

UNIVERSIDAD DE GRANADA



**ESTIMACIÓN EFICIENTE DE ALGUNOS
INDICADORES DE POBREZA**

TESIS DOCTORAL

Directores

Juan Francisco Muñoz Rosas

Encarnación Álvarez Verdejo

Doctorando

Leocadio José Barroso Bravo

**DEPARTAMENTO DE MÉTODOS CUANTITATIVOS PARA LA
ECONOMÍA Y LA EMPRESA**

Programa de Doctorado en Ciencias Económicas y Empresariales

Granada, 2015

Editorial: Universidad de Granada. Tesis Doctorales

Autor: Leocadio José Barroso Bravo

ISBN: 978-84-9125-211-5

URI: <http://hdl.handle.net/10481/40664>

AGRADECIMIENTOS

Gracias de corazón a mis tutores Encarnación Álvarez Verdejo y Juan Francisco Muñoz Rosas por su apoyo continuo, ánimo, paciencia, dedicación y confianza. Gracias por haber facilitado siempre los medios suficientes para llevar a cabo las actividades propuestas durante el desarrollo de esta tesis. Vuestro apoyo y confianza en mi trabajo ha sido una aportación invaluable, no solamente en la investigación de esta tesis, sino también en mi formación como investigador. Gracias por compartir vuestra profesionalidad y rigurosidad, las cuales han sido clave del buen trabajo que hemos realizado juntos. Gracias por vuestra orientación mediante llamadas por Skype, e intercambios de innumerables correos electrónicos.

Merece mención personal mi agradecimiento al Dr. Juan Francisco Muñoz Rosas por haber sido un pilar básico en el desarrollo de mis 4 últimos años, tanto académico, como profesional y por último personal. Para mi es un grandísimo honor que en su día aceptase guiarme en mi doctorado, el cual comencé gracias a su confianza y motivación a seguir formándome académicamente. Este trabajo no se puede concebir sin su siempre oportuna participación. Gracias.

Igualmente, gracias Dra. Encarnación Álvarez por toda su ayuda en los trabajos tan complejos y meticulosos de simulación Monte Carlo, trabajos en el programa R, etc. No cabe duda que su participación ha perfeccionado el trabajo realizado, llevando a la tesis a un nivel superior. También agradezco a Rosa María García Fernández su disposición para facilitarnos las poblaciones de la Encuesta Europea sobre Ingresos y Condiciones de Vida consideradas en esta tesis, así como sus aportaciones en el artículo conjunto que hemos publicado junto con los directores de tesis.

Y por encima de todo, gracias a los míos por vuestro apoyo incondicional. Gracias Papá y Mamá, Shannon, Mónica y Jose, Ángel Luis y María y Tito Paco por motivarme todos los días, darme fuerzas para seguir, todos vuestros consejos, han sido fundamentales en el desarrollo de este trabajo. Sobre todo gracias por vuestra paciencia en momentos difíciles en los que parece que nada va bien, pero que siempre he tenido la suerte de haber contado con vuestro apoyo para levantarme y seguir adelante. Me considero una persona muy afortunada de tener gente tan maravillosa como vosotros, donde el cariño no escasea.

Por último, me gustaría agradecer a mis amigos los cuales también han sido de ayuda, para poder desconectar y recargar pilas con energía positiva.

ÍNDICE GENERAL

1. Introducción

1.1. Objetivos

1.2. Medición de la Pobreza

1.2.1. Introducción

1.2.2. Encuestas Familiares

1.2.3. Enfoques de la pobreza

1.2.4. Elección de variable económica y escalas de equivalencia

1.3. Algunos indicadores de pobreza

1.3.1. Introducción

1.3.2. Líneas de pobreza

1.3.3. Cuantiles en la población de pobres

1.3.4. Proporción de pobres

1.3.5. Índice de brecha de pobreza

1.3.6. Índice de pobreza severa

1.3.7. Otros indicadores de pobreza y desigualdad

2. Estimación de la función de distribución y cuantiles en la población de pobres

2.1. Introducción

2.2. Antecedentes de estimadores de la función de distribución y cuantiles

2.2.1. Estimadores basados en el diseño

2.2.2. Estimador de Chambers y Dunstan

2.2.3. Estimadores de Rao, Kovar y Mantel

2.2.4. Estimador de Silva y Skinner

2.2.5. Algunos estimadores de cuantiles

- 2.3. Función de distribución y cuantiles en la población de pobres
 - 2.3.1. Notación y definición de parámetros
 - 2.3.2. Estimación bajo un diseño muestral general
 - 2.3.3. Estimación bajo muestreo aleatorio simple
- 2.4. Nuevos estudios de simulación Monte Carlo

3. Estimación de la proporción de pobres

- 3.1. Introducción
- 3.2. Definición y estimación de la proporción de pobres
- 3.3. Métodos de estimación propuestos
 - 3.3.1. Transformaciones propuestas
 - 3.3.2. Estimadores de tipo razón
 - 3.3.3. Estimadores de tipo regresión
 - 3.3.4. Estimación mediante intervalos de confianza
 - 3.3.5. Extensión a un diseño muestral general
- 3.4. Estudios de simulación Monte Carlo

4. Estimación de los índices de brecha de pobreza y pobreza severa

- 4.1. Introducción
- 4.2. Los índices de brecha de pobreza y pobreza severa.
- 4.3. Estimador de tipo razón
- 4.4. Estimación mediante intervalos de confianza
- 4.5. Estudios de simulación Monte Carlo

5. Conclusiones y futuras líneas de investigación

- 5.1. Conclusiones
- 5.2. Limitaciones y futuras líneas de investigación.

Referencias

Apéndice I. Descripción de las poblaciones en estudio

Capítulo 1

Introducción

1.1. Objetivos

La importancia de la medición de la pobreza y/o la desigualdad en una sociedad está justificada porque de ello dependerá el poder dar soluciones a un importante problema social. Por ejemplo, según Ravallion (1998), *“medir la pobreza correctamente es un sólido instrumento para centrar la atención de los políticos en las condiciones de vida de los individuos clasificados como pobres”*. Sin amplias bases de datos muestrales sobre las características de los pobres, o bien sin estimaciones eficientes de las medidas o indicadores que explican la distribución e intensidad de la población de pobres, sería imposible, entre otros objetivos:

- a) Crear políticas destinadas a reducir los índices de pobreza.
- b) Identificar a qué segmentos poblacionales hay que destinar medidas correctoras.
- c) Predecir los efectos de políticas y programas de ayuda a los pobres.
- d) Evaluar la actuación de las políticas sociales de los gobiernos.

El objetivo principal de esta tesis es la estimación de distintos indicadores y medidas de pobreza mediante técnicas cuantitativas de estimación que hagan un buen uso de toda la información muestral disponible. De forma específica, se analiza en esta tesis la estimación de la proporción de pobres, los índices de brecha de pobreza y pobreza severa, y la función de distribución y cuantiles de distintos órdenes en la población considerada como pobre. En el caso de la estimación de cuantiles en la población de pobres, se

Capítulo 1. Introducción

trata de una medida o técnica utilizada para analizar la distribución de los individuos clasificados como pobres. Algunas referencias relevantes que describen las medidas de pobreza abarcadas en esta memoria, así como otros aspectos relacionados con la medición de la pobreza son Foster et al. (1984), Ravallion (1992), Foster (1998), Ravallion (1998), World Bank (2000), Jolliffe (2001), Zheng (2001), Berger y Skinner (2003), Khandker (2005), Haughton y Khandker (2009), Molina y Rao (2010), Muñoz et al. (2014), etc.

Las encuestas por muestreo juegan un papel muy importante en la mayoría de las investigaciones y estudios desarrollados por empresas, instituciones, oficinas de estadística, organismos nacionales, etc. Así, por ejemplo, los principales indicadores demográficos, económicos o sociales de un país están basados en datos muestrales obtenidos a partir de diseños muestrales complejos. A modo de ejemplo, podemos citar en este sentido la Encuesta Europea sobre Ingresos y Condiciones de Vida (véase, por ejemplo, Van Kerm 2007, Goedemé 2010, Molina y Rao 2010, Berger y Torres 2012, Muñoz et al. 2014, etc). En adelante, denominaremos a esta encuesta como EU-SILC, con motivo de su acrónimo en la terminología inglesa (European Union Survey on Income and Living Conditions), y dado que EU-SILC es cómo popularmente se denomina a esta encuesta. En las mencionadas encuestas no sólo se recoge información relacionada con la variable de estudio. Por el contrario, es habitual obtener información auxiliar que permita mejorar la estimación de un determinado parámetro en estudio. En otras palabras, en las encuestas por muestreo es común disponer de variables adicionales relacionadas con la variable de interés objeto de estudio. Este también es el caso de la EU-SILC, tal como puede comprobarse en el Apéndice I de esta memoria. A tales variables adicionales se les suele denominar como variables auxiliares en el contexto del muestreo en poblaciones finitas.

La importancia de la información auxiliar en la fase de estimación ha propiciado que las investigaciones de muestreo en poblaciones finitas de los

Capítulo 1. Introducción

últimos años se hayan centrado en mejorar la estimación de numerosos parámetros en presencia de información auxiliar. Estas investigaciones se justifican tanto por la mejor precisión por parte de las nuevas técnicas de estimación basadas en información auxiliar, como por la reducción de costes en investigaciones o estudios sociales que pueden obtenerse a través de estas nuevas técnicas de estimación, puesto que para obtener el mismo error de estimación que otras técnicas más básicas será preciso, en general, un menor tamaño muestral.

En general, son dos las metodologías seguidas para la obtención de estimaciones más precisas. Por un lado, el uso apropiado de diseños muestrales complejos puede producir importantes beneficios en comparación con esquemas muestrales simples tal como el muestreo aleatorio simple sin reemplazamiento (en adelante denominado con la siglas MAS). Algunas referencias relevantes que describen distintos diseños muestrales son Chaudhuri y Vos (1988), Särndal et al. (1992), Cochran (1997), Singh (2003), Chaudhuri y Stenger (2005), etc. En segundo lugar, el uso de información auxiliar en la etapa de estimación también puede producir mejores estimaciones en comparación con métodos de estimación más simples, especialmente cuando la relación entre la variable objeto de estudio y las variables auxiliares sea elevada. Algunos ejemplos muy conocidos de esta segunda metodología son los conocidos estimadores de tipo regresión (Casell et al. 1976, Casell et al. 1977, Särndal et al. 1992 pg.219, etc) y estimadores de razón (Särndal et al. 1992 pg.229) para la estimación de la media o el total en una población. Más recientemente, el método de calibración (Deville y Särndal 1992) y el método de verosimilitud pseudo empírica (Chen y Qin 1993 y Chen y Sitter 1999) son nuevas técnicas cuantitativas de estimación para el problema de la estimación de la media poblacional en presencia de información auxiliar. Otras referencias relacionadas con la estimación de la media poblacional en presencia de variables auxiliares son Wu (2002), Wu (2003), Berger et al. (2009), etc. Para el problema de la estimación de funciones de distribución y cuantiles, algunos

Capítulo 1. Introducción

métodos de estimación relevantes están descritos por Chambers y Dunstan (1986), Kuk y Mak (1989), Rao et al. (1990), Francisco y Fuller (1991), Dorfman y Hall (1993), Mak y Kuk (1993), Silva y Skinner (1995), Chen y Wu (2002), Dorfman (2009), etc. Finalmente, Molina y Rao (2010) y Muñoz et al. (2014) proponen estimadores de algunas medidas de pobreza basadas en información auxiliar.

Con el fin mejorar la precisión, entre otras propiedades, de algunos indicadores y medidas de pobreza, en esta tesis se utilizará, en general, la segunda de las dos metodologías descritas en el párrafo anterior, es decir, consideraremos técnicas cuantitativas de estimación basadas en información auxiliar con el objetivo de obtener estimaciones más precisas para la proporción de personas consideradas como pobres, y para los índices de brecha de pobreza y pobreza severa. Respecto a la estimación de la función de distribución y cuantiles en la población de pobres, completaremos los estudios de simulación realizados por Álvarez et al. (2012). Destacamos que este estudio puede considerarse como el inicio de esta investigación, y su realización supuso una formación y una base importante para la realización del resto de capítulos de esta tesis, en los cuales si se plantearon nuevas metodologías de estimación, se debatieron algunas propiedades teóricas y se completaron con los correspondientes estudios de simulación. Teniendo en cuenta las líneas anteriores y después de realizar una revisión exhaustiva de la bibliografía más relevante relacionada con los métodos de estimación basados en información auxiliar así como las técnicas de estimación utilizadas en los indicadores de pobreza objeto de análisis en esta memoria, los principales objetivos que se han perseguido en la elaboración de esta tesis han sido los siguientes:

- a) Plantear y definir nuevos estimadores puntuales para la proporción de pobres e índices de brecha de pobreza y pobreza severa. Se persigue la metodología del uso de la información auxiliar disponible en la etapa de estimación, la cual es una técnica muy conocida y utilizada que puede

mejorar el comportamiento de los estimadores propuestos en comparación con los estimadores existentes. En general, deseamos que los estimadores propuestos sean más eficientes y tengan sesgos empíricos más pequeños. Destacamos que en el caso de la estimación de la función de distribución y cuantiles en la población de pobres se completaron los estudios de simulación realizados por Álvarez et al. (2012), y para ello se usaron las nuevas poblaciones descritas en el Apéndice I de este trabajo.

- b) Plantear y estudiar otros temas relacionados con la estimación puntual de parámetros, tal como puede ser la estimación de las varianzas de los estimadores propuestos o la construcción de intervalos de confianza.
- c) Completar el estudio teórico de los estimadores, varianzas e intervalos de confianza propuestos con estudios empíricos basados en estudios de simulación Monte Carlo que avalen los beneficios obtenidos por los métodos propuestos en comparación con los existentes actualmente en la literatura. Por una parte, los estudios de simulación Monte Carlo se basarán en poblaciones simuladas con el fin de abarcar distintos escenarios que pueden presentarse en la práctica. Por ejemplo, consideraremos valores pequeños y elevados para los distintos indicadores de pobreza, correlaciones entre las variables, etc. Por otra parte, con el fin de analizar los métodos propuestos en situaciones reales, los estudios de simulación se basarán en datos reales extraídos de la EU-SILC en el año 2011 y para distintos países de la Unión Europea (Bélgica, Bulgaria, Eslovenia, España, Italia, Lituania, Polonia y Reino Unido).

