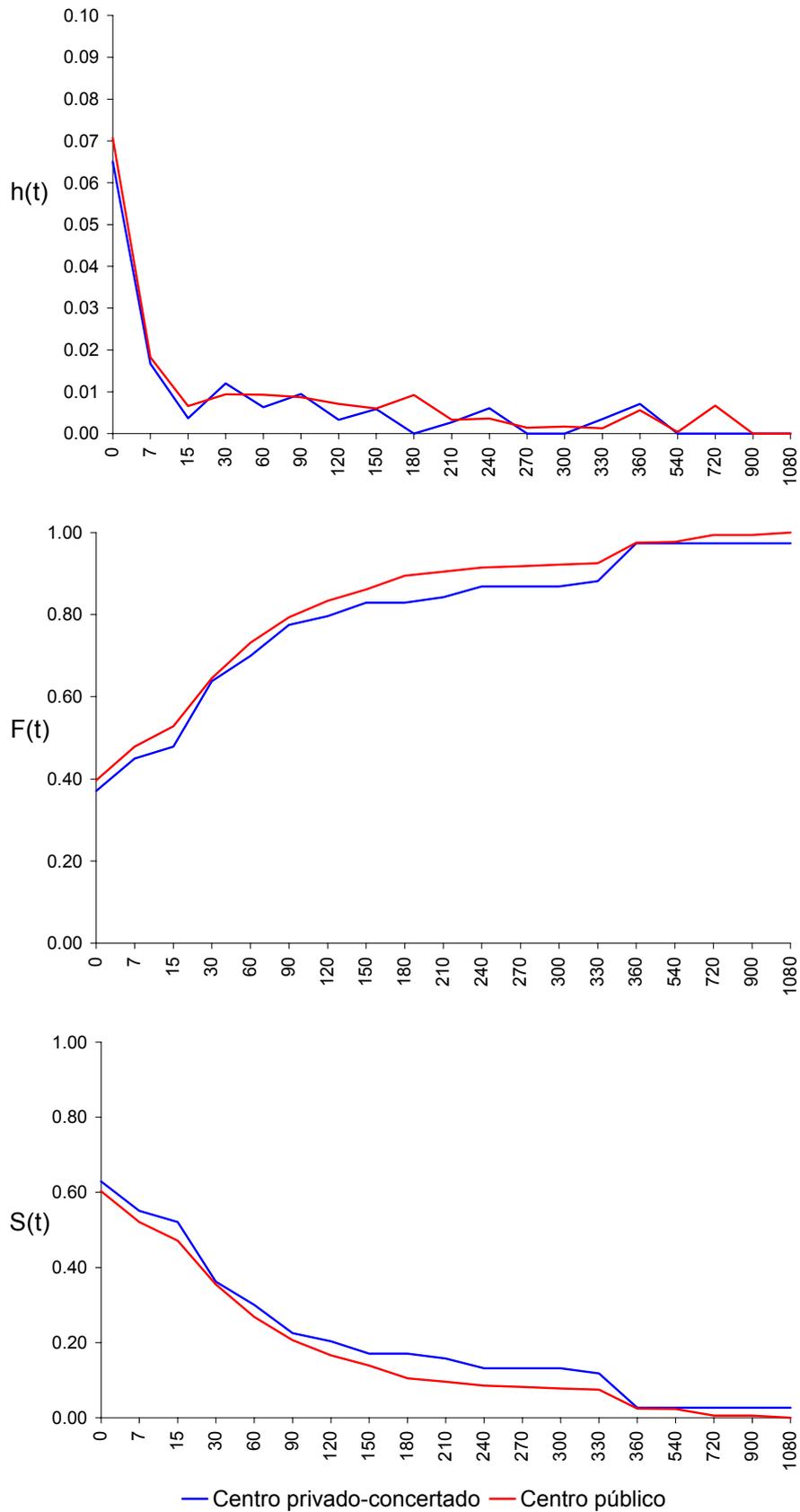
Ilustración IV.8: Función de riesgo, de distribución y de supervivencia de la duración del desempleo previo al primer empleo por género



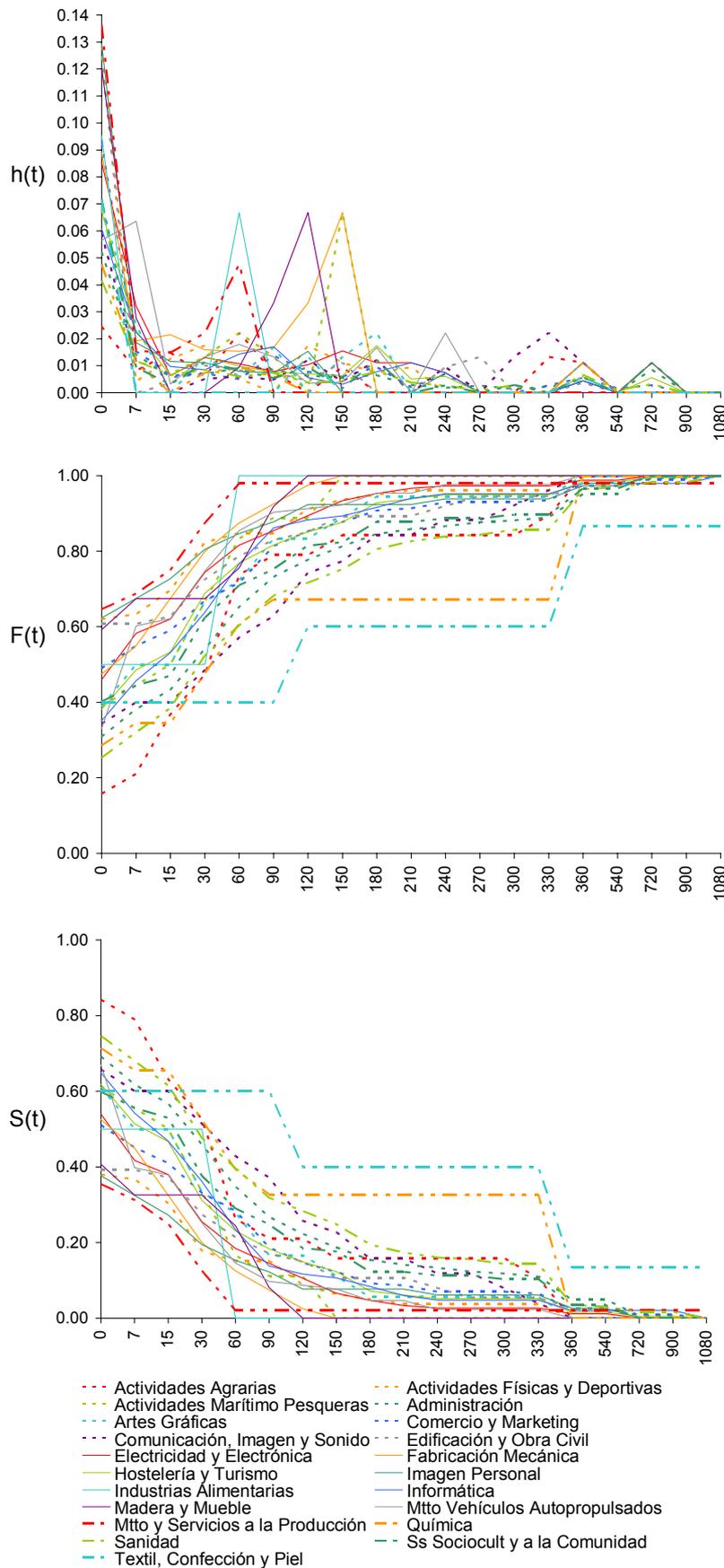
**Ilustración IV.9: Función de riesgo, de distribución y de supervivencia de la duración del desempleo previo al primer empleo por nivel del CF**



**Ilustración IV.10: Función de riesgo, de distribución y de supervivencia de la duración del desempleo previo al primer empleo por titularidad del centro**



**Ilustración IV.11: Función de riesgo, de distribución y de supervivencia de la duración del desempleo previo al primer empleo por Familia Profesional**



De la mera observación de dichas ilustraciones, parece existir diferencias en cada una de ellas, puesto que las curvas graficadas no se superponen unas a otras. Sin embargo resulta conveniente realizar previamente un análisis de homogeneidad de los estratos que nos confirme la existencia de diferencias significativas entre cada uno de los diferentes colectivos.

#### 4.1.2.1. Contraste de homogeneidad de estratos

Como ya comentamos en el capítulo segundo, y posteriormente recordamos en el capítulo tercero, a la hora de realizar cualquier tipo de estimaciones con nuestra muestra, hemos de ponderar la misma, puesto que el diseño muestral se realizó de manera que el colectivo de los titulados en 'Administración' quedaba infra-representado en comparación con el resto de Familias Profesionales. Por tanto, a la hora de realizar la estimación no paramétrica de la duración del primer período de desempleo, nuestra muestra ha sido nuevamente ponderada<sup>124</sup>, lo que limita la utilización de algunos de los contrastes de homogeneidad habituales.

Dada la existencia de censura en los datos, los contrastes de igualdad de poblaciones independientes de Wilcoxon y Mann-Whitney no son aplicables, pero si a esto añadimos que la muestra está ponderada, el único contraste de homogeneidad aplicable es el contraste de 'Cox' (StataCorp., 2001)<sup>125</sup>, el cual, a su vez, está basado en el log-rank test propuesto por Mantel y Haenszel (1959).

A continuación vamos a mostrar cómo se lleva a cabo dicho contraste de manera general, para posteriormente poder analizar los resultados obtenidos al aplicarlo a las diferentes estratificaciones sugeridas previamente.

Sea  $t_1 < t_2 < \dots < t_k$  las diferentes duraciones observadas en la muestra. Llamemos  $d_j$  al número de observaciones que finalizan tras una duración  $t_j$  y  $n_j$  al conjunto de riesgo justo antes del momento  $t_j$ . Añadiendo el subíndice 'i' haremos referencia a cada uno de los diferentes grupos, siendo  $i = 1, 2, \dots, r$ .

Estamos interesados en contrastar la siguiente hipótesis nula

$$H_0 : h_1(t) = h_2(t) = \dots = h_r(t) \quad [\text{Eq. IV.80}]$$

donde  $h(t)$  es la función de riesgo en el momento  $t$ , frente a una hipótesis alternativa que sugiera que al menos una de las  $h_i(t)$  es diferente en algún  $t_j$ .

<sup>124</sup> De nuevo se ha realizado la ponderación con el comando 'pweight' del programa Stata. Los valores de dicha ponderación son los recogidos en la Tabla II.4 (capítulo II).

<sup>125</sup> El término de contraste de 'Cox' ha sido acuñado en el manual de referencias del programa Stata, puesto que se trata de un contraste de log-rank que utiliza regresiones de Cox.

En caso de que la hipótesis nula fuera cierta, el número esperado de finalizaciones del grupo  $i$  en el momento  $t_j$  sería  $\hat{d}_{ij} = n_{ij}d_j/n_j$ , obteniendo el siguiente estadístico de contraste

$$u' = \sum_{j=1}^r W(t_j) (d_{1j} - \hat{d}_{1j}, \dots, d_{rj} - \hat{d}_{rj}). \quad [\text{Eq. IV.81}]$$

Donde  $W(t_j)$  es una función de ponderación positiva que toma el valor cero cuando  $n_{ij}$  vale cero.

Este estadístico, que es genérico para cualquier contraste de homogeneidad, tomará diferentes especificaciones dependiendo de la función de ponderación elegida. Así, en el caso concreto del log-rank test de Mantel y Haenzel, se asigna un peso igual a la unidad para cada momento de finalización  $t_j$ , pues se supone que la función de riesgo, si no es igual, es proporcional entre los diferentes grupos, por lo que la contribución de cada duración sobre el conjunto del estadístico debe ser la misma.

La matriz de varianzas covarianzas  $V$  para el estadístico  $u$  estará compuesta por los elementos

$$V_{il} = \sum_{j=1}^r \frac{W(t_j)^2 n_{ij} d_j (n_j - d_j)}{n_j (n_j - 1)} \left( \omega_{il} - \frac{n_{ij}}{n_j} \right) \quad [\text{Eq. IV.82}]$$

donde  $\omega_{il} = 1$  si  $i = l$  y  $\omega_{il} = 0$  en otro caso.

En caso de tener una muestra no estratificada, el estadístico  $u'V^{-1}u$  se distribuirá como una  $\chi^2$  con  $r-1$  grados de libertad.

El contraste de 'Cox' se realiza estimando un modelo de riesgo proporcional de Cox con  $r-1$  variables índice. Cada una de estas variables hará referencia a cada uno de los diferentes grupos, quedando uno de ellos omitido como referencia. El valor del contraste  $\chi^2$  se corresponderá entonces con el valor del contraste de Wald o con el del contraste de razón de verosimilitudes.

Se trata, por tanto, de comparar la probabilidad de supervivencia empírica con la estimada. Para ello se calcula para cada intervalo, tomando en consideración el global de la muestra, el número de individuos  $n_j$  que conforman el conjunto de riesgo y el número de observaciones  $d_j$  que finalizan tras una duración  $t_j$ . Para  $r$ -grupos, se calculará para cada intervalo:

$$n_{jT} = n_{j1} + n_{j2} + \dots + n_{jr} \quad \text{y} \quad d_{jT} = d_{j1} + d_{j2} + \dots + d_{jr} \quad [\text{Eq. IV.83}]$$

A continuación, para cada grupo, en cada intervalo, se calcula el número de finalizaciones predichas

$$\hat{d}_{jr} = d_{jT} \times \frac{n_{jr}}{n_{jT}} \quad [\text{Eq. IV.84}]$$

donde  $r$  identifica al grupo en cuestión.

Finalmente, sumando los valores predichos y observados para cada grupo se puede calcular el estadístico  $\chi^2$

$$\Sigma d_1 = \sum d_{j1} \quad (\text{la suma de todas las } d_j \text{ del grupo 1}) \quad [\text{Eq. IV.85}]$$

$$\Sigma \hat{d}_1 = \sum \hat{d}_{j1} \quad (\text{la suma de todas las } \hat{d}_j \text{ del grupo 1}) \quad [\text{Eq. IV.86}]$$

etc.

$$\chi^2 = \frac{(\Sigma d_1 - \Sigma \hat{d}_1)^2}{\Sigma \hat{d}_1} + \frac{(\Sigma d_2 - \Sigma \hat{d}_2)^2}{\Sigma \hat{d}_2} + \dots + \frac{(\Sigma d_r - \Sigma \hat{d}_r)^2}{\Sigma \hat{d}_r} \sim \chi_{r-1}^2 \quad [\text{Eq. IV.87}]$$

En la realización del contraste de 'Cox' obtenemos un valor del riesgo relativo para cada uno de los grupos. Éste no es más que la exponencial del coeficiente de una regresión de Cox re-normalizada, aunque dicha re-normalización no juega ningún papel en el cálculo del estadístico de contraste, puesto que ésta se elige de forma tal que la media ponderada de los coeficientes de regresión para el número esperado de finalizaciones de cada grupo valga cero, lo que equivale a decir que el riesgo valga uno.

Por tanto, sean  $b_i$ , con  $i=1,2,\dots,r-1$ , los coeficientes estimados y sea  $b_r=0$ . La constante  $K$  se calculará como

$$K = \sum_{i=1}^r \frac{\hat{d}_i b_i}{d} \quad [\text{Eq. IV.88}]$$

donde  $\hat{d}_i = \sum_j \hat{d}_{ij}$  es el número esperado de finalizaciones para el grupo  $i$ ,  $d$  es el número total de finalizaciones entre todos los grupos y  $r$  el número de grupos. El riesgo relativo se calculará, finalmente, como  $e^{(b_i-K)}$ .

Aplicando el contraste de homogeneidad de 'Cox' a las diferentes estratificaciones propuestas a través de la Ilustración IV.8 a la Ilustración IV.11, obtenemos los siguientes resultados:

**Tabla IV.4: Contrastes de homogeneidad de estratos para el análisis de la duración del desempleo previo al primer empleo**

Estratos	Estadístico	Probabilidad
Género	47,03	0,0000
Nivel del Ciclo Formativo	4,88	0,0271
Titularidad del centro	0,81	0,3685
Familias Profesionales	155,64	0,0000

De los valores observados de los estadísticos de prueba podemos concluir que existe evidencia que permite rechazar la hipótesis nula de igualdad de las funciones de riesgo (y, por tanto, de supervivencia y distribución), en el caso de la estratificación de la muestra por género, nivel del Ciclo Formativo y Familias Profesionales. En cambio, en lo que respecta a la diferenciación por titularidad del centro, el contraste de homogeneidad de 'Cox' no aporta evidencias que nos permita rechazar dicha hipótesis.

La Ilustración IV.8 muestra las funciones de riesgo, distribución y supervivencia diferenciando entre hombres y mujeres<sup>126</sup>. Como vemos, en todos los períodos de duración, la probabilidad condicionada de abandonar la situación de desempleo es superior para los hombres que para las mujeres, al igual que lo es la probabilidad no condicionada de salir de esa situación de desempleo. Se observa la misma relación de dependencia negativa entre la función de riesgo y el tiempo, tanto para el caso de los hombres, como para el de las mujeres, sufriendo esta función pequeños repuntes, siempre superiores entre los hombres, en los períodos de 1 mes, 3 meses y 6 meses de desempleo.

Tanto los hombres, como las mujeres, presentan una alta probabilidad de salida del desempleo en un período inferior a una semana, siendo ésta del 42.6% para los hombres y del 37% para las mujeres. A partir de esa duración, hasta llegar al año, las diferencias en la probabilidad no condicionada de conseguir un primer empleo, son acusadas entre hombres y mujeres, con una clara ventaja, de en torno al 10%, para los primeros. Al cabo del año las diferencias se acortan, pasando a ser tan sólo del 2.3%. Mientras que, al llegar a los 2 años de desempleo, prácticamente no existen diferencias por género.

En la Ilustración IV.9 recogemos las funciones de riesgo, distribución y supervivencia para técnicos y técnicos superiores. Las diferencias entre ambos colectivos en la probabilidad de salida del desempleo son bastante menos acusadas que las diferencias detectadas por género, aunque al observar la distribución de la probabilidad no

<sup>126</sup> Al tratarse de la representación de una tabla de vida en la cual las duraciones han sido agrupadas en intervalos, el valor que aparece representado al comienzo de cada intervalo, debe considerarse como un valor estable a lo largo del mismo, hasta el comienzo del siguiente. La representación gráfica exacta de dicha tabla debería mostrar gráficas escalonadas, en lugar de continuas.

condicionada de salida  $F(t)$  o de permanencia  $S(t)$ , comprobamos cómo los técnicos superiores presentan una cierta ventaja sobre los técnicos en períodos inferiores al año. Las mayores diferencias entre estos dos colectivos aparecen tras los 4 y 6 meses de desempleo, momento en el cual los técnicos superiores experimentan mayores probabilidades de salida en su función de riesgo.

La Ilustración IV.10 presenta las funciones de riesgo, distribución y supervivencia diferenciando a los titulados en función de la titularidad de su centro de estudios. Estas gráficas parecen presentar diferencias más acusadas que las recogidas en la Ilustración IV.9. Sin embargo, la realización del contraste de homogeneidad de 'Cox' nos indicaba que no parecían existir diferencias significativas. De la observación de la función de distribución y de supervivencia, podríamos suponer que los titulados de centros privados se enfrentan a peores oportunidades en el mercado de trabajo, sin embargo, si nos fijamos en la función de riesgo, la pauta de comportamiento no es tan clara, pues las curvas aparecen entrelazadas a lo largo de todos los períodos de duración.

Finalmente, la Ilustración IV.11 recoge las curvas de riesgo, distribución y supervivencia diferenciando entre las 21 Familias Profesionales. En este caso, las diferencias son bastante acusadas. Ya de entrada, basta comprobar, cómo la función de riesgo durante la primera semana de desempleo toma valores tan dispares como el 2.5% para 'Actividades Agrarias' o el 13.6% de 'Mantenimiento y Servicios a la Producción', cuando el valor medio de la muestra para esa duración es del 7%.

La probabilidad condicionada de abandono del desempleo durante la segunda semana es, con diferencia, para los titulados de 'Mantenimiento de Vehículos Autopropulsados', la mayor de todas (6.4% frente a probabilidades inferiores al 3% para el resto). Esta situación se repite, para la probabilidad condicionada de salida durante el segundo mes de desempleo, en el caso de los titulados en 'Industrias Alimentarias' (6.6%) y 'Mantenimiento y Servicios a la Producción' (4.8%). En el cuarto mes, la ventaja es para los titulados en 'Madera y Mueble' (6.6%). Mientras que, en el quinto mes, la ventaja es para los de 'Fabricación Mecánica' (6.5%) y 'Actividades Marítimo Pesqueras' (6.6%).

Las funciones de distribución y supervivencia nos muestran cómo la Familia Profesional que se enfrenta a peores oportunidades laborales en el mercado de trabajo, parece ser la de 'Textil, Confección y Piel', puesto que, al cabo de tres años, aún mantiene una proporción de desempleo del 13.3%, cuando en el resto de las familias, este porcentaje no llega al 3%. Por el lado contrario, la Familia Profesional que se enfrenta a las mejores oportunidades laborales es la de 'Mantenimiento y Servicios a la Producción', puesto que, al cabo de los dos meses, presenta una proporción de desempleados de tan sólo el 2%.

En líneas generales se puede observar que, transcurridos tres meses desde el comienzo de la búsqueda de empleo, la probabilidad de mantenerse en situación de

desempleo es inferior al 50% para todas las Familias Profesionales (desde el 40% para 'Textil, Confección y Piel' hasta el 0% para 'Madera y Mueble'). Transcurridos seis meses, esta probabilidad es inferior al 20% para todas las familias, con la excepción de 'Textil, Confección y Piel' (40%) y 'Química' (32.7%). Y, finalmente, transcurrido un año, la probabilidad de continuar desempleado es inferior al 5% para todas las Familias Profesionales, con la excepción de los titulados en 'Textil, confección y Piel' (13%).

## 4.2. ESTIMACIÓN PARAMÉTRICA

Una vez finalizado el análisis no paramétrico, pasamos a analizar cuáles son los rasgos más significativos que rigen el período de transición entre la finalización de los estudios y el acceso al mercado de trabajo. Para ello, al igual que hemos hecho en la estimación no paramétrica, nos centraremos exclusivamente en el período de búsqueda de empleo, obviando cualquier posible período de inactividad previo.

De esta manera trataremos de comprobar qué características individuales, así como qué factores del entorno económico, inciden sobre la probabilidad de encontrar el primer empleo tras la finalización de los estudios. El objetivo de este apartado, por tanto, radica en la determinación de los factores que facilitan el acceso al mercado de trabajo, en términos de su incidencia sobre el período transcurrido desde que los titulados comienzan a buscar su primer empleo, tras la finalización de la Formación Profesional Específica, hasta que consiguen acceder al mismo.

### 4.2.1. Factores condicionantes de la duración del desempleo

El marco teórico del análisis se encuadra en el modelo de búsqueda de empleo de Holzer (1988), según el cual los individuos desempleados actuarán conforme a un proceso de búsqueda de empleo que trate de maximizar la suma de sus utilidades actuales y futuras.

La utilidad actual dependerá, por un lado, de los ingresos menos el coste monetario de la búsqueda de empleo, y por otro lado, de la diferencia entre el valor del tiempo de ocio y el coste del tiempo dedicado a la búsqueda. Por su parte, la utilidad futura dependerá de la probabilidad de estar empleado o desempleado en el futuro, lo que, a su vez, es función de la distribución probabilística de la oferta de empleos y de la oferta salarial.

Formalmente, bajo el modelo de Holzer (1988), los desempleados tratan de maximizar la siguiente función de utilidad:

$$U_t = V\left(Y - \sum c_i B_i, L - \sum B_i\right) + \pi(B_1, B_2) [1 - f(\bar{w})] E\psi\left[\frac{w}{\bar{w}}\right] + \left\{1 - \pi(B_1, B_2) [1 - f(\bar{w})]\right\} U_{t+1} \quad [\text{Eq. IV.89}]$$

donde  $\bar{w}$  es el salario de reserva;  $\pi$  es la función de densidad de la oferta de empleos;  $f(w)$  es la distribución de la oferta salarial;  $B_i$  la intensidad de búsqueda del método  $i$ -ésimo, que presenta un coste monetario asociado  $c_i$ ;  $V$  es la utilidad actual;  $\psi$  es la función de utilidad del empleo en el momento  $t+1$ ;  $Y$  es la renta y  $L$  es el ocio.

Bajo esta especificación, hay que tener en cuenta que la productividad de cada método de búsqueda de empleo disminuye cuando su coste aumenta, siendo diferentes entre individuos, tanto el coste como la productividad, en función de sus características socioeconómicas.

Este modelo, por tanto, concuerda con la aportación de Nickel (1979), quien desarrolló un sencillo modelo de búsqueda de empleo a través del cual concluyó que la probabilidad de abandonar el desempleo era una constante que dependía de un vector de características individuales  $x$  y de otras variables que afectaban al salario de reserva.

Por otro lado, el análisis de los determinantes de la duración del desempleo no se puede llevar a cabo sin contemplar las aportaciones a los modelos de búsqueda de empleo de Mortensen (1986) y Devine y Kiefer (1991), quienes comentan que, la tasa de salida del desempleo se puede obtener, mediante la multiplicación de la probabilidad de obtener una oferta de empleo por la probabilidad de que dicha oferta sea aceptable. La tasa de llegada de ofertas, según estos autores, será función de la demanda de trabajo para las características del individuo considerado y para su esfuerzo de búsqueda.

Por tanto, la especificación del modelo causal paramétrico, que explica la duración del primer período de desempleo, se realizará, incorporando como covariables, aquellas que emanen de las aportaciones teóricas de los autores mencionados previamente.

En la línea argumental de Mortensen (1986) y Devine y Kiefer (1991), Congregado-Ramírez y García-Pérez (2002) consideran que, el componente de demanda de trabajo, que afecta a la tasa de llegada de ofertas, puede ser captado mediante la utilización de variables agregadas, como pueden ser la tasa de crecimiento del PIB o la tasa de desempleo de la zona de residencia del individuo. Igualmente consideran que, la tasa de llegada de ofertas, depende también de la red de contactos del individuo, optando por introducir como indicador de dicha red, el número de contactos que puedan ofrecerles sus progenitores, a través de la introducción del nivel de formación y la situación profesional de los mismos, como *proxy* para dicha red. Por último, como factor del lado de la demanda de trabajo que puede tener un efecto significativo sobre la probabilidad de recibir una oferta de empleo, consideran la utilidad de introducir variables de localización como indicadores de la tasa de desempleo regional.

En lo que respecta a la disposición de los individuos a aceptar una oferta laboral, ésta puede venir determinada por las expectativas que tenga el trabajador sobre las posibles

ofertas futuras, así como por sus preferencias por el trabajo, en la medida en que estos factores influyen en la determinación de su salario de reserva. No obstante, dado que, estas variables no son directamente observables, se puede utilizar como aproximación a las mismas la edad, el nivel educativo o el lugar de residencia (Congregado-Ramírez y García-Pérez, 2002).

Sin embargo, estas características personales de edad y nivel educativo, como componentes del lado de la oferta de trabajo, no presentan de antemano un efecto claro sobre la probabilidad de la duración del desempleo (Lázaro-Pardo et al., 1995). De hecho, la educación aumenta *-caeteris paribus-* la probabilidad de recibir una oferta de empleo, pero también disminuye la probabilidad de aceptar una determinada oferta salarial, puesto que el salario de reserva aumenta con el nivel educativo. Igualmente, la edad disminuirá la probabilidad de recibir ofertas de empleo, a la vez que disminuirá el salario de reserva y, por tanto, aumentará la probabilidad de aceptar la oferta salarial. Por ello, introducir dichas covariables en el modelo de duración, no conlleva unas expectativas previas claras en cuanto a su influencia sobre la duración del primer período de desempleo.

En lo que respecta al salario de reserva, que, como hemos comentado influye directamente en la duración del período de desempleo, éste también se ve afectado por la percepción o no de una prestación por desempleo (Lázaro-Pardo et al., 1995 y Bover et al., 1997), en el sentido de que si se está cobrando dicha prestación, el salario de reserva aumenta, reduciendo, por tanto, la probabilidad de aceptar un empleo y aumentando así la duración del desempleo. No obstante, dado que nuestro modelo se centra en el período de desempleo previo al primer trabajo de los técnicos y técnicos superiores, no cabe la posibilidad de que éstos estén cobrando prestación por desempleo, puesto que aún no han conseguido su primer trabajo<sup>127</sup>. De todas formas, Han y Hausman (1990) comprobaron que, a pesar de que dicha prestación tenía un efecto importante en la determinación de la duración del período de desempleo, el efecto de las variables sociodemográficas sobre dicha duración era considerablemente superior.