Una vez expuestos los objetivos y líneas de investigación que se van a seguir en esta memoria, a continuación se expone brevemente cómo está estructurada el resto de esta tesis doctoral.

Capítulo 1. Introducción

En este Capítulo 1, además de describir los principales objetivos ya expuestos, se recogen algunas indicaciones y se presentan algunas definiciones relacionadas con los distintos capítulos de esta memoria. De este modo, en la Sección 1.2 se describen algunos aspectos básicos relacionados con la medición de la pobreza, lo cual supone una introducción a este problema social. Por su parte, en la Sección 1.3 se introducen los indicadores de pobreza más conocidos y/o utilizados, de entre los cuales se encuentran los indicadores de pobreza discutidos en esta memoria.

Una vez definida y establecida la línea o umbral de pobreza que clasifica la población objeto de estudio en pobres y no pobres, en el Capítulo 2 se plantea la estimación del primer indicador de pobreza. En este caso se trata de la estimación de cuantiles en la población clasificada como pobre. Dada la relación directa entre un cuantil y una función de distribución, en este Capítulo 2 también se analiza desde un punto de vista empírico la estimación de la función de distribución en la población de pobres. Tal como se debatirá en este capítulo, el uso de cuantiles de distintos órdenes puede resultar útil como técnica para analizar la distribución de los pobres. Las aportaciones realizadas en este capítulo se hacen extensibles al caso de muestras extraídas mediante un diseño muestral general, de modo que los resultados de simulación Monte Carlo llevados a cabo están basados tanto en MAS como en un esquema de muestreo con probabilidades desiguales denominado método de Midzuno (Midzuno 1952, Singh 2003).

En los Capítulos 3 y 4 se plantea la estimación eficiente de varios indicadores de pobreza que surgen como casos particulares de la familia de indicadores de pobreza propuesta por Foster et al. (1984). En particular, nos referimos a la proporción de pobres y los índices de brecha de pobreza y pobreza severa.

Capítulo 1. Introducción

Respecto a la proporción de pobres (Capítulo 3), se plantea la estimación eficiente de este parámetro basándonos en una transformación inicial de la variable auxiliar cuantitativa. También se plantea la estimación mediante intervalos de confianza y la extensión de la metodología planteada a un diseño muestral general. Las aportaciones teóricas aportadas están avaladas por una serie de estudios de simulación Monte Carlo.

En el Capítulo 4 se aborda la estimación de los índices de brecha de pobreza y de pobreza severa. En primer lugar, se plantea la estimación puntual de estos parámetros apoyándonos en la predicción de un modelo de regresión. En segundo lugar, se propone la estimación de la varianza de los estimadores propuestos mediante la técnica Bootstrap (véase, por ejemplo, Kovar et al. 1988, Stine 1989, Efron y Tibshirani 1993, Shao y Tu 1995, Biewen 2002, Moran 2006, Wolter 2007), comúnmente utilizada para la estimación de varianzas de estimadores. En particular, proponemos usar el método de bootstrap reescalado propuesto por Rao et al. (1992) para la estimación de estos parámetros. A partir de las varianzas sugeridas se resuelve también el problema de la estimación de estos parámetros mediante intervalos de confianza. Al igual que en los casos anteriores, las aportaciones teóricas estarán avaladas por los resultados empíricos obtenidos a partir de distintos estudios de simulación basados tanto en poblaciones reales extraídas de encuestas sobre ingresos y condiciones de vida como en poblaciones artificiales con distintas características.

Las poblaciones artificiales en los estudios de simulación Monte Carlo se utilizan con el fin de evaluar el comportamiento de los distintos estimadores planteados en esta memoria bajo distintos escenarios posibles en la práctica. En particular, las poblaciones artificiales tienen distintas características, tal como distintos valores para los distintos indicadores de pobreza, o bien distintos coeficientes de correlación lineal entre la variable de interés y la variable auxiliar. Por su parte, las poblaciones reales están basadas en datos muestrales recogidos a través de la EU-SILC del 2011 en varios países de la Unión Europea.

Capítulo 1. Introducción

En el Apéndice I de esta memoria se realiza un análisis descriptivo completo de las distintas poblaciones utilizadas en esta memoria. Como criterio para la comparación de los distintos estimadores puntuales se utilizarán medidas comúnmente utilizadas como el sesgo relativo y el error cuadrático medio relativo, entre otras. Tanto las comparaciones teóricas como las empíricas muestran que los estimadores propuestos tienen un buen comportamiento en términos de sesgo y error cuadrático medio. Respecto a los estimadores de las varianzas también se considera el correspondiente sesgo relativo, mientras que para evaluar empíricamente los intervalos de confianza se utilizará la cobertura empírica de tales intervalos de confianza construidos con un nivel de confianza del 95%.

En el Capítulo 5 se resumen las principales conclusiones derivadas de este trabajo, y se plantean futuras líneas de investigación que puede surgir a partir de las aportaciones e investigaciones realizadas en esa tesis doctoral.

1.2. Medición de la Pobreza

1.2.1. Introducción

El primer paso antes comenzar a tratar la medición de la pobreza es intentar dar una definición de este problema social. Según World Bank (2000), Khandker (2005), Haughton y Khandker (2009), *“la pobreza es la privación pronunciada del bienestar”*. Aunque esta definición es bastante simple y puede ser relativamente clara, no lo es tanto la definición de “bienestar”, puesto que se trata de un concepto complejo y cuya definición puede depender de la situación económica, del país donde se aplique, o incluso cada persona puede tener un concepto diferente del bienestar. La complejidad de la definición de pobreza hace que este fenómeno sea considerado como multidimensional, es decir, la pobreza se suele considerar como un fenómeno que tiene que medirse a través

Capítulo 1. Introducción

de índices más complejos tal como el Índice de Desarrollo Humano (Kelly 1991, Anand y Sen 1992, McGillivray 1991, Trabold-Nübler 1991, UNDP 1997, etc.). Además, la propia línea o umbral de pobreza que clasifica la población en pobres y no pobres puede obtenerse desde una perspectiva objetiva, o bien desde un enfoque relativo. Destacamos que la discusión entre el uso de líneas de pobreza relativas o absolutas, o bien el uso entre indicadores de pobreza unidimensionales frente a indicadores multidimensionales de pobreza, no son objetivos a seguir de esta tesis doctoral. El uso entre unas u otras técnicas tienen sus propias ventajas e inconvenientes, por lo que la decisión de considerar una determinada técnica como medición de la pobreza dependerá del objetivo del análisis. Algunas referencias relevantes que debaten a fondo la definición de la pobreza y versan sobre la elección entre medidas de pobreza relativas frente a las medidas absolutas son: Friedman (1965), Watts (1968), Sen (1979), Thon (1979), Blackorby y Donaldson (1980), Chakravarty (1983), Colosanto et al. (1984), Hagenaars (1986), Atkinson (1987), Vaughan (1987), De Vos y Garner (1991), Fernández (1992), Townsend (1992), Citro y Michael (1995), Foster (1998), Madden (2000), Jolliffe (2001), Khandker (2005) pp. 43-52, Domínguez y Martín (2006), Haughton y Khandker (2009), etc. Como se adelantaba en la Sección 1.1, el objetivo principal de este trabajo es la estimación eficiente de algunos indicadores ya existentes de pobreza, quedando completamente fuera de nuestro objetivo la discusión de los temas anteriormente comentados. No obstante, en esta Sección 1.2 se comentan algunas observaciones relacionadas con la medición de la pobreza.

Una de las razones para medir la pobreza es la de poder realizar comparaciones de pobreza entre países, regiones, etc. Ravallion (1992) afirma que, en primer lugar, se debe reconocer que la pobreza existe, y a continuación, debemos preguntarnos cuánta pobreza existe en nuestra población en estudio, dónde están estos individuos considerados como pobres y por qué son pobres. Estas preguntas serán muy importantes para muchos propósitos, incluyendo el análisis y diseño de políticas destinadas a que los individuos clasificados como

Capítulo 1. Introducción

pobres dejen de serlo. Por ello, Ravallion identifica tres preguntas claves a responder a la hora de estudiar la pobreza:

- a) ¿Qué es el bienestar?
- b) ¿A qué nivel de bienestar consideramos que una persona no es pobre?
- c) ¿Cómo agregamos indicadores individuales de bienestar en una medida de pobreza?

Las dos primeras preguntas hacen referencia a la “*identificación del problema*” (qué individuos son pobres y cuáles no). Para ello suele emplearse lo que se denomina línea o umbral de pobreza. Uno de los objetivos de esta tesis no es la estimación de esta línea de pobreza, puesto que este indicador es una medida de pobreza estándar que generalmente viene fijada a través de una cantidad oficial. En otras palabras, asumiremos que el umbral de pobreza viene establecido por la autoridad o el organismo correspondiente, tal como la Oficina de Estadística Europea (EUROSTAT). Por otra parte, la tercera pregunta anterior se relaciona con lo que se denomina “*agregación del problema*” (cuánta pobreza hay). En esta tesis nos centraremos en la agregación del problema con el objetivo de mejorar los distintos indicadores de pobreza abarcados en esta tesis (proporción de pobres, índices de brecha de pobreza y pobreza severa, y cuantiles en la población de pobres para analizar la distribución de éstos).

Al igual que otros indicadores o estudios sociales y económicos, la medición de la pobreza está basada en muestras representativas de una población objeto de estudio. El tamaño de tales muestras no debería ser demasiado pequeño con el fin de poder obtener indicadores de pobreza más precisos, y las muestras se deberían de repetir con una determinada periodicidad para poder conocer en cada momento el fenómeno de la pobreza. En varios países de la Unión Europea, esta encuesta la desarrolla EUROSTAT a través de la EU-SILC. Estos argumentos son un ejemplo de la dificultad que puede suponer la medición de la pobreza.

Capítulo 1. Introducción

A pesar de los inconvenientes que se pueden presentar a la hora de medir la pobreza, existen también numerosas razones por las cuales necesitamos medirla. Como se adelantaba en la Sección 1.1, la medición de la pobreza es un tema de interés relevante en la actualidad por diferentes motivos, de entre los que destacamos:

- a) La creación de políticas destinadas a reducir los índices de pobreza. Para ello, es necesario contar con amplias bases de datos que proporcionen información detallada sobre las características de la población de pobres. Recordamos las palabras de Ravallion (1998), el cual afirmaba que *“medir la pobreza correctamente es un sólido instrumento para centrar la atención de los políticos en las condiciones de vida de los pobres”*. En otras palabras, se podría decir que es fácil ignorar a la población de pobres si ésta no aparece en las estadísticas.
- b) Conocer a quién hay que dirigir las medidas correctoras. Para llevarlo a cabo, se establece un perfil de pobres que permite identificar las características de éstos y estudiar el comportamiento de la pobreza; cómo varía por geografía, por áreas y por características de los hogares. De este modo, el uso de un perfil de pobres redundará en un mejor conocimiento de la población objeto de estudio y por tanto un mayor alcance de las medidas de desarrollo en áreas y hogares más desfavorecidos. Puede verse un buen ejemplo en las encuestas socioeconómicas de Camboya (Prescott y Menino 1997).
- c) Predecir los efectos de políticas y programas de ayuda a los pobres gracias a determinados análisis que se desarrollen en profundidad, los cuales pueden mejorar el diseño de los proyectos o erradicar aquellos programas que no sean viables.
- d) Evaluar a los gobiernos e instituciones, puesto que no se puede determinar que un determinado gobierno esté combatiendo la

Capítulo 1. Introducción

pobreza de manera efectiva a menos que se disponga de la información necesaria.

A modo de ejemplo, el Banco Mundial realiza informes estratégicos para la reducción de la pobreza. A través de los mencionados informes se lleva a cabo la medición de la pobreza y el posterior análisis de la misma para dirigir medidas y programas públicos que tengan un fuerte impacto en la población de pobres, que definan unos indicadores que midan el progreso y que controlen el cambio de una forma sistemática. En resumen, tanto la medición de la pobreza como el análisis y valoración de los programas son fundamentales para establecer políticas de reducción de la pobreza, y en este escenario, la disponibilidad de indicadores precisos contruidos a partir de la información muestral también juega un papel muy importante.

En conclusión, se podría decir que la medición de la pobreza es fundamental, pero no es suficiente en el sentido de que nuevos métodos de análisis podrían paliar o mejorar las condiciones de la población considerada como pobre. En este sentido, resulta imprescindible disponer, a partir de los datos muestrales, de indicadores con mejores propiedades que expliquen de la mejor forma posible la distribución e intensidad real de los pobres así como cuál es la proporción real de pobres. Estos son algunos de los principales objetivos de esta tesis doctoral.

Como se mencionaba con anterioridad, en varios países de la Unión Europea se utiliza la EU-SILC para recoger distintas características de la población, especialmente en el ámbito económico, y poder identificar la población de pobres. En la siguiente sección se describen algunos conceptos relacionados con las encuestas familiares sobre ingresos y/o condiciones de vida.

1.2.2. Encuestas familiares

Para la medición de la pobreza se suelen emplear una serie de indicadores económicos, y para la construcción de tales indicadores necesitamos conocer determinadas condiciones económicas de los individuos o familias de la población objeto de estudio. Por razones obvias, la recogida de esta información a través de un censo resulta inviable, por lo que la mejor opción es recurrir a una muestra. Para el estudio del fenómeno de la pobreza se suelen emplear datos procedentes de encuestas familiares sobre ingresos y condiciones de vida. Por ejemplo, EUROSTAT utiliza para este fin la EU-SILC, y sus resultados son aplicados y están basados en distintos países miembros de la Unión Europea. La realización de la EU-SILC permite poner a disposición de la Comisión Europea de un instrumento estadístico de primer orden para el estudio de la pobreza y la desigualdad, el seguimiento de la cohesión social en el territorio de su ámbito, el estudio de las necesidades de la población y del impacto de las políticas sociales y económicas sobre los hogares y las personas, así como para el diseño de nuevas políticas. En este trabajo utilizaremos datos procedentes de la EU-SILC del 2011. Recordamos que el Apéndice I contiene un estudio descriptivo completo de los datos de esta encuesta y para 8 países de la Unión Europea (Bélgica, Bulgaria, Eslovenia, España, Italia, Lituania, Polonia y Reino Unido).