Otro factor que influye sobre el salario de reserva es la situación económica en el momento de la búsqueda de empleo, así como las expectativas sobre el futuro. Concretamente, cabría esperar que el salario de reserva fuera menor en períodos de recesión o en zonas con peor situación económica. No obstante, a la hora de estimar nuestro modelo, no podremos hacer uso de la variable 'salario de reserva' puesto que solamente disponemos de dicha información para el caso de los '*titulados del grupo 4*', pero desconocemos dicha información en el caso de los '*titulados del grupo 3*'<sup>128</sup>.

<sup>127</sup> Este es otro de los motivos que nos llevó a eliminar de este análisis a los '*titulados de los grupos 1 y 2*'

<sup>128</sup> La encuesta planteaba la pregunta del salario de reserva únicamente a los titulados que se encontraban desempleados en el momento de responder a la encuesta, o estaban ocupados pero buscaban un nuevo empleo (Anexo A.1, pregunta P.83). Pero se planteaba esta pregunta como requisito para aceptar un futuro empleo y nunca como requisito de aceptación de un empleo ya disfrutado.

La aceptación o no de una oferta laboral vendrá también determinada por la situación familiar del individuo (Bratberg y Nilsen, 1998), por lo que también tendremos que incluir variables que hagan referencia a dicha situación en el modelo de duración. Es más, Ahn y Ugidos-Olazábal (1995) comentaron, en su estudio sobre el efecto del subsidio por desempleo y de las características familiares sobre la duración del desempleo en España, que, dado que dicho subsidio tenía una duración limitada, un factor que revelaba tener una gran importancia a la hora de explicar la alta tasa de desempleo española eran, precisamente, las características familiares, y más concretamente, el apoyo económico familiar.

#### 4.2.2. Elección del mejor modelo

A continuación procedemos a la estimación paramétrica de nuestro modelo de duración. La variable endógena del modelo es el tiempo transcurrido entre el comienzo de búsqueda del primer empleo y la consecución del mismo, para los *'titulados de los grupos 3 y 4'*, es decir, para los que buscaron su primer empleo tras la finalización del Ciclo Formativo. Ésta es una variable aleatoria continua  $T$ , que presenta censura por la derecha, puesto que, los *'titulados del grupo 4'*, en el momento de responder a la encuesta, aún no habían encontrado su primer empleo, por lo que su duración completa de desempleo resulta desconocida.

Para la estimación de un modelo que explique la variabilidad de dicha duración, en función de un conjunto de covariables, como hemos visto previamente, disponemos de un amplio abanico de distribuciones de probabilidad. El problema radica en la elección de la distribución más apropiada a los datos.

Kiefer (1988) comenta al respecto que, para la elección del modelo paramétrico más adecuado a los datos, resulta útil recurrir a la información aportada por las estimaciones no paramétricas de las funciones de riesgo. En nuestro caso, la estimación no paramétrica de la duración del período de desempleo previo al primer empleo, parecía reflejar, en líneas generales, un decrecimiento monótono (Ilustración IV.8), por lo que los modelos más adecuados resultan ser el modelo paramétrico de Weibull y el de Gompertz. De hecho, estos dos modelos se suelen utilizar en la modelización de datos que presentan tasas de riesgo monótonas, tanto crecientes como decrecientes (StataCorp., 2001).

Tras las consideraciones teóricas realizadas en el apartado anterior, observamos que, el efecto esperado de algunas de las posibles variables explicativas a incorporar en el modelo puede resultar ambiguo, debido a que las mismas pueden afectar tanto a la probabilidad de recibir una oferta laboral, como a la probabilidad de aceptar la misma. Por tanto, será el análisis empírico, el que resuelva dicha ambigüedad, proporcionándonos un resultado en función del efecto que prevalezca.

Las variables independientes que hemos seleccionado para explicar la duración del desempleo, las podemos agrupar en cuatro bloques:

1. Características personales y familiares:

En este bloque recogemos aquellos factores propios del individuo que, al poder influir en la probabilidad de recibir ofertas laborales, o en la probabilidad de aceptar las mismas, pueden influir sobre la duración de su período de desempleo, como es el caso de las variables:

SEXO: variable dicotómica que toma valor 1 si el titulado es hombre

EDAD1: edad del titulado al comenzar a trabajar en su primer empleo

EXPPREV: variable dicotómica que toma valor 1 si el titulado reconoce haber tenido algún empleo previo a la finalización del CF

Igualmente introducimos características de su entorno familiar, como medida de aproximación a las obligaciones o cargas familiares que tenía el titulado al finalizar el CF, y que, por tanto, pueden afectar a la determinación del salario de reserva y, en consecuencia, a la mayor o menor duración del primer período de desempleo. Introducimos así las variables:

HIJOS1: variable dicotómica que toma valor 1 si el titulado tenía hijos al finalizar sus estudios

NTRABAJ1: número de miembros de la familia que estaban trabajando cuando el titulado finalizó sus estudios

INGRHOG1: ingresos del hogar, mientras el titulado realizaba en Ciclo Formativo, en Euros constantes de 2001

Y, por último, al igual que hicieron Theeuwes (1989), Biggeri et al. (2001) y Congregado-Ramírez y García-Pérez (2002), incorporamos, como explicativa de la duración del desempleo del titulado, el nivel formativo del cabeza de familia, así como su situación laboral, pues consideramos que estos factores influyen sobre la probabilidad de recepción de ofertas laborales del titulado. Estos dos factores vienen recogidos mediante las variables siguientes:

ESTCABFAS: variable dicotómica que toma valor 1, cuando el cabeza de familia posee estudios superiores, y el valor 0, en otro caso

CABFAOCUPA1: variable dicotómica que toma valor 1, si el cabeza de familia trabajaba cuando el titulado finalizó sus estudios, y el valor 0, en otro caso

## 2. Características educativas:

En este bloque recogemos aquellas características que se refieren a la formación académica del titulado y que, en consecuencia, esperamos que influyan sobre la probabilidad de recibir ofertas laborales, afectando así a la duración del período de desempleo. Estas características son las siguientes:

NIVELCF: variable dicotómica que toma el valor 1, si el titulado cursó un Ciclo Formativo de Grado Superior, y el valor 0, en caso de haber cursado uno de Grado Medio.

FAMILIAP: recoge un conjunto de variables dicotómicas que hacen referencia a la Familia Profesional a la que pertenece el Ciclo Formativo cursado por el titulado. En el caso de FAMILIAP1 el CF pertenece a la Familia de Actividades Agrarias; en FAMILIAP2 el CF pertenece a la Familia de Actividades Físicas y Deportivas; en FAMILIAP3 el CF pertenece a la Familia de Actividades Marítimo Pesqueras; en FAMILIAP5 el CF pertenece a la Familia de Artes Gráficas; en FAMILIAP6 el CF pertenece a la Familia de Comercio y Marketing; en FAMILIAP7 el CF pertenece a la Familia de Comunicación, Imagen y Sonido; en FAMILIAP8 el CF pertenece a la Familia de Edificación y Obra Civil; en FAMILIAP9 el CF pertenece a la Familia de Electricidad y Electrónica; en FAMILIAP10 el CF pertenece a la Familia de Fabricación Mecánica; en FAMILIAP11 el CF pertenece a la Familia de Hostelería y Turismo; en FAMILIAP12 el CF pertenece a la Familia de Imagen Personal; en FAMILIAP13 el CF pertenece a la Familia de Industrias Alimentarias; en FAMILIAP14 el CF pertenece a la Familia de Informática; en FAMILIAP15 el CF pertenece a la Familia de Madera y Mueble; en FAMILIAP16 el CF pertenece a la Familia de Mantenimiento de Vehículos Autopropulsados; en FAMILIAP17 el CF pertenece a la Familia de Mantenimiento y Servicios a la Producción; en FAMILIAP18 el CF pertenece a la Familia de Química; en FAMILIAP19 el CF pertenece a la Familia de Sanidad; en FAMILIAP20 el CF pertenece a la Familia de Servicios Socioculturales y a la Comunidad; y, finalmente, FAMILIAP21, que indica si el CF pertenece a la Familia de Textil, Confección y Piel. Hemos dejado como referencia los Ciclos Formativos pertenecientes a la Familia de Administración.

ACCDIREC: variable dicotómica que toma valor 1 cuando el titulado reconoce haber accedido al Ciclo Formativo a través de lo que hemos llamado una 'vía estructurada', es decir, cuando accedió a un CF2 tras la ESO, o a un CF3 tras el Bachillerato o un CF2 de la misma Familia Profesional.

PUB\_PRI: variable dicotómica que toma valor 1, si el titulado cursó el Ciclo Formativo en un centro público, y toma el valor 0, en caso contrario.

MALESTUD: variable dicotómica que toma valor 1 cuando el titulado reconoce haber optado por cursar un Ciclo Formativo por considerarse mal estudiante

FIN: recoge un conjunto de variables dicotómicas que hacen referencia al año de finalización del Ciclo Formativo. FIN98 indica que el titulado finalizó el CF en el año 1998; FIN 99 indica que lo finalizó en el año 1999; FIN00 indica que lo finalizó en el año 2000 o posteriormente. La categoría de referencia, en este caso, se refiere a los que finalizaron el CF en el año 1997 o antes.

TARDAPREV: recoge un conjunto de variables dicotómicas que hacen referencia al tiempo que tardan los titulados en finalizar su titulación. TARDAPREV2 toma el valor 1 cuando el titulado ha tardado el tiempo previsto, mientras que TARDAPREV3 toma el valor 1 cuando ha tardado menos de lo previsto. La categoría de referencia la constituyen los titulados que han tardado más de lo previsto en finalizar sus estudios.

NOTASUP: variable dicotómica que toma valor 1, cuando el titulado finaliza el CF con una nota media superior a la de su especialidad en el año de finalización, y toma el valor 0, en caso contrario.

### 3. Variables de entorno:

Con el fin de captar la influencia de las condiciones del mercado de trabajo sobre la duración del período de búsqueda del primer empleo, hemos optado por introducir variables que hagan referencia a la zona geográfica en la que el titulado estudió<sup>129</sup>, así como variables que recojan el efecto del ciclo económico. Estas últimas las recogemos a través de la introducción de la tasa de desempleo juvenil, en la fecha en la que el titulado consigue el empleo<sup>130</sup>, así como mediante la introducción de dicotómicas que recojan el año en el que los titulados consiguen su primer empleo. Las variables que introducimos en este bloque son:

RUR\_URB: variable dicotómica que toma el valor 1, cuando el titulado cursó el CF en un centro de estudios de fuera de la capital, y toma el valor 0 si los cursó en la ciudad de Las Palmas de Gran Canaria

<sup>129</sup> Suponemos que si el titulado estudió en un centro educativo de la ciudad de Las Palmas de Gran Canaria, buscará su primer empleo en esa misma ciudad; mientras que si estudió fuera de la ciudad, buscará su primer empleo fuera de ella.

<sup>130</sup> En el caso de los titulados que aún no habían conseguido su primer empleo, tomamos la tasa de desempleo existente en el mercado al responder la encuesta.

TASADESJ: tasa trimestral de desempleo juvenil (16 a 25 años) en Canarias en la fecha en la que el titulado consiguió su primer empleo

AÑOCOM\_: recoge un conjunto de variables dicotómicas que hacen referencia al año en el que los titulados consiguen su primer empleo. AÑOCOM\_98 toma el valor 1 cuando el titulado consigue su empleo en el año 1998; AÑOCOM\_99 toma el valor 1 cuando lo consiguen en el año 1999; AÑOCOM\_00 toma el valor 1 cuando lo consigue en el año 2000; y, finalmente, AÑOCOM\_01 toma el valor 1 cuando lo consiguen en el año 2001. Se deja como referencia a los que aún no han conseguido su primer empleo.

#### 4. Características del proceso de búsqueda:

Finalmente, el cuarto bloque de variables explicativas hace referencia al proceso de búsqueda de empleo, puesto que este proceso estará directamente relacionado con la probabilidad de obtener ofertas de trabajo por parte de los titulados. Dentro de este bloque se encuentran las variables:

NMEDBUS1: recoge el número de medios de búsqueda diferentes utilizados por el titulado

CURRIC1: variable dicotómica que toma el valor 1, cuando el titulado ha utilizado el envío de currículum como medio de búsqueda de su primer empleo

INEMC1: variable dicotómica que toma el valor 1, cuando el titulado se ha registrado en el INEM o la ACE (Agencia Canaria de Empleo)

ETT1: variable dicotómica que toma el valor 1, cuando el titulado ha recurrido a los servicios de una Agencia de Trabajo Temporal para buscar su primer empleo

PRENSA1: variable dicotómica que toma el valor 1, cuando el titulado ha respondido a anuncios en prensa como medio de búsqueda de su primer empleo

CONTACT1: variable dicotómica que toma el valor 1, cuando el titulado ha acudido a contactos personales como medio de búsqueda de su primer empleo

FCT1: variable dicotómica que toma el valor 1, cuando el titulado ha acudido a los responsables de la FCT como medio de búsqueda de su primer empleo

BOLSAS1: variable dicotómica que toma el valor 1, cuando el titulado se ha apuntado en alguna bolsa de trabajo para conseguir su primer empleo

Aún introduciendo este conjunto de variables explicativas en nuestro modelo, siempre podrán existir otras variables que no hayamos observado y que influyan en la función de riesgo, pudiendo causar, de esta manera, un problema de heterogeneidad inobservable en los modelos estimados.

El hecho de no controlar por la posible heterogeneidad durante el proceso de estimación, nos llevaría a un sesgo en las inferencias acerca de las relaciones de dependencia (Lancaster, 1992), por lo que deberíamos estudiar la posible existencia de heterogeneidad inobservable en nuestros datos.

Según comenta Lancaster (1992), la existencia de heterogeneidad inobservable entre individuos es un problema habitual que puede venir motivado por tres factores: a) error en la medición de las duraciones, b) error en la medición de las covariables y c) omisión de variables relevantes.

En la estimación de nuestro modelo, aunque quepa la posibilidad de tener errores de medición de la duración, puesto que ésta se basa en el tiempo que los encuestados 'recuerdan' haber dedicado a la búsqueda, es más probable que, en caso de tener heterogeneidad inobservable, ésta sea debida a la existencia de regresores no observables y, en consecuencia, omitidos. Éste puede ser el caso del propio salario de reserva de los titulados, o de la posible percepción de subsidio de desempleo, por parte de los titulados que han reconocido haber disfrutado de un empleo previo a la finalización de sus estudios.

Lancaster (1992) comenta también que la heterogeneidad puede derivarse igualmente de las condiciones de incertidumbre en las que se realiza el proceso de búsqueda por parte de los desempleados, y, por tanto, de las diferencias en el éxito de dichos procesos de búsqueda. No obstante, al introducir las variables enumeradas previamente en el cuarto bloque, recogemos, de alguna manera, el esfuerzo de búsqueda realizado por los egresados de nuestra muestra, con lo que deberíamos reducir la heterogeneidad causada por este motivo.

La forma habitual de controlar el fenómeno de la heterogeneidad inobservable, consiste en asumir, que el término de perturbación aleatoria del modelo, proviene de una función de distribución paramétrica determinada, teniendo que estimar así, los parámetros de la función de distribución que genera estos errores.

La teoría no proporciona una pauta clara que nos permita decidir qué función de distribución debe utilizarse para el término de heterogeneidad, aunque son muchos los estudios empíricos que se decantan por la heterogeneidad gamma (Lancaster, 1979; Theeuwes, 1989; Han y Hausman, 1990; Melkersson, 1999; Petrongolo, 2001; Gámez-Amián y García-Pérez, 2002; Kettunen, 2002).

Nosotros optaremos por estimar modelos de duración Weibull y Gompertz introduciendo, como posible heterogeneidad inobservada, la derivada de una variable Gamma de media 1 y varianza  $\theta$  (ver Lancaster, 1992 o Klein y Moeschberger, 1997), así como la derivada de una variable Gaussiana Inversa de media 1 y varianza  $1/\theta$  (ver StataCorp., 2001). Los resultados de estas estimaciones son los que se recogen en la Tabla IV.5.

**Tabla IV.5: Estimación de modelos de duración tipo Weibull y Gompertz, del primer período de desempleo, con corrección de heterogeneidad Gamma y Gaussiana Inversa**

	Weibull con heterogeneidad gamma		Weibull con heterogeneidad gaussiana inversa		Gompertz con heterogeneidad gamma		Gompertz con heterogeneidad gaussiana inversa	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z
(Constante)	-3,057	1,41	-2,501	1,44	-3,578 *	1,70	-3,001	1,61
<i>CARACTERÍSTICAS PERSONALES Y FAMILIARES:</i> (Referencia: mujer, sin experiencia laboral previa, sin hijos, cabeza de familia con estudios no superiores, cabeza de familia desempleado)								
SEXO	0,117	0,89	0,113	1,05	0,143	1,16	0,136	1,18
EDAD1	-0,021	0,99	-0,042 **	2,53	-0,026	1,25	-0,052 ***	2,87
EXPPREV	0,281 ***	2,64	0,296 ***	3,42	0,271 ***	2,72	0,336 ***	3,68
HIJOS1	-0,542	0,69	-0,191	0,47	-0,330	0,48	-0,103	0,25
NTRABAJ1	-0,066	1,27	-0,063	1,53	-0,053	1,11	-0,061	1,41
INGRHOG1	0,000	1,26	0,000 *	1,77	0,000	1,04	0,000 *	1,85
ESTCABFS	0,206	0,95	0,004	0,02	0,185	0,93	0,006	0,03
CABFAOCUPA1	0,052	0,39	-0,033	0,29	0,050	0,40	-0,056	0,47
<i>CARACTERÍSTICAS FORMATIVAS:</i> (Referencia: Ciclo Formativo de grado medio; Familia de Administración; Eligió el CF por motivos diferentes a considerarse mal estudiante; Finalizó los estudios en 1997 o antes; Tardó más de lo previsto en finalizar los estudios; nota final inferior a la media de su especialidad el año de su promoción)								
NIVELCF	0,355 **	2,34	0,263 **	2,24	0,416 ***	2,97	0,309 **	2,48
FAMILIAP1	-0,157	0,35	0,146	0,45	0,668 **	2,37	0,200	0,58
FAMILIAP2	0,767 **	2,54	0,415	0,09	-0,649	0,86	0,407	1,31
FAMILIAP3	-0,795	0,90	-0,536	1,23	-0,147	0,33	-0,486	1,10
FAMILIAP5	0,032	0,05	0,274	0,68	0,070	0,13	0,338	0,81
FAMILIAP6	0,389 **	2,08	0,426 ***	2,82	0,369 **	2,05	0,455 ***	2,82
FAMILIAP7	-0,526	1,09	-0,390	1,26	-0,527	1,17	-0,415	1,29
FAMILIAP8	0,601 **	2,09	0,722 ***	2,65	0,546* **	2,00	0,751 ***	2,58
FAMILIAP9	0,237	1,11	0,323 *	1,88	0,225	1,16	0,347 *	1,90
FAMILIAP10	0,347	1,05	0,566 **	2,14	0,381	1,32	0,645 **	2,32
FAMILIAP11	0,161	0,82	0,104	0,62	0,144	0,79	0,060	0,32
FAMILIAP12	0,827 ***	3,06	0,663 **	2,51	0,797 ***	3,17	0,654 **	2,25
FAMILIAP14	0,102	0,39	0,220	0,98	0,120	0,50	0,209	0,86
FAMILIAP15	1,048 ***	3,10	1,435 ***	3,52	0,938 ***	3,24	1,519 ***	3,52
FAMILIAP16	-0,003	0,01	0,064	0,29	0,074	0,29	0,090	0,38
FAMILIAP17	0,993 ***	3,46	1,148 ***	4,43	0,944 ***	3,64	1,232 ***	4,55
FAMILIAP18	0,359	0,49	0,374	0,87	0,255	0,34	0,418	0,94
FAMILIAP19	0,061	0,30	0,030	0,19	0,014	0,07	0,038	0,23
FAMILIAP20	0,483	1,62	0,071	0,31	0,498 *	1,88	0,028	0,11
FAMILIAP21	-0,065	0,04	-1,420	0,10	-0,294	0,21	0,180	0,19
ACCDIREC	0,096	0,65	0,123	1,02	-0,015	0,13	0,142	1,06
PUB_PRI	0,044	0,18	-0,076	0,38	0,008	0,03	-0,130	0,60
MALESTUD	-0,146	0,67	-0,090	0,43	-0,087	0,44	-0,111	0,48
FIN98	0,435 **	2,06	0,445 ***	2,61	0,422 **	1,97	0,477 ***	2,61
FIN99	0,486 **	2,08	0,486 **	2,52	0,514 **	2,14	0,530 **	2,53
FIN00	0,588	0,96	0,607	1,23	0,559	0,89	0,708	1,38
TARDAPREV2	0,463	1,02	0,207	0,57	0,631	1,23	0,298	0,69

TARDAPREV3	0,190	0,39	-0,083	0,21	0,391	0,72	0,024	0,05
NOTASUP	0,236 **	2,33	0,293 ***	3,47	0,219 **	2,32	0,338 ***	3,79
<i>VARIABLES DE ENTORNO:</i> (Referencia: Estudió en un centro de la ciudad de Las Palmas de Gran Canaria; aún no ha conseguido su primer empleo)								
RUR_URB	0,131	1,06	0,108	1,08	0,155	1,32	0,123	1,14
TASADESJ	0,055	0,99	0,056	1,29	0,072	1,35	0,075	1,62
AÑOCOM_98	-0,555	0,96	-0,642	1,45	-0,491	0,95	-0,692	1,51
AÑOCOM_99	-0,919	1,52	-0,906 **	1,98	-0,415	0,64	-1,004 **	2,12
AÑOCOM_00	-1,206 *	1,95	-1,221 ***	2,60	-0,378	0,48	-1,347 ***	2,76
AÑOCOM_01	-1,376 **	2,17	-1,454 ***	3,04	-0,318	0,34	-1,646 ***	3,29
<i>CARACTERÍSTICAS DEL PROCESO DE BÚSQUEDA:</i> (Referencia: No utilizó el envío de currículos, no se apuntó en el INEM, no acudió a una ETT, no respondió anuncios en prensa, no contactó con los encargados de la FCT, no se apuntó en bolsas de trabajo)								
NMEDBUS1	-0,468 **	2,11	-0,332 *	1,87	-0,534 **	2,20	-0,360 *	1,86
CURRIC1	-1,454 ***	4,06	-0,721 ***	3,41	-1,074 ***	3,53	-0,702 ***	3,07
INEM1	-0,350	1,28	-0,231	1,07	-0,181	0,64	-0,230	0,97
ETT1	-0,341	1,22	-0,151	0,74	-0,271	0,95	-0,136	0,61
PRENSA1	-0,339	1,16	-0,242	1,12	-0,306	0,99	-0,261	1,10
CONTACT1	-0,154	0,59	-0,207	1,03	-0,037	0,13	-0,181	0,83
FCT1	0,955 ***	3,73	0,878 ***	4,22	0,930 ***	3,32	0,941 ***	4,22
BOLSAST1	0,440 *	1,65	0,253	1,16	0,479 *	1,67	0,284	1,21
$\alpha$	1,237 ***	12,10	0,986 ***	56,00				
$\gamma$					0,009 ***	8,55	0,002 ***	3,53
$\sigma^2$	1,389 ***	6,10	2,664 ***	9,55	1,474 ***	16,03	1,276 ***	11,97
Chi <sup>2</sup> <sub>51</sub>	393,880	p=0,000	1060,070	p=0,000	2040,640	p=0,000	1275,370	p=0,000
Log. Verosimilitud	-4990,639		-5038,526		-4873,841		-5019,127	
AIC	10085,279		10181,051		9851,681		10142,253	
$H_0 : \sigma^2 = 0$	$\bar{\chi}_1^2 = 227,98$ p= 0,000		$\bar{\chi}_1^2 = 155,63$ p= 0,000		$\bar{\chi}_1^2 = 1004,36$ p= 0,000		$\bar{\chi}_1^2 = 811,60$ p= 0,000	
Nº observaciones	1691		1691		1691		1691	

Estadísticos z robustos en valor absoluto

\* significativo al 10%; \*\* significativo al 5%; \*\*\* significativo al 1%

Al final de la Tabla IV.5 incluimos el resultado del contraste de hipótesis de heterogeneidad inobservable, cuya hipótesis nula plantea la ausencia de dicho problema, pudiendo concluir, en los cuatro casos estimados, que el término estimado de  $\sigma^2$  es estadísticamente significativo y que, en consecuencia, la heterogeneidad inobservable resulta relevante.

Resulta razonable plantearse cómo seleccionar el modelo paramétrico más adecuado, dado que tenemos varias alternativas a elegir. Si los modelos hubiesen estado anidados, podríamos utilizar un contraste de ratio de verosimilitudes o un contraste de Wald para discriminar entre ellos. Este tipo de contrastes sería válido en caso de tener que optar entre una distribución Weibull y una Exponencial, o entre una distribución Gamma, una Weibull o una Log-normal<sup>131</sup>. Sin embargo, cuando los modelos no están anidados, como es nuestro caso, la discriminación entre ellos no se puede realizar de esta manera.

<sup>131</sup> Basta recordar que la distribución Weibull incorpora a la Exponencial como caso especial, cuando  $\alpha = 1$ ; y que las distribuciones Weibull y Log-normal son casos específicos de distribuciones Gamma generalizadas, cuando  $m = 1$  y  $m = 0$ , respectivamente.

Un acercamiento común a este tipo de decisiones consiste en utilizar el Criterio de Información de Akaike (AIC). Este criterio propone penalizar cada logaritmo de verosimilitud, de manera que refleje el número de parámetros que se están estimando en cada modelo y, entonces, comparar los mismos. En este tipo de modelos, el valor de AIC vendrá definido por

$$AIC = -2(\log\_verosimilitud) + 2(c + p + 1) \quad [\text{Eq. IV.90}]$$

donde  $c$  es el número de covariables y  $p$  el número de parámetros auxiliares del modelo (1 en la distribución Weibull y 1 en la Gompertz).

Aunque el modelo que presenta un mejor ajuste es siempre aquél que tiene un mayor valor en su logaritmo de verosimilitud, siempre será preferible el que presente un menor valor de AIC. Por tanto, según se desprende de los resultados de la Tabla IV.5, la mejor estimación resulta ser la del modelo Gompertz con heterogeneidad Gamma, puesto que, su logaritmo de verosimilitud es el mayor de los cuatro, y su valor AIC el menor de ellos.

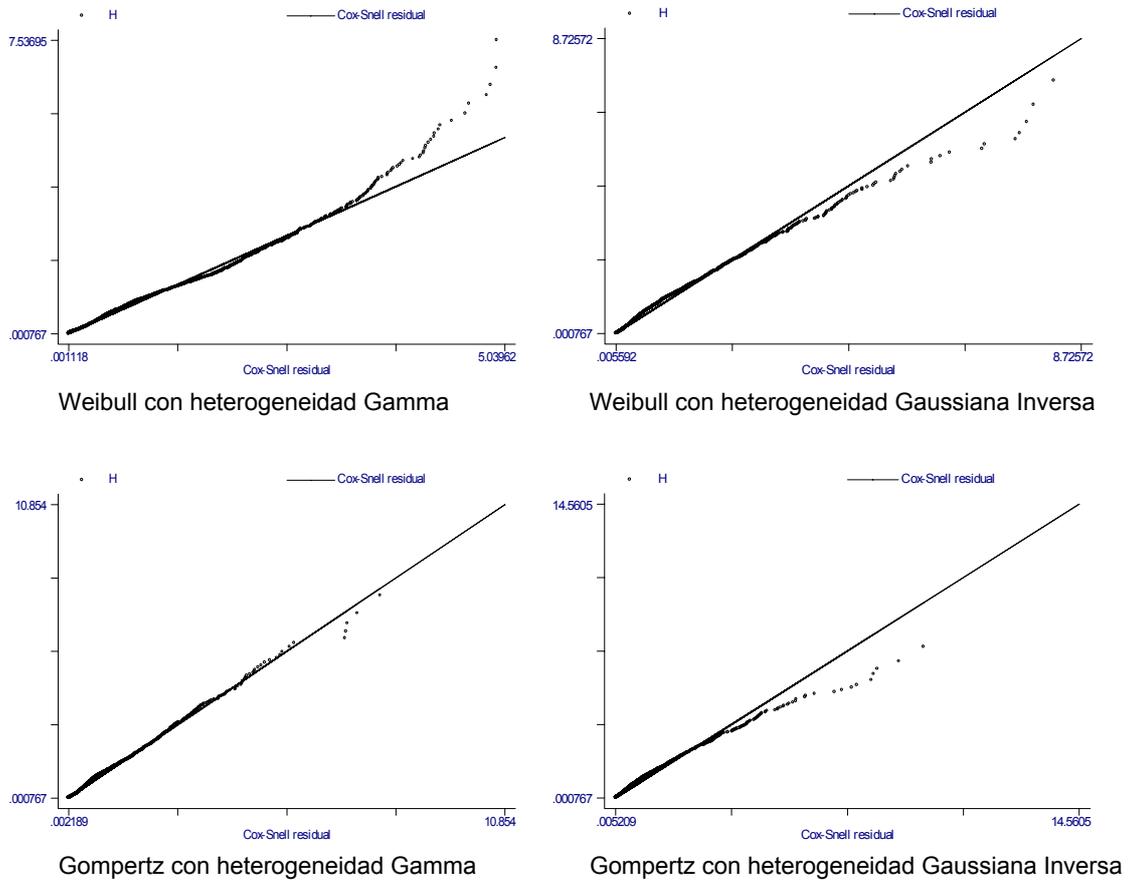
Otra alternativa para seleccionar el mejor modelo consiste en analizar el comportamiento de los residuos generados a través de la estimación y quedarse con aquel modelo, cuyos residuos presenten un mejor comportamiento (Kiefer, 1988). El método de análisis de residuos más sencillo consiste en la estimación de los residuos generalizados de Cox-Snell, puesto que, si el modelo ha sido especificado correctamente, estos deberían comportarse como una función exponencial unitaria.

Kiefer (1988) propone verificar el ajuste del modelo, una vez obtenidos los residuos generalizados de Cox-Snell, calculando una estimación empírica de la función integrada de riesgo, basada por ejemplo en los estimadores de Kaplan-Meier, tomando los residuos de Cox-Snell como la nueva variable de duración, pero manteniendo la misma variable indicadora de la existencia de censura. De esta manera, se podrá graficar dicha función integrada de riesgo, poniendo en el eje de abscisas los residuos. Si el modelo se ajusta bien a los datos, entonces el gráfico debería mostrar una línea con pendiente igual a uno.

La representación gráfica de los residuos de Cox-Snell obtenidos a partir de los cuatro modelos estimados, son los que se recogen en la Ilustración IV.12.

Como vemos en la Ilustración IV.12, los residuos cuya función integrada de riesgo se asemeja más a una distribución exponencial unitaria, son los generados a través de la estimación Gompertz con heterogeneidad Gamma. Por lo que, sin duda, ésta es la estimación que mejor se ajusta a nuestros datos.

**Ilustración IV.12: Función integrada de riesgo de los residuos generalizados de Cox-Snell para la estimación de la duración del primer período de desempleo**



### 4.2.3. Interpretación de resultados

Finalmente, en este apartado, presentaremos los resultados obtenidos, en el análisis de la duración del período de paro al que se enfrentan los titulados antes de obtener su primer empleo, a través de la estimación de un modelo Gompertz que controla la existencia de heterogeneidad de tipo Gamma. Estos resultados son los que se presentan en la Tabla IV.6.

En este modelo, un coeficiente estimado positivo, indica que la variable en cuestión provoca un incremento en la tasa de riesgo de salir del desempleo hacia la ocupación y, por tanto, provoca una disminución en la permanencia del individuo en el paro.

Hemos optado por estimar el modelo diferenciando la muestra entre hombres y mujeres, aparte de realizar la estimación conjunta para toda ella, puesto que cabe esperar que la probabilidad de abandonar el desempleo difiera entre géneros y, además, que la influencia de algunas de las variables consideradas sea diferente en función del colectivo de que se trate.

Tabla IV.6: Estimación del modelo de duración tipo Gompertz con heterogeneidad Gamma para el primer período de desempleo, diferenciando por género

	Todos		Mujeres		Hombres	
	Coef.	z	Coef.	z	Coef.	z
(Constante)	-3,578 *	1,70	-5,949 **	2,11	-1,009	0,46
<b>CARACTERÍSTICAS PERSONALES Y FAMILIARES:</b> (Referencia: mujer, sin experiencia laboral previa, sin hijos, cabeza de familia con estudios no superiores, cabeza de familia desempleado)						
SEXO	0,143	1,16				
EDAD1	-0,026	1,25	-0,052 *	1,68	-0,012	0,40
EXPPREV	0,271 ***	2,72	0,343 **	2,31	0,228	1,57
HIJOS1	-0,330	0,48	-0,827	0,73	-0,220	0,27
NTRABAJ1	-0,053	1,11	-0,034	0,50	-0,048	0,71
INGRHOG1	0,000	1,04	0,000	0,91	0,000	0,19
ESTCABFS	0,185	0,93	0,020	0,07	0,541 *	1,76
CABFAOCUPA1	0,050	0,40	-0,030	0,16	-0,023	0,13
<b>CARACTERÍSTICAS FORMATIVAS:</b> (Referencia: Ciclo Formativo de grado medio; Familia de Administración; Accedió por 'ruta desestructurada'; Estudió en centro privado; Eligió el CF por motivos diferentes a considerarse mal estudiante; Finalizó los estudios en 1997 o antes; Tardó más de lo previsto en finalizar los estudios; nota final inferior a la media de su especialidad el año de su promoción)						
NIVELCF	0,416 ***	2,97	0,728 ***	3,52	0,178	0,88
FAMILIAP1	0,668 **	2,37	0,881 **	2,15	0,458	1,16
FAMILIAP2	-0,649	0,86	1,184	1,37	-2,013 ***	2,80
FAMILIAP3	-0,147	0,33	-1,426 *	1,79	0,304	0,66
FAMILIAP5	0,070	0,13	0,164	0,31	-0,306	0,26
FAMILIAP6	0,369 **	2,05	0,677 ***	3,17	-0,376	1,07
FAMILIAP7	-0,527	1,17	-0,744	1,10	-0,403	0,63
FAMILIAP8	0,546 **	2,00	0,630 *	1,70	0,456	1,12
FAMILIAP9	0,225	1,16	0,257	0,25	0,101	0,39
FAMILIAP10	0,381	1,32	1,211 ***	4,26	0,115	0,31
FAMILIAP11	0,144	0,79	-0,001	0,00	0,145	0,45
FAMILIAP12	0,797 ***	3,17	1,002 ***	3,60	1,373 ***	3,61
FAMILIAP14	0,120	0,50	-0,220	0,42	0,068	0,23
FAMILIAP15	0,938 ***	3,24			0,801 **	2,14
FAMILIAP16	0,074	0,29			-0,111	0,36
FAMILIAP17	0,944 ***	3,64	1,049 ***	2,78	0,809 **	2,45
FAMILIAP18	0,255	0,34	-0,245	0,19	0,710	1,25
FAMILIAP19	0,014	0,07	0,133	0,56	-0,111	0,27
FAMILIAP20	0,498 *	1,88	0,364	1,20	1,204 ***	2,79
FAMILIAP21	-0,294	0,21	-0,158	0,12	-1,160	
ACCDIREC	-0,015	0,13	-0,065	0,35	-0,048	0,32
PUB_PRI	0,008	0,03	0,086	0,30	-0,364	1,00
MALESTUD	-0,087	0,44	-0,494 *	1,67	0,383	1,60
FIN98	0,422 **	1,97	0,376	1,34	0,473	1,39
FIN99	0,514 **	2,14	0,649 **	2,14	0,188	0,48
FIN00	0,559	0,89	0,067	0,10	2,236 *	1,93
TARDAPREV2	0,631	1,23	1,020	1,42	-0,085	0,10
TARDAPREV3	0,391	0,72	0,939	1,24	-0,599	0,68
NOTASUP	0,219 **	2,32	0,350 **	2,53	0,132	0,99
<b>VARIABLES DE ENTORNO:</b> (Referencia: Estudió en centro de la ciudad de Las Palmas de Gran Canaria; aún no ha conseguido su primer empleo)						
RUR_URB	0,155	1,32	0,255	1,51	0,071	0,43
TASADESJ	0,072	1,35	0,149 **	2,10	-0,005	0,06
AÑOCOM_97					0,758	0,43

AÑOCOM_98	-0,491	0,95	-1,153	1,50	0,954	0,72
AÑOCOM_99	-0,415	0,64	-0,314	0,40	0,316	0,42
AÑOCOM_00	-0,378	0,48	-0,108	0,11	0,294	0,68
AÑOCOM_01	-0,318	0,34	0,257	0,22		
<b>CARACTERÍSTICAS DEL PROCESO DE BÚSQUEDA:</b> (Referencia: No utilizó el envío de currículos, no se apuntó en el INEM, no acudió a una ETT, no respondió anuncios en prensa, no contactó con los encargados de la FCT, no se apuntó en bolsas de trabajo)						
NMEDBUS1	-0,534 **	2,20	-0,864 ***	3,61	-0,789 ***	3,25
CURRIC1	-1,074 ***	3,53	-0,790 **	2,17	-0,808 **	2,22
INEM1	-0,181	0,64	0,076	0,21	0,282	0,89
ETT1	-0,271	0,95	0,190	0,59	-0,171	0,51
PRENSA1	-0,306	0,99				
CONTACT1	-0,037	0,13	0,284	0,91	0,254	0,77
FCT1	0,930 ***	3,32	1,201 ***	4,14	1,293 ***	4,08
BOLSAST1	0,479 *	1,67	1,005 ***	2,73	0,580 *	1,78
OPOCIC1			0,369	0,96	0,133	0,25
$\gamma$	0,009 ***	8,55	0,009 ***	7,01	0,009 ***	4,39
$\sigma^2$	1,474 ***	16,03	1,571 ***	11,25	1,202 ***	9,67
Chi <sup>2</sup>	1004,360	p=0,000	583,320	p=0,000	310,780	p=0,000
Log. Verosimilitud	-4873,841		-2762,602		-2059,541	
Nº observaciones	1691		926		765	

Estadísticos z robustos en valor absoluto

\* significativo al 10%; \*\* significativo al 5%; \*\*\* significativo al 1%

Como vemos a partir de los resultados recogidos en la Tabla IV.6, la heterogeneidad inobservable resulta significativa, tanto en el conjunto de la muestra, como en el colectivo de hombres y mujeres por separado, puesto que la estimación del parámetro  $\sigma^2$  resulta estadísticamente significativa en todos los casos. Esta heterogeneidad resulta ser más acusada entre las mujeres, en cuya muestra se obtiene un mayor valor para dicho parámetro.

En lo que respecta a la influencia de las *características personales y familiares*, la única variable significativa para el colectivo de la muestra es la que hace referencia a la experiencia previa, mostrando un efecto positivo de la misma sobre la probabilidad de encontrar empleo<sup>132</sup>. Biggeri et al. (2001) detectan un efecto similar de la experiencia sobre la probabilidad de abandonar el desempleo, estimando un modelo discreto de duración multinivel, que explica la probabilidad de salida del desempleo de los titulados universitarios que finalizaron sus estudios en Italia en 1992.

El resto de las variables incluidas en este primer bloque resultan no significativas en el conjunto de la muestra. Así, por ejemplo, el número de hijos no influye sobre la probabilidad de abandono del desempleo, al igual que sucedió en los estudio de Ahn y Ugidos-Olazábal (1995) y Kettunen (2002); y el nivel educativo del cabeza de familia, así como el hecho de que éste tenga o no un empleo, tampoco influye sobre dicha

<sup>132</sup> Para obtener la magnitud de su influencia habría que calcular  $(e^{\beta} - 1) \cdot 100$ , lo que nos proporcionaría el incremento porcentual que sufriría la tasa de riesgo de abandono del desempleo.

probabilidad, como ya sucedió en el estudio de Theeuwes (1989) sobre la población entre 15 y 26 años desempleada en los Países Bajos.

A pesar de que, en la muestra conjunta, la variable sexo no resulte estadísticamente significativa, la estimación diferenciada por género nos revela una influencia desigual de las características personales sobre la duración del período de desempleo. De hecho, así como la edad no parece tener ninguna influencia sobre la probabilidad de abandono del desempleo entre los varones, ésta presenta un efecto negativo en el caso de las mujeres, de forma que a mayor edad, menor resulta su tasa de riesgo de abandono del desempleo. Biggeri et al. (2001) y Kettunen (2002) constataron este mismo efecto de la edad para el colectivo global de su muestra.

Igualmente, la influencia positiva de la experiencia laboral previa sobre la probabilidad de consecución del primer empleo detectada para el conjunto de la muestra, al estimar el modelo diferenciando por género, resulta solamente significativa en el colectivo de las mujeres. Por otro lado, el pertenecer a una familia cuyo cabeza de familia posea estudios superiores, conlleva un aumento en la probabilidad de salida del desempleo únicamente entre los titulados varones.

En lo que respecta a las *características formativas* de los titulados observamos que la duración del período de paro disminuye por haber cursado un Ciclo Formativo Superior. Este resultado es consistente con la estimación no paramétrica realizada previamente (ver Ilustración IV.9), en donde veíamos que la función de supervivencia de los técnicos se mantenía continuamente por encima de la de los técnicos superiores. De la misma manera, este resultado es consistente con los resultados de los estudios de Theeuwes (1989), Ahn y Ugidos-Olazábal (1995), Bratberg y Nilsen (1998) y Congregado-Ramírez y García-Pérez (2002), quienes observan que los individuos con mayor nivel educativo salen antes del desempleo que los de menor nivel formativo.

Diferenciando por sexo, sin embargo, vemos que este efecto no es significativo en el caso de los hombres, aunque sí lo es en el de las mujeres. Ahn y Ugidos-Olazábal (1995), en su estudio a partir de la Encuesta de Condiciones de Vida y Trabajo de 1985, constataron que la influencia de la educación era más pronunciada entre mujeres que entre hombres, aunque para estos últimos seguía siendo significativa. Sin embargo, Petrongolo (2001), al igual que nosotros, detecta una influencia positiva del nivel educativo únicamente entre las mujeres.

Tomando como referencia las titulaciones de la Familia Profesional de 'Administración', observamos que, la duración del período de desempleo, disminuye en el caso de haber cursado algún Ciclo Formativo de las Familias de 'Actividades Agrarias', 'Comercio y Marketing', 'Edificación y Obra Civil', 'Imagen Personal', 'Madera y Mueble', 'Mantenimiento y Servicios a la Producción' o 'Servicios Socioculturales y a la Comunidad'. Siendo los titulados de 'Mantenimiento y Servicios a la Producción' los que

presentan una mayor probabilidad de consecución del primer empleo, al igual que ya detectamos mediante la estimación no paramétrica (ver la Ilustración IV.11). Diferenciando por género, observamos que, curiosamente, la titulación que presenta una menor duración del desempleo entre las mujeres, es la de 'Fabricación Mecánica', mientras que, entre los hombres, es la de 'Imagen Personal'.

El acceder a los estudios de Formación Profesional Específica a través de una ruta estructurada (ACCDIREC), la titularidad del centro de estudios (PUB\_PRI) o el tiempo que tardan los titulados en finalizar el Ciclo Formativo (TARDAPREV2 a TARDAPREV3), según nuestras estimaciones, no influyen en la duración del período de desempleo, de donde se podría deducir que los empleadores no tienen en cuenta estos aspectos a la hora de ofertar un empleo.

Sin embargo, el año en el que los titulados finalicen sus estudios sí que influye sobre la duración del período de desempleo. De hecho, para el colectivo global de la muestra, se observa que las cohortes más recientes de titulados presentan una menor duración de su período de desempleo, lo que nos vuelve a presentar indicios de la mayor apreciación de la Formación Profesional en el mercado de trabajo.

Por último, en cuanto a las características formativas de los titulados, observamos cómo las mujeres que han optado por realizar un Ciclo Formativo por considerarse malas estudiantes (MALESTUD), presentan una menor probabilidad de abandonar el desempleo, lo que podría ser debido a una posible falta de autoestima de estas tituladas o a que, sencillamente, estas personas carecen de las habilidades demandadas por el mercado de trabajo para ocupar un empleo. Y, finalmente, observamos que, la nota media obtenida en el Ciclo Formativo (NOTASUP) sólo influye en la duración del desempleo entre las mujeres, al igual que sucedía en el estudio de Biggeri et al. (2001). De manera que, aquellas tituladas que presentan una nota media superior a la de su cohorte en su especialidad, tendrán una mayor probabilidad de abandonar el desempleo que las que tienen una nota media inferior.

En cuanto a la influencia que ejercen las características del *entorno económico*, no se aprecia efecto alguno en el conjunto de la muestra, ni en el colectivo de los hombres. Entre las mujeres, sin embargo, resulta llamativo que la tasa de desempleo juvenil canario tenga un efecto positivo sobre la tasa de riesgo de salida del desempleo de éstas.

Finalmente, en lo que respecta a las *características del proceso de búsqueda*, observamos que, cuantos más medios de búsqueda diferentes utiliza el titulado, menor resulta su tasa de riesgo de salida, es decir, mayor su permanencia en el desempleo. Es posible que este fenómeno responda a que, los titulados que se consideran con menos probabilidades de conseguir empleo, utilicen el mayor número posible de medios a su alcance para lograr su primer empleo. Mientras que, por el lado contrario, aquellos

titulados que se consideran más 'capaces' utilizan un menor número de medios, pero que les resultan más efectivos.

Entrando en el análisis de la efectividad de los medios de búsqueda, observamos que, el envío de currículos, curiosamente, parece reducir la tasa de riesgo de salida, mientras que, el acudir a los responsables de la FCT y el apuntarse a las bolsas de trabajo, reduce considerablemente la duración del período de desempleo. El resto de los medios de búsqueda utilizados por los titulados, resultan no ser estadísticamente significativos.

Por último, observamos que el valor estimado del parámetro auxiliar del modelo de duración tipo Gompertz,  $\gamma$ , resulta significativo y toma un valor positivo. Este valor nos indica una dependencia temporal de la duración del desempleo positiva. Es decir, que la función de riesgo  $y$ , por tanto, la probabilidad condicionada de abandonar el desempleo, aumenta ligeramente con la duración del episodio de paro.

A la luz de este resultado podemos concluir que, aunque al prolongarse la búsqueda, la tasa de llegada de ofertas laborales pueda disminuir, por considerar los empleadores que esto es una señal de la baja productividad de los individuos  $y$ , en consecuencia, del deterioro de su capital humano, el efecto en la probabilidad de aceptación de los individuos, como consecuencia del deterioro de su nivel de exigencia y de la disminución de su salario de reserva, llega a compensar dicha disminución, con lo que, el efecto final resulta ser positivo y la probabilidad de abandonar el desempleo aumenta con su duración.

Por tanto, la estimación señala hacia una función de supervivencia que va disminuyendo hacia una asíntota que toma el valor cero, implicando así que, a largo plazo, prácticamente todos los titulados conseguirían su primer empleo. No obstante, el hecho de que el valor estimado de  $\gamma$  resulte tan cercano al cero, nos indica que la relación de dependencia positiva entre la función de riesgo y el tiempo es débil y, por tanto, el crecimiento de dicha función será moderado<sup>133</sup> (ver Ilustración IV.7).

---

<sup>133</sup> Al estimar el modelo sin la corrección de heterogeneidad inobservable, el valor estimado de  $\gamma$  muestra una relación de dependencia negativa, aunque de nuevo vuelve a tomar un valor muy cercano a cero (-0.003).

## 5. ESTIMACIÓN DE LA DURACIÓN DEL PERÍODO DE EMPLEO

El estudio de la duración del período de ocupación, con el fin de detectar relaciones entre los factores que influyen en la probabilidad de abandono del empleo y la duración en el mismo, ha sido abordado, entre otros, por Sicherman y Galor (1990) para Estados Unidos, Taylor (1999) y Battu et al. (2002) para Gran Bretaña y Bratberg y Nilsen (1998), Nilsen y Risa (1999) y Nilsen et al. (2000) para Noruega. En España, éste aspecto ha sido tratado recientemente por Espinal-Berenguer, 1992, Toharia-Cortés et al. (1998), Ibáñez-Pascual (1999), Salas-Velasco (2000), Alba-Ramírez y Álvarez-Llorente (2001), Gámez-Amián y García-Pérez (2002).

Aún así, nuevamente detectamos la ausencia de estudios al respecto en el marco de la economía canaria, por lo que consideramos pertinente contribuir a salvar parte de dicho vacío, mediante una aportación acerca de la duración del período de empleo de los titulados que conforman nuestra muestra.

Dedicaremos un primer apartado al análisis de la duración del primer empleo que poseen los titulados al finalizar el Ciclo Formativo y, a continuación, nos centraremos en el estudio de la duración del último empleo ocupado por estos titulados. En ambos casos, la duración en el empleo estará medida como una variable continua, puesto que hemos obtenido la misma a partir de la diferencia entre la fecha exacta de comienzo y de finalización en el mismo.

Al igual que ya hicimos en el estudio de la duración del primer período de desempleo de los titulados, en este caso, realizaremos, en primer lugar, un análisis no paramétrico de duración y, a continuación, procederemos a la realización del análisis paramétrico, con la finalidad de estimar el efecto de determinadas covariables sobre esta duración.

### 5.1. DURACIÓN DEL PRIMER EMPLEO

En el capítulo segundo del presente trabajo, al analizar el proceso de búsqueda de empleo, dedicamos un apartado a la duración en el primer puesto de trabajo conseguido por los titulados. Entonces centramos nuestra atención únicamente en los '*titulados del grupo 3*', es decir, en aquellos egresados que antes de finalizar el Ciclo Formativo aún no disfrutaban de su primer empleo y que consiguieron su primer puesto de trabajo tras la finalización de los estudios. Sin embargo, en este momento hemos preferido no centrarnos exclusivamente en este colectivo y ampliarlo también a los '*titulados de los grupos 1 y 2*', es decir, a aquellos titulados que, al finalizar el Ciclo Formativo, ya disponían de un empleo, y que continuaron durante un tiempo trabajando en el mismo. De esta manera, al realizar el análisis paramétrico, podremos comprobar si, la obtención de una nueva titulación tiene algún efecto sobre la probabilidad de abandono del empleo.

La Tabla IV.7 nos muestra cuál es la duración media en el primer empleo de los titulados seleccionados, así como su desviación típica. La tabla se ha confeccionado, al igual que hicimos para el análisis descriptivo de la muestra, diferenciando a los titulados por género, por nivel del Ciclo Formativo cursado, por titularidad del centro de estudios y por Familia Profesional.

De esta manera comprobamos cómo parecen existir grandes diferencias en la duración en el primer empleo por género y titularidad del centro de estudios. De hecho, los hombres permanecen en su primer empleo en torno a 2 meses menos que las mujeres, mientras que los titulados de centros públicos permanecen en él unos cuatro meses y medio más que los titulados de centros privados. Sin embargo, las diferencias en función del nivel del Ciclo Formativo cursado son prácticamente inexistentes, siendo de tan sólo una semana y media más a favor de los técnicos.

**Tabla IV.7: Duración en su primer empleo de los ‘titulados de los grupos 1 a 3’**

	Duración del empleo (días)	
	Media	Desv. Típ.
Hombre	489,9	650,8
Mujer	548,6	844,8
Grado medio (CF2)	520,7	797,7
Grado superior (CF3)	511,4	682,6
Centro privado-concertado	388,3	421,1
Centro público	522,2	755,8
Actividades Agrarias	692,0	1256,3
Actividades Físicas y Deportivas	481,8	839,7
Actividades Marítimo Pesqueras	693,8	944,6
Administración	523,0	705,5
Artes Gráficas	612,4	769,5
Comercio y Marketing	477,3	675,1
Comunicación, Imagen y Sonido	365,5	522,4
Edificación y Obra Civil	664,6	1258,7
Electricidad y Electrónica	556,6	717,5
Fabricación Mecánica	576,3	853,8
Hostelería y Turismo	547,0	1060,9
Imagen Personal	412,3	347,7
Industrias Alimentarias	305,0	272,2
Informática	684,4	798,3
Madera y Mueble	511,5	426,2
Mantenimiento de Vehículos Autopropulsados	360,9	302,4
Mantenimiento y Servicios a la Producción	421,4	530,3
Química	485,5	405,6
Sanidad	511,2	613,0
Servicios Socioculturales y a la Comunidad	418,6	464,9
Textil, Confección y Piel	762,0	794,0
TOTAL	516,2	744,5

Finalmente, al distinguir por Familias Profesionales, comprobamos que las diferencias vuelven a ser bastante acusadas. Así, nos encontramos con Familias Profesionales como las de 'Textil, Confección y Piel' y 'Actividades Agrarias', cuyos titulados permanecen en su primer empleo una media de 2 años, o como la Familia de 'Informática', cuyos titulados permanecen en dicho empleo en torno al año y 10 meses; frente a Familias Profesionales como la de 'Industrias Alimentarias', cuyos titulados no permanecen en su primer empleo más de 10 meses

No debemos perder de vista que los datos recogidos en la Tabla IV.7, a pesar de mostrarnos la duración en el primer empleo, también incorporan observaciones que adolecen de un problema de censura por el lado derecho. De hecho, de los 2540 titulados que conforman la muestra seleccionada, 983 aún seguían trabajando en su primer empleo el día que respondieron a la encuesta, por lo que la muestra utilizada presenta censura en el 38.7% de los casos<sup>134</sup>.

### 5.1.1. Estimación no paramétrica

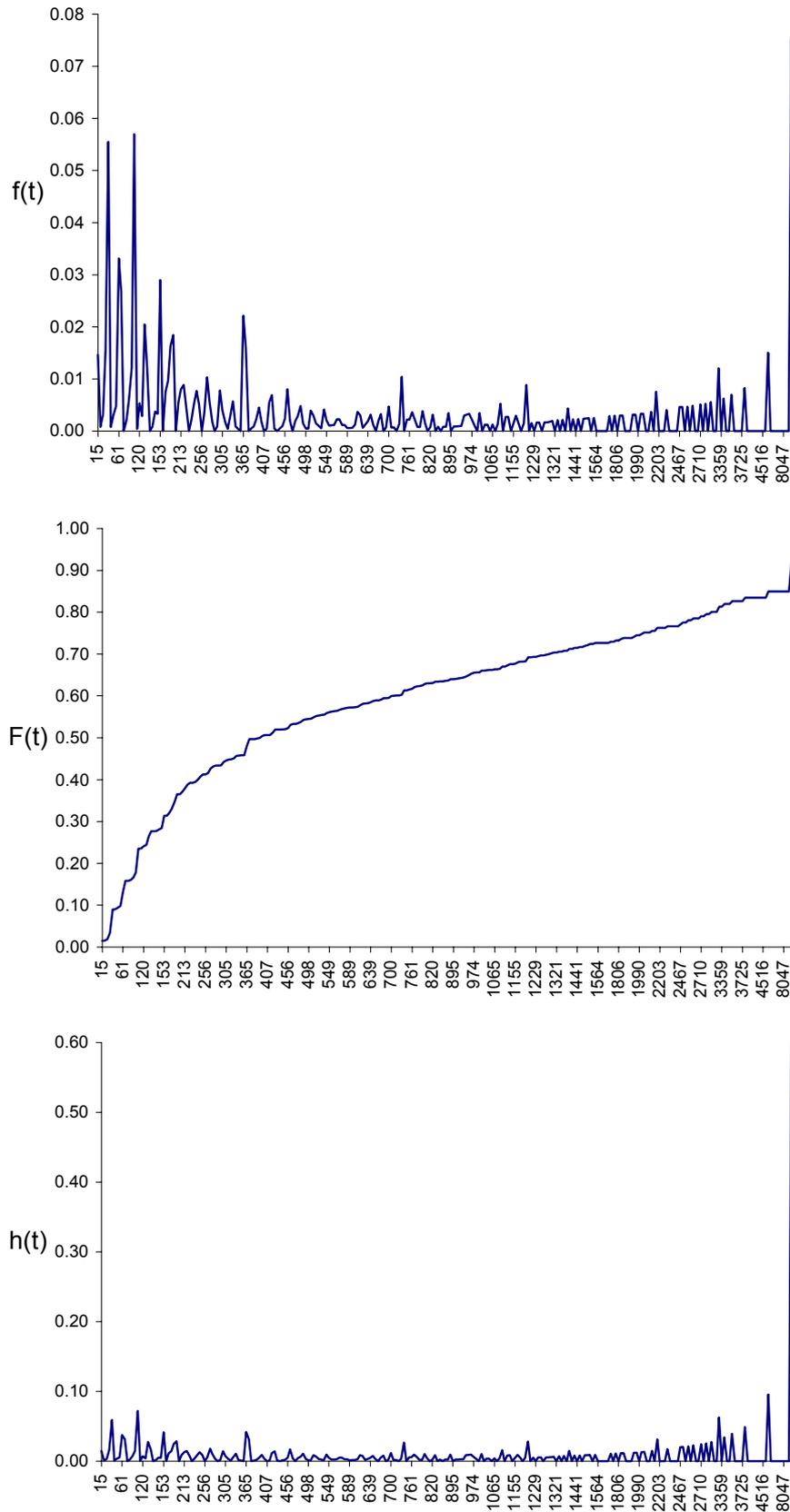
Comenzamos la estimación no paramétrica de la duración en el primer puesto de trabajo a través de la representación gráfica de las funciones empíricas de densidad, distribución y riesgo según la estimación de Kaplan y Meier (1958). De esta manera, la Ilustración IV.13 recoge estas tres funciones, sin agregar los datos en intervalos, pudiendo observar así, cómo las duraciones de empleo manifestadas por los titulados de la muestra oscilan entre 15 días y más de 22 años<sup>135</sup>.