El principal antecedente de la EU-SILC es el Panel de Hogares de la Unión Europea (PHOGUE), la cual era una encuesta de ingresos en todos los países de la UE que recababa anualmente información detallada sobre personas, hogares y sus ingresos y características socioeconómicas. Con el objetivo de mejorar el PHOGUE, se creó la nueva fuente estadística, la encuesta que estamos denominando EU-SILC. La EU-SILC asegura un mayor nivel de armonización de los datos de la encuesta y permite una mejor medición de la pobreza. Todo esto asegura calidad y comparabilidad entre los países de la

Unión Europea. En España, la EU-SILC recibe el nombre de Encuesta de Condiciones de Vida (ECV).

Teniendo en cuenta que las encuestas son realizadas en los hogares, merece la pena destacar que con esta metodología se está excluyendo a una parte de la población de pobres. En particular, se excluyen a las personas y familias sin hogar. Existe una gran diferencia entre individuos que son “*pobres*” y aquellos que son “*indigentes*” (De Tocqueville 1997). Estos últimos son fácilmente identificados en una población, y a partir de este problema puede decirse que procede el concepto actual de exclusión social.

Por otra parte, también hay que tener en cuenta que en algunos casos se requiere una contribución ponderada de las diferentes áreas y tipos de poblaciones, ya que se tiende a realizar más encuestas (sobre-encuestar) en las poblaciones poco pobladas (áreas rurales), y a realizar una menor proporción de encuestas en poblaciones muy densas. Por ejemplo, tal como recoge Vietnam (2000), en la encuesta del nivel de vida de Vietnam de 1997-1998 (Vietnam Living Standards Survey VLSS98), se recogió una muestra muy elevada de los hogares residentes en áreas rurales y poco pobladas, mientras que por otro lado, se recogió una muestra pequeña del delta del río Rojo, un área densamente poblada.

1.2.3. Enfoques de la pobreza

Las encuestas familiares son la fuente de datos más importante para realizar comparaciones de pobreza, es más, es la única fuente que directamente puede proporcionar la distribución de las condiciones de vida de una sociedad, como cuántas familias no superan los niveles de consumo, etc. Sin embargo, se ha de tener precaución a la hora de interpretar estos datos. En esta sección se presentarán los aspectos más importantes a tener en cuenta.

La pobreza es un fenómeno multidimensional que no tiene una única definición y puede ser objeto de múltiples mediciones, interpretaciones y definiciones dependiendo del punto de vista adoptado, y de ahí la dificultad del enfoque del estudio de la misma. Recordamos la definición de pobreza dada por World Bank (2000): *“la pobreza es una privación pronunciada del bienestar”*. Sen (1987) sostiene que el bienestar es la capacidad de sobrevivir en una determinada sociedad, y de ahí que se defiende que la pobreza surge cuando los individuos carecen de necesidades básicas y poseen bajos ingresos, mala educación, inseguridad, ausencia de libertades, bajo acceso a la sanidad, etc. Por esta razón existen múltiples métodos y definiciones para enfocar el estudio de la pobreza. El objetivo de este apartado es describir distintos enfoques de la pobreza y proporcionar una visión general de la medición de ésta.

Se pueden realizar diversos análisis de pobreza dependiendo del punto de vista adoptado y de los aspectos que interesen resaltar. Para enfocar de una manera correcta la pobreza se exponen algunos tipos de estudios de pobreza:

- a) Según la procedencia de la información distinguimos entre estudios objetivos y estudios subjetivos. La fuente de la información es fundamental a la hora de investigar la pobreza. Los estudios de pobreza objetivos utilizan información recogida mediante variables cuya medición proviene de la observación directa del investigador, lo cual implica un alto grado de objetividad (las variables más usadas son el ingreso y el gasto de los hogares). Por otra parte, los análisis de pobreza subjetiva se basan en la percepción que tienen los propios individuos de su situación. Esta manera de estudiar la pobreza introduce la visión subjetiva que los hogares tienen de su situación económica, frente al enfoque objetivo que utilizan solamente variables medibles y observables.

- b) Utilizando el enfoque objetivo, se pueden realizar tanto análisis de pobreza absolutos como relativos. La pobreza absoluta está altamente relacionada con la miseria y con la carencia de bienes y servicios básicos. Según este concepto de pobreza, un individuo será clasificado como pobre por igual en cualquier país o sociedad del mundo. La pobreza relativa está basada en la sociedad objeto de estudio, es decir, bajo este enfoque una persona será considerada pobre cuando está en clara desventaja (económica y social) con respecto al resto de personas de su entorno. En otras palabras, podemos decir que si bien la pobreza absoluta está ligada a la miseria, la pobreza relativa está ligada con la desigualdad.
- c) También son relevantes los análisis de pobreza persistente o de larga duración. En este caso, el objetivo es realizar estudios dinámicos de pobreza que analicen poblaciones en periodos largos de tiempo y no aisladamente. Según EUROSTAT, un individuo es considerado pobre de forma persistente si es clasificado como pobre el último año y al menos dos de los tres anteriores.

Según el enfoque de pobreza relativa, la clasificación entre pobres y no pobres depende del grado de desarrollo de la sociedad objeto de estudio, y por tanto esta clasificación no puede ser trasladada a otra sociedad diferente. En el estudio de pobreza relativa interviene el concepto de línea o umbral de pobreza que permite clasificar a los individuos como pobres o no pobres dependiendo de a qué lado de la línea de pobreza se encuentran. Para la elaboración del umbral de pobreza es posible utilizar tanto variables basadas en ingresos como basadas en gastos familiares, tal como se explicará detalladamente a continuación en la Sección 1.2.4.

Cada una de estas formas de concebir y medir la pobreza ofrece una perspectiva diferente del mismo fenómeno, proporcionando información muy

rica y variada que debe ser combinada para obtener una visión general lo más completa posible. Por tanto, la combinación de medidas absolutas y relativas puede proporcionar una mejor comprensión de la pobreza.

La pobreza está relacionada con los conceptos de desigualdad y vulnerabilidad. La desigualdad se centra en la distribución de variables como el ingreso o el gasto en una determinada población. La vulnerabilidad se define como el riesgo de caer en la pobreza en el futuro. Este concepto afecta al comportamiento de los individuos en términos de inversión, consumo, etc.

Algunas referencias importantes que definen y comentan con detalle algunos indicadores económicos para la medición de la pobreza así como otros estudios relevantes relacionados con este fenómeno pueden consultarse en Kakwani (1980), Atkinson (1987), Foster (1998), Madden (2000), Jolliffe (2001), Pérez-Mayo y Fajardo (2003), Bárcena y Cowell (2006), Domínguez y Martín (2006), Foster (2007), Silber (2007), Gradín et al. (2008), Núñez (2009), Haughton y Khandker (2009), Pérez-Moreno (2009), etc.

1.2.4. Elección de variable económica y escalas de equivalencia

Otro aspecto importante en el estudio del fenómeno de la pobreza es la elección de la variable a utilizar para la medición de la pobreza. Hay diversos métodos para medir la pobreza. El enfoque utilitarista (Sen 1979) tiene como objetivo la medición de la utilidad de las familias, utilizando variables monetarias como son el gasto y los ingresos de los individuos. Dichas variables son consideradas como contribuciones para generar utilidad. El enfoque de capacidades (Sen, 1979) insiste en el aspecto multidimensional del bienestar, que defiende que la maximización de la utilidad del individuo (enfoque utilitarista) no proporciona información suficiente en términos de grado de

Capítulo 1. Introducción

bienestar, por lo que el enfoque de capacidades centra su atención en si las familias alcanzan los niveles mínimos de alimentación, sanidad, seguridad, etc.

Esta tesis utiliza variables monetarias para la medición de la pobreza. Lo más habitual es recurrir a los ingresos o al gasto familiar. Tanto los ingresos como los gastos presentan inconvenientes a la hora de utilizarlos como variables monetarias en la medición de la pobreza. Siguiendo INE (2005a), a continuación se indican las ventajas, inconvenientes y condiciones por las cuales se prefiere una u otra variable económica.

Por una parte, el ingreso anual refleja la capacidad económica del hogar, aunque se excluyen los bienes y activos que posee el hogar y que también pueden formar parte de la riqueza total del hogar. Tales bienes y activos pueden influir en el nivel de vida que los hogares pueden sostener. Además, los ingresos pueden variar en gran medida de un año a otro sin que esto suponga un cambio en las condiciones de vida. Por ejemplo, éste puede ser el caso de familias que tengan acceso a líneas de crédito, ahorros, etc. Por otra parte, la medición de los ingresos relacionados con actividades agrarias o ganancias en capital es una labor más complicada, puesto que se debería también valorar el incremento del valor de los animales de una granja, los cambios de valor en bienes de capital como una casa, coche, etc. En la mayoría de los casos, el ingreso se infravalora en sociedades con amplia actividad agraria.

Por otro lado, la variable es más estable cuando se toma de referencia el gasto familiar, puesto que, en general, los hogares no modificarían sustancialmente sus hábitos de consumo en caso de una disminución puntual de los ingresos, es decir, que el gasto depende más de la renta permanente, que de los ingresos actuales. La pobreza a su vez está relacionada con la renta permanente y por tanto el gasto sería una buena variable para medirla.

Capítulo 1. Introducción

La elección del gasto como variable monetaria, también presenta desventajas. Se sabe que los hábitos de consumo de los hogares dependen en gran medida del entorno en el que se viva y de las costumbres adquiridas a lo largo del tiempo. Este hecho implicaría que, en muchos casos, los gastos no tienen una relación directa con las características del hogar. Por otra parte, los hogares tienden a no declarar los gastos en productos tal como el alcohol, tabaco, etc. Como ejemplo destacamos que con la encuesta de presupuestos de los hogares de 1972-1973 de los Estados Unidos, los hogares declararon la mitad del gasto en alcohol en relación con la cantidad que publicó la industria del alcohol (Carlson 1974).

En cualquier caso, es preciso tener en cuenta que ambas variables, ingreso y gasto, están sujetas a errores de medida. Se ha comprobado que las cifras de estas variables infravaloran los datos reales. Por tanto, la estimación de la pobreza está sobrevalorada. Esto produce sesgos en la información final utilizada para realizar análisis de pobreza. La mayoría de los países desarrollados usan los ingresos para medir la pobreza, mientras que los países en desarrollo utilizan el gasto. En Europa se viene utilizando el ingreso como variable oficial para la producción de estadísticas sobre la pobreza y exclusión social. Un ejemplo es la EU-SILC llevada a cabo por EUROSTAT, cuyos datos para varios países de la Unión Europea son los utilizados en esta tesis.

En resumen, podemos afirmar que la elección de una variable económica para la medición de la pobreza es un aspecto muy importante, puesto que dicha elección afecta en última instancia a las medidas de pobreza que se proporcionan.

Por otra parte, las comparaciones entre hogares no son perfectas ya que los hogares no solo se diferencian en las variables de gasto e ingreso, sino también en variables como el número de individuos, precios, tiempo destinado al ocio y tiempo libre, etc. Con el fin de poder tratar a los hogares de la misma

Capítulo 1. Introducción

forma, es habitual corregir las variables de ingreso o gasto mediante indicadores tal como el gasto e ingreso per cápita o por unidad de consumo. De este modo, la variable que se suele utilizar en las encuestas elaboradas por los organismos oficiales de estadística es el ingreso por unidad de consumo. La razón de la elección del ingreso por unidad de consumo se debe a que éste tiene en cuenta las economías de escala y la existencia de unidades de consumo equivalentes en el hogar.

En otras palabras, es habitual seguir los siguientes pasos para la obtención de ingresos por unidad de consumo. En primer lugar se obtiene el ingreso total de cada hogar y se escoge la unidad de análisis a utilizar, como por ejemplo individuos. Aunque sean los individuos las unidades a analizar, también se buscará reconocer la influencia del hogar en los individuos. De esta forma, se adjudicará a todas las personas del hogar un ingreso que depende del ingreso total del hogar. A todos los miembros se les adjudica un mismo ingreso, y este ingreso es llamado ingreso por unidad de consumo o ingreso equivalente. Este ingreso es el ingreso total del hogar dividido entre el número de unidades de consumo del hogar. Tal como se comentó anteriormente, el ingreso por unidad de consumo es la variable utilizada en la medición de la pobreza en la EU-SILC.

Para la obtención de las unidades de consumo se utilizan las llamadas escalas de equivalencia. El objetivo de las escalas de equivalencia es determinar qué parte del ingreso del hogar corresponde a cada uno de sus miembros, con el fin de calcular coherentemente un ingreso medio por individuo, reflejando de esta forma la realidad de los hogares. Por ello, se tiene en cuenta las economías de escala, ya que un aumento de miembros de un hogar no implica un mismo aumento proporcional de ingresos para conservar el mismo bienestar. Además, las teorías de unidades de consumo equivalentes sostienen que las pautas de consumo de un niño son diferentes a las un adulto, y por tanto esta diferencia debe reflejarse en el número de unidades de consumo de un hogar.

Existen múltiples técnicas para la obtención de las unidades de consumo. En la Unión Europea se suele utilizar la escala de la Organización para la Cooperación y el Desarrollo Económico (OCDE) modificada. En esta escala, el número de unidades de consumo de un hogar se calcula como la suma del peso que se adjudica a cada miembro. Los pesos que se asignan a cada miembro del hogar están descritos en la Tabla 1.1.

Miembro del hogar	Peso
Primer adulto	1.0
Segundo adulto y siguientes	0.5
Menores de 14 años	0.3

Tabla 1.1. Pesos asignados a los distintos miembros del hogar en la escala OCDE modificada

En otras palabras, el número de unidades de consumo en la escala OCDE modificada viene determinada a través de la siguiente expresión $1 + (a - 1) \times 0.5 + b \times 0.3$, donde a es el número de adultos en el hogar y b es el número de menores de 14 años.

1.3. Algunos indicadores de pobreza

1.3.1. Introducción

El objetivo de esta sección es describir algunos de los indicadores de pobreza más conocidos y utilizados actualmente. Dada la relación con algunas medidas de pobreza, también se describen algunos indicadores de desigualdad. La idea de esta sección es introducir algunas medidas de pobreza y desigualdad, que sirvan de preámbulo a las aportaciones realizadas en este

trabajo, así como justificar o introducir posibles líneas de investigación a partir del resto de medidas no investigadas en este trabajo (véase Capítulo 5).

Como se comentó con anterioridad, el objetivo principal de este trabajo es obtener estimadores más precisos para algunos de los indicadores de pobreza que se van a describir en esta sección. En general, los estimadores propuestos utilizarán información auxiliar en la etapa de estimación con el fin de obtener estimaciones más eficientes. El estudio de las nuevas técnicas de estimación se completará mediante el análisis empírico del comportamiento de los estimadores propuestos y para los distintos indicadores de pobreza. En concreto, en este trabajo se plantearán y definirán nuevos estimadores para la proporción de pobres, para un determinado cuantil en la población de pobres, y para los índices de brecha de pobreza y pobreza severa.

1.3.2. Líneas de pobreza

Las líneas o umbrales de pobreza son indicadores económicos utilizados para clasificar a las personas como pobres o no pobres, dependiendo de a qué lado de la línea se encuentra un individuo, es decir, aquellas personas que tengan un ingreso por unidad de consumo (o unidad de gasto si fuese ésta la variable seleccionada para la determinación del umbral de pobreza) por debajo de la línea de pobreza serán consideradas pobres. Destacamos que las líneas de pobreza se utilizan para la obtención de una parte importante de medidas de pobreza.

En otras palabras, el objetivo de las líneas de pobreza es establecer el gasto o ingreso mínimo que un hogar necesita para salir de la pobreza. Sin embargo, la determinación de una línea de pobreza es una técnica arbitraria, en el sentido de que distintas opciones para la determinación pueden ser factibles y todas ellas pueden ser difíciles de definir. Haughton y Khandker (2009)

Capítulo 1. Introducción

argumentan que las líneas de pobreza pueden verse como un punto de inflexión, en el que en numerosas ocasiones no es relevante. Por ejemplo, en aquellas ocasiones en las que un individuo se encuentra una unidad por debajo de la línea de pobreza, puesto que en este caso, dicho individuo puede pasar de estar en la pobreza a tener un nivel mínimo de bienestar por el simple hecho de situar su nivel de ingresos justo por encima del umbral de pobreza.

A continuación se describen distintos tipos de líneas de pobreza, las cuales se pueden clasificar como absolutas y relativas. La ventaja de las líneas de pobreza absolutas es la obtención de resultados sensibles al desarrollo económico. Es decir, en el caso que la renta en una sociedad aumente, aunque dicho aumento esté distribuido homogéneamente en la población, el porcentaje de pobres calculado con la línea de pobreza absoluta disminuirá. El objetivo de este trabajo no es debatir el uso de un tipo de líneas de pobreza frente a otras, aunque parece relevante mencionar este hecho y aportar las referencias más importantes existentes en la literatura y relacionada con esta discusión.

1.3.2.1. Líneas de pobreza absolutas

Las líneas de pobreza absolutas surgen con el objetivo de analizar el coste de adquirir una cesta de productos esenciales que permitan alcanzar unos niveles mínimos de satisfacción en términos de bienestar y necesidades básicas.

Como ejemplo de línea de pobreza absoluta, el Banco Mundial usa la medida de ingreso de 1.25 dólares estadounidenses per cápita al día como el valor de los recursos mínimos que se necesitan para no considerar que una persona esté inmersa en la pobreza. Esta línea de pobreza se puede utilizar en el contexto mundial en aquellos países con un determinado nivel de vida, entendiéndose por tanto que cualquier persona que viva en tales países con menos de 1.25 dólares al día se puede considerar como pobre. En las Figuras 1.1

Capítulo 1. Introducción

y 1.2 se pueden observar mapas de pobreza mundial en los que se reflejan las proporciones de pobres basados en líneas de pobreza absolutas (fijadas, respectivamente, en 1 y 1.25 dólares americanos diarios).

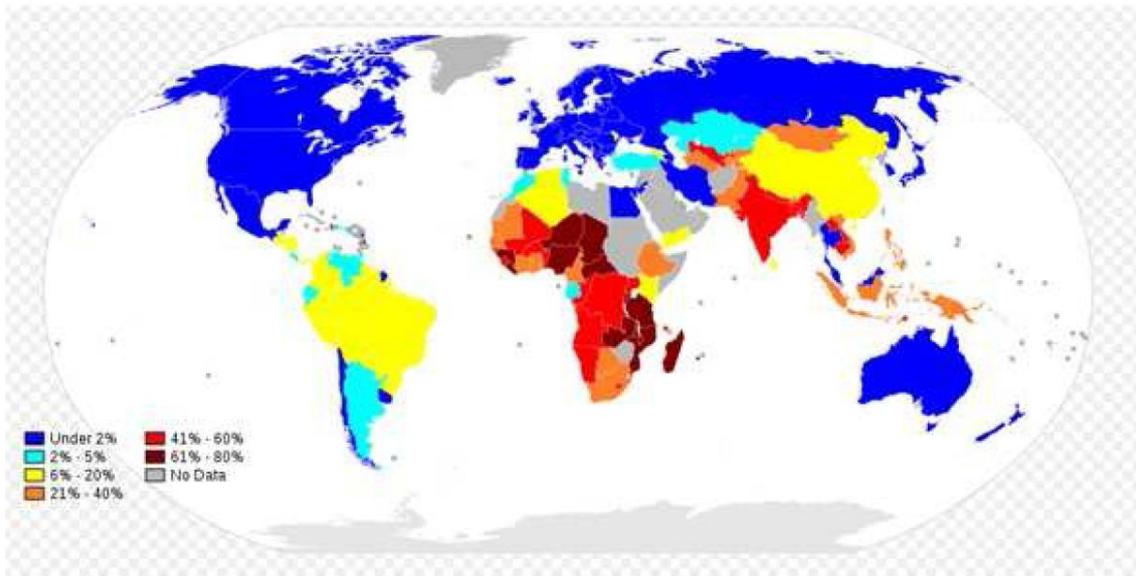


Figura 1.1. Mapa de la pobreza mundial por país, mostrando los porcentajes de población que viven con un dólar americano diario o menos. Basado en el informe de desarrollo humano de la ONU. Fuente: <https://es.wikipedia.org>.

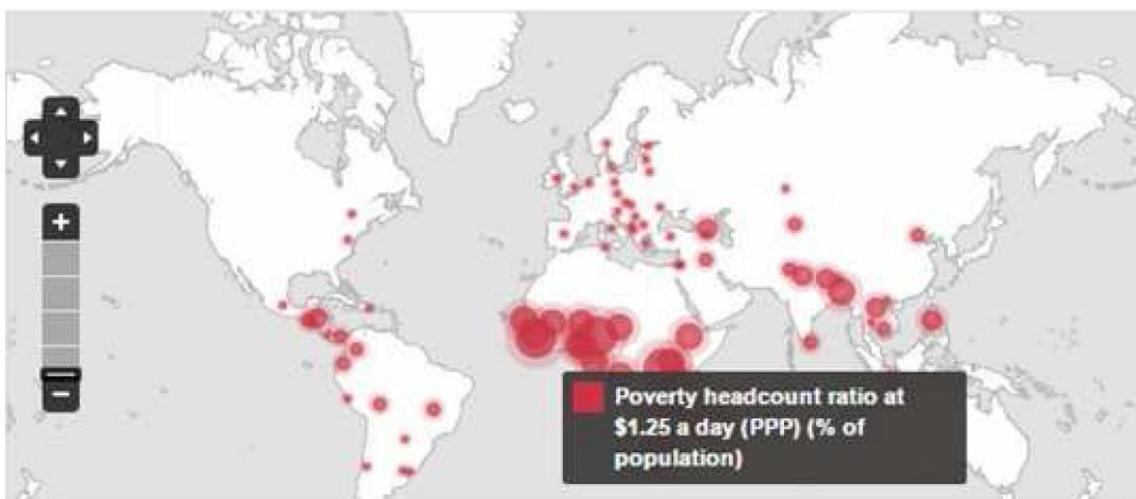


Figura 1.2. Mapa de la pobreza mundial por país, mostrando los porcentajes de población que viven con 1.25 dólares americanos diarios o menos. Fuente: <http://www.bancomundial.org>.

Capítulo 1. Introducción

La construcción de una línea de pobreza absoluta válida para diferentes sociedades y países es bastante difícil. Se han construido muchas de ellas, pero ninguna cumple con los requisitos de una medida de pobreza absoluta pura.

Una de las líneas de pobreza absolutas más conocida es la determinada por Rowntree (1941), el cual definió una línea de pobreza a partir de una cesta de productos compuesta por aquellos bienes y servicios imprescindibles para satisfacer las necesidades mínimas de sustento de los hogares. El umbral de pobreza se establece en el valor monetario de esa cesta más una cantidad fija de dinero destinada a cubrir otro tipo de gastos como la gasolina o el alquiler. Todo hogar que tenga ingresos inferiores a esa cifra sería considerado como pobre.

La línea de pobreza de Rowntree ha sido objeto de numerosas críticas a lo largo de los años. Esto se debe a que a pesar de que las necesidades mínimas alimentarias están bastante consensuadas, el resto de bienes y servicios a incluir en la cesta no lo están tanto. La elección de los productos tiende a depender de la forma de vida de una sociedad concreta, y por lo tanto este enfoque introduce cierta relatividad en la medición de la pobreza.

Destacamos que las líneas absolutas son de interés limitado en los países desarrollados. En los países subdesarrollados y en los países en vías de desarrollo es donde las líneas de pobreza absolutas encuentran más aceptación y son más utilizadas.

1.3.2.2. Líneas de pobreza relativas

Las líneas de pobreza relativas también clasifican a la población de una determinada sociedad en dos grupos: las que se encuentran más desfavorecidas y a las que se denominan pobres; y el resto. A diferencia de las líneas de

pobreza absolutas, si se produce un aumento homogéneo del nivel de ingresos, estas líneas relativas muestran las mismas tasas de pobres antes que después del mencionado aumento. De este modo, aunque el umbral de pobreza sea mayor, la proporción de personas que se encuentran por debajo del umbral de pobreza permanecerá invariable. Para que disminuyan los porcentajes de pobres calculados con este tipo de líneas es necesario que se produzcan cambios en la distribución de la renta. Las líneas de pobreza relativa son muy utilizadas para dirigir programas de ayuda a los más desfavorecidos.

En la práctica, los países más ricos o con un mayor consumo tienen generalmente líneas de pobreza mayores que los países más pobres, tal como se puede observar en la Figura 1.3 elaborada por Chen y Ravallion (2008).

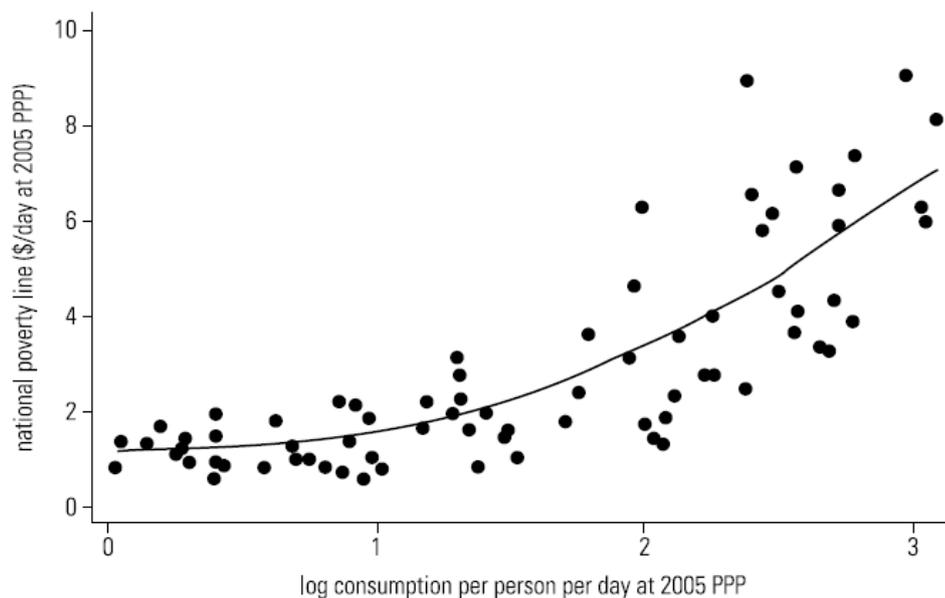


Figura 1.3. Líneas de pobreza relativas (Eje de ordenadas) de distintos países clasificados según el consumo por persona y día (Eje de abscisas). Fuente: Chen y Ravallion (2008).

Capítulo 1. Introducción

En general, los indicadores de pobreza para los que plantearemos nuevos estimadores en esta tesis dependen directamente de un umbral de pobreza relativa, por lo que a continuación se explica el procedimiento que suelen seguir los organismos oficiales de estadística, tal como EUROSTAT, para la determinación de este indicador económico. Como se comentaba anteriormente, esta línea de pobreza relativa es fundamental para el cálculo de la mayoría de indicadores de pobreza.

En primer lugar, asumamos que la población en estudio viene dada por U y contiene N individuos. Una vez elegidas tanto la escala de equivalencia como la variable económica utilizada para la medición de la pobreza (véase la Sección 1.2.4), se selecciona de U una muestra s , con tamaño n , donde se observa dicha variable económica, la cual denominaremos variable de interés y la denotaremos como y . En general, esta variable de interés utilizada para la determinación, entre otros indicadores económicos, de la línea de pobreza relativa es el ingreso por unidad de consumo adjudicado a cada uno de los miembros del hogar, tal como ocurre en la EU-SILC.