La primera gráfica de la Ilustración IV.13 nos muestra la distribución porcentual de la duración del primer empleo, es decir, su función de densidad empírica. De la observación de la misma, observamos cómo presenta unos 'picos' más acusados en los instantes correspondientes a 1 mes, 2, meses, 3 meses, 5 meses y un año. Por lo que la finalización del primer empleo, según los datos de nuestra muestra, tiene lugar, mayormente, tras el primer mes (5.6%), tras 2 meses de contrato (3.3%), tras llevar 3 meses contratado (5.7%), tras 5 meses (3%) o tras un año (2.2%). No obstante, se observa igualmente un 'pico' que alcanza el 7.5% una vez transcurridos 28 años trabajando en el primer empleo.

<sup>134</sup> La muestra seleccionada de 'titulados de los grupos 1 a 3' estaba compuesta por un total de 2540 egresados. No obstante, 18 de ellos no facilitaron la fecha de comienzo o de finalización en su primer empleo, por lo que no pudimos generar la variable de duración y tuvimos que desecharlos del análisis, quedándonos finalmente con una muestra de 2522 titulados, de los cuales 977 presentaban un problema de censura.

<sup>135</sup> Obviamente, esta duración en el primer empleo sólo la pueden presentar aquellos titulados que ya poseían dicho trabajo antes de la realización del Ciclo Formativo.

Ilustración IV.13: Funciones empíricas de densidad, distribución y riesgo de la duración en el primer empleo (Kaplan-Meier)



La segunda gráfica de la Ilustración IV.13, que recoge la función de distribución empírica, nos muestra que, prácticamente un cuarto de la muestra (23.5%), tuvo un primer empleo con una duración inferior a 3 meses. Igualmente observamos cómo, en torno a un tercio de la muestra (32.1%), disfrutó de un primer puesto de trabajo durante menos de 6 meses; cerca de la mitad de la misma (49.6%) trabajó en su primer empleo durante menos de un año; y, finalmente, más de dos tercios de la misma (67%), trabajó durante menos de tres años en su primer puesto de trabajo.

Por último, la tercera gráfica de la Ilustración IV.13 muestra la función de riesgo empírica, por lo que representa la probabilidad condicional de abandono del empleo, una vez que el individuo ha estado ocupado en el mismo durante un período  $t$ . Como vemos, dicho riesgo es, en todo momento, inferior al 10%.

Con la finalidad de observar de manera más clara la relación de dependencia que mantiene la función de riesgo con el tiempo, optamos nuevamente por realizar el análisis no paramétrico a través de cálculo actuarial de la duración medida en días. Para ello hemos agrupado las diferentes duraciones en el primer empleo en un total de 28 intervalos. Estos permanecen abiertos por la derecha y cerrados por la izquierda. Comenzamos estableciendo un primer intervalo de quince días para aquellas duraciones en el primer empleo de entre 15 y 30 días. A partir de ahí, establecimos intervalos de mes en mes hasta llegar al año. Entre uno y tres años, establecimos intervalos de 3 meses. A continuación, y hasta los 10 años, los intervalos son de 6 meses. Entre los 10 y los 20 años, los intervalos son de dos años. Y, finalmente, dejamos un último intervalo abierto, para duraciones superiores a los 20 años.

La estructura de la Tabla IV.8 es prácticamente la misma que la ya comentada para la Tabla IV.3, con la excepción de la amplitud de los intervalos. Si nos fijamos en la columna que recoge el número de sucesos que presentan censura, como resulta lógico, dado que hablamos de titulados entre 1997 y 2000, observamos que, cuando la duración en el empleo es superior al año, el número de observaciones censuradas llega a superar al número de observaciones cuya duración es conocida

.La quinta columna de la Tabla IV.8, que recoge una estimación de la proporción de sucesos terminales, es decir, de la probabilidad de que un titulado que llegue a dicha duración finalice su contrato en ese intervalo, nos muestra que los intervalos que presentan una mayor proporción de dichos sucesos son los que van de 3 a 4 meses (9%), de 1 año a 1 año y tres meses (18%), de 3 años a 3 años y medio (15%), de 9 años a 9 años y medio (13%) y, por último, el intervalo abierto para más de 20 años de duración (33.3%).

Tabla IV.8: Tabla de vida de la duración en el primer empleo

Intervalo	Total	Finalizan	Censura	q	p	Función de Supervivencia	f	Función de Distribución	Función de Riesgo
15 30	2522	47	0	0,019	0,981	0,981	0,001	0,019	0,001
30 60	2475	189	13	0,077	0,923	0,906	0,003	0,094	0,003
60 90	2273	167	27	0,074	0,926	0,839	0,002	0,161	0,003
90 120	2079	187	23	0,090	0,910	0,763	0,003	0,237	0,003
120 150	1869	100	22	0,054	0,946	0,722	0,001	0,278	0,002
150 180	1747	89	17	0,051	0,949	0,685	0,001	0,315	0,002
180 210	1641	124	11	0,076	0,924	0,633	0,002	0,367	0,003
210 240	1506	64	21	0,043	0,957	0,606	0,001	0,394	0,001
240 270	1421	47	8	0,033	0,967	0,586	0,001	0,414	0,001
270 300	1366	49	10	0,036	0,964	0,565	0,001	0,435	0,001
300 330	1307	34	16	0,026	0,974	0,550	0,000	0,450	0,001
330 360	1257	23	11	0,018	0,982	0,540	0,000	0,460	0,001
360 540	1223	210	149	0,183	0,817	0,441	0,001	0,559	0,001
540 720	864	78	210	0,103	0,897	0,396	0,000	0,604	0,001
720 900	576	52	124	0,101	0,899	0,356	0,000	0,644	0,001
900 1080	400	22	121	0,065	0,935	0,333	0,000	0,667	0,000
1080 1440	257	32	97	0,153	0,847	0,282	0,000	0,718	0,000
1440 1800	128	7	30	0,062	0,938	0,264	0,000	0,736	0,000
1800 2160	91	7	17	0,085	0,915	0,242	0,000	0,758	0,000
2160 2520	67	4	13	0,066	0,934	0,226	0,000	0,774	0,000
2520 2880	50	5	6	0,106	0,894	0,202	0,000	0,798	0,000
2880 3240	39	1	5	0,027	0,973	0,196	0,000	0,804	0,000
3240 3600	33	4	5	0,131	0,869	0,171	0,000	0,829	0,000
3600 4320	24	1	9	0,051	0,949	0,162	0,000	0,838	0,000
4320 5040	14	1	4	0,083	0,917	0,148	0,000	0,852	0,000
5040 5760	9	0	1	0,000	1,000	0,148	0,000	0,852	0,000
5760 7200	8	0	3	0,000	1,000	0,148	0,000	0,852	0,000
7200 .	5	1	4	0,333	0,667	0,099	.	0,901	.

En cambio, si nos fijamos en la última columna, que recoge la estimación de la función de riesgo, es decir, la estimación de la probabilidad de abandonar el empleo por unidad de tiempo supuesto que el titulado ha permanecido en él hasta el comienzo del intervalo, observamos que dicho riesgo no llega al 1% en ningún caso. Por tanto, la probabilidad condicional de abandono del primer empleo en nuestra muestra es, en todo momento, prácticamente nula. De nuevo observamos que la estimación del riesgo mediante al cálculo actuarial es inferior a la estimación del mismo mediante la metodología de Kaplan y Meier. No obstante, como ya hemos comentado previamente, esto es debido, tanto a las correcciones que lleva a cabo el cálculo actuarial ante la presencia de censura en los datos, como a la necesaria agrupación de datos en intervalos.

Al igual que hicimos en el análisis de la duración del desempleo, hemos optado por graficar las funciones de riesgo, distribución y supervivencia, obtenidas mediante el cálculo actuarial, diferenciando la muestra en diversos colectivos. Así, la Ilustración IV.14 recoge las tres funciones diferenciando entre hombres y mujeres, la Ilustración

IV.15 hace la diferenciación en función del nivel del Ciclo Formativo, la Ilustración IV.16 diferencia en función de la titularidad del centro de estudios y, finalmente, en la Ilustración IV.17 recogemos las tres funciones diferenciando por Familias Profesionales.

Observando cada una de estas ilustraciones, podría parecer que tan sólo en la Ilustración IV.16 y en la Ilustración IV.17 aparecen diferencias significativas entre los grupos. De hecho, en las otras dos ilustraciones, las gráficas aparecen entrelazadas a lo largo de todas las duraciones consideradas por lo que no se distingue una pauta de comportamiento claramente definida para cada subgrupo considerado.

Al disponer nuevamente de censura en los datos y seguir trabajando con la muestra ponderada, de nuevo hemos optado por la realización del contraste de homogeneidad de 'Cox' (StataCorp., 2001), cuyos resultados vienen recogidos en la Tabla IV.9.

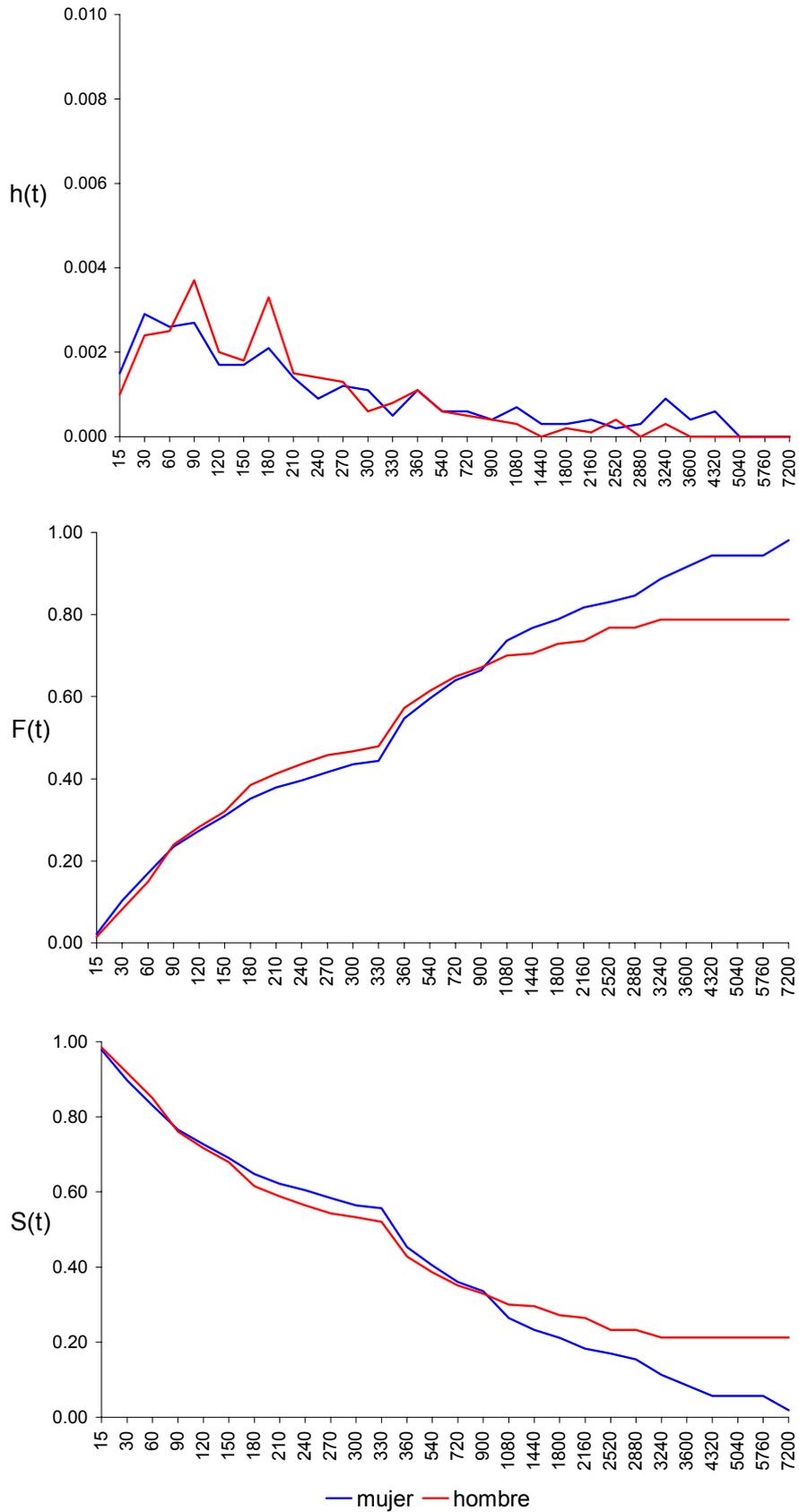
**Tabla IV.9: Contrastes de homogeneidad de estratos para el análisis de la duración en el primer empleo**

Estratos	Estadístico	Probabilidad
Género	0,00	0,9878
Nivel del Ciclo Formativo	0,96	0,3276
Titularidad del centro	0,71	0,3992
Familias Profesionales	35,32	0,0185

Los resultados de dicho contraste nos muestran que no existen evidencias que nos lleven a rechazar la hipótesis nula de igualdad de funciones de riesgo en los tres primeros casos. Debido a esto, no tiene mayor interés ahondar en el análisis de las funciones de riesgo, distribución y supervivencia, diferenciando por género, nivel del Ciclo Formativo o titularidad del centro de estudios. En cambio, el contraste de homogeneidad realizado con estratificación de la muestra por Familias Profesionales, sí que nos lleva a rechazar la hipótesis nula de igualdad de funciones de riesgo, por lo que conviene analizar con mayor detenimiento las gráficas recogidas en la Ilustración IV.17.

Efectivamente, basta echar un vistazo al eje de ordenadas de la primera gráfica de la Ilustración IV.17, para comprobar la disparidad existente, en cuanto a funciones de riesgo, entre Familias Profesionales. Como ya mostramos, mediante la Tabla IV.8, la estimación del riesgo para el conjunto de la muestra, en cada uno de los intervalos considerados, no superaba en ningún momento el 0.3%. En cambio, al graficar la estimación de dicho riesgo diferenciando por Familias Profesionales, nos hemos encontrado con valores que oscilaban entre el 0% el 2.2%.

Ilustración IV.14: Función de riesgo, de distribución y de supervivencia de la duración en el primer empleo por género



**Ilustración IV.15: Función de riesgo, de distribución y de supervivencia de la duración en el primer empleo por nivel del CF**

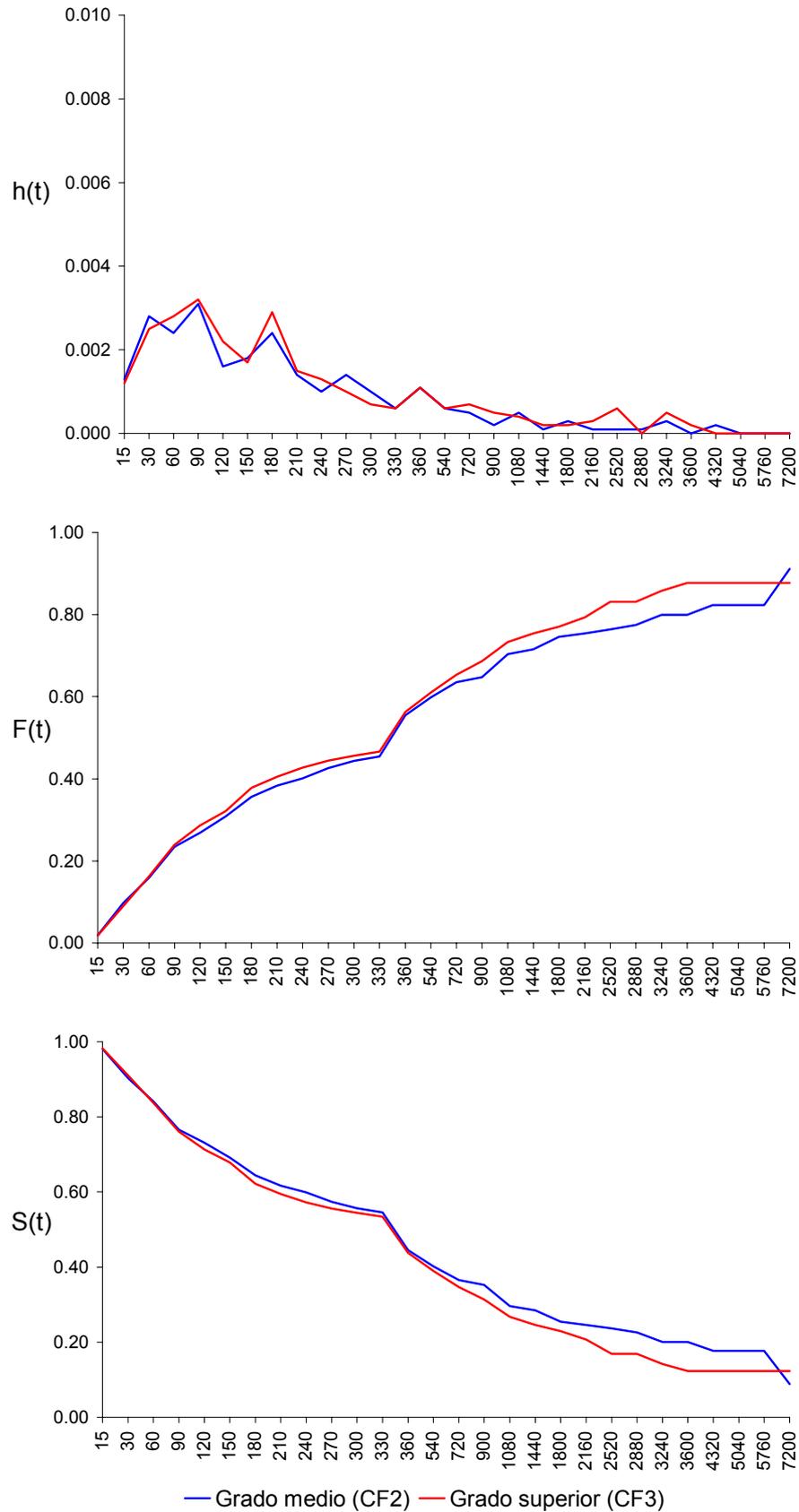


Ilustración IV.16: Función de riesgo, de distribución y de supervivencia de la duración en el primer empleo por titularidad del centro

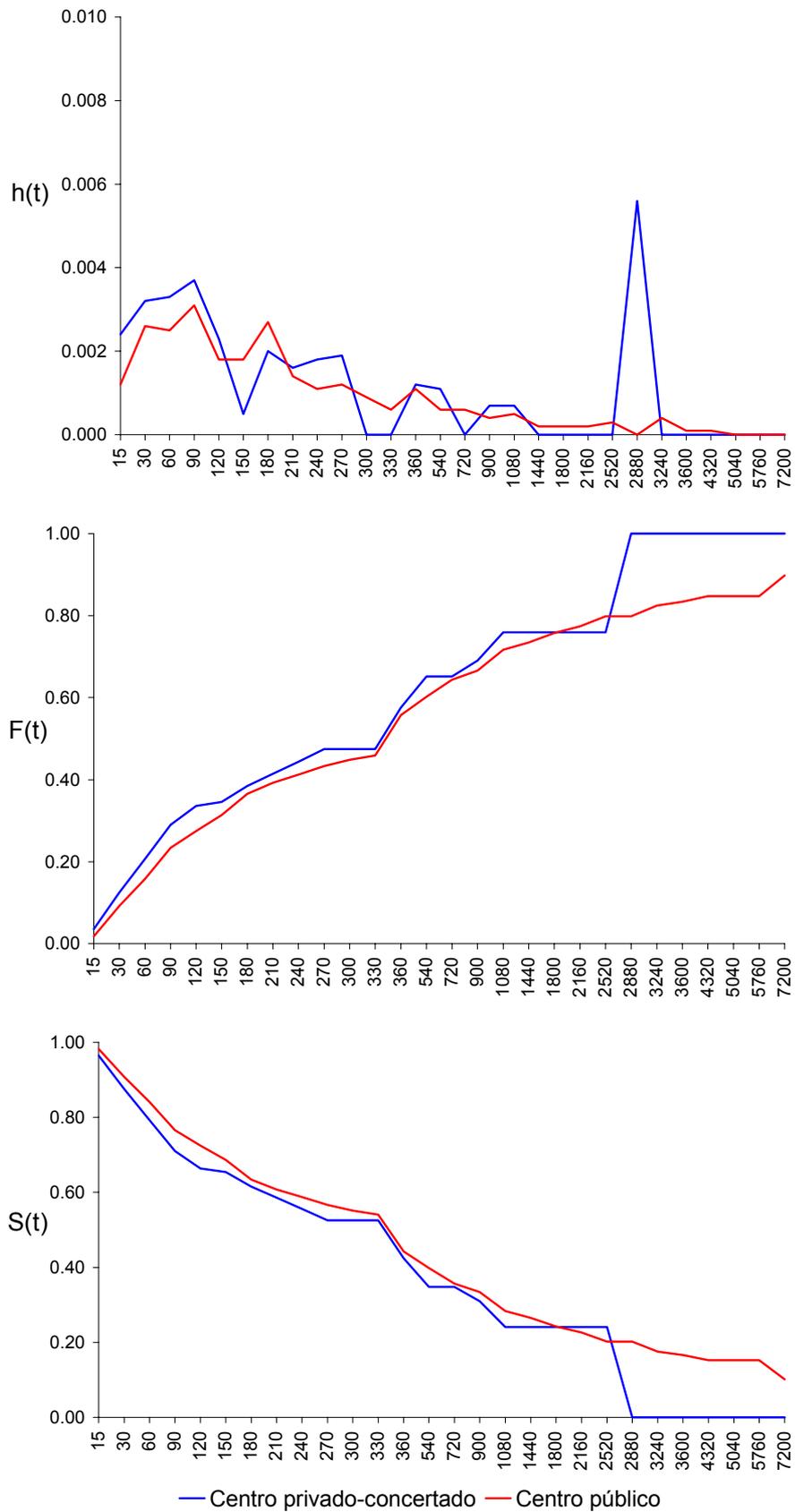
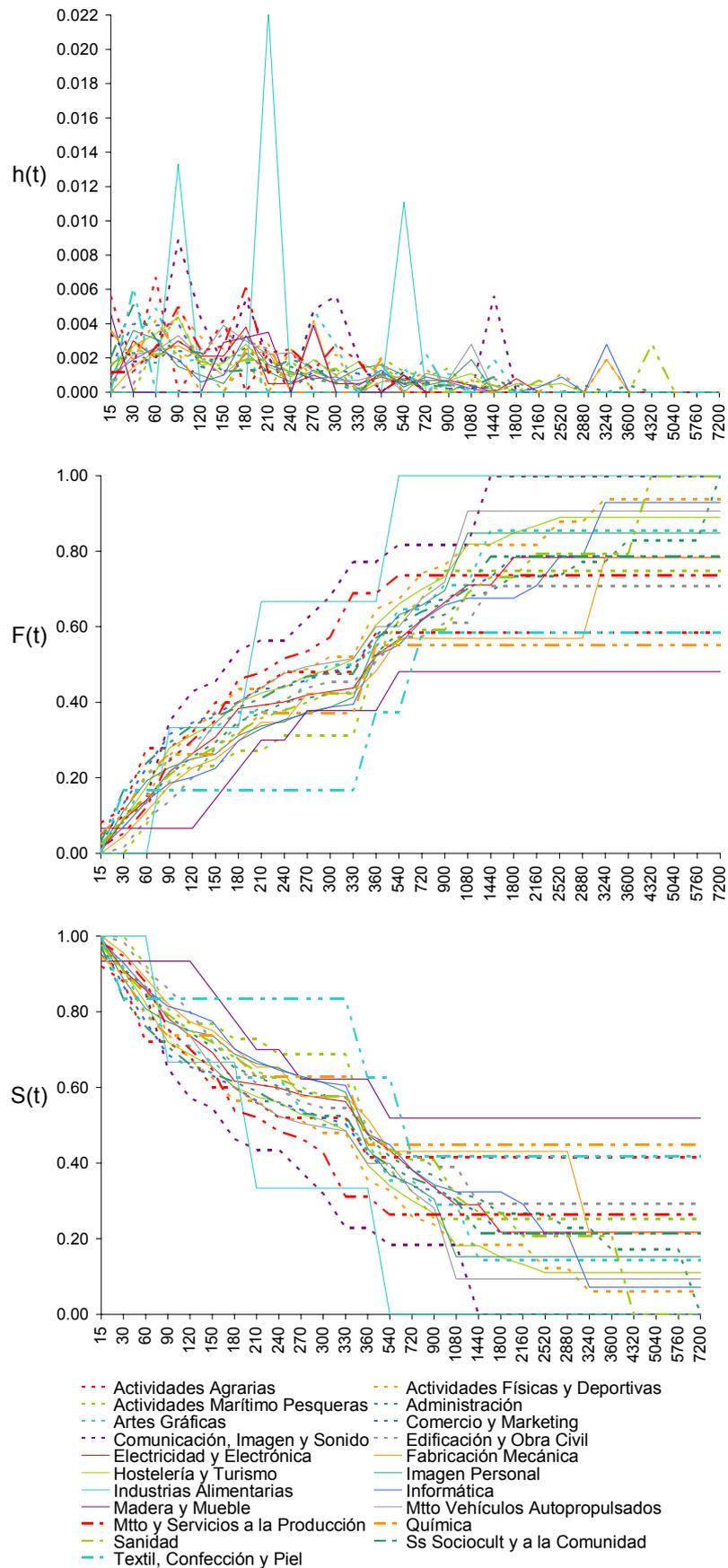


Ilustración IV.17: Función de riesgo, de distribución y de supervivencia de la duración en el primer empleo por familia profesional



La Familia Profesional que presenta una mayor probabilidad condicionada de abandono del empleo es, con diferencia, la de 'Industrias Alimentarias', cuyos titulados presentan una probabilidad del 1.3% al cabo de 3 meses, del 2.2% al cabo de 7 meses y del 1.1%, al cabo de año y medio. En segundo lugar, se encuentran los titulados de 'Comunicación, Imagen y Sonido', quienes presentan una probabilidad condicionada del 0.9% al cabo de 3 meses, del 0.6% al cabo de 10 meses y del 0.5% al cabo de 4 años. Y en tercer lugar los de 'Actividades Agrarias', quienes presentan una probabilidad condicionada del 0.7% tras dos meses de empleo y del 0.6% tras 6 meses.

Las funciones de distribución y supervivencia nos muestran cómo las Familias Profesionales, cuyos titulados se enfrentan a primeros empleos con una menor duración, son las de 'Comunicación, Imagen y Sonido' e 'Industrias Alimentarias'. De hecho, al cabo de un año, por ejemplo, tan sólo el 22.8% de los titulados en 'Comunicación, Imagen y Sonido' y el 33.3% de los titulados en 'Industrias Alimentarias' permanecían aún en su primer empleo. Por el lado contrario, nos encontramos con el caso de los titulados de 'Textil, Confección y Piel' y 'Madera y Mueble', quienes, en el 62.5% de los casos, aún permanecían en su primer empleo al cabo de un año. Estos últimos, seguidos de los titulados en 'Química', son los que mantienen una mayor probabilidad de supervivencia en comparación con el resto de las titulaciones, pues el 52% de sus titulados (45% de los de 'Química') permanecen trabajando en su primer empleo una vez transcurridos 2 años.

### 5.1.2. Estimación paramétrica

Finalizado el análisis no paramétrico, pasamos a analizar cuáles son los rasgos más significativos a la hora de explicar la duración del primer empleo entre los técnicos y técnicos superiores de la isla de Gran Canaria.