EUROSTAT, entre otros organismos oficiales de estadística, define el umbral de pobreza como un porcentaje α de un cuantil de orden β de la variable de interés y , es decir, la línea de pobreza relativa (L) definida a nivel de la población finita U puede expresarse como

$$L = \alpha Y_\beta$$

donde

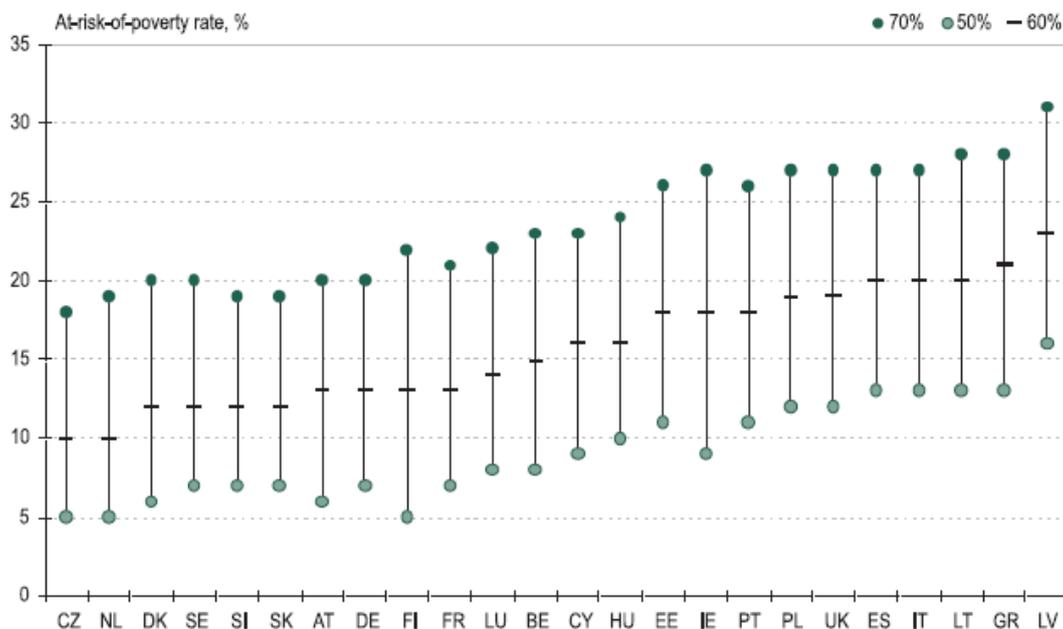
$$Y_\beta = \inf \{t : F(t) \geq \beta\}$$

es el cuantil poblacional de orden β de la variable de interés y , y donde

$$F(t) = \frac{1}{N} \sum_{i=1}^N \delta(y_i \leq t)$$

es la función de distribución de la variable y evaluada en el argumento t . Finalmente, $\delta(\cdot)$ es la función indicadora que toma el valor 1 si su argumento es cierto y dicha función toma el valor 0 en caso contrario.

El Instituto Nacional de Estadística o EUROSTAT consideran $\alpha = 0.6$ y $\beta = 0.5$ para la determinación del umbral de pobreza, es decir, consideran el 60% de la mediana de ingresos por unidad de consumo. En la Figura 1.4 se puede observar un estudio sobre la sensibilidad de la proporción de pobres cuando la línea de pobreza relativa se obtiene a partir de distintos porcentajes α . Podemos observar cómo en algunos países la proporción de pobres es similar a la proporción de pobres de otros países para un determinado porcentaje, pero en el caso de utilizar otros porcentajes, las proporciones de pobres entre tales países puede ser diferente. Véase, por ejemplo, los casos de Finlandia (FI) y Francia (FR).



Source: Eurostat New Cronos database (<http://epp.eurostat.ec.europa.eu/>). Access date: June 2008

Figura 1.4. Estudio de la sensibilidad en la proporción de pobres cuando las líneas de pobreza relativas se toman en los valores $\alpha = \{0.5, 0.6, 0.7\}$. Fuente: Lelkes et al. (2009) y Eurostat New Cronos database (<http://epp.eurostat.ec.europa.eu/>).

Capítulo 1. Introducción

Una vez determinado el umbral de pobreza L , se considera que un individuo es pobre o pertenece a la población de pobres si su ingreso por unidad de consumo es menor o igual que el mencionado umbral de pobreza, es decir, el ingreso por unidad de consumo del i -ésimo individuo verifica $y_i \leq L$. Una vez determinada la línea de pobreza L así como la población que es considerada pobre, es posible determinar los indicadores de pobreza objeto de estudio en esta tesis. En particular, tales indicadores de pobreza son cuantiles de distintos órdenes en la población de pobres, la propia proporción de pobres en la población, y los índices de brecha de pobreza y pobreza severa, etc. Tales indicadores de pobreza se describen brevemente a continuación.

La estimación eficiente de la línea de pobreza L definida anteriormente es un tema relativamente sencillo de analizar a partir de la literatura existente, puesto que para ello, y siguiendo la definición de L , lo único que se requiere para obtener estimaciones más precisas para la línea de pobreza L es utilizar estimadores más eficientes para un cuantil. Destacamos que este tema se ha estudiado intensamente en los últimos años. A su vez y siguiendo la definición de un cuantil, destacamos que para la estimación eficiente de cuantiles se requieren estimadores eficientes de la función de distribución. Algunas referencias relevantes sobre la estimación eficiente de cuantiles y funciones de distribución son Chamber y Duntan (1986), Kuk (1988), Rao et al. (1990), Kuk y Mak (1994), Silva y Skinner (1995), Wang y Dorfman (1996), Rueda et al. (2007), Singh et al. (2007, 2008, 2011), Berger y Muñoz (2015). Una reciente revisión de los distintos estimadores existentes para la función de distribución y cuantiles puede consultarse en Dorfman (2009). Por las razones que se acaban de comentar, en esta tesis no se plantea la estimación eficiente de la línea de pobreza L , dada la abundante literatura existente sobre este tema.

1.3.3. Cuantiles en la población de pobres

Una vez determinada la línea de pobreza es posible la obtención de la una gran mayoría de indicadores de pobreza, los cuales pueden proporcionar un conocimiento más exhaustivo del fenómeno de la pobreza. Uno de estos indicadores es la obtención de una serie de cuantiles en la población de pobres.

Además de conocer las personas que son o no pobres, también es importante conocer cómo de pobres son las personas pobres, es decir, analizar la profundidad de la pobreza. En otras palabras, el hecho de utilizar medidas de pobreza, tal como la proporción de pobres, no nos aporta información del grado de pobreza de los pobres, por lo que resulta necesario complementarlas con medidas de pobreza que analicen, por ejemplo, la profundidad de la pobreza.

Las medidas de intensidad de la pobreza nos informan de cómo afecta la pobreza a la población y se centran en el grado de pobreza que sufren las personas más que el número de individuos considerados pobres. La intensidad de la pobreza es uno de los factores que más influyen en la gravedad del fenómeno de la pobreza. Es por tanto necesario utilizar conjuntamente con las medidas relativas algún indicador de la profundidad de la pobreza que proporcione información sobre la situación económica de los pobres y sus diferencias con el resto de la población.

Mediante la utilización conjunta de las medidas de incidencia e intensidad de la pobreza se puede describir con más detalle lo que ocurre en una sociedad. Por ejemplo, pueden existir sociedades con muchos pobres cerca del umbral de pobreza y sociedades con pocos pobres pero muy alejados del umbral de pobreza. Todas estas medidas son indispensables para analizar en profundidad la pobreza.

Una forma de analizar la profundidad de la pobreza, es a través del cálculo de distintos cuantiles en la población de pobres. Por ejemplo, en INE (2005b) utilizan los deciles para analizar la profundidad de la pobreza. En particular, este tipo de estudio tiene el objetivo de proporcionar una idea de la intensidad de la pobreza y de la desigualdad existente entre los individuos pobres.

Según INE (2005b) y con datos de la Encuesta de Condiciones de Vida del año 2004, los deciles de los ingresos por unidad de consumo de la población clasificada como pobre pueden observarse en la Figura 1.5. A partir de estos resultados puede derivarse conclusiones relevantes para la población pobre. Por ejemplo, cabe destacar que para el mencionado año un 20% de los pobres (aproximadamente un 4% de la población total) están situados por debajo de la mitad del umbral de pobreza, es decir, tienen unos ingresos por unidad de consumo inferiores a 3139.4 euros anuales, lo que equivale a 259.7 euros al mes. El resto de los pobres (el 80% restante) se reparte casi en su totalidad y de forma más o menos homogénea en la segunda mitad del intervalo (0, 6278.7), es decir, tienen unos ingresos por unidad de consumo superiores a 3139.4 euros anuales y por debajo de la línea de pobreza situada en 6278.7.

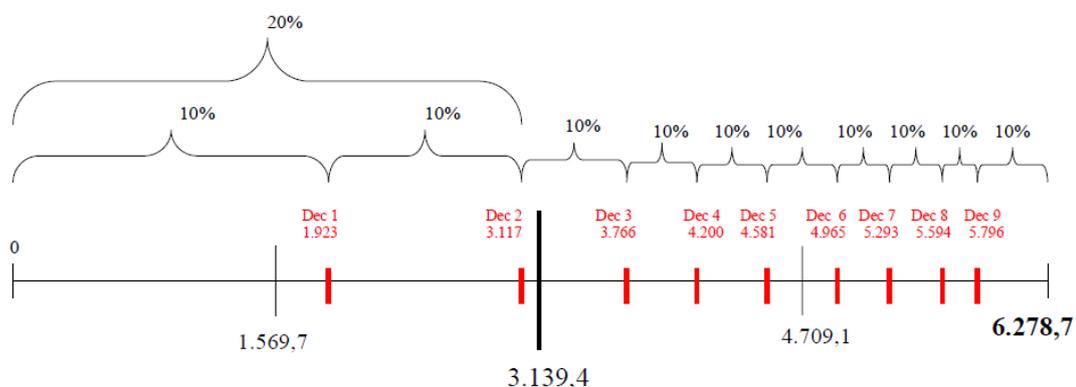


Figura 1.5. Deciles de los ingresos por unidad de consumo de la población clasificada como pobre. Fuente: INE (2005b).

Este estudio de los deciles en la población pobre realizado en INE (2005b) nos permite hacernos una idea de la importancia de los cuantiles en el estudio del fenómeno de la pobreza, y en particular, en los estudios relacionados con la distribución y la intensidad de la pobreza. Destacamos que éste es el objetivo que se persigue en el Capítulo 2 de la presente memoria. En particular, en dicho capítulo se realizan nuevos estudios de simulación para distintos estimadores de cuantiles en la población considerada como pobre. El problema de la estimación de la función de distribución en esta población de pobres también se aborda en este capítulo. Destacamos que esta investigación fue la primera llevada a cabo de entre las distintas aportaciones de este trabajo, y aunque tan sólo se llevaron a cabo nuevos estudios de simulación, tales aportaciones fueron fundamentales para el desarrollo del resto de esta tesis doctoral.

1.3.4. Proporción de pobres

El siguiente indicador de pobreza que resulta obvio obtener es la proporción de individuos que se encuentran por debajo de la línea de pobreza relativa L . A este indicador se le conoce como proporción de pobres (P_0) o tasa de pobreza. Este indicador de pobreza es muy popular porque es muy fácil de interpretar y medir. En el lado opuesto, el inconveniente de la proporción de pobres es que no indica el nivel de pobreza de la población. En el Capítulo 3 de esta tesis se analiza el problema de la estimación de la proporción pobres mediante nuevas técnicas cuantitativas. No obstante, con el fin de aportar en este momento una definición simple que permita entender el concepto de proporción de pobres, podemos decir que este indicador puede expresarse mediante la siguiente expresión

$$P_0 = \frac{N_p}{N},$$

Capítulo 1. Introducción

donde N_p denota el número de individuos considerados como pobres en toda la población U , y N se definió anteriormente como el tamaño total de la población en estudio. El objetivo del Capítulo 3 será la estimación del parámetro P_0 . Las Figuras 1.1, 1.2 y 1.4 descritas con anterioridad son ejemplos de representaciones gráficas que aportan información relacionada con la proporción de pobres.

La tasa de pobreza presenta algunos inconvenientes. Por ejemplo, la tasa de pobreza no tiene en cuenta la intensidad de pobreza, tal como puede comprobarse en un simple ejemplo con los datos proporcionados en la Tabla 1.2. El objetivo de este ejemplo es ilustrar y comparar distintas medidas de pobreza en distintas situaciones extremas o bajo distintos escenarios.

Poblaciones					
A	B	C	D	E	F
100	400	100	100	400	100
100	400	250	100	400	400
100	400	400	100	400	700
750	750	750	7500	7500	7500
1000	1000	1000	10000	10000	10000
1000	1000	1000	10000	10000	10000
1000	1000	1000	10000	10000	10000

Tabla 1.2. Ingresos por unidad de consumo, en euros, en distintas poblaciones con tamaño $N = 7$ en cada una de ellas.

La Tabla 1.2 presenta datos ficticios de ingresos por unidad de consumo en distintas poblaciones con tamaño poblacional $N = 7$ en cada una de ellas. Siguiendo la definición de línea de pobreza dada por EUROSTAT, las líneas de pobreza relativas coinciden en el primer grupo de poblaciones (A, B y C), y vienen dadas por:

$$L_A = L_B = L_C = 0.6 \times Y_{0.5} = 0.6 \times 750 = 450$$

Por su parte, en el segundo grupo de poblaciones (D, E y F), las líneas de pobreza también coinciden, y ésta se sitúa en

$$L_D = L_E = L_F = 0.6 \times Y_{0.5} = 0.6 \times 7500 = 4500$$

Recordamos que se trata de situaciones extremas con el propósito de analizar el valor de los distintos indicadores de pobreza analizados en esta tesis doctoral, y poder realizar comparaciones entre las distintas poblaciones.