La literatura proporciona una serie de posibles factores que pueden influir en la determinación de la antigüedad de un trabajador en su empleo. Estos factores son los que se derivan, fundamentalmente, de la teoría del capital humano (Becker, 1962) y de la teoría del *job-matching* (Jovanovic, 1979). La primera de ellas enfatiza aquellos aspectos que están relacionados con la oferta laboral, mientras que la segunda permite una interacción más explícita entre componentes de la demanda y de la oferta laboral.

Desde el punto de vista de la *oferta laboral*, la teoría del capital humano considera que un trabajador abandonará su empleo si la utilidad esperada de abandonar el mismo, es mayor que la utilidad alternativa de permanecer, menos el coste del cambio. Dado que los trabajadores reciben formación específica en su empleo, y que ésta es recompensada por parte de la empresa a través de salarios más altos pero, sin embargo, la misma no resulta útil para otras empresas, la formación específica recibida en el puesto, tendrá el efecto de reducir la rotación laboral.

Desde la perspectiva de la teoría del *job-match*, la calidad del ajuste entre empleador y empleado es lo que determina que la relación entre estos continúe o finalice. De esta manera, se espera que los ajustes de 'buena calidad' sean duraderos, mientras que los que resultan insatisfactorios, se espera que finalicen antes. La teoría del *job-match* afirma que los trabajadores más jóvenes muestran una mayor pauta de movilidad, con el fin de adquirir un mayor conocimiento del mercado, que les permita determinar qué empleo / empleador es el que mejor se ajusta a sus habilidades. Esta propensión al cambio entre los trabajadores más jóvenes implica así una correlación positiva entre la edad y la antigüedad en el empleo.

Por otro lado, desde el punto de vista de la *demanda laboral*, las empresas estarán interesadas en despedir a un trabajador, si el beneficio esperado de su continuidad es inferior al beneficio esperado tras el despido. La disminución del beneficio esperado se produce cuando la productividad marginal del trabajador llega a ser inferior a su salario. No obstante, dado que las empresas no pueden disminuir los salarios, en caso de una disminución de la productividad marginal, el reajuste se realizará vía disminución del número de empleados. Teniendo en cuenta que, lo habitual es que las empresas sigan una política basada en el método LIFO<sup>136</sup>, los últimos trabajadores contratados, en caso de una reestructuración de plantilla, serán los primeros en ser despedidos, por lo que serán los trabajadores más jóvenes los que presenten una mayor probabilidad de abandono de la empresa.

Aparte de esto, existe un consenso entre los investigadores acerca de la influencia del tamaño de la empresa en la duración de los empleos. Así, se observa una mayor duración de los empleos en las empresas de mayor tamaño. Esto se debe, quizás, a que las grandes empresas invierten más en la formación de sus trabajadores, por lo que les resultaría más costoso perderlos y tratarán de reducir la rotación.

Finalmente, Battu et al. (2002) hablan de otra serie de factores que igualmente pueden influir en la duración del empleo. Estos son los que ellos llaman 'factores institucionales', como la propia legislación laboral -que puede restringir al empleador las causas de disolución de una relación contractual- o, de manera más general, las propias cláusulas contractuales.

Teniendo en cuenta las aportaciones de estas teorías, procedemos a la especificación de nuestro modelo. En nuestro caso, la variable dependiente a estudiar es la duración del primer episodio de ocupación de los técnicos y técnicos superiores que constituyen nuestra muestra<sup>137</sup>. Ésta se obtuvo como diferencia entre la fecha de inicio y finalización

---

<sup>136</sup> LIFO son las siglas de *Last In First Out*

<sup>137</sup> La muestra, por tanto, no incluye a los '*titulados del grupo 4*'. El hecho de analizar solamente una selección muestral, no introduce un problema de sesgo de selección en los resultados. No obstante, como comentan Nilsen y Risa (1999), hemos de tener en cuenta que, los coeficientes estimados, se interpretan como el efecto de la covariable en cuestión sobre la duración en el primer empleo, condicionado a la consecución de un primer empleo.

en el primer puesto de trabajo, excepto en el caso de los titulados que, en el momento de responder a la encuesta, aún trabajaban en su primer empleo, para quienes se sustituye la fecha de finalización (que es desconocida), por la fecha de realización de la encuesta.

La muestra constará, por tanto, de titulados para los que conocemos la duración total de su primer período de ocupación, así como de titulados para los que la duración completa resulta desconocida, teniendo así un problema de censura en los datos. Este problema afectará al 38.7% de las observaciones.

La elección del modelo de duración paramétrico más afín al comportamiento de nuestros datos, parte del análisis de la función de riesgo estimada de manera no paramétrica. Ésta, como refleja la Ilustración IV.14, presenta un crecimiento hasta los tres meses de duración, a partir del cual comienza a decrecer. Por tanto, hemos considerado conveniente ajustar un modelo log-normal y otro log-logístico, puesto que estos modelos se suelen utilizar cuando los datos muestran tasas de riesgo no monótonas con crecimiento al comienzo y decrecimiento al final (StataCorp., 2001).

No obstante, si en lugar de observar la función de riesgo de la Ilustración IV.14, observamos la de la Ilustración IV.13, podemos plantearnos la posibilidad de ajustar un modelo que recoja un decrecimiento monótono de la función de riesgo, por lo que decidimos estimar también un modelo de duración tipo Gompertz y otro tipo Weibull.

En cualquier caso, las variables independientes que hemos introducido en cada uno de los modelos son las mismas. Éstas las podemos agrupar en 5 grandes bloques:

1. Características personales y familiares:

En este bloque incluimos aquellas variables que, al referirse a las características intrínsecas del titulado, recogen aspectos del lado de la oferta laboral. Estas variables son:

SEXO: variable dicotómica que toma valor 1 si el titulado es hombre

EDAD1: edad del titulado al comenzar a trabajar en su primer empleo

EDAD12 edad del titulado al comenzar a trabajar en su primer empleo al cuadrado

EXPPREV: variable dicotómica que toma valor 1 si el titulado reconoce tener experiencia previa por haber trabajado antes de cursar el CF

HIJOS1<sup>138</sup>: variable dicotómica que toma valor 1 si el titulado tenía hijos antes de comenzar a trabajar en su primer empleo

NTRABAJ1<sup>138</sup>: número de miembros de la familia que estaban trabajando cuando el titulado finalizó sus estudios

INGRHOG1<sup>138</sup>: ingresos del hogar, mientras el titulado realizaba en Ciclo Formativo, en Euros constantes de 2001

CABFAOCUPA1<sup>138</sup>: variable dicotómica que toma valor 1, si el cabeza de familia trabajaba cuando el titulado finalizó sus estudios, y el valor 0, en otro caso

## 2. Características educativas:

En este segundo bloque hacemos referencia a la formación académica del titulado como componente de la oferta laboral que puede afectar a la duración del período de empleo. Estas características son las siguientes:

NIVELCF: variable dicotómica que toma el valor 1, si el titulado cursó un Ciclo Formativo de Grado Superior, y el valor 0, en caso de haber cursado uno de Grado Medio.

FAMILIAP: recoge un conjunto de variables dicotómicas que hacen referencia a la Familia Profesional a la que pertenece el Ciclo Formativo cursado por el titulado. En el caso de FAMILIAP1 el CF pertenece a la Familia de Actividades Agrarias; en FAMILIAP2 el CF pertenece a la Familia de Actividades Físicas y Deportivas; en FAMILIAP3 el CF pertenece a la Familia de Actividades Marítimo Pesqueras; en FAMILIAP5 el CF pertenece a la Familia de Artes Gráficas; en FAMILIAP6 el CF pertenece a la Familia de Comercio y Marketing; en FAMILIAP7 el CF pertenece a la Familia de Comunicación, Imagen y Sonido; en FAMILIAP8 el CF pertenece a la Familia de Edificación y Obra Civil; en FAMILIAP9 el CF pertenece a la Familia de Electricidad y Electrónica; en FAMILIAP10 el CF pertenece a la Familia de Fabricación Mecánica; en FAMILIAP11 el CF pertenece a la Familia de Hostelería y Turismo; en FAMILIAP12 el CF pertenece a la Familia de Imagen Personal; en FAMILIAP13 el CF pertenece a la Familia de Industrias Alimentarias; en FAMILIAP14 el CF pertenece a la Familia de Informática; en FAMILIAP15 el CF pertenece a la Familia de Madera y Mueble; en FAMILIAP16 el CF pertenece a la Familia de Mantenimiento de Vehículos Autopropulsados; en FAMILIAP17 el CF pertenece a la Familia de Mantenimiento y Servicios a la Producción; en FAMILIAP18 el CF pertenece a la Familia de Química; en FAMILIAP19 el CF pertenece a la Familia de

<sup>138</sup> Esta variable realmente se refiere a la situación del titulado al finalizar sus estudios. No obstante, dado que el tiempo que transcurre entre la finalización del CF y el comienzo en el primer empleo es de 7 meses en promedio (período de inactividad + período de desempleo), suponemos que el valor que toma la variable al final del CF coincide con el que tomaría al comienzo del primer empleo.

Sanidad; en FAMILIAP20 el CF pertenece a la Familia de Servicios Socioculturales y a la Comunidad; y, finalmente, FAMILIAP21, que indica si el CF pertenece a la Familia de Textil, Confección y Piel. Hemos dejado como referencia los Ciclos Formativos pertenecientes a la Familia de Administración.

### 3. Variables de entorno:

En el tercer bloque de variables recogemos aquellos aspectos del entorno económico que pueden influir en la demanda de trabajo, afectando así a la duración del primer empleo. Introducimos así las siguientes variables:

TASADESJ: tasa trimestral de desempleo juvenil (16 a 25 años) en Canarias en la fecha en la que el titulado consiguió su primer empleo

AÑOCOM\_: recoge un conjunto de variables dicotómicas que hacen referencia al año en el que los titulados consiguen su primer empleo. AÑOCOM\_95 toma el valor 1 cuando el titulado consigue su empleo en el año 1995; AÑOCOM\_96 toma el valor 1 cuando el titulado consigue su empleo en el año 1996; AÑOCOM\_97 toma el valor 1 cuando el titulado consigue su empleo en el año 1997; AÑOCOM\_98 toma el valor 1 cuando el titulado consigue su empleo en el año 1998; AÑOCOM\_99 toma el valor 1 cuando lo consiguen en el año 1999; AÑOCOM\_00 toma el valor 1 cuando lo consigue en el año 2000; y, finalmente, AÑOCOM\_01 toma el valor 1 cuando lo consiguen en el año 2001. Se deja como referencia a los que consiguieron su primer empleo en 1994.

### 4. Características del empleo:

El cuarto bloque de variables explicativas hace referencia a las características específicas del primer puesto de trabajo ocupado por los titulados. Éstas son:

MEDEFEC1: variable dicotómica que toma el valor 1, si el medio que resultó efectivo en la consecución de ese primer empleo era un medio formal, y que toma el valor 0, en caso de ser un medio de búsqueda informal<sup>139</sup>.

EMPRIV\_: recoge un conjunto de variables dicotómicas que hacen referencia al tipo de empresa. EMPRIV\_A toma el valor 1 cuando se trata de una empresa privada grande (más de 250 trabajadores); EMPRIV\_B toma el valor 1 cuando se trata de una PYME privada (entre 11 y 250 trabajadores); EMPRIV\_C toma el valor 1 cuando se trata de una microempresa privada (menos de 10 trabajadores); EMPRIV\_D toma

---

<sup>139</sup> Consideramos como medios de búsqueda informales los que se refieren al uso de contactos personales, así como los contactos con los encargados de la FCT, mientras que los procesos de búsqueda formales serán: enviar currículos, apuntarse al INEM o la ACE, acudir a ETT, responder a anuncios en prensa, apuntarse en bolsas de trabajo o preparar oposiciones.

el valor 1 cuando se trata de un trabajo por cuenta propia. Se toma como referencia a las empresas públicas.

TIPCONT\_: dicotómicas que hacen referencia al tipo de contrato en el primer empleo. TIPCONT\_2 se refiere al contrato laboral temporal; TIPCONT\_3 se refiere a los contratos en formación; TIPCONT\_4 se refiere a becas o similar; TIPCONT\_5 recoge a los que trabajan sin contrato; TIPCONT\_6 se refiere a los que trabajan por cuenta propia con o sin licencia fiscal; TIPCONT\_7 se refiere a las ayudas familiares; y TIPCONT\_8 recoge a cualquier otro tipo de contrato, como puede ser el de soldado profesional. La categoría de referencia es el contrato laboral indefinido.

SUELHOR1: sueldo horario en el primer empleo en Euros constantes de 2001

CATLAB1: variable dicotómica que toma el valor 1, si la categoría laboral en el primer empleo es de directivo o mando medio, y que toma el valor 0, en caso de ser un operario.

TCOMPLET1: es una variable dicotómica que toma el valor 1 cuando el titulado trabajaba en el primer empleo con un contrato a tiempo completo, es decir, de 40 o más horas semanales.

#### 5. Atributos laborales:

Finalmente, en el quinto bloque de variables explicativas incluimos una serie de variables que recogen las características laborales de los propios trabajadores y que, a su vez, sirven como 'medidas de ajuste'. Éstas tratan de recoger el grado de compatibilidad existente entre el individuo y su puesto de trabajo. Dentro de este bloque introducimos las siguientes variables:

TBUSCEMP1: tiempo de búsqueda, previo a la consecución del primer empleo, medido en días.

TOKTRA1: tiempo, medido en días, que la persona ha tardado en desempeñar eficientemente su trabajo.

TIT\_: dicotómicas que hacen referencia al colectivo al que pertenece el titulado. TIT\_2 se refiere a los '*titulados del grupo 2*'; TIT\_3 se refiere a los '*titulados del grupo 3*'. Se ha dejado como referencia a los '*titulados del grupo 1*'.

HABAD\_: conjunto de variables dicotómicas que recogen el grado de desajuste en término de habilidades. HABAD\_E toma el valor 1 cuando el titulado considera estar sobrehabilitado; HABAD\_I toma el valor 1 cuando el titulado se considera subhabilitado; HABAD\_N toma el valor 1 cuando el titulado considera que sus

habilidades no guardan relación con las necesidades del puesto. Tomamos como referencia a los titulados que consideran estar adecuadamente habilitados.

CONADE\_: conjunto de variables dicotómicas que recogen el grado de desajuste en término de conocimientos. CONADE\_S toma el valor 1 cuando el titulado considera tener conocimientos superiores a los necesarios para el puesto; CONADE\_I toma el valor 1 cuando el titulado considera tener conocimientos inferiores a los necesarios para el puesto. Tomamos como referencia a los titulados que consideran tener el nivel adecuado de conocimientos.

SAT\_: conjunto de variables dicotómicas que recogen el grado de satisfacción de los titulados con respecto a una serie de características del puesto de trabajo. SATSUEL1 toma el valor 1 si el titulado está poco / nada satisfecho con su sueldo; SATESTA1 toma el valor 1 si el titulado está poco / nada satisfecho con la estabilidad en su puesto de trabajo; SATPROM1 toma el valor 1 si el titulado está poco / nada satisfecho con las posibilidades de promoción; SATRIES1 toma el valor 1 si el titulado está poco / nada satisfecho con la prevención de riesgos laborales; SATHORA1 toma el valor 1 si el titulado está poco / nada satisfecho con la duración y flexibilidad de su horario de trabajo.

Para la estimación de los 4 modelos propuestos consideramos, en primer lugar, la posibilidad de existencia de heterogeneidad inobservable de tipo Gamma o de tipo Gaussiana Inversa. Los resultados del contraste de existencia de dicha heterogeneidad son los que vienen reflejados en la Tabla IV.10.

**Tabla IV.10: Contrastes de existencia de heterogeneidad en los modelos de duración del primer empleo**

	Heterogeneidad Gamma				Heterogeneidad Gaussiana Inversa			
	Log-Normal	Log-Logístico	Gompertz	Weibull	Log-Normal	Log-Logístico	Gompertz	Weibull
$\bar{\chi}_1^2$	0,19	0,82	21,81	117,24	0,23	1,10	17,51	105,63
p	0,332	0,182	0,000	0,000	0,317	0,147	0,000	0,000

Como se deduce de esta tabla, las estimaciones de los modelos de duración log-normal y log-logístico no presentan problemas de heterogeneidad, mientras que las de los modelos Gompertz o Weibull sí presentan este problema. Por tanto, a la hora de estimar los modelos para finalmente optar por el más adecuado a nuestros datos, hemos de estimar los dos primeros, sin considerar la existencia de heterogeneidad, y los dos últimos, incorporando la consideración de heterogeneidad en sus dos variantes.

En la Tabla IV.11 se recogen diferentes medidas de bondad del ajuste que nos ayudan a la elección del mejor de ellos.

**Tabla IV.11: Medidas de bondad del ajuste de diversos modelos de duración para el primer período de empleo**

Distribución	Heterogeneidad	Chi <sup>2</sup> <sub>67</sub>	Log. Verosimilitud	AIC	Nº observ.
Log-Normal	-	1227,15 p=0,000	-7117,5961	14373,192	2020
Log-Logística	-	1131,27 p=0,000	-7145,7570	14429,514	2020
Gompertz	Gamma	635,40 p=0,000	-3801,4808	7740,962	2020
	Gaussiana Inversa	1092,56 p=0,000	-2476,2786	5090,557	2020
Weibull	Gamma	377,00 p=0,000	-3731,9717	7601,943	2020
	Gaussiana Inversa	904,75 p=0,000	-3741,2271	7620,454	2020

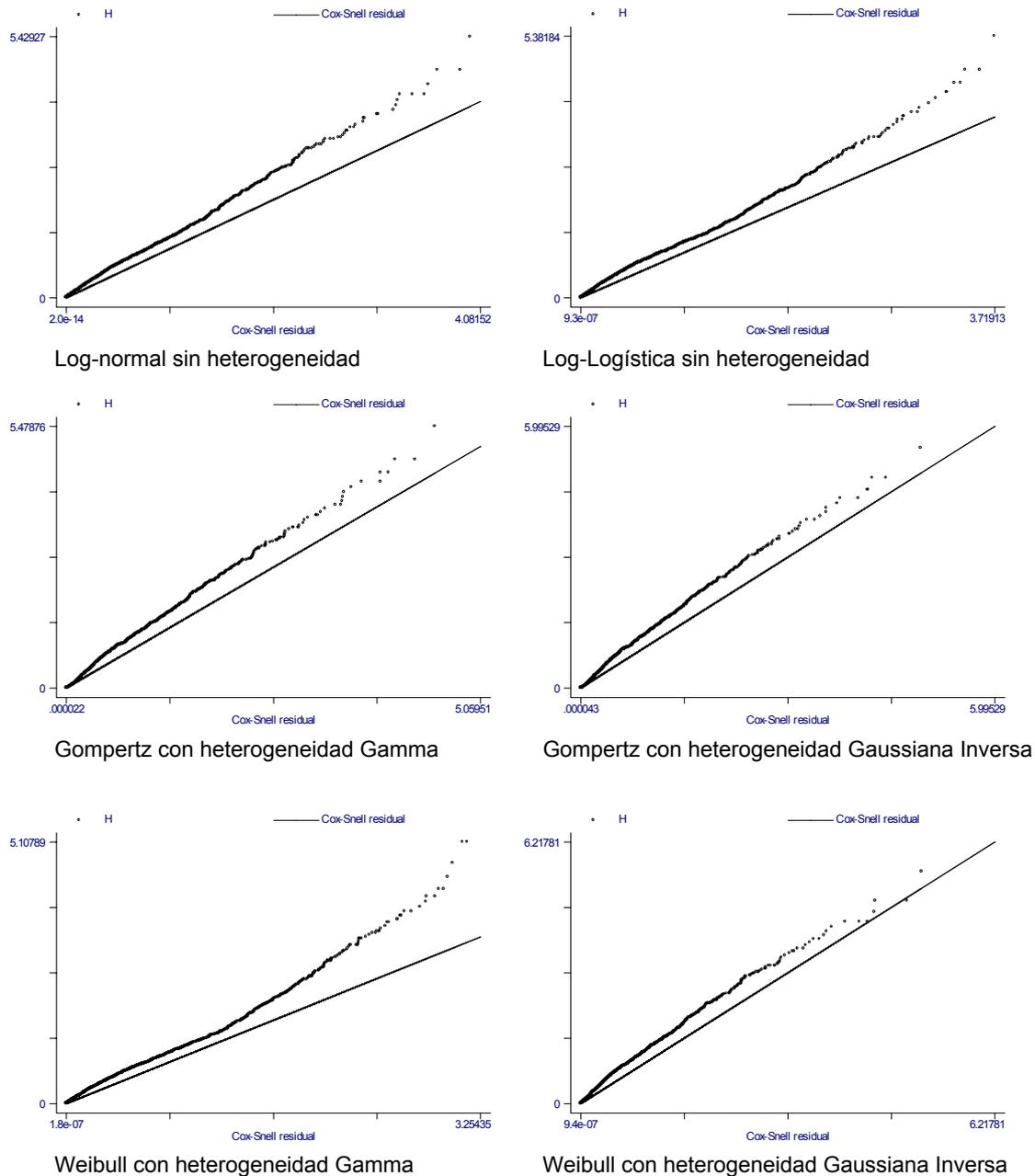
Como vemos, los contrastes Chi-cuadrado de significación global nos permiten aceptar cualquiera de los 6 modelos, puesto que nos llevan a rechazar la hipótesis nula de falta de significación global. En lo que respecta a la medida del Logaritmo de Verosimilitud, los resultados apuntan hacia una preferencia por las especificaciones tipo Gompertz o Weibull, puesto que éstas son las que proporcionan, con diferencia, un mayor valor de dicha magnitud. Finalmente, el Criterio de Información de Akaike apunta hacia una preferencia por la estimación tipo Gompertz con heterogeneidad Gaussiana Inversa, puesto que ésta es la que proporciona un menor valor para el AIC.

No obstante, antes de optar por cualquiera de las especificaciones, al igual que hicimos en la estimación paramétrica del primer período de desempleo, debemos observar lo que sucede con los residuos de Cox-Snell generados por cada uno de estos modelos, puesto que Kiefer (1988) recomienda optar por el modelo que presente un mejor comportamiento de estos residuos.

La Ilustración IV.18 nos muestra la representación gráfica de la función integrada de riesgo de los residuos de Cox-Snell. Según apunta Kiefer (1988), el modelo que presenta un mejor ajuste a los datos, será aquél, cuyos residuos se asemejen más a una distribución Exponencial unitaria o, lo que es lo mismo, cuya representación gráfica de la función integrada de riesgo de sus residuos se ajuste más a una línea con pendiente unitaria. Por tanto, siguiendo este último criterio, la mejor estimación corresponde a la obtenida a través del modelo Weibull con heterogeneidad Gaussiana Inversa<sup>140</sup>.

<sup>140</sup> Este modelo, según el Criterio de Información de Akaike, se hubiese posicionado en el tercer lugar.

**Ilustración IV.18: Función integrada de riesgo de los residuos generalizados de Cox-Snell para la estimación de la duración del primer empleo**



Una vez más, hemos optado por estimar el modelo de duración, diferenciando nuestra muestra entre hombres y mujeres, aparte de estimarlo para todo el colectivo, puesto que consideramos que, probablemente, el comportamiento de las variables explicativas seleccionadas difiera en función del género del trabajador. Los resultados de estas tres estimaciones son los que recoge la Tabla IV.12.

Tabla IV.12: Estimación del modelo de duración tipo Weibull con heterogeneidad Gaussiana Inversa para la duración del primer empleo, diferenciando por género

	Todos		Mujeres		Hombres	
	Coef.	z	Coef.	z	Coef.	z
(Constante)	-13,870 ***	5,26	-7,149 **	2,04	-22,103 ***	5,97
<b>CARACTERÍSTICAS PERSONALES Y FAMILIARES:</b> (Referencia: mujer, sin experiencia laboral previa, sin hijos, cabeza de familia desempleado)						
SEXO	0,005	0,04				
EDAD1	0,099	0,71	-0,329 *	1,73	0,269	1,54
EDAD12	-0,325	1,09	0,651	1,58	-0,726 **	2,06
EXPPREV	0,379 ***	3,75	0,384 ***	2,65	0,424 ***	2,97
HIJOS1	-0,388	0,89	-0,419	0,82	-0,444	0,57
NTRABAJ1	-0,136 ***	2,70	-0,170 **	2,13	-0,124 *	1,89
INGRHOG1	0,000	1,35	0,000	0,85	0,000	1,37
CABFAOCUPA1	0,008	0,06	0,059	0,35	-0,029	0,13
<b>CARACTERÍSTICAS FORMATIVAS:</b> (Referencia: Ciclo Formativo de grado medio; Familia Profesional de Administración)						
NIVELCF	-0,048	0,38	-0,071	0,39	0,172	0,91
FAMILIAP1	0,080	0,30	-0,017	0,04	0,059	0,16
FAMILIAP2	0,206	0,41	0,425	0,50	-0,158	0,22
FAMILIAP3	0,512	0,88	-1,072	0,97	1,004	1,53
FAMILIAP5	0,196	0,40	-0,539	0,76	1,159 **	1,98
FAMILIAP6	0,342 *	1,84	0,391 *	1,74	0,343	1,00
FAMILIAP7	0,664 *	1,82	0,463	0,91	0,945 *	1,81
FAMILIAP8	0,061	0,21	-0,661	1,50	0,366	0,93
FAMILIAP9	-0,018	0,08	-1,301	0,90	0,043	0,16
FAMILIAP10	-0,104	0,26	1,657 ***	5,05	-0,054	0,13
FAMILIAP11	0,676 ***	3,71	0,668 ***	2,84	1,018 ***	2,89
FAMILIAP12	0,102	0,39	-0,149	0,48	0,864 *	1,66
FAMILIAP13	0,531	1,04	2,202 ***	4,31	0,220	0,50
FAMILIAP14	0,089	0,37	0,345	0,83	0,085	0,26
FAMILIAP15	-1,387 ***	2,64			-1,343 **	2,44
FAMILIAP16	0,224	0,87			0,353	1,18
FAMILIAP17	0,228	0,67	3,820 ***	7,46	0,153	0,41
FAMILIAP18	-0,187	0,30	-0,375	0,40	-0,028	0,05
FAMILIAP19	-0,412 **	2,17	-0,572 **	2,44	-0,375	1,00
FAMILIAP20	0,230	0,92	0,179	0,64	0,886	1,48
FAMILIAP21	0,115	0,13	0,066	0,08	-0,160	
<b>VARIABLES DE ENTORNO:</b> (Referencia: Trabaja en el sector primario; consiguió su trabajo en 1994)						
SECTSEC	0,505	1,01	1,712 **	1,98	-0,258	0,43
SECTTERC	0,292	0,61	1,626 **	1,98	-0,676	1,16
TASADESJ	0,021	0,46	-0,022	0,36	0,120 *	1,73
AÑOCOM_95	0,570	0,80	-0,173	0,17	0,845	0,71
AÑOCOM_96	-0,403	0,55	-1,510	1,50	1,205	1,10
AÑOCOM_97	1,235 *	1,80	0,858	0,99	2,310 **	2,14
AÑOCOM_98	1,529 **	2,08	1,019	1,10	2,587 **	2,17
AÑOCOM_99	1,984 **	2,01	1,278	1,01	3,624 **	2,33
AÑOCOM_00	2,052 *	1,76	1,067	0,71	4,176 **	2,29
AÑOCOM_01	2,190	1,63	0,943	0,54	4,721 **	2,30
<b>CARACTERÍSTICAS DEL EMPLEO:</b> (Referencia: Medio de búsqueda efectivo de tipo informal; empresa pública; contrato laboral indefinido; operario; trabajo a tiempo parcial)						
MEDEFEC1	0,134	1,28	-0,040	0,26	0,363 **	2,40
EMPRIV_A	0,499 **	2,13	0,303	1,07	0,640	1,56

EMPRIV_B	0,425 *	1,88	0,059	0,21	0,897 **	2,21
EMPRIV_C	0,731 ***	3,17	0,534 *	1,90	1,074 ***	2,58
EMPRIV_D	-0,022	0,04	-0,412	0,61	0,730	0,75
TIPCONT_2	2,702 ***	16,99	2,618 ***	12,35	3,020 ***	11,58
TIPCONT_3	2,716 ***	11,89	2,580 ***	7,96	3,016 ***	9,08
TIPCONT_5	2,897 ***	11,76	2,655 ***	7,54	3,304 ***	9,35
TIPCONT_6	-0,165	0,20	-0,791	0,68	0,910	0,68
TIPCONT_7	-0,645	0,54	-17,717 ***	24,75	1,160	0,95
TIPCONT_8	2,965 ***	2,77	2,798 **	2,10	4,336 ***	3,09
SUELHOR1	-0,108 ***	3,30	-0,172 ***	4,23	-0,043	0,78
CATLAB1	0,060	0,14	0,435	0,72	-1,163	1,55
TCOMPLET1	-0,078	0,62	-0,078	0,47	-0,080	0,40
<b>ATRIBUTOS LABORALES:</b> (Referencia: Titulados del grupo 1; habilidades adecuadas; conocimientos adecuados; muy / bastante satisfecho con el sueldo, con la estabilidad, con las posibilidades de promoción, con la prevención de riesgos laborales y con el horario)						
TBUSCEMP1	-0,001 **	1,97	-0,001 **	2,02	-0,001	1,17
TOKTRA1	-0,011 ***	5,87	-0,015 ***	3,88	-0,009 ***	4,59
TIT_2	0,244	0,74	-0,265	0,59	1,178 **	2,14
TIT_3	0,663 **	2,01	0,258	0,59	1,557 ***	2,85
HABAD1_E	0,483	1,39	0,825 *	1,75	0,003	0,01
HABAD1_I	0,005	0,03	-0,315	1,57	0,315 *	1,72
HABAD1_N	0,117	0,96	0,039	0,23	0,276	1,58
CONADE_S	0,261 **	2,22	0,343 **	2,03	0,144	0,83
CONADE_I	-0,282 **	2,09	-0,418 **	2,04	-0,110	0,61
SATSUEL1	0,017	0,16	0,116	0,78	-0,105	0,68
SATESTA1	-0,738 ***	6,89	-0,788 ***	5,12	-0,606 ***	3,92
SATPROM1	-0,536 ***	4,49	-0,436 **	2,51	-0,696 ***	4,24
SATRIES1	-0,107	0,93	-0,435 ***	2,72	0,294 *	1,75
SATHORA1	-0,579 ***	5,39	-0,587 ***	3,95	-0,576 ***	3,62
$\alpha$	1,573 ***	40,45	1,619 ***	27,81	1,616 ***	28,06
$\sigma^2$	4,567 ***	7,92	6,011 ***	4,98	3,357 ***	5,20
Chi <sup>2</sup>	904,750		1844,940			
Log. Verosimilitud	-3741,227		-2053,006		-1609,245	
Nº observaciones	2020		1090		930	

Estadísticos z robustos en valor absoluto

\* significativo al 10%; \*\* significativo al 5%; \*\*\* significativo al 1%

En primer lugar, en lo que respecta al primer bloque de variables explicativas (*características personales y familiares*), observamos que, para el conjunto de la muestra, sólo resultan significativas las variables que hacen referencia a la experiencia previa y al número de personas que trabajan en la unidad familiar. La primera de estas covariables, con coeficiente positivo, nos indica que la tasa de riesgo de finalización del empleo aumenta cuando el titulado manifiesta tener algún tipo de experiencia laboral previa. Esto puede deberse a dos motivos diferentes. En primer lugar, es posible que el titulado posea un nivel de información del mercado laboral superior al de cualquier otro titulado sin experiencia previa, por lo que, probablemente, su nivel de exigencia será mayor. También es posible que, aquellos titulados que manifiesten tener experiencia previa (EXPPREV), presenten mayores dificultades de adaptación a un nuevo puesto de trabajo, puesto que se incorporan al nuevo puesto con determinados hábitos, de los que

resulta difícil deshacerse y que, pueden dificultar el adecuado desempeño de las labores del nuevo puesto de trabajo.

Por otro lado, el número de miembros de la unidad familiar que están trabajando (NTRABJ1), con coeficiente negativo, indican una menor tasa de riesgo de finalización cuanto mayor sea este número. Esta variable puede estar actuando como *proxy* de la clase social del titulado, puesto que, cuando en una misma familia son muchos los miembros ocupados, suele ser porque la aportación económica del cabeza de familia no resulta suficiente para mantener a la unidad familiar. Por tanto, resultan necesarias las aportaciones económicas de otros miembros de la misma, por lo que la duración del período de empleo de este tipo de titulados, será más alta que la de aquellos otros que no tengan la ‘necesidad’ real de permanecer en él.

El comportamiento de estas dos covariables se mantiene igual en cada uno de los dos colectivos en los que se ha desagregado la muestra. Sin embargo, en la estimación realizada para el colectivo de tituladas, observamos que, la variable que recoge la influencia de la edad, toma significación, apuntando hacia una mayor duración en el primer empleo, cuando mayor sea la edad de la titulada al comenzar a trabajar en él. Entre los hombres, sin embargo, la edad no condiciona la duración en el empleo. Este comportamiento resulta lógico, puesto que, cuanto mayor sea una mujer, menor probabilidad tendrá de conseguir un nuevo empleo (Gámez-Amián y García-Pérez, 2002), por lo que preferirá, en caso de ser posible, permanecer en el que ya posee.

En lo que respecta a las *características formativas*, observamos la falta de significatividad estadística de la variable que recoge el nivel del Ciclo cursado (NIVELCF), por lo que no podemos hablar de una influencia clara del nivel formativo del titulado en la duración de su primer empleo. Bratberg y Nilsen (1988) llegan a este mismo resultado en su estudio para la población ocupada de Noruega. No obstante, sí que se observan diferencias en función de la Familia Profesional a la que pertenezca el Ciclo Formativo cursado. Así, aquellos titulados que hayan realizado un Ciclo de la Familia de ‘Comercio y Marketing’, ‘Comunicación, Imagen y Sonido’ u ‘Hostelería y Turismo’, presentan una menor duración de su primer empleo, en comparación con los titulados de ‘Administración’, mientras que los que han cursado un Ciclo de la Familia de ‘Madera y Mueble’ o ‘Sanidad’ presentan una mayor duración de su primer empleo que los de ‘Administración’.

Diferenciando por género, se puede apreciar la clara segmentación del mercado laboral en puestos de trabajo eminentemente femeninos y puestos de trabajo claramente masculinos, hasta el punto de que, quienes hayan cursado un Ciclo Formativo ‘atípico’ para su género, finalmente conseguirán permanecer menos tiempo en su primer empleo. Así, por ejemplo, observamos que las tituladas de ‘Mantenimiento y Servicios a la Producción’, ‘Industrias Alimentarias’ y ‘Fabricación Mecánica’ presentan una duración

en su primer empleo claramente inferior a las tituladas en Administración<sup>141</sup>, al igual que sucede con los titulados en 'Artes Gráficas' e 'Imagen Personal'.

Adentrándonos en el análisis de las *variables de entorno*, observamos que, al igual que sucedió en el estudio de Ibáñez-Pascual (1999), el sector de actividad resulta sólo significativo en el caso de las mujeres, mostrando además una tasa de riesgo de finalización del empleo mayor para las que trabajan en los sectores secundario y terciario, en comparación con las que trabajan en el primario. Espinal-Berenguer (1992) detectó igualmente una mayor probabilidad de cambio de empleo entre los que trabajaban en los sectores de la Construcción y de la Hostelería.

El resto de las variables de entorno tienen influencia sobre la duración del primer empleo, únicamente en el caso de los varones. Para este colectivo se observa una tasa de riesgo de finalización creciente con la tasa de desempleo juvenil, así como con el año de consecución del primer empleo. Parece claro que la duración de los primeros empleos resulta ser cada vez menor en comparación con la duración de los empleos conseguidos en 1994.

En cuanto a las variables recogidas en el bloque de *características del empleo* observamos, en primer lugar, que, al igual que en el estudio de Ibáñez-Pascual (1999), el medio de búsqueda efectivo influye únicamente sobre la duración del primer empleo en el caso de los hombres. Sin embargo, el efecto resulta ser justo el inverso, puesto que, en nuestro caso, conseguir el primer empleo a través de un medio formal, conlleva, finalmente, una menor duración del primer empleo.

En el colectivo global de la muestra se observa que aquellos que consiguieron un primer empleo como asalariados en una empresa privada, presentan una mayor tasa de riesgo de finalización del mismo que los que consiguieron su primer empleo en una empresa pública, siendo la diferencia mayor cuanto menor es el tamaño de la empresa privada<sup>142</sup>. En el colectivo de las mujeres, las diferencias sólo son apreciables cuando se trata de empleadas de PYMES.

El tipo de contrato resulta ser significativo a la hora de explicar las diferencias en la duración del primer empleo. En primer lugar observamos, en el colectivo global de la muestra, que, excepto los que trabajan por cuenta propia y los que trabajan como ayuda familiar, cuyos coeficientes no son estadísticamente significativos, todos los demás tipos de contratos conllevan una menor duración del primer período de empleo, en

---

<sup>141</sup> Para obtener la magnitud de su influencia habría que calcular  $(e^{\beta} - 1) \cdot 100$ , lo que nos proporcionaría el incremento porcentual que sufriría la tasa de riesgo de abandono del empleo.

<sup>142</sup> Dado que en las grandes empresas su tamaño permite cambios de puestos u ocupaciones dentro de la organización (especialmente movilidad ascendente) ésta es otra de las razones por las que se observa una menor rotación (movilidad externa) en estas empresas (Mincer, 1993).

comparación con los contratos indefinidos. Sin embargo, resulta llamativo que, en el caso de las mujeres, cobre significatividad la ayuda familiar como relación contractual, de manera que, finalmente, conlleva una mayor estabilidad en el empleo el trabajar como ayuda familiar, que la consecución de un contrato indefinido.

Por último, en cuanto a las características del empleo, observamos que un aumento en el sueldo horario provoca una disminución en la tasa de riesgo de salida en el colectivo global de la muestra, así como en el de las mujeres. Sin embargo, este efecto no resulta significativo entre los hombres. Nilsen y Risa (1999) detectaron una influencia del salario sobre la duración del empleo en el mismo sentido que nosotros, aunque con significatividad estadística en ambos colectivos.

El último bloque de covariables (*atributos laborales*) nos muestra, al igual que sucedió en el estudio de Bratberg y Nilsen (1998), una ligera influencia negativa de la duración del período de búsqueda sobre la tasa de salida del empleo. No obstante, este efecto no resulta significativo en el caso de los hombres. De la misma manera, observamos una influencia negativa sobre dicha tasa de salida, del tiempo necesario para la eficiente realización de las tareas laborales.

Por otro lado, la estimación de las variables de control que hacen referencia al grupo de pertenencia del titulado, nos muestra que, para el colectivo de hombres, la duración del primer empleo resulta ser inferior en el caso de los '*titulados de los grupos 2 y 3*' en comparación con los '*titulados del grupo 1*', mientras que, en el caso de las mujeres, no resultan ser estadísticamente significativas.

Las variables de desajuste educativo aportan evidencias que apoyan las predicciones de las teorías del '*job matching*' y de la '*movilidad ocupacional*', permitiéndonos considerar el desajuste como un fenómeno transitorio, tal y como señala la teoría neoclásica. De hecho, la existencia de desajuste de habilidades por exceso, en el caso de las mujeres, o por defecto, en el de los hombres, conlleva una menor duración del primer empleo, mientras que las mujeres sobreeducadas muestran una menor duración de su primer empleo, así como las subeducadas muestran una mayor permanencia en el mismo (siempre en comparación con las adecuadamente educadas).

En este último bloque de variables, introducimos finalmente aquellas covariables que nos hablan del grado de satisfacción de los titulados con su puesto de trabajo. Excepto la variable que recoge el grado de satisfacción con el sueldo, el resto de ellas resultan significativas en las tres regresiones realizadas, mostrando una mayor duración en el empleo, curiosamente, cuando el individuo muestra una mayor insatisfacción con él.

Finalmente observamos que, el coeficiente que recoge el efecto de la heterogeneidad inobservable,  $\sigma^2$ , resulta ser significativo, tanto en el conjunto de la muestra, como en la estimación diferenciada por género, tomando un valor más alto en la estimación del

modelo de duración de las mujeres. Este resultado apunta hacia una mayor heterogeneidad en este último colectivo.

## **5.2. DURACIÓN DEL ÚLTIMO EMPLEO**

Finalmente, centramos nuestra atención en la duración del último episodio de ocupación de los titulados que componen nuestra muestra. Concretamente nos centraremos en el colectivo compuesto por aquellos '*titulados de los grupos 1, 2 y 3*', que han sufrido al menos un episodio de movilidad laboral y que, por tanto, han disfrutado de más de un puesto de trabajo tras la realización del Ciclo Formativo. Para ellos analizaremos el tiempo que permanecieron en su último puesto de trabajo. Éste es el caso de 1220 titulados, 30 de los cuales, no facilitaron la fecha de comienzo en su último puesto de trabajo, por lo que no hemos podido reconstruir la duración de su último período de ocupación y, por tanto, hemos tenido que omitirlos en el presente análisis, quedándonos finalmente con una muestra de 1190 titulados.

De estos 1190 titulados, 950 estaban aún trabajando en su último empleo en el momento de responder a la encuesta, por lo que su duración estará censurada por la derecha, mientras que 240 se encontraban desempleados o inactivos, por lo que disponemos de la duración completa de su último período de ocupación. Nos enfrentamos, por tanto, a una muestra que adolece de un problema de censura en el 79.8% de sus casos.

En el capítulo segundo ya dedicamos un apartado al análisis descriptivo del tiempo que los titulados permanecieron en su último empleo previo a la realización de la encuesta. A partir de los resultados obtenidos entonces, comprobamos que las mujeres permanecían en torno a dos semanas menos que los hombres en su último puesto de trabajo; que la permanencia de los titulados de grado superior en el último puesto de trabajo era superior a la de los titulados de grado medio en más de un mes; que los titulados de centros públicos mostraban permanencias superiores a los titulados de centros privados, que llegaban a superar el mes y medio; y, finalmente, que las Familias Profesionales cuyos titulados mostraban permanencias superiores en su último empleo eran las de '*Actividades Marítimo Pesqueras*' (14 meses) e '*Informática*' (13 meses y medio), frente a los titulados de '*Madera y Mueble*' (5 meses y medio) y '*Actividades Agrarias*' (6 meses), que eran quienes presentaban una menor duración en su último empleo.

A continuación, en el próximo epígrafe, centraremos aún más el estudio de estos datos de duración, mediante la utilización de herramientas paramétricas y no paramétricas. Éstas últimas serán, justamente, las que nos ayuden a determinar qué factores pueden estar influyendo en que la duración observada en el último período de empleo sea mayor o menor, sin olvidar la realidad patente en nuestros datos de la existencia de censura.

### 5.2.1. Estimación no paramétrica

Comenzamos el estudio de la duración del último episodio de ocupación de nuestros titulados mediante la aproximación no paramétrica y, más concretamente, mediante la obtención de los estimadores de Kaplan y Meier (1958). La representación gráfica de dichos estimadores viene recogida en la Ilustración IV.19, a través de la cual reflejamos las funciones empíricas de densidad, distribución y riesgo sin agregar previamente los datos en intervalos. De esta manera observamos cómo, las duraciones manifestadas por los egresados varían desde los 11 días y a los poco más de cuatro años.

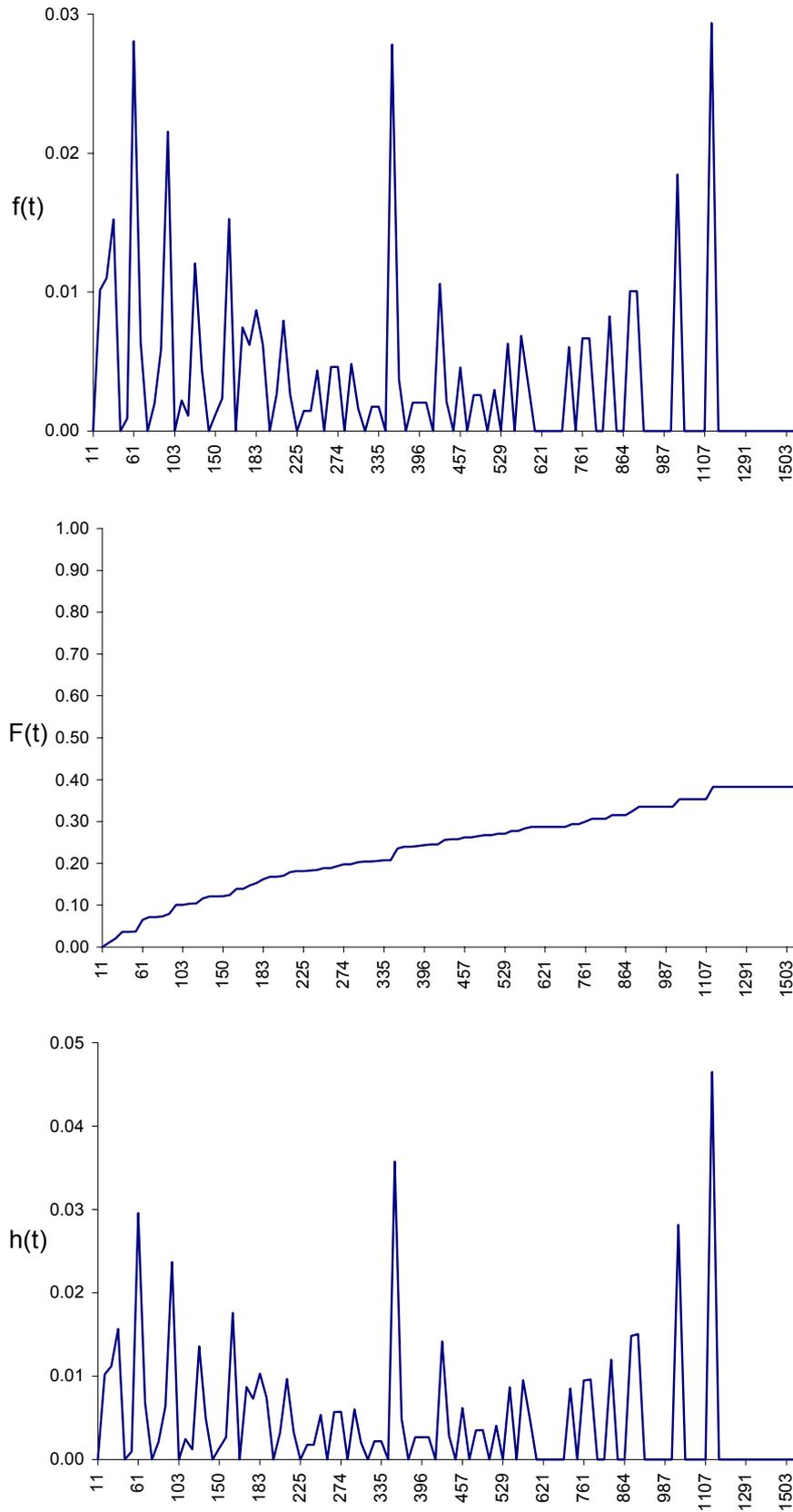
La primera gráfica de la Ilustración IV.19 es la de la función de densidad, de la que se desprende que las duraciones del último empleo más repetidas en la muestra, son las de 61 días (2.8%), 92 días (2.2%), 365 días (2.8%) y 1127 días (2.9%). No obstante, dado que no hemos agregado los datos en intervalos y que las duraciones las hemos obtenido mediante la diferencia entre la fecha de inicio y finalización del contrato<sup>143</sup>, la densidad de las duraciones contractuales de un mes puede quedar escondida en la gráfica a través de la desagregación de la misma entre duraciones de 28 a 31 días. De la misma manera, las duraciones contractuales de dos meses aparecerán desagregadas en la gráfica entre las duraciones de 59 a 62 días, al igual que las de 3 meses se diluyen entre los 89 y los 92 días. Por tanto, tomando en consideración éste aspecto, podemos afirmar que las duraciones más repetidas en la muestra son las de 1 mes (2.6%), 2 meses (3.5%), 3 meses (3%), 6 meses (2.8%), 1 año (2.8%) y 3 años (2.9%).

La segunda gráfica de la Ilustración IV.19 recoge la función de distribución de la duración en el último empleo. Como vemos, ésta no llega a alcanzar el 100%, sino que se mantiene en un máximo del 38%, por lo que los estimadores de Kaplan-Meier nos hablan de una proporción de supervivencia en el último puesto de trabajo de en torno al 62%. De la observación más detenida de esta función de distribución podemos entresacar que el 16.2% de la muestra permaneció en su último empleo un máximo de medio año, que el 23.6% de la misma no permaneció más de un año y, finalmente, que el 29.3% de la muestra abandonó su último empleo antes de transcurridos 2 años.

Por último, la tercera gráfica de la Ilustración IV.19 nos muestra, mediante la función de riesgo empírica, la probabilidad condicionada de finalización del último empleo en un plazo muy corto de tiempo, dado que el individuo ya ha llegado a una duración  $t$ . Observamos así, que los períodos de mayor riesgo de abandono del empleo se corresponden, precisamente, con los períodos de mayor densidad muestral, es decir con los períodos de 61 días (3%), 92 días (2.4%), 365 días (3.6%) y 1127 días (4.7%).

<sup>143</sup> O la diferencia entre la fecha de inicio y la fecha de realización de la encuesta, para aquellas observaciones con censura.

Ilustración IV.19: Funciones empíricas de densidad, distribución y riesgo de la duración en el último empleo (Kaplan-Meier)



Una vez finalizado el análisis de los estimadores de Kaplan-Meier, pasamos al análisis de los datos de duración mediante la metodología actuarial. Para ello hemos agrupado los datos en 18 intervalos abiertos por la izquierda y cerrados por la derecha. La amplitud de los dos primeros intervalos es de una semana y quince días, respectivamente. A partir de ahí, y hasta llegar al año, establecimos intervalos de un mes. Finalmente, construimos intervalos de tres meses, para recoger las duraciones que oscilaban entre uno y tres años, mientras que el último intervalo abierto, recoge las duraciones superiores a los tres años.

Una vez agrupados los datos en intervalos obtuvimos la tabla de vida que figura en la Tabla IV.13. Como se puede observar, en la inmensa mayoría de los intervalos es mayor el número de observaciones censuradas que el de observaciones completas, por lo que la estimación de la proporción de sucesos terminales 'q' es bastante baja a lo largo de toda la muestra. De hecho, los intervalos que encierran una mayor probabilidad de que una observación que entra en el intervalo finalice dentro del mismo, presentan todos una probabilidad inferior al 10%. Estos son, concretamente, los intervalos que van de 2 a 3 meses (4%), de 1 año a 1 año y medio (9.1%), de 2 años a 2 años y medio (6.9%) y el intervalo abierto para las duraciones superiores a los tres años (9.1%).

Como es lógico, los valores estimados de la probabilidad de supervivencia 'p' son, en todos los intervalos, superiores al 90%, ya que estos no son más que la función complementaria de 'q'. A raíz de esos valores tan altos de 'p', la estimación de la función de supervivencia para el conjunto de la muestra presenta también valores altos. De hecho, la probabilidad de supervivencia al final del intervalo 330 a 360, aún es del 78.5%, indicándonos así cuál es la proporción de titulados que permanecen en su último empleo durante un período superior al año.

Al proporcionar la función de supervivencia valores tan altos en todos los intervalos considerados, la disminución de un intervalo al siguiente es bastante baja, por lo que, al calcular la densidad de probabilidad, obtenemos que la probabilidad diaria de finalización de la duración del último empleo alcanza escasamente el 0.1%. De la misma manera, la estimación de la función de riesgo, en cada instante de tiempo considerado, proporciona probabilidades condicionadas de finalización instantánea de la duración del último empleo, inferiores al 0.14%.

Tabla IV.13: Tabla de vida de la duración en el último empleo

Intervalo	Total	Finalizan	Censura	q	p	Función de Supervivencia	f	Función de Distribución	Función de Riesgo
7	15	1190	0	11	0,000	1,000	1,000	0,000	0,0000
15	30	1179	12	0	0,010	0,990	0,990	0,001	0,0007
30	60	1167	31	74	0,027	0,973	0,963	0,001	0,0009
60	90	1062	41	78	0,040	0,960	0,924	0,001	0,0014
90	120	943	28	98	0,031	0,969	0,895	0,001	0,0011
120	150	817	18	53	0,023	0,977	0,875	0,001	0,0008
150	180	746	16	40	0,022	0,978	0,855	0,001	0,0007
180	210	690	23	42	0,034	0,966	0,826	0,001	0,0012
210	240	625	10	54	0,017	0,983	0,812	0,000	0,0006
240	270	561	5	30	0,009	0,991	0,805	0,000	0,0003
270	300	526	6	24	0,012	0,988	0,795	0,000	0,0004
300	330	496	4	38	0,008	0,992	0,789	0,000	0,0003
330	360	454	2	28	0,005	0,995	0,785	0,000	0,0002
360	540	424	31	164	0,091	0,909	0,714	0,000	0,0005
540	720	229	5	109	0,029	0,971	0,694	0,000	0,0002
720	900	115	6	55	0,069	0,931	0,646	0,000	0,0004
900	1080	54	1	32	0,026	0,974	0,629	0,000	0,0001
1080	.	21	1	20	0,091	0,909	0,572	.	0,428

Nuevamente hemos optado por realizar la representación gráfica de la tabla de supervivencia, diferenciando la muestra en función de diversos colectivos. De esta manera podemos ir anticipando si la duración en el último empleo diverge entre los mismos, con lo cual podremos detectar, qué posibles variables afectan a la mayor o menor duración del último episodio de empleo.

La Ilustración IV.20 muestra así las funciones de riesgo, distribución y supervivencia en el último empleo, diferenciando la muestra entre hombres y mujeres. La Ilustración IV.21 recoge estas mismas tres funciones, pero distinguiendo entre Ciclo Formativo de Grado Medio y de Grado Superior. La Ilustración IV.22 hace la distinción, en función de la titularidad del centro de estudios, mientras que, la Ilustración IV.23 diferencia estas tres funciones para cada una de las 21 Familias Profesionales.

De la mera observación de las tres primeras ilustraciones, parece claro que, en líneas generales, la muestra presenta una mayor duración en su último puesto de trabajo por parte de los hombres que de las mujeres, al igual que ésta es superior en el caso de los técnicos superiores que en el de los técnicos, y en el de los titulados de centros públicos que en el de los privados.