El hecho de que las líneas de pobreza coincidan para determinadas poblaciones implica que la proporción de pobres en cada población también sea la misma. En particular, en cada población habría 3 individuos clasificados como pobres, lo que supone una proporción de pobres del 42.9% ($P_{0,A} = P_{0,B} = \dots = P_{0,F} = 0.429$). Sin embargo, a raíz de los datos de la Tabla 1.2 y comparando las 3 primeras poblaciones, puede comprobarse fácilmente que la pobreza es mayor en la población A en comparación con las poblaciones B y C, es decir, los individuos pobres en la población A tienen menos recursos económicos que los pobres de las poblaciones B y C. La misma situación ocurre en las poblaciones D, E y F, donde los pobres de la población D están en desventaja económica en comparación con los pobres de las poblaciones E y F. Comparando ambos grupos de poblaciones, se puede observar fácilmente que los pobres del segundo grupo (poblaciones D, E y F) tienen una mayor desventaja económica en comparación con los no pobres, es decir, esta diferencia entre pobres y no pobres tiene un menor impacto en las poblaciones A, B y C. En resumen, la tasa de pobreza o proporción de pobres, por sí misma, no es medida apropiada como herramienta cuantitativa para la medición del bienestar, puesto que no permite conocer la situación o la distribución de los pobres. En otras palabras, la tasa de pobreza no cumple con el principio de

Capítulo 1. Introducción

transferencias propuesto por Dalton (1920), el cual expone que si hay solo dos receptores de renta, y se produce una transferencia del más rico al más pobre, la desigualdad tendría que disminuir.

Siguiendo con los datos de la Tabla 1.2, el uso de cuantiles en la población de pobres, lo cual se trata en el Capítulo 2, si permite conocer mejor la distribución de los pobres. En la población A, todos los cuantiles (deciles por ejemplo) que utilizemos en la población de pobres tendrán un valor de 100 €, cantidad alejada del umbral de pobreza situado en 450 €. Por su parte, cualquier cuantil en la población de pobres en la sociedad B tendrá un valor de 400 €, lo cual indica que aunque la proporción de pobres sea la misma en las dos poblaciones, la situación de los pobres en la población B es considerablemente mejor. Un gráfico de deciles para pobres tal como el ilustrado en la Figura 1.5 nos permitiría poder realizar una mejor comparación entre las distintas poblaciones. También puede resultar útil la interpretación de los cuantiles obtenidos para la población de pobres (véase más adelante la Tabla 1.9) en cada una de las poblaciones. Los índices de brecha de pobreza y pobreza severa son medidas de pobreza que también permiten conocer mejor la situación de los pobres. Estas dos medidas de pobreza se tratan en el Capítulo 4 de esta tesis.

Puesto que la tasa de pobreza no mide el nivel de pobreza de la población de pobres, dicha tasa de pobreza no cambia si la población que vive por debajo del umbral de pobreza pasa a ser más pobre (véase por ejemplo los ingresos de las poblaciones A y B). En relación con este último argumento, una forma simple de disminuir la tasa de pobreza es beneficiar a los individuos más cercanos a la línea de pobreza. Sin embargo, aunque este procedimiento nos haría disminuir la proporción de pobres, en realidad los que más necesitan ayudas sociales son aquellos que están más alejados de ella. Aunque el hecho de dar ayudas sociales a los más pobres no implica una disminución de la proporción de pobres, si es cierto que la situación de éstos mejoraría y el problema de la pobreza sería menor. En definitiva, la proporción de pobres es

Capítulo 1. Introducción

un buen indicador económico para la pobreza, aunque para un análisis más completo de la pobreza es imprescindible que esté acompañado de otros indicadores de pobreza complementarios.

El Capítulo 3 de esta memoria está destinado a la obtención de estimadores más precisos para la proporción de pobres. Los estimadores propuestos asumen que existe información auxiliar relacionada con la variable de interés y , y que dicha información puede utilizarse en la etapa de estimación con el objetivo de obtener estimaciones más precisas de este indicador de pobreza. Tanto la estimación de la varianza del estimador propuesto como la estimación de la tasa de pobreza mediante intervalos de confianza son objetivos que también se persiguen en este Capítulo 3. Finalmente, todos los métodos de estimación se evalúan numéricamente mediante estudios de simulación Monte Carlo basados en poblaciones reales extraídas de encuestas de ingresos y condiciones de vida, así como basados en poblaciones artificiales con distintas características. En el Apéndice I de esta tesis se describen todas las poblaciones utilizadas en este trabajo y se proporcionan los valores de las distintas medidas de pobreza para estas poblaciones.

Con el fin de resumir los últimos datos disponibles de la proporción de pobres en regiones y países de la Unión Europea, en la Figura 1.6 se presentan Los valores de este indicador económico para el año 2013 (último año disponible antes de finalizar la realización de esta tesis doctoral). A partir de esta tabla podemos observar cómo Bulgaria es el país con una mayor proporción de pobres en torno al 48%. En el lado opuesto se encontrarían países como Islandia, Noruega y la República Checa con proporciones de pobres por debajo del 15%.

Capítulo 1. Introducción

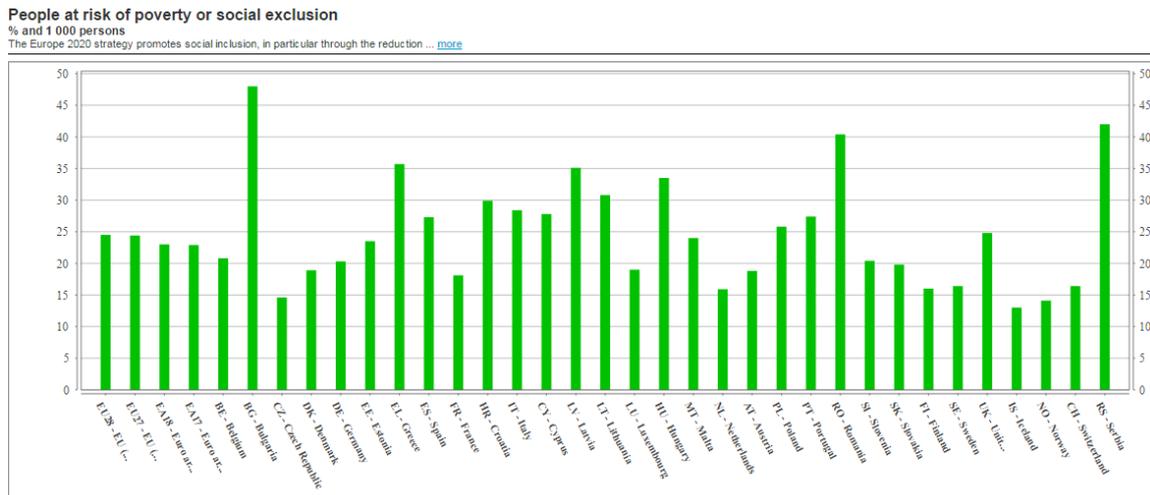


Figura 1.6. Proporción de pobres en distintas regiones y países de la Unión Europea en el año 2013. Representación gráfica generada desde EUROSTAT

(<http://ec.europa.eu/eurostat>)

1.3.5. Índice de brecha de pobreza

La brecha de pobreza es un indicador económico que mide la distancia de los individuos clasificados como pobres al umbral de la pobreza. Dicho de otra manera, mide la cantidad media de dinero que necesitarían los individuos para dejar de ser pobres. Este indicador nos permite analizar la profundidad y la intensidad de la pobreza.

La brecha de pobreza para un individuo se define como el umbral de pobreza L menos el ingreso del individuo perteneciente a la población pobre y_i . Hay que tener en cuenta que la brecha es considerada cero para aquellos que no pertenezcan a la población de pobres. La brecha de pobreza se suele expresar en términos porcentuales, por lo que la distancia anterior habría que dividirla entre el umbral de pobreza. El índice de brecha de pobreza para una

Capítulo 1. Introducción

población, la cual denominaremos como P_1 , se define por tanto como la media de estos porcentajes, es decir,

$$P_1 = \frac{1}{N} \sum_{i=1}^N \left(\frac{L - y_i}{L} \right) \delta(y_i \leq L)$$

Destacamos que en la expresión que define el indicador P_1 se incluye la variable indicadora $\delta(\cdot)$ con el fin de discriminar a los pobres entre los que no lo son.

Siguiendo con el ejemplo expuesto en la Tabla 1.2, a continuación se obtienen los índices de brecha de pobreza en las distintas poblaciones descritas en la Tabla 1.2. El objetivo es poder analizar y comparar los índices de brecha de pobreza en distintas situaciones extremas.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L} \right) \delta(y_i \leq L)$
100	350	0.778
100	350	0.778
100	350	0.778
750	0	0
1000	0	0
1000	0	0
1000	0	0
Media (Índice de brecha de pobreza)		0.333

Tabla 1.3. Obtención del índice de brecha de pobreza en la población A de la Tabla 1.2.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L}\right)\delta(y_i \leq L)$
400	50	0.111
400	50	0.111
400	50	0.111
750	0	0
1000	0	0
1000	0	0
1000	0	0
Media (Índice de brecha de pobreza)		0.048

Tabla 1.4. Obtención del índice de brecha de pobreza en la población B de la
Tabla 1.2.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L}\right)\delta(y_i \leq L)$
100	350	0.778
250	200	0.444
400	50	0.111
750	0	0
1000	0	0
1000	0	0
1000	0	0
Media (Índice de brecha de pobreza)		0.190

Tabla 1.5. Obtención del índice de brecha de pobreza en la población C de la
Tabla 1.2.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L}\right)\delta(y_i \leq L)$
100	4400	0.978
100	4400	0.978
100	4400	0.978
7500	0	0
10000	0	0
10000	0	0
10000	0	0
Media (Índice de brecha de pobreza)		0.419

Tabla 1.6. Obtención del índice de brecha de pobreza en la población D de la
Tabla 1.2.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L}\right)\delta(y_i \leq L)$
400	4100	0.911
400	4100	0.911
400	4100	0.911
7500	0	0
10000	0	0
10000	0	0
10000	0	0
Media (Índice d brecha de pobreza)		0.390

Tabla 1.7. Obtención del índice de brecha de pobreza en la población E de la
Tabla 1.2.

y_i	$(L - y_i)\delta(y_i \leq L)$	$\left(\frac{L - y_i}{L}\right)\delta(y_i \leq L)$
100	4400	0.978
400	4100	0.911
700	3800	0.844
7500	0	0
10000	0	0
10000	0	0
10000	0	0
Media (Índice de brecha de pobreza)		0.390

Tabla 1.8. Obtención del índice de brecha de pobreza en la población F de la Tabla 1.2.

A partir de los resultados obtenidos en las tablas anteriores podemos observar, como parece razonable a partir de la correspondiente definición, que valores del índice de brecha de pobreza próximos a 0 indican que los ingresos de los pobres están próximos al umbral de pobreza. En este sentido, los pobres de la población B son los que se encuentran en una mejor situación en comparación con el resto de poblaciones. Los ingresos de los pobres de la población E son los mismos que los ingresos de la población B, pero evidentemente la brecha de pobreza no puede ser la misma al tratarse la población E de una sociedad donde el umbral de pobreza está mucho más alto, es decir, el nivel de vida es mayor. Como parece razonable, la peor situación de los pobres se encuentra en la población D. Este hecho lo recoge la brecha de pobreza, dado que es en esta población donde se registra el mayor valor para este indicador de pobreza (41.9%). En otras palabras, a medida que aumenta el índice de brecha de pobreza, peor será la situación de los pobres en comparación con la población considerada como no pobre.

El índice de brecha de pobreza puede considerarse un indicador del ahorro potencial para la reducción de la pobreza, es decir, cuanto menor sea el ratio de la brecha de pobreza, mayor potencial tendrá una economía para identificar las características de los pobres, y por tanto para dirigir medidas y políticas.

El índice de la brecha de pobreza también incumple el principio de transferencias de Dalton, lo cual puede observarse a partir de las poblaciones E y F. En tales poblaciones, el índice de la brecha de pobreza es del 39%, pero puede pensarse que el problema de la pobreza en la población F es mayor debido al hecho de tener un individuo más sumido en la pobreza y en comparación con la población E. El índice de pobreza severa que describimos brevemente a continuación si permite discriminar entre la situación de los pobres en las poblaciones E y F.

1.3.6. Índice de pobreza severa

La brecha de pobreza severa es una medida de pobreza que tiene en cuenta la desigualdad en la población de pobres. Se construye elevando la brecha de pobreza de cada individuo al cuadrado. Con esta operación se pretende dar más peso en las observaciones que están más alejadas del umbral de pobreza. Destacamos que el índice de brecha de pobreza pondera por igual todas las observaciones.

El índice de pobreza severa, el cual denotaremos como P_2 , para una determinada población U con N individuos puede definirse a través de la siguiente expresión

$$P_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{L - y_i}{L} \right)^2 \delta(y_i \leq L).$$

Resaltamos que tanto la proporción de pobres como los índices de brecha de pobreza y pobreza severa pertenecen a la familia de indicadores de pobreza propuesta por Foster, Greer y Thorbecke (1984), también conocida como la familia de indicadores de pobreza FGT, y la cual tiene la siguiente expresión

$$P_\gamma = \frac{1}{N} \sum_{i=1}^N \left(\frac{L - y_i}{L} \right)^\gamma \delta(y_i \leq L),$$

donde la proporción de pobres P_0 es un caso particular de esta familia de medidas de pobreza. En efecto, se puede comprobar que P_0 viene definida por P_γ cuando $\gamma=0$. El índice de brecha de pobreza se define cuando $\gamma=1$, mientras que al asignar el valor $\gamma=2$ en P_γ obtendremos el índice de pobreza severa.