Ilustración IV.20: Función de riesgo, de distribución y de supervivencia en el último empleo por género

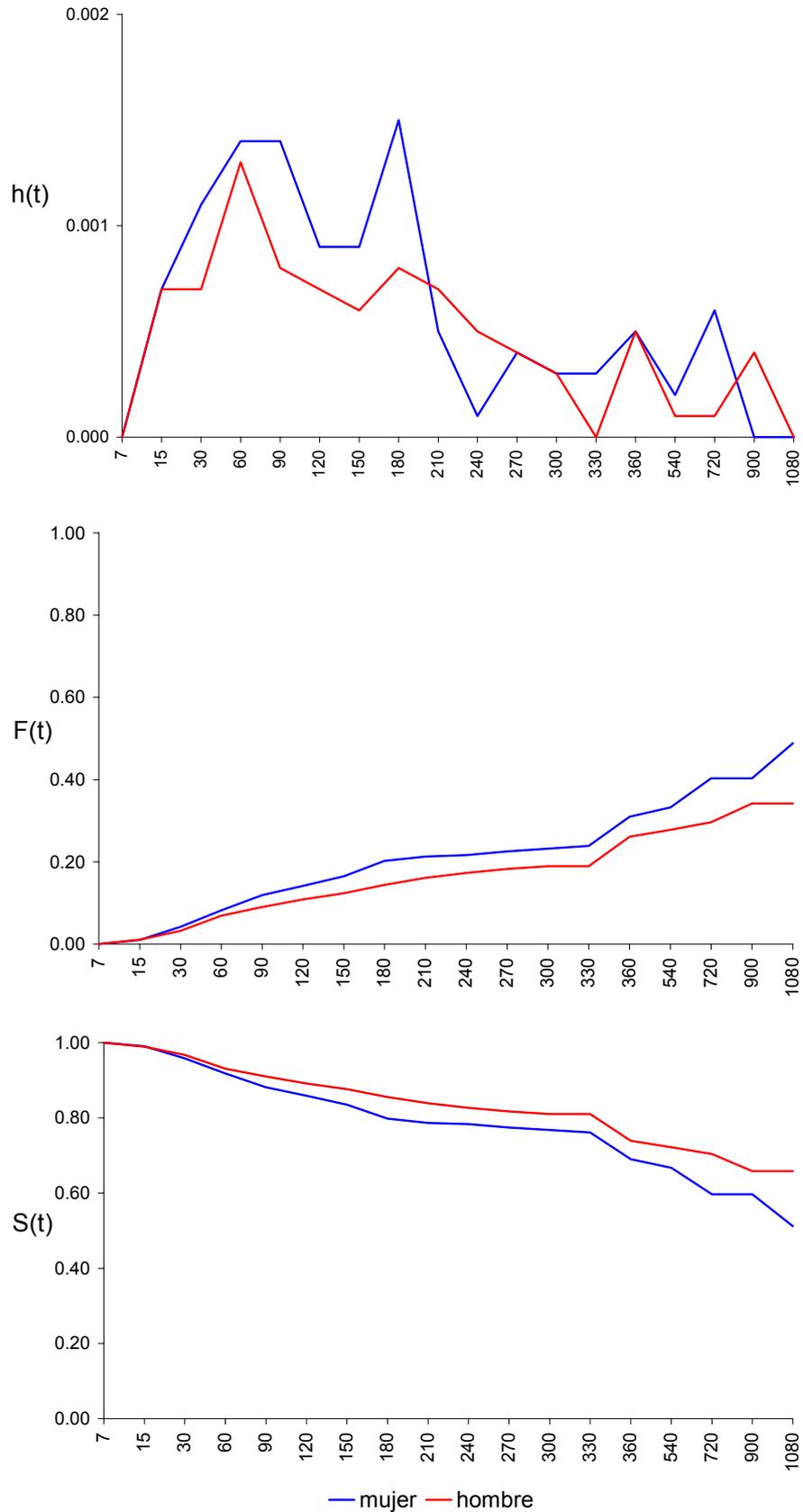


Ilustración IV.21: Función de riesgo, de distribución y de supervivencia en el último empleo por nivel del CF

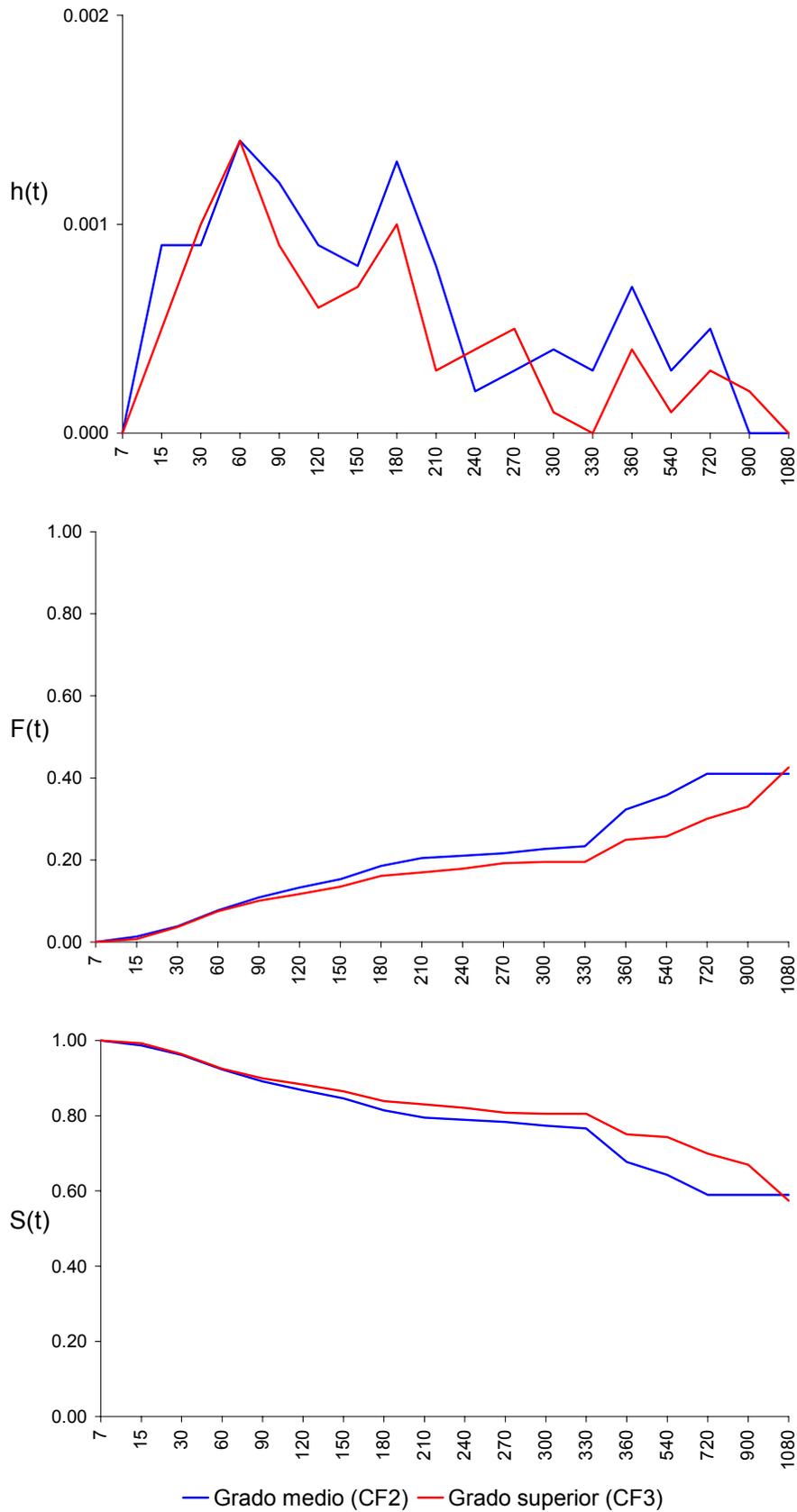


Ilustración IV.22: Función de riesgo, de distribución y de supervivencia en el último empleo por titularidad del centro

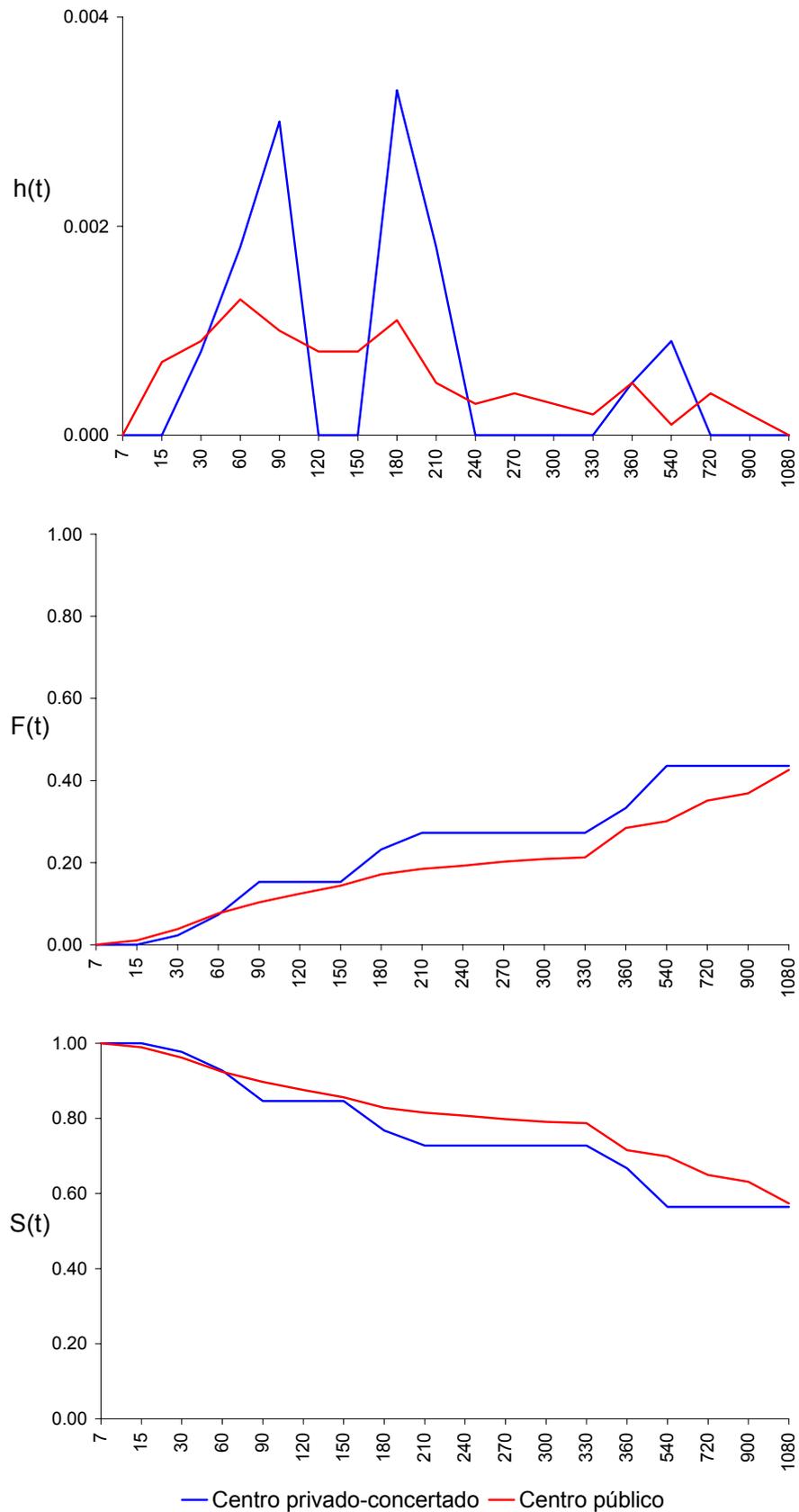
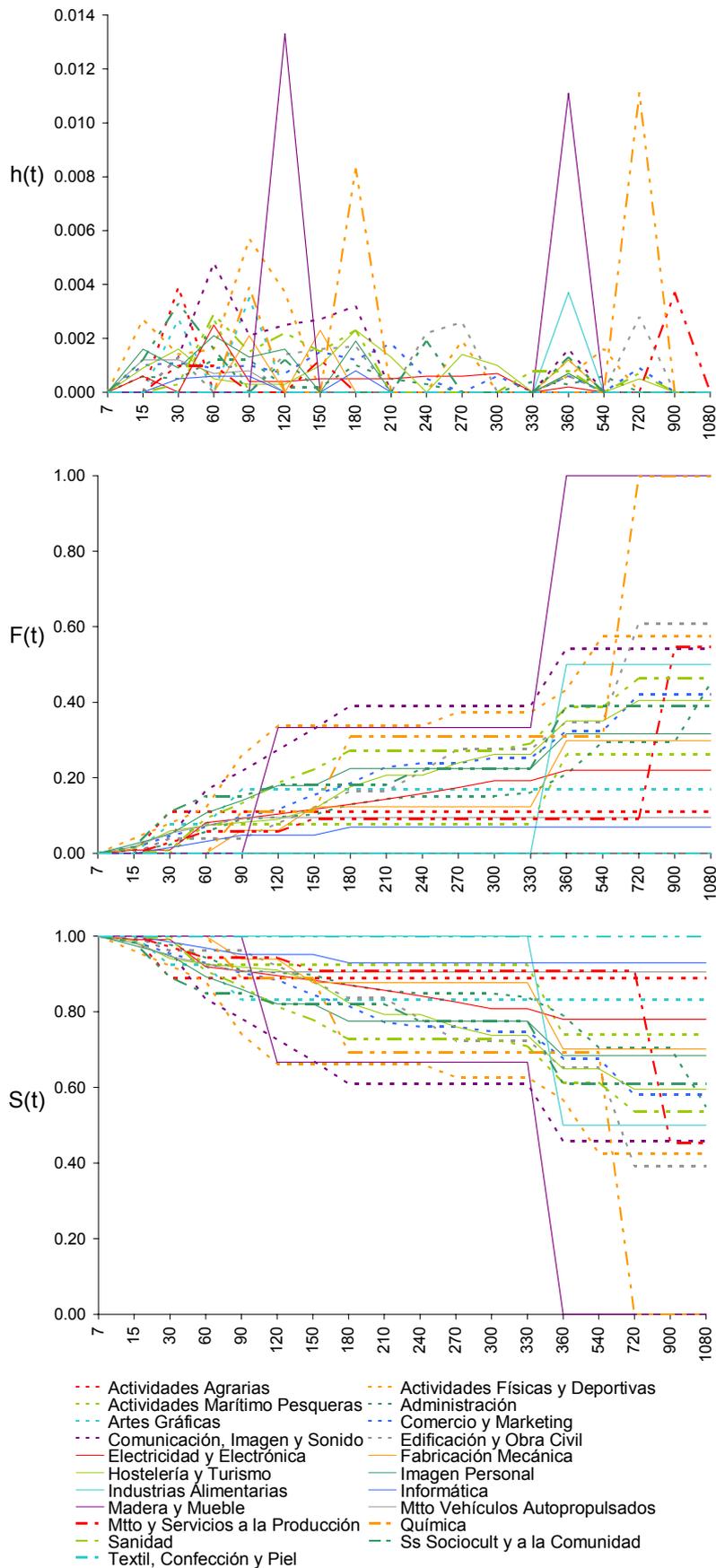


Ilustración IV.23: Función de riesgo, de distribución y de supervivencia en el último empleo por familias profesionales



Aún así, hemos optado por realizar nuevamente el contraste de homogeneidad de 'Cox' (StataCorp., 2001), con la finalidad de contrastar la existencia de diferencias significativas en la función de riesgo de cada uno de estos subgrupos. Los resultados de dicho contraste vienen recogidos en la Tabla IV.14., donde vemos que no existen evidencias que nos lleven a rechazar la hipótesis de igualdad de funciones de riesgo en el caso de estratificación de la muestra por género y titularidad del centro. Por otro lado, en la estratificación por nivel del Ciclo Formativo y por Familia Profesional, vuelven a detectarse diferencias estadísticamente significativas. Por tanto, a continuación nos centraremos exclusivamente en el análisis de la Ilustración IV.21 y de la Ilustración IV.23., puesto que son éstas las que recogen las funciones de riesgo, distribución y supervivencia, diferenciando por nivel del Ciclo Formativo y por Familia Profesional.

**Tabla IV.14: Contrastes de homogeneidad de estratos para el análisis de la duración en el último empleo**

Estratos	Estadístico	Probabilidad
Género	2,67	0,1022
Nivel del Ciclo Formativo	3,56	0,0591
Titularidad del centro	0,95	0,3289
Familias Profesionales	1157,47	0,0000

En lo que respecta a la diferenciación por nivel del Ciclo Formativo, de la primera gráfica de la Ilustración IV.21 deducimos que, a pesar de que en ambos colectivos la función de riesgo toma valores extremadamente bajos<sup>144</sup>, estos son mayoritariamente superiores entre los técnicos. Por lo que estos presentan una mayor probabilidad condicionada de abandono del último empleo que los técnicos superiores. De hecho, al observar la función de distribución (de supervivencia) de los técnicos, comprobamos que ésta, a partir de las duraciones superiores a los dos meses, se mantiene continuamente por encima (por debajo) de la función de distribución (supervivencia) de los técnicos superiores. Las diferencias entre las funciones de ambos colectivos se acentúan para aquellas duraciones superiores a los 11 meses, reduciéndose finalmente las mismas, para aquellos titulados, cuya duración en el último empleo, ha sido de 3 o más años.

Centrándonos, por último, en la Ilustración IV.23, observamos que la mayoría de las Familias Profesionales presentan una función empírica de riesgo cuyos valores no llegan, en ningún intervalo, al 0.2%<sup>144</sup>. No obstante, también nos encontramos con el caso de los titulados en 'Madera y Mueble', quienes, en el intervalo de 4 a 5 meses, presentan una probabilidad condicionada de salida del empleo del 1.3%, y en el intervalo de 1 año a 1 año y medio, presentan una probabilidad condicionada del 1.1%. Otra Familia Profesional que destaca, en cuanto a los valores de la función de riesgo

<sup>144</sup> No perdamos de vista que el máximo valor que toma la función de riesgo en el conjunto de la muestra es del 0.14%.

empírica de sus titulados es la de 'Química', puesto que presenta valores del 0.8% y 1.1% para los intervalos de 6 a 7 meses y de 2 a 2 años y medio respectivamente.

Precisamente, los titulados de la Familia de 'Madera y Mueble', junto con los de 'Comunicación, Imagen y Sonido' y los de 'Actividades Físicas y Deportivas', son los que parecen presentar una menor duración en su último empleo. De hecho, el 67% de ellos permanecieron en su último empleo un mínimo de 6 meses; el 67% de los titulados de 'Madera y Mueble', el 63% de los titulados de 'Actividades Físicas y Deportivas' y el 60% de los titulados de 'Comunicación, Imagen y Sonido', consiguieron permanecer al menos 1 año; y el 46% de los de 'Comunicación, Imagen y Sonido', el 42% de los de 'Actividades Físicas y Deportivas' y el 0% de los de 'Madera y Mueble' consiguieron permanecer un mínimo de 2 años. Justo en la situación contraria se encuentran los titulados de 'Informática', 'Textil, Confección y Piel' e 'Industrias Alimentarias', puesto que el 100% de los titulados de estas dos últimas Familias Profesionales, así como el 93% de los titulados de 'Informática', consiguió permanecer en su último empleo un mínimo de 2 años.

### 5.2.2. Estimación paramétrica

Tras el análisis no paramétrico de la duración del último episodio de ocupación de los titulados, pasamos a la detección de los factores que influyen en la mayor o menor duración de dicho episodio. No obstante, hemos de tener en cuenta en todo momento el grave problema de censura del que adolecen nuestros datos<sup>145</sup>, puesto que esta realidad va a condicionar sobremanera la calidad de nuestras estimaciones.

En este caso, la variable dependiente a estudiar es la duración del último episodio de ocupación de aquellos técnicos y técnicos superiores que han disfrutado de más de un empleo. Esta duración se ha obtenido, para los titulados que al responder a la encuesta estaban desempleados, como diferencia entre la fecha de inicio y finalización en su último puesto de trabajo; mientras que, en el caso de los titulados que al responder a la encuesta aún estaban ocupados, se obtuvo como diferencia entre la fecha de inicio y la fecha de realización de la encuesta.

Las aportaciones teóricas, en cuanto a las covariables que deberíamos considerar en la especificación de este nuevo modelo de duración, no difieren de las comentadas en el estudio de la duración del primer período de empleo. Por lo tanto, teniendo en cuenta esto, las variables explicativas que decidimos introducir en el mismo, resultan ser prácticamente las mismas, aunque referidas todas al último período de ocupación. Las únicas diferencias, con respecto a las variables comentadas en el apartado 5.1.2., son las siguientes:

---

<sup>145</sup> Recordemos que el 79.8% de la muestra presenta problemas de censura.

1. En el bloque de *características personales y familiares*:

- Las variables ya no se refieren a la situación personal del encuestado en el momento de finalizar sus estudios, sino en el momento de responder a la encuesta.
- En esta ocasión no incluimos la variable dicotómica EXPPREV, puesto que todos los titulados considerados en la presente estimación disponen de experiencia previa.
- No podemos introducir una variable que recoja los ingresos del hogar, puesto que sólo conocemos dicha información para el momento de finalización de los estudios.
- Introducimos una nueva variable dicotómica que toma el valor 1 en el caso de que el titulado se encuentre aún viviendo con sus padres, y el valor 0 en otro caso (CONVIV2)<sup>146</sup>.

2. En el bloque de *atributos laborales*:

- No diferenciamos el colectivo al que pertenece el titulado, puesto que todos ellos han obtenido su último empleo tras la finalización del CF y, por tanto, la valoración que puedan hacer acerca del grado de ajuste de la formación recibida a las necesidades del puesto es igualmente válida para todos ellos.
- No podemos introducir la variable de desajuste de habilidades, puesto que no planteamos esa pregunta en el último empleo.
- Introducimos una nueva variable dicotómica que recoge la posibilidad de que el titulado, a pesar de estar trabajando, esté buscando un nuevo empleo (BUSCAEMP).

La función de riesgo estimada de forma no paramétrica (Ilustración IV.21), mostraba al comienzo un perfil creciente, a partir de cuyo máximo, comenzaba a decrecer. De ahí que consideremos la estimación de un modelo log-normal y de otro log-logístico. No obstante, la observación de la estimación no paramétrica recogida en la Ilustración IV.20, podría llevarnos a considerar un decrecimiento monótono inicial de la función de riesgo, por lo que también optamos por realizar la estimación Gompertz y la Weibull. Finalmente, viendo que los márgenes de oscilación de dicha función de riesgo son tan reducidos, nos planteamos la posibilidad de estimar un modelo con una tasa de riesgo constante, por lo que decidimos estimar un modelo Exponencial y otro Gamma Generalizado<sup>147</sup>.

<sup>146</sup> Esta variable también fue introducida inicialmente en la estimación de la duración del primer empleo, aunque referida a la situación del titulado al finalizar el CF, pero hubo de ser eliminada puesto que provocaba problemas de colinealidad

<sup>147</sup> Como vimos en el apartado 3.2.3.3. del presente capítulo, la distribución Gamma generalizada es tan flexible que permite casi cualquier comportamiento de la función de riesgo, incluida la de tasa de riesgo constante.

En primer lugar, tal y como hemos hecho en el caso de los anteriores modelos de duración, nos planteamos la posibilidad de existencia de heterogeneidad inobservable. En este caso, a diferencia de los anteriores, todos los modelos presentaban una probabilidad asociada al estadístico de contraste del 100%, lo que nos llevaba a la aceptación de la hipótesis nula de homogeneidad en la muestra.

Es más que probable que este resultado se deba a la alta incidencia de censura en la muestra, puesto que esto mismo sucede en el estudio que realizan García-Serrano y Malo-Ocaña (1996) con la Encuesta de Estructura, Conciencia y Biografía de Clase realizada en España en 1991. En su estimación de la longitud del episodio de ocupación estos autores trabajan con una muestra que presenta censura en el 90% de los datos, lo que les lleva a afirmar que *“esto conlleva a un problema en la identificación de la relación de dependencia y en el efecto de la heterogeneidad”*.

Al estimar los seis modelos propuestos, sin contemplar la presencia de heterogeneidad, con la finalidad de analizar las diferentes medidas de bondad del ajuste obtenemos los resultados recogidos en la Tabla IV.15.

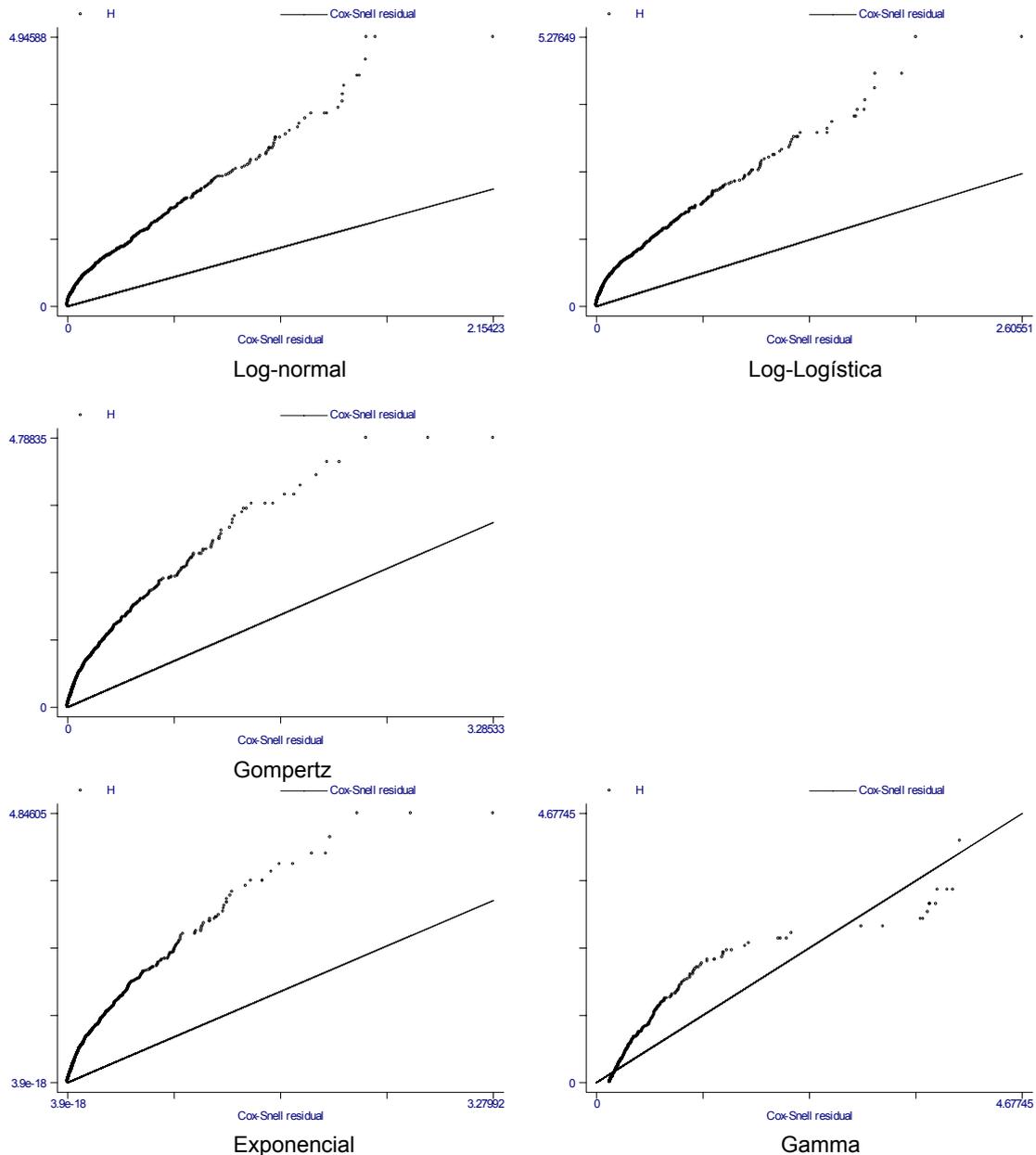
**Tabla IV.15: Medidas de bondad del ajuste de diversos modelos de duración para el último período de empleo**

Distribución	Chi <sup>2</sup> <sub>58</sub>	Log. Verosimilitud	AIC	Nº observ.
Log-Normal	-	-1296,9408	2715,881	1027
Log-Logística	-	-1287,9328	2697,866	1027
Gompertz	-	-1288,7751	2699,550	1027
Weibull	-	-	-	1027
Exponencial	6772,59 p=0,000	-788,6259	1699,252	1027
Gamma	-	-723,6865	1569,373	1027

En la mayoría de los casos, la existencia de una censura tan alta conlleva que las estimaciones no proporcionen los resultados de los contrastes de razón de verosimilitud Chi-cuadrado. No obstante, a excepción de la distribución Weibull, que no pudo ser estimada, el resto proporcionaba una estimación del logaritmo de verosimilitud a partir del cual poder obtener el valor del Criterio de Información de Akaike (AIC). A la vista de sus resultados, la mejor estimación resulta ser la Gamma generalizada, puesto que es la que proporciona un menor valor de AIC.

Graficando la función integrada de riesgo de los residuos generalizados de Cox-Snell (Ilustración IV.24), detectamos que, como ya advertían García-Serrano y Malo-Ocaña (1996), el ajuste de los modelos, como era previsible, no resulta ser demasiado bueno. Aún así, nuevamente es la estimación Gamma la que parece presentar un mejor comportamiento de los residuos, por lo que finalmente nos decantamos por la elección de dicho modelo.

**Ilustración IV.24: Función integrada de riesgo de los residuos generalizados de Cox-Snell para la estimación de la duración del último empleo**



Finalmente, en la Tabla IV.16 recogemos los resultados de la estimación del modelo Gamma generalizado tanto para el colectivo global de la muestra seleccionada, como diferenciando la misma entre hombres y mujeres.

Como vemos, el problema de censura de la muestra nos lleva a la obtención de estimaciones para las cuales, en muchos casos, no resulta posible la obtención de un estadístico de contraste que nos permita hablar de su significatividad. Por ello, las conclusiones a las que lleguemos a partir de esta estimación, deben ser tomadas con cautela.

Tabla IV.16: Estimación del modelo de duración tipo Gamma generalizado para la duración del último empleo, diferenciando por género

	Todos		Mujeres		Hombres	
	Coef.	z	Coef.	z	Coef.	z
(Constante)	29,444	(.)	19,553	(.)	5,588 *	1,67
<b>CARACTERÍSTICAS PERSONALES Y FAMILIARES:</b> (Referencia: mujer, no convive con sus padres, sin hijos, cabeza de familia desempleado)						
SEXO	-0,203	1,26				
EDAD4	0,233 ***	345,86	0,209	(.)	-0,008	0,06
EDAD42	-0,350	(.)	-0,305	(.)	0,202	0,92
CONVIV2	-0,886 ***	3,14	-0,934 ***	2,92	-0,879 **	2,23
HIJOS2	-0,323	0,98	-0,554	(.)	-0,510	0,83
NTRABAJ2	0,563 ***	4,21	0,609 ***	3,48	0,235 ***	2,66
CABFAOCUPA2	-0,518	(.)	-0,598	(.)	-0,595 ***	2,79
<b>CARACTERÍSTICAS FORMATIVAS:</b> (Referencia: Ciclo Formativo de grado medio; Familia Profesional de Administración)						
NIVELCF	0,190	(.)	0,154	0,97	0,683 ***	2,76
FAMILIAP1	0,090	0,12			-1,292	1,60
FAMILIAP2	-0,452	1,43	-0,291	35,41	-1,754 ***	3,57
FAMILIAP3	3.6E+05	(.)	2.3E+06	(.)	2.6E+02	(.)
FAMILIAP5	0,675	1,29	-1,585	(.)	5,116 ***	6,18
FAMILIAP6	-0,204	1,05	-0,505 **	2,14	1,547 ***	3,00
FAMILIAP7	-0,326	1,22	-1,150 ***	2,93	-0,855	1,17
FAMILIAP8	0,095	0,10	0,146	0,22	0,103	0,18
FAMILIAP9	0,446 *	1,75	-3,399	(.)	1,027 ***	2,80
FAMILIAP10	0,890	0,82	49,692	(.)	0,920	1,63
FAMILIAP11	-0,012	(.)	-0,082	0,29	-0,211	0,53
FAMILIAP12	0,211	0,72	0,097	0,21	203,781	(.)
FAMILIAP13	-1,769 *	1,77			-0,990	1,37
FAMILIAP14	0,733	1,36	0,452	0,05	0,792	1,63
FAMILIAP15	-0,237	0,63			2,315 ***	4,29
FAMILIAP16	0,915 *	1,68			0,682 *	1,67
FAMILIAP17	-0,145	0,24	47,986	(.)	0,356	0,84
FAMILIAP18	-0,899 ***	3,50	-1,281 ***	5,27	1,737 ***	3,00
FAMILIAP19	0,138	0,64	0,356	1,35	-0,591	1,28
FAMILIAP20	0,136	0,43	0,074	0,27	-0,315	0,67
FAMILIAP21	2.0E+05	(.)	-0,610	(.)	-3,570	
<b>VARIABLES DE ENTORNO:</b> (Referencia: Trabaja en el sector primario; consiguió su trabajo en 1997)						
SECTSEC4	-9,483 ***	29,89	-4,279	(.)	-2,465 ***	3,92
SECTTERC4	-9,877	(.)	-5,193 ***	11,92	-2,253 ***	3,89
TASADESJ4	-0,014	(.)	0,048	(.)	0,139	1,60
AÑOCOM98	-12,324 ***	26,04	-9,347 ***	21,31	11,579	(.)
AÑOCOM99	-12,990 ***	41,56	-9,624 ***	24,86		
AÑOCOM00	-13,547 ***	73,72	-9,873 ***	36,92	0,205	0,37
AÑOCOM01	-14,275	(.)	-10,405	(.)	0,681	0,81
<b>CARACTERÍSTICAS DEL EMPLEO:</b> (Referencia: Medio de búsqueda efectivo de tipo informal; empresa pública; contrato laboral indefinido; operario; trabajo a tiempo parcial)						
MEDEFEC4	-0,238	(.)	-0,216	0,73	-0,818 ***	3,13
EMPRIV_A	-0,608 ***	3,08	-0,460 *	1,66	0,269	0,62
EMPRIV_B	-0,458 **	2,38	-0,143	(.)	-0,019	0,04
EMPRIV_C	-0,572	(.)	-0,256	0,82	-0,368	0,82
EMPRIV_D	-0,455	1,59	3,4E+06	(.)	-1,012	0,92
TIPCONT_2	-1,284	(.)	-1,124	(.)	-1,603 ***	7,03
TIPCONT_3	-0,930 **	2,50	-0,931 *	1,96	-2,473 ***	3,96

TIPCONT_5	-1,460 ***	4,50	-0,854	0,82	-1,416 ***	2,64
TIPCONT_6	23,878	(.)	55,368	(.)	3,531 ***	5,72
TIPCONT_7	3,3E+10	(.)	30,451	(.)	0,304	0,39
TIPCONT_8	1,4E+07	(.)			1,742	(.)
SUELHOR4	0,044	(.)	0,052	(.)	-0,154 ***	2,74
CATLAB4	-1,147 ***	3,07	-1,002 *	1,86	-0,329	0,48
TCOMPLETE4	-0,317	(.)	-0,386 *	1,80	-0,519 *	1,72
<b>ATRIBUTOS LABORALES:</b> (Referencia: Conocimientos adecuados; satisfecho con el sueldo, con la estabilidad, con las posibilidades de promoción, con la prevención de riesgos laborales y con el horario; no está buscando empleo)						
TBUSCEMP4	0,001	1,49	0,002 ***	2,86	0,002	1,02
TOKTRA4	0,008	1,14	0,012 ***	4,41	0,011 ***	4,37
CONADE_S	0,143	0,92	0,054	0,17	-0,486 **	2,01
CONADE_I	0,047	(.)	-0,088	(.)	0,224	0,92
SATSUEL4	-0,096	(.)	-0,145	(.)	-0,335	1,56
SATESTA4	0,476	(.)	0,890	(.)	0,420 **	2,04
SATPROM4	0,082	(.)	-0,032	0,16	0,914 ***	4,53
SATRIES4	0,021	(.)	-0,288	(.)	0,305	1,42
SATHORA4	0,077	0,48	0,129	(.)	0,543 **	2,50
BUSCAEMP	-0,890 ***	5,49	-0,810 ***	5,19	-0,688 ***	3,13
$\alpha$	0,055	0,44	0,041	0,20	1,497 ***	12,42
$m$	18,925 ***	67,13	31,277 ***	15,73	-1,936	(.)
Log. Verosimilitud	-723,687		-318,057		-338,730	
Nº observaciones	1027		533		494	

Estadísticos z robustos en valor absoluto

\* significativo al 10%; \*\* significativo al 5%; \*\*\* significativo al 1%

En lo que respecta a las variables de *características personales y familiares*, observamos un coeficiente positivo de las variables EDAD4 y NTRABAJ2 para el colectivo global de la muestra. Por lo tanto, parece que, cuanto mayor sea la edad del titulado y cuanto mayor sea el número de miembros ocupados en la unidad familiar, mayor será la tasa de riesgo de salida del último empleo. Este efecto negativo de la edad sobre la duración del empleo lo detectaron también Bratberg y Nilsen (1998), argumentando que esto se podía deber a un aumento de las expectativas laborales a medida que aumentaba la edad del individuo.

La variable dicotómica CONVIV2, sin embargo, muestra un coeficiente negativo, señalando así una mayor duración en su último empleo para aquellos titulados que aún conviven con sus padres, en comparación con los que viven solos o en pareja. Este resultado nos puede estar hablando de titulados que contribuyen económicamente al sustento de la unidad familiar y que, por tanto, 'no se pueden permitir' el abandono del empleo que poseen.

Finalmente, en este primer bloque de covariables observamos que, el hecho de que el cabeza de familia esté ocupado (CABFAOCUPA2), influye positivamente en la duración del último empleo del titulado varón, en comparación con los titulados cuyo cabeza de familia está desempleado.

En lo que respecta a las *características formativas* observamos que, entre los hombres, el haber cursado un Ciclo Formativo de Grado Superior conlleva una mayor tasa de riesgo de salida del empleo que el haber realizado uno de Grado Medio (NIVELCF). Taylor (1999) llegó a un resultado similar en su análisis de la duración del empleo con datos del *British Household Panel Survey*, donde comprobó que el poseer un *A-level* o más, aumentaba la tasa de salida de los hombres.

Asimismo se observan diferencias significativas en cuanto a la influencia que ejerce la especialización sobre la duración del último empleo. Así, en el caso de las mujeres, aquellas que hayan cursado un Ciclo Formativo de la Familia de 'Comercio y Marketing', 'Comunicación, Imagen y Sonido' o 'Química', presentan una mayor duración en su último empleo que las que hayan cursado un CF de 'Administración'. Por otro lado, entre los hombres, los que hayan cursado un CF de la Familia de 'Actividades Agrarias', 'Artes Gráficas', 'Comercio y Marketing', 'Madera y Mueble', 'Mantenimiento de Vehículos Autopropulsados' o 'Química', por el contrario, presentan una menor duración en su último empleo que los que hayan cursado uno de la rama de 'Administración'.

Adentrándonos en el análisis de las *variables de entorno* observamos, en primer lugar, que el sector primario resulta ser el que presenta empleos de menor duración (SECTSEC4 y SECTTERC4). Este resultado coincide con el obtenido por Taylor (1999), quien hablaba de una mayor tasa de salida entre los trabajadores de la agricultura.

Por otro lado, al contrario de lo que sucedía en la estimación de la duración del primer empleo, en este caso, el año de consecución del último empleo (AÑOCOM98 a AÑOCOM01) resulta significativo únicamente entre el colectivo femenino. Esta influencia, además, presenta un signo negativo, de forma que, las que consiguen su último empleo en los últimos años, son las que presentan una menor tasa de riesgo de salida, lo que puede indicar un aumento de la apreciación de la capacitación técnico-profesional femenina en el mercado de trabajo.

Del análisis de las variables recogidas en el cuarto bloque (*características del empleo*), observamos nuevamente que la influencia del medio de búsqueda utilizado (MEDEFEC4), tan solo afecta al colectivo masculino. Sin embargo, en esta ocasión, la influencia apunta hacia una mayor duración del empleo cuando se ha utilizado un medio de búsqueda formal, coincidiendo así plenamente con los resultados de Ibáñez-Pascual (1999).

Por otro lado, las mujeres contratadas en grandes empresas de más de 250 trabajadores (EMPRIV\_A a EMPRIV\_D), presentan una tasa de riesgo de salida de ese empleo inferior a las contratadas en empresas públicas, mientras que, entre los hombres, el tamaño y titularidad de la empresa, no presenta influencia estadísticamente significativa sobre la duración de su último empleo.

La influencia del tipo de contrato sobre la duración del último empleo (TIPCONT\_2 a TIPCONT\_8) muestra un signo contrario al esperado, puesto que en el caso de los hombres, por ejemplo, tener un contrato temporal, de formación o no tener contrato alguno, conlleva una relación más duradera en el empleo que poseer un contrato indefinido.

La influencia del sueldo por hora (SUELHOR4), en cambio, sí muestra el resultado esperado, aunque en este caso, a diferencia de lo que sucedía en la Tabla IV.12, la influencia sólo es significativa entre el colectivo de los hombres. Así, cuanto mayor sea la retribución horaria, menor será la tasa de riesgo de salida del empleo y, por tanto, mayor será la duración del mismo. García-Serrano y Malo-Ocaña (1996) obtuvieron igualmente diferencias significativas en el efecto de los salarios sobre la movilidad en función del género detectando un efecto mayor en el caso de los varones.

La categoría laboral (CATLAB4), que no resultaba significativa en la Tabla IV.12, sí resulta significativa cuando estamos hablando del último empleo, de manera que las mujeres que ostentan puestos directivos o de mando medio en su último empleo (el 3% de ellas), presentan una menor tasa de salida del mismo que las que ocupan un puesto de operarias. Este mismo patrón se reproduce en lo que respecta a la parcialidad del empleo (TCOMPLET4) puesto que esta covariable, que no resultaba significativa en la Tabla IV.12, sí lo es en la Tabla IV.16, mostrando una mayor duración del empleo, tanto en hombres como en mujeres, cuando el trabajo que poseen es a tiempo completo en vez de a tiempo parcial.

Por último, en el bloque de *atributos laborales*, observamos un signo positivo en la estimación de la variable TBUSCEMP4 para las mujeres. Este signo, que implica un aumento de la tasa de salida del empleo cuanto mayor haya sido el tiempo dedicado a la búsqueda del mismo, coincide con los resultados de Ibáñez-Pascual (1999) y los de Taylor (1999). De la misma forma, el efecto del tiempo necesario para el desempeño adecuado de las tareas (TOKTRA4), implica una menor duración del empleo para quienes hayan necesitado un mayor período de formación, por lo que, en el mercado, parece existir una 'penalización' para los titulados (hombres y mujeres) que demuestren tener una menor velocidad de aprendizaje.

En lo que respecta a las variables de desajuste de conocimientos (CONADE\_S y CONADE\_I), las estimaciones para el colectivo femenino, a pesar de no ser estadísticamente significativas, presentan el signo esperado, indicando una mayor rotación entre las sobreeducadas y una menor rotación entre las subeducadas. Sin embargo, el único valor significativo es el de la sobreeducación de los varones, indicando una mayor duración de estos en su último empleo, en comparación con los que presentan un nivel de conocimientos adecuado.

Las variables de satisfacción, en el caso del último empleo (SATSUEL4 a SATHORA4), sólo muestran significación estadística en la estimación realizada para los hombres. En este caso, cuando el nivel de satisfacción con la estabilidad laboral y con las posibilidades de promoción es bajo, como era de esperar, la tasa de riesgo de salida de ese empleo resulta ser mayor que cuando la satisfacción es alta.

Finalmente, la variable que indicaba si el titulado se encontraba buscando un nuevo empleo a pesar de estar ocupado (BUSCAEMP), resulta ser significativa en ambos colectivos, aunque, nuevamente, el signo de la estimación parece estar en desacuerdo con las expectativas previas. De hecho, el resultado de nuestra estimación muestra cómo los titulados que dicen estar buscando un empleo, presentan una menor tasa de riesgo de salida y, en consecuencia, una mayor duración del último empleo.

## 6. RECAPITULACIONES

El objetivo que nos marcamos en el presente capítulo consistía en el análisis del proceso de inserción laboral de los técnicos y técnicos superiores de la isla de Gran Canaria, así como en la trayectoria laboral subsiguiente. Para ello nos centramos en tres grandes momentos dentro de la historia laboral de los titulados que constituyen nuestra muestra: 1. Primer período de desempleo; 2. Primer período de ocupación; 3. Último período de ocupación.

Comenzamos el capítulo, por tanto, delineando aquellas teorías que, de alguna manera, contribuyen a la explicación de la existencia de estos tres estados en la trayectoria laboral de los titulados. En primer lugar nos centramos en aquellas teorías de competencia imperfecta que contribuyen a la explicación de la existencia de un volumen de desempleo involuntario en el mercado laboral; para, a continuación, centrarnos en aquellas otras, que introducen argumentos para la explicación de la existencia de movilidad laboral.

Dentro de las teorías de competencia imperfecta, hicimos referencia a la *teoría de 'salarios de eficiencia'*, desarrollada por Weiss en 1980. Ésta hablaba del proceso de contratación de trabajadores en un mercado heterogéneo, en donde el salario de aceptación del trabajador es función creciente de su habilidad, y en el que las empresas tienen información imprecisa acerca de los atributos laborales de cada trabajador. Dicha teoría otorga importancia a las compensaciones monetarias para reducir la rotación voluntaria, sugiriendo que las empresas pueden estar dispuestas a pagar salarios más altos que los del mercado, con el fin de evitar despidos (Weiss, 1980).

Introducimos también, como referencia a las teorías de competencia imperfecta, la *Teoría 'insider-outsider'*, desarrollada originalmente por Lindbeck y Snower en 1988. Ésta enfatiza el poder de los trabajadores ocupados, a los que denominan 'insiders', para elevar sus salarios y evitar despidos, distinguiendo a estos de los 'outsiders'. Estos últimos estarían dispuestos a trabajar cobrando un salario inferior al de los 'insiders', de forma tal, que la diferencia compensara a la empresa por sus menores habilidades de trabajo cooperativo, pero, la capacidad negociadora de los 'insiders', impide dicha contratación.

En lo que respecta a las teorías de movilidad entre empresas, introdujimos en primer lugar la *Teoría del 'job matching'* (Jovanovic, 1979). Ésta afirmaba que los beneficios no salariales del puesto, como pueden ser las posibilidades de promoción, el ajuste educación-empleo, la seguridad en el puesto de trabajo, etc., también juegan un papel importante en la movilidad externa. De esta manera, sugiere que la probabilidad de separación permanente, dada la antigüedad, será menor entre aquellos individuos que presenten un buen ajuste entre la educación poseída y la requerida por el puesto ('buen

emparejamiento'), mientras que, si el emparejamiento no es el óptimo, el trabajador se moverá hacia otro empleo con la finalidad de mejorarlo.

Finalmente, hablamos de la *Teoría de la movilidad ocupacional* (Rosen, 1972; Sicherman y Galor, 1990), como Teoría que aporta una explicación a la existencia de la sobreeducación como fenómeno temporal. De hecho, puede explicar el porqué algunos individuos prefieren, al comienzo de su vida laboral, trabajar en puestos para los que presentan un exceso de formación. Según argumenta, los individuos permanecerán en la ocupación actual (aún estando sobreeducados) adquiriendo experiencia y formación que les permita moverse a ocupaciones mejores (más remuneradas o con mayores expectativas de promoción), dentro o fuera de la empresa. Las personas sobreeducadas, por tanto, ocupan empleos para los que son necesarios menos requerimientos educativos, puesto que, de esta manera, tienen una mayor probabilidad de moverse hacia ocupaciones de mayor nivel.

Tras el repaso a estas teorías y a la metodología de modelos de duración, como herramienta de trabajo fundamental en el presente capítulo, procedimos al estudio de los tres momentos vitales en la trayectoria ocupacional de los titulados. Para cada uno de estos momentos, resaltaremos a continuación aquellos aspectos, obtenidos de nuestro estudio, que consideramos de mayor relevancia.

#### Resultados del primer período de desempleo:

El análisis del primer período de desempleo lo realizamos para el 74% de nuestra muestra que ya había logrado su primer empleo tras la finalización del CF y para el 2.8% que aún estaba buscándolo<sup>148</sup>. Para estos últimos, que son los que presentan un problema de censura, la duración media de su periodo de desempleo, a día de la encuesta, oscilaba en torno a los 6 meses y medio, y tan sólo el 5%, presentaba duraciones de desempleo superiores al año y medio. Por lo que prácticamente no nos encontramos ante ningún caso de paro de larga duración.

La realización de contrastes de homogeneidad de estratos, nos aportó evidencias acerca de la divergencia en las funciones de riesgo en función del género, del nivel del CF y de la Familia Profesional. Del estudio no paramétrico de las mismas pudimos observar que:

- Tanto los hombres, como las mujeres, presentan una alta probabilidad de salida del desempleo en un período inferior a una semana, siendo ésta del 42.6% para los hombres y del 37% para las mujeres.
- Los técnicos superiores presentan una cierta ventaja sobre los técnicos en períodos inferiores al año.

---

<sup>148</sup> El resto de la muestra había logrado su primer empleo antes de finalizar el CF (17.8%) o se declaraba inactivo (5.4%).

- La Familia Profesional que se enfrenta a peores oportunidades laborales en el mercado de trabajo, parece ser la de 'Textil, Confección y Piel', puesto que, al cabo de tres años, aún mantiene una proporción de desempleo del 13.3%, cuando en el resto de las familias, este porcentaje no llega al 3%.
- La Familia Profesional que se enfrenta a las mejores oportunidades laborales es la de 'Mantenimiento y Servicios a la Producción'.
- Transcurridos tres meses desde el comienzo de la búsqueda de empleo, la probabilidad de mantenerse en situación de desempleo es inferior al 50% para todas las Familias Profesionales.

El análisis de los datos de duración nos presenta la evidencia de que la transición hacia el primer empleo es un fenómeno complejo en el que intervienen muchos y diversos factores. Los puntos más relevantes que se deducen del estudio, a través de la estimación de un modelo Gompertz con heterogeneidad Gamma, son los siguientes:

En la estimación para el colectivo de *hombres* observamos que:

- Los factores que provocan una *disminución* en la duración del primer período de desempleo, tras la finalización del CF, son: tener un cabeza de familia con estudios superiores; cursar un CF de la rama de 'Imagen Personal'; pertenecer a las cohortes más recientes de titulados; acudir a los responsables de la FCT y apuntarse a bolsas de trabajo, como medio de búsqueda de empleo
- Los factores que provocan un *aumento* en la duración del primer período de desempleo, tras la finalización del CF, son: utilizar muchos medios de búsqueda diferentes; utilizar el envío de currículos como medio de búsqueda

Por otra parte, en la estimación para el colectivo de *mujeres*, observamos que:

- Los factores que provocan *disminución* son: demostrar tener experiencia laboral previa a la realización del CF; cursar un CF3; cursar un CF de la rama de 'Fabricación Mecánica'; pertenecer a las cohortes más recientes de titulados; tener una nota media superior a la media de su promoción y especialidad; que la tasa de desempleo juvenil sea alta; acudir a los responsables de la FCT y apuntarse a bolsas de trabajo, como medio de búsqueda de empleo
- Los factores que provocan *aumento* son: la edad; optar por la Formación Profesional Específica por considerarse mala estudiante; utilizar muchos medios de búsqueda diferentes; utilizar el envío de currículos como medio de búsqueda

En cambio, las responsabilidades familiares, como tener hijo, y la capacidad demostrada en el sistema educativo, medida en términos de tiempo necesario para obtener la titulación, no resultan ser significativos. Al igual que tampoco lo es el optar por un CF a través de una 'ruta desestructurada'.

### Resultados de la duración del primer empleo:

Este análisis lo realizamos para el 91.8% del total de la muestra, ya que estos son los que, en algún momento, lograron su primer empleo. De estos, el 38.7% (977), aún se encontraba trabajando en ese primer empleo al responder a la encuesta, por lo que presentan un problema de censura. Aún así, la duración media en el primer puesto de trabajo es de, en torno al año y 5 meses, aunque ésta presenta una gran dispersión.

La realización de contrastes de homogeneidad de estratos, en este caso, apunta hacia una diferencia únicamente entre Familias Profesionales pero no por género, titularidad del centro o nivel del CF. Del estudio no paramétrico de la muestra estratificada por Familias Profesionales pudimos observar que:

- La estimación de la función de riesgo es prácticamente nula en todo momento, por lo que la probabilidad condicional de salida del primer empleo resulta ser muy baja.
- Las Familias Profesionales, cuyos titulados se enfrentan a primeros empleos con una menor duración, son las de 'Comunicación, Imagen y Sonido' e 'Industrias Alimentarias'.
- Por el lado contrario nos encontramos con el caso de los titulados de 'Textil, Confección y Piel' y 'Madera y Mueble', quienes, en el 62.5% de los casos, aún permanecían en su primer empleo al cabo de un año.

El análisis paramétrico, a través de la estimación de un modelo Weibull con heterogeneidad Gaussiana Inversa, nos muestra la influencia de determinadas covariables sobre la duración del primer empleo. De la estimación diferenciada por género, destacamos los siguientes resultados:

En la estimación para el colectivo de *hombres* observamos que:

- Los factores que provocan una *disminución* en la duración del primer empleo son: demostrar tener experiencia laboral previa; ser titulado en 'Artes Gráficas' o 'Imagen Personal'; que la tasa de desempleo juvenil sea grande; haber conseguido el empleo a través de un medio de búsqueda formal; estar contratado en una empresa privada; conseguir el primer empleo durante o tras finalizar el CF; presentar subhabilitación en el puesto
- Los factores que provocan un aumento en la duración del primer empleo son: que el número de miembros de la unidad familiar ocupados sea grande; un aumento en el tiempo necesario para la eficiente realización de las tareas; que el nivel de satisfacción con la estabilidad, las posibilidades de promoción y el horario sea bajo, o que la insatisfacción con la prevención de riesgos laborales sea alta.

Por otro lado, en la estimación para el colectivo de *mujeres*, observamos que:

- Los factores que provocan *disminución* en la duración del primer empleo son: demostrar tener experiencia laboral previa; ser titulada en 'Mantenimiento y Servicios a la Producción', 'Industrias Alimentarias' o 'Fabricación Mecánica'; trabajar en el sector secundario o terciario; estar contratada en una empresa privada; mostrar sobrehabilitación o sobreeducación
- En cambio, los que provocan un *aumento* en la duración del primer empleo son: que el número de miembros de la unidad familiar ocupados sea grande; un aumento en la edad; trabajar como ayuda familiar; un aumento en el sueldo por hora; que el tiempo de búsqueda de ese empleo haya sido grande; un aumento en el tiempo necesario para la eficiente realización de las tareas; demostrar subhabilitación; que el nivel de satisfacción con el puesto de trabajo, en general, sea bajo

#### Resultados de la duración del último empleo:

Por último, el análisis de la duración del último período de empleo lo realizamos para el 42.8% del total de la muestra, que son quienes han cambiado, al menos una vez, de empleo. En este caso, el 79.8% de estos (950) aún se encontraba trabajando en ese último empleo al responder a la encuesta, por lo que presentan un problema de censura. Aún así, la duración media en ese puesto de trabajo es de 10 meses y medio, aunque presenta una gran dispersión (con duraciones que varían desde los 11 días hasta los poco más de cuatro años).

En este caso, el contraste de homogeneidad de estratos nos apunta hacia una diferencia por nivel del CF y por Familias Profesionales. Del estudio no paramétrico de la muestra destacamos que:

- Las titulados de 'Madera y Mueble', 'Comunicación, Imagen y Sonido' y 'Actividades Físicas y Deportivas', parecen presentar una menor duración de su último empleo.
- Los de 'Informática', 'Textil, Confección y Piel' e 'Industrias Alimentarias' son los que presentan una mayor duración de su último empleo.
- En cualquier caso, la estimación no paramétrica nos vuelve a mostrar una función de riesgo que vuelve a ser prácticamente nula en todo momento, por lo que la probabilidad condicional de salida del primer empleo es de nuevo muy baja.

El análisis de la duración, a través de la estimación paramétrica de tipo Gamma Generalizado sin heterogeneidad inobservable, nos muestra las siguientes relaciones de dependencia. No obstante, estas deben ser interpretadas con cautela puesto que la presencia de una censura tan alta en los datos lleva a una escasa bondad de ajuste:

En la estimación para el colectivo de *hombres* observamos que:

- Los factores que provocan una *disminución* en la duración del último empleo son: número de miembros ocupados en la unidad familiar; haber cursado un CF3; ser

titulado en 'Actividades Agrarias', 'Artes Gráficas', 'Comercio y Marketing', 'Madera y Mueble', 'Mantenimiento de Vehículos Autopropulsados' o 'Química'; trabajar en el sector primario; un aumento en el tiempo necesario para desempeñar adecuadamente las tareas; un bajo nivel de satisfacción con el puesto de trabajo

- Mientras que, los que provocan un *aumento* en la duración del último empleo son: convivir con los padres; que el cabeza de familia esté ocupado; utilizar un medio de búsqueda de empleo formal; un aumento del sueldo por hora; que el trabajo sea a tiempo parcial; considerarse sobreeducado; estar buscando un nuevo empleo

En cambio, en la estimación para el colectivo de *mujeres* observamos que:

- Los factores que provocan *disminución* en la duración del último empleo son: número de miembros ocupados en la unidad familiar; trabajar en el sector primario; un aumento en el tiempo dedicado a la búsqueda del empleo; un aumento del tiempo necesario para desempeñar adecuadamente las tareas
- Mientras que los que provocan un *aumento* en la duración son: convivir con los padres; ser titulado en 'Comercio y Marketing', 'Comunicación, Imagen y Sonido' o 'Química'; que la consecución del empleo sea posterior a 1997; estar contratadas en grandes empresas de más de 250 trabajadores; ocupar un puesto directivo o de mando medio; que el trabajo sea a tiempo parcial; estar buscando un nuevo empleo.

Como vemos, nuestras estimaciones parecen confirmar las previsiones que hacía la teoría de 'salarios de eficiencia', en cuanto a la compensación monetaria como método de reducción de la rotación voluntaria. Este efecto queda patente en la estimación del modelo de duración para el primer empleo de las mujeres, y en la del último empleo de los hombres, puesto que en ellas se aprecia cómo un aumento en el salario conlleva una mayor duración del empleo. No obstante, al desconocer el salario de reserva de los desempleados, no hemos podido verificar la teoría '*insider-outsider*' para nuestra muestra.

De la misma manera, mediante las estimaciones de los modelos de duración del primer empleo, comprobamos cómo, en nuestra muestra, hay evidencias que apoyan la teoría de la 'movilidad ocupacional'. Ésta afirmaba que, cuando el emparejamiento no era el óptimo, los individuos abandonaban el empleo en busca de uno más afín a su formación y, de hecho, en la estimación realizada para la submuestra de hombres, observamos que, los que mostraban estar subhabilitados, presentaban una menor duración de su empleo, al igual que sucedía en la estimación para la submuestra de mujeres, con las que manifestaban estar sobrehabilitadas o sobreeducadas. Este último fenómeno, precisamente, apoya también la teoría de la 'movilidad ocupacional', la cual consideraba la sobreeducación como un fenómeno transitorio, que desaparecía con la movilidad hacia mejores ocupaciones.