En la Tabla 1.9 pueden observarse todos los indicadores de pobreza que se analizan en esta memoria y para las 6 poblaciones del ejemplo de la Tabla 1.2, es decir, se obtienen los cuartiles en la población de pobres $(Y_{0.25}^*, Y_{0.5}^*, Y_{0.75}^*)$, la proporción de pobres P_0 , el índice de brecha de pobreza P_1 , y el índice de pobreza severa P_2 . Al igual que en el caso del índice de brecha de pobreza, valores del índice de pobreza severa próximos a 0 indicarán una situación económica más favorable para los individuos clasificados como pobres. En este sentido, la población donde los pobres gozan de unas mejores condiciones es la población B, donde el índice de pobreza severa se sitúa en el 0.5%. Merece la pena destacar que, a diferencia del índice de brecha de pobreza, el índice de pobreza severa si realiza una discriminación entre los pobres de la población E y los pobres de la población F, es decir, según el índice de brecha de pobreza, la situación de los pobres en las poblaciones E y F sería la misma, puesto que en

ambos casos este indicador de pobreza se sitúa en el 39%. Sin embargo, según el índice de pobreza severa, la situación de los pobres en la población F es ligeramente peor que la situación de los pobres en la población E, puesto que el correspondiente valor del índice de pobreza severa en la población F es ligeramente mayor. Estas conclusiones son coherentes con los ingresos de los pobres de las poblaciones E y F que pueden observarse en la Tabla 1.2.

Población	$Y_{0.25}^*$	$Y_{0.5}^*$	$Y_{0.75}^*$	$P_0(\%)$	$P_1(\%)$	$P_2(\%)$
A	100	100	100	42.9	33.3	25.9
B	400	400	400	42.9	4.8	0.5
C	100	250	400	42.9	19.0	11.6
D	100	100	100	42.9	41.9	41.0
E	400	400	400	42.9	39.0	35.6
F	100	400	700	42.9	39.0	35.7

Tabla 1.9. Valores de los cuartiles en la población considera como pobre ($Y_{0.25}^, Y_{0.5}^*, Y_{0.75}^*$), proporción de pobres, brecha de pobreza y pobreza severa para las distintas poblaciones de la Tabla 1.2.*

1.3.7. Otros indicadores de pobreza y desigualdad

La función principal de las medidas de pobreza descritas anteriormente es conocer la incidencia o intensidad de la pobreza. Sin embargo, existen otras técnicas cuantitativas que permiten completar el estudio del fenómeno de la pobreza. En su mayoría se trata de indicadores de pobreza más complejos y más difíciles de interpretar que las medidas de pobreza investigadas con profundidad en este documento. La particularidad de la mayoría de medidas de pobreza es que tratan de proporcionar información de los tres factores

esenciales en la pobreza: su incidencia, su intensidad y la desigualdad entre los pobres. Las medidas de desigualdad también son muy utilizadas en estudios de pobreza. De hecho, algunos indicadores de pobreza dependen de medidas de desigualdad tal como el índice de Gini. A continuación se presentan algunas medidas de pobreza y desigualdad también utilizadas en muchos estudios de pobreza, desigualdad, exclusión social, etc. Dado que estas medidas no son objeto de estudio en esta tesis, debemos tener en cuenta que la intención no es ni describirlas en detalle ni aportar una revisión exhaustiva de todas las medidas y enfoques de la pobreza. Para ello, se pueden consultar referencias como Kakwani (1980), Atkinson (1987), Foster (1998), Madden (2000), Jolliffe (2001), Pérez-Mayo y Fajardo (2003), Bárcena y Cowell (2006), Domínguez y Martín (2006), Foster (2007), Silber (2007), Gradín et al. (2008), Núñez (2009), Haughton y Khandker (2009), Pérez-Moreno (2009), etc.

1.3.7.1. Brecha de ingreso

La brecha de ingreso es una medida utilizada para la intensidad de la pobreza, la cual viene definida por

$$\begin{aligned}
 I &= \frac{P_1}{P_0} = \frac{1}{NP_0} \sum_{i=1}^N \left(\frac{L - y_i}{L} \right) \delta(y_i \leq L) = \frac{1}{NP_0 L} \sum_{i=1}^N (L - y_i) \delta(y_i \leq L) = \\
 &= \frac{1}{N^* L} \left(\sum_{i=1}^N L \delta(y_i \leq L) - \sum_{i=1}^N y_i \delta(y_i \leq L) \right) = \frac{L \sum_{i=1}^N \delta(y_i \leq L)}{N^* L} - \frac{\sum_{i=1}^N y_i \delta(y_i \leq L)}{N^* L} = \\
 &= 1 - \frac{\bar{y}^*}{L}
 \end{aligned}$$

donde

$$N^* = N \times P_0 = \sum_{i=1}^N \delta(y_i \leq L)$$

denota el número de pobres en la población y

$$\bar{y}^* = \frac{1}{N^*} \sum_{i=1}^N y_i \delta(y_i \leq L)$$

es la media aritmética de los ingresos de la población de pobres o bien los ingresos mínimos, en media, que tendrían que tener los pobres para dejar de serlos. Para más información de este indicador de pobreza se puede consultar INE (2005b).

1.3.7.2. Brecha de pobreza utilizada en EUROSTAT

EUROSTAT define la brecha de pobreza BP_E como la diferencia, en términos relativos, entre el umbral de pobreza y la mediana de los ingresos de la población clasificada como pobre, es decir,

$$BP_E = \frac{L - Y_{0.5}^*}{L}$$

En la Figura 1.7 podemos observar los valores de la medida BP_E en el año 2013 para distintas regiones y países de la Unión Europea. A partir de esta figura se observa que Bulgaria y Serbia tienen una mayor brecha de pobreza, lo cual indica que en estos países hay una mayor distancia entre el correspondiente umbral de pobres y la mediana de los ingresos por unidad de consumo de los individuos clasificados como pobres. En el lado opuesto, los países donde la población pobre parece estar en una mejor situación son Holanda, Finlandia y Noruega, en los cuales la brecha de pobreza BP_E se sitúa por debajo del 13%.

Capítulo 1. Introducción

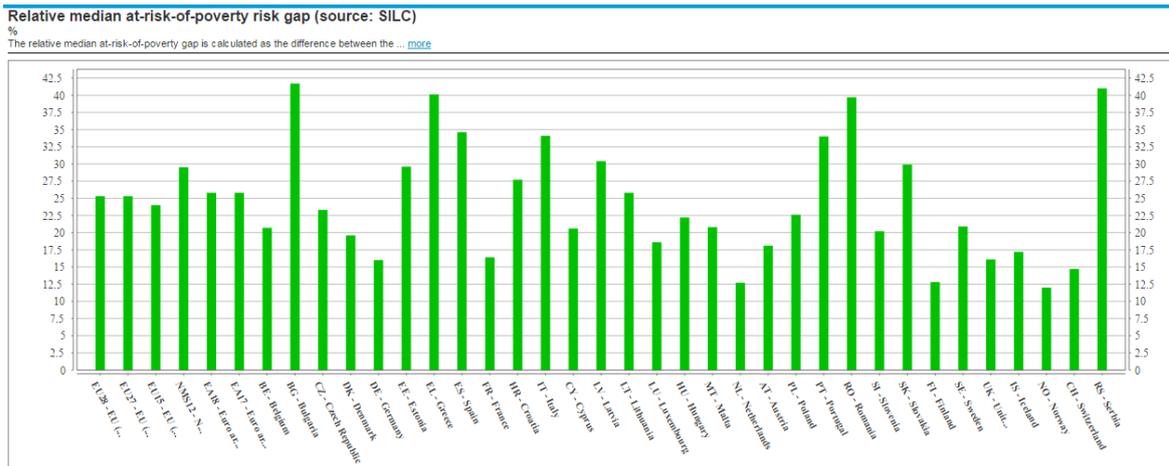


Figura 1.7. Brecha de pobreza en distintas regiones y países de la Unión Europea en el año 2013. Representación gráfica generada desde EUROSTAT (<http://ec.europa.eu/eurostat>)

1.3.7.3. Ratio o razón entre dos cuantiles

Los cuantiles proporcionan información muy valiosa y detallada en el análisis de la distribución de ingresos o gastos de una determinada población. Por ejemplo, en la Sección 1.3.3, y en particular en la Figura 1.5, se describe un ejemplo de la aplicación del uso de cuantiles para analizar la distribución de los pobres de una población. En ese sentido, también es común el uso del cociente entre dos cuantiles para analizar las diferencias entre dos grupos de la población. El ratio o razón entre dos cuantiles de órdenes α_1 y α_2 (véase Dickens y Manning 2004, EUROSTAT 2000, Rueda y Muñoz 2011, etc) se define como

$$r_{\alpha_1, \alpha_2} = \frac{Y_{\alpha_1}}{Y_{\alpha_2}}$$

Por ejemplo, el ratio $r_{0.8, 0.2}$ compara cuánto más rico es el individuo que delimita el 20% de la población más rica con el individuo que delimita el 20% de la población más pobre. En la Figura 1.W8 se presentan los valores de la

Capítulo 1. Introducción

medida $r_{0,8,0,2}$ para distintas regiones y países de la Unión Europea. A modo de ejemplo, observamos que en la República Checa, Islandia y Noruega habría una baja desigualdad entre estas subpoblaciones, dado que el 20% de los más ricos de estos países tienen, aproximadamente, unos ingresos 3.5 veces mayores que el 20% de los más pobres de cada correspondiente país. En el lado opuesto, destacamos que este cociente se sitúa en 8.6 en Serbia, seguido de Bulgaria, Grecia y Rumania, con valores de este indicador económico en torno a 6.5.

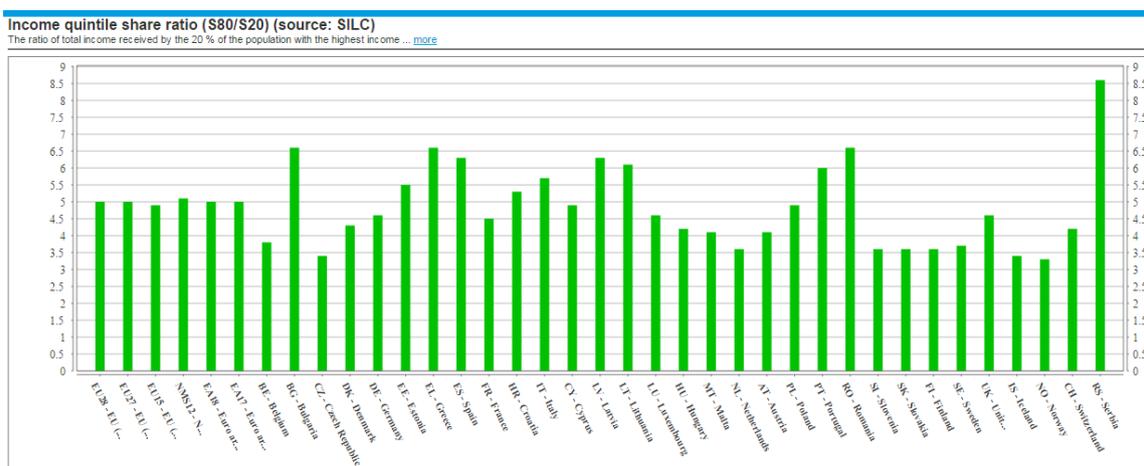


Figura 1.8. Ratio entre el cuarto y el primer quintil ($r_{0,8,0,2}$) en distintas regiones y países de la Unión Europea en el año 2013. Representación gráfica generada desde EUROSTAT (<http://ec.europa.eu/eurostat>)

Otra aplicación o uso de este indicador económico la hemos encontrado en un estudio sobre la variación de la desigualdad publicado en el diario “El País” el 5 de diciembre de 2014 (véase la Figura 1.9), o bien la web de procedencia de este estudio:

http://economia.elpais.com/economia/2014/12/05/actualidad/1417770534_603645.html

En este caso observamos que se utiliza el ratio $r_{0,9,0,1}$ para analizar la desigualdad de distintos países, con la finalidad de estudiar posteriormente la variación de esta medida entre los años 2006 y 2010. A partir de esta figura

Capítulo 1. Introducción

podemos observar cómo la mayor variación se ha producido en España y en Estados Unidos. En el lado opuesto, se ha reducido la desigualdad, según esta medida y en el periodo mencionado, en Rumania y Bulgaria. En este estudio también se han investigado las razones principales de este aumento de desigualdad en España, destacando como principal razón el cambio en los salarios.

Como consecuencia de los resultados y conclusiones derivados del último estudio, se ha incluido en esta sección la evolución del ratio $r_{0.8,0.2}$ entre los años 2005 y 2013 y para distintos países de la Unión Europea (véase Figura 1.10). En efecto, se observa cómo España ha sufrido una mayor variación de este indicador económico en el mencionado periodo y entre los distintos países considerados en esta comparación. En este caso, la variación absoluta de la medida $r_{0.8,0.2}$ en España y durante el periodo 2005 y 2010 ha sido de 1.7 (ha pasado de un valor de 5.5 en el año 2005 a un valor de 7.2 en el año 2010). En términos relativos, esta medida ha variado un 30.9% entre 2005 y 2010. En los años 2011 y 2012 esta medida ha permanecido prácticamente constante, mientras que en el año 2013 ha sufrido una ligera disminución.

VARIACIÓN DE LA DESIGUALDAD

Cambio en la desigualdad entre el 10% más rico y el 10% más pobre, de 2006 a 2010



► Causas del cambio en España



Fuente: Organización Internacional del Trabajo (OIT)

EL PAÍS

Figura 1.9. Estudio sobre la variación de la desigualdad basado en un ratio entre el decil 9 y el decil 1 y publicado en el diario "El País" el 5 de diciembre de 2014.

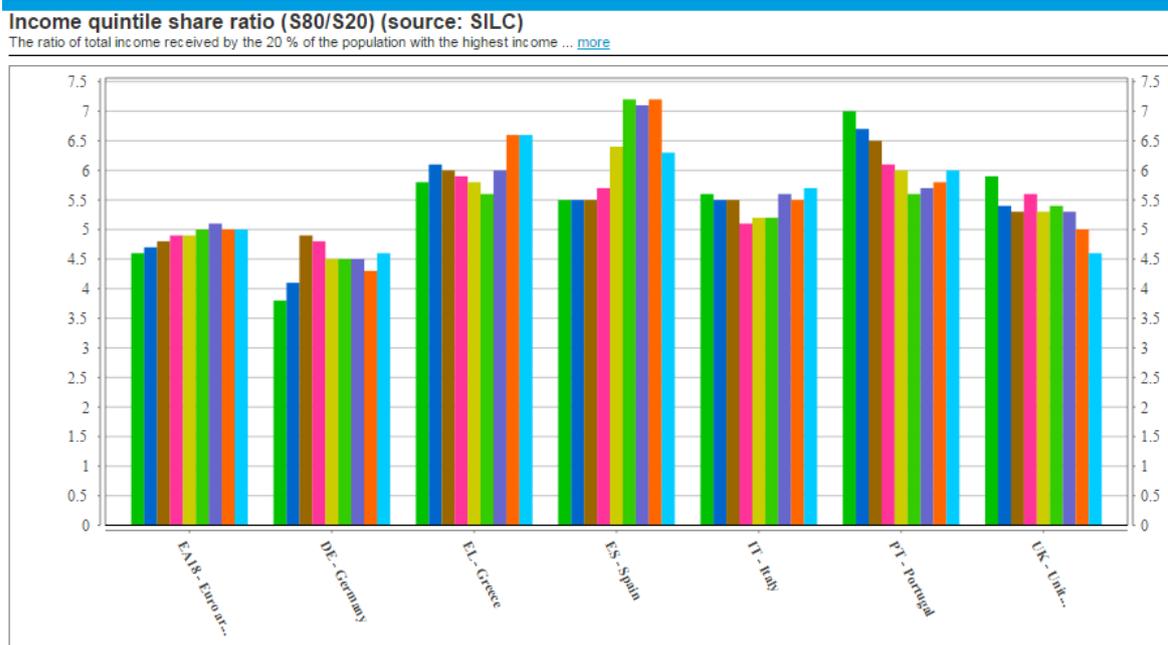


Figura 1.10. Evolución del ratio entre el cuarto y el primer quintil ($r_{0.8,0.2}$) en distintas regiones y países de la Unión Europea entre los años 2005 y 2013 (las series están ordenadas cronológicamente). Representación gráfica generada desde EUROSTAT (<http://ec.europa.eu/eurostat>)

1.3.7.4. Índice de Gini

Posiblemente, el índice de Gini (Gini 1914) es la medida más conocida para medir la desigualdad de la renta o ingresos de un conjunto de individuos. Según Glasser (1962) y Berger (2008), el índice de Gini se puede definir como

$$I_G = \frac{1}{Y} \sum_{i=1}^N (2F(y_i) - 1)y_i$$

donde

Capítulo 1. Introducción

$$Y = \sum_{i=1}^N y_i$$

es el total poblacional de la variable de interés y , y $F(\cdot)$ es la función de distribución ya definida en la Sección 1.3.2. Es bastante conocido que el índice de Gini oscila entre 0 y 1, donde $I_G = 0$ en el caso de una distribución perfecta o concentración mínima de la renta, y $I_G = 1$ en caso contrario. Algunas referencias de interés sobre el índice de Gini y su estimación a partir de datos muestrales son Anand (1983), Sandström (1983), Nygard y Sandström (1985), Sandström et al. (1985), Sandström et al. (1988), Lerman y Yitzhaki (1989), Yitzhaki (1991), Deville (1997), Ogwang (2000), Giles (2004), Ogwang (2004), Frick et al. (2006), Modarres y Castwirth (2006).

La Figura 1.11 incluye los valores del Índice de Gini en el año 2013 para distintas regiones y países de la Unión Europea. A partir de esta figura podemos observar que Serbia, Bulgaria y Letonia son los países con una mayor desigualdad de ingresos según este indicador económico. En el lado opuesto encontramos países como Noruega.

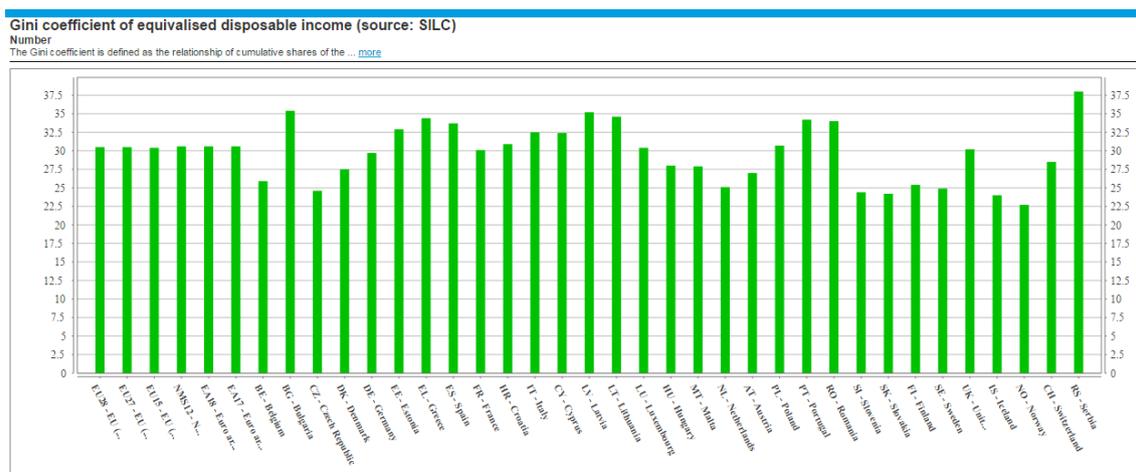


Figura 1.11. Índice de Gini en distintas regiones y países de la Unión Europea en el año 2013. Representación gráfica generada desde EUROSTAT

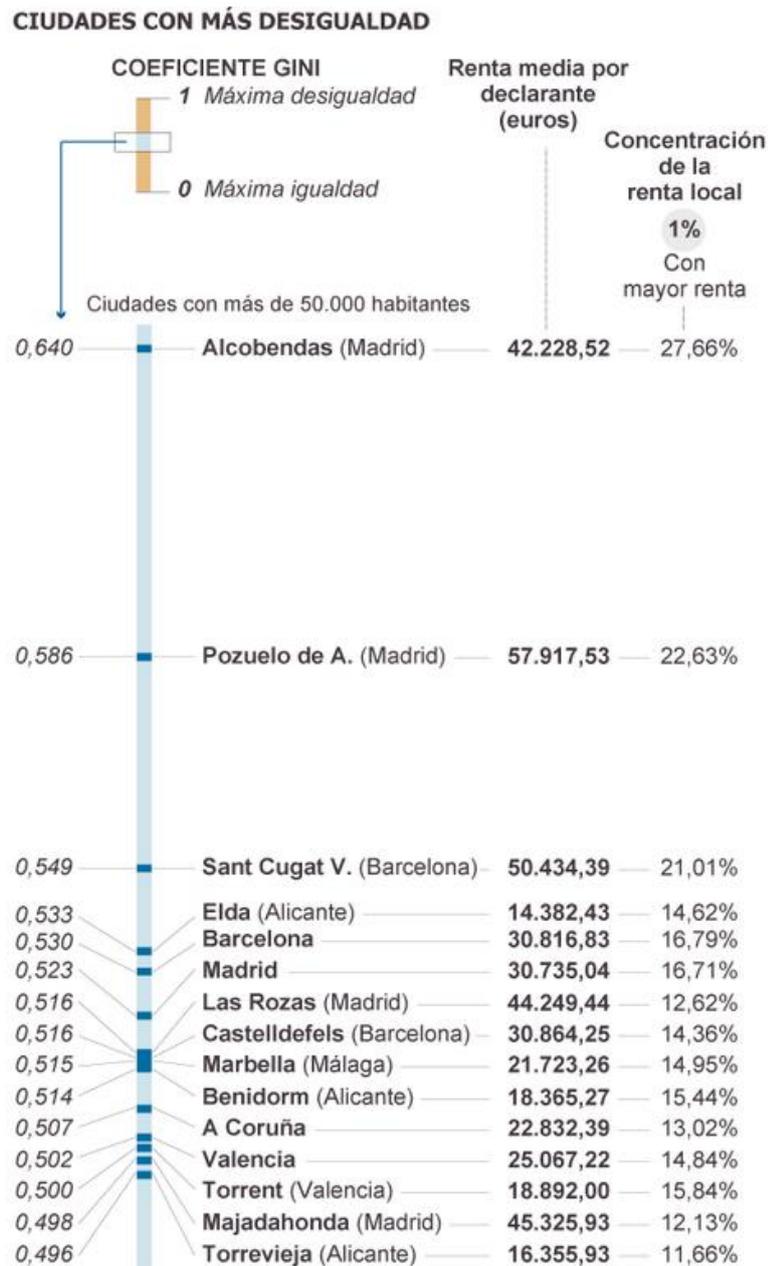
(<http://ec.europa.eu/eurostat>)

Capítulo 1. Introducción

Además de los valores del índice de Gini de la Figura 1.11, a nivel nacional destacamos el estudio publicado en el diario “El País” el 29 de noviembre de 2014, el cual puede consultarse en la Figura 1.12 o bien en la siguiente dirección web:

http://economia.elpais.com/economia/2014/11/29/actualidad/1417288172_845340.html

En el mencionado estudio se clasifican las ciudades con más de 50.000 habitantes según la desigualdad de las mismas en términos del índice de Gini. Según este estudio, Alcobendas es la ciudad con mayor desigualdad. En esta ciudad, el 1% de la población con mayor renta acapara el 27.66% de la concentración.



Fuente: Fundación de Estudios de Economía Aplicada (Fedea).

EL PAÍS

Figura 1.12. Estudio sobre las ciudades españolas con mayor desigualdad basado en el índice de Gini y publicado en el diario "El País" el 29 de noviembre de 2014.

1.3.7.5. Índice de Sen

Sen (1976) propuso un indicador de pobreza que busca combinar los efectos del número de pobres, la profundidad de su pobreza, y la distribución de la pobreza dentro de su grupo. Esta medida de pobreza puede definirse a través de la expresión

$$I_S = P_0 \left(1 - (1 - I_G) \frac{\bar{y}^*}{L} \right).$$

Una expresión alternativa para el índice de Sen viene dada por la media ponderada de la proporción de pobres P_0 y el índice de brecha de pobreza P_1 , donde los pesos viene dados por el índice de Gini, es decir,

$$I_S = I_G P_0 + (1 - I_G) P_1$$

Según afirman Haughton y Khandker (2009), el índice Sen tiene en cuenta la distribución de los ingresos de los pobres, por lo que puede considerarse una medida más apropiada o que aporta más información que otros indicadores de pobreza. Sin embargo, el índice de Sen apenas se utiliza fuera del mundo académico, quizás debido al hecho de depender del índice Gini, y por tanto compartir sus inconvenientes. Una segunda razón a su escaso uso en la práctica es el hecho de no ser un indicador tan intuitivo como otras medidas de pobreza más simples. Deaton (1997) presenta otros inconvenientes de este indicador de pobreza.

1.3.7.6. Índice de Sen-Shorrocks-Thon

El índice de Sen-Shorrocks-Thon I_{SST} es una de las varias versiones modificadas del índice de Sen, y cuya definición viene dada por

$$I_{SST} = P_0 P_1^* (1 + I_{GG})$$

donde P_1^* es el índice de brecha de pobreza aplicado a la población de pobres, y I_{GG} es el índice de Gini aplicado a la variable G con valores dados por

$$G_i = (L - y_i) \delta(y_i \leq L), \quad i = \{1, \dots, N\}$$

Para más información del índice de Sen-Shorrocks-Thon se puede consultar, por ejemplo, Haughton y Khandker (2009) y Osberg and Xu (1999).

1.3.7.7. Índice de Watts

El índice de Watts fue propuesto por Watts (1968), y su expresión es la siguiente:

$$I_w = \frac{1}{N} \sum_{i=1}^N \ln \left(\frac{L}{y_i} \right) \delta(y_i \leq L)$$

Para más información sobre este índice se puede consultar Zheng (1993).

1.3.7.8. Índice de Hagenaars

El índice de Hagenaars (véase también INE 2005b) viene definido por:

$$I_H = P_o \left(\frac{\log(L) - \log(\bar{y}_g^*)}{\log(L)} \right)$$

donde \bar{y}_g^* denota la media geométrica de ingresos de la población pobre.

1.3.7.9. Curva de Lorenz y curvas TIP de pobreza

Para finalizar esta breve revisión de indicadores de pobreza y desigualdad, destacamos que además de las propias medidas de pobreza y desigualdad, también es común utilizar mapas o curvas en los estudios sobre el fenómeno de la pobreza. Algunos ejemplos son la conocida curva Lorenz utilizada para analizar gráficamente la desigualdad de los ingresos, y las curvas TIP de pobreza (véase INE 2005b). Estas curvas de pobreza se obtienen de forma similar a la curva de Lorenz, donde en el eje de abscisas se representan los porcentajes de pobres y en el eje de ordenadas se representan las brechas de pobreza acumuladas de los porcentajes de pobres indicados en el eje de abscisas. Las curvas TIP de pobreza son útiles para medir las tres dimensiones más comunes de la pobreza (incidencia, intensidad y desigualdad) entre la población clasificada como pobre.

Capítulo 2

Estimación de la función de distribución y cuantiles en la población de pobres

2.1. Introducción

En este capítulo se investiga la primera de las medidas de pobreza de este trabajo y se realizan las primeras aportaciones, las cuales consisten en realizar nuevos estudios de simulación basados en poblaciones reales con datos de ingresos por unidad de consumo. Una vez determinada la población de pobres, es decir, los individuos con unas características económicas que lo sitúan por debajo del umbral de pobreza, se puede analizar la situación o distribución de estos pobres mediante el cálculo de una serie de cuantiles basados exclusivamente en esta población de pobres. Por ejemplo, es común utilizar los deciles o los cuartiles de la variable objeto de estudio y para la subpoblación de pobres. La estimación de estos cuantiles a partir de los datos muestrales es por tanto la primera medida de pobreza abordada en este trabajo. De forma adicional, también se analiza la estimación de la función de distribución para la población de pobres, dada la relación directa entre ambos parámetros. En el Capítulo 3 se plantea la estimación de la proporción de pobres, mientras que en el Capítulo 4 se estudia la estimación de los índices de brecha de pobreza y pobreza severa.

Un cuantil es una función no lineal que depende o se define a través de la función de distribución, por lo que se trata de una medida más compleja que parámetros más comunes tal como una media, un total o una proporción. En la Sección 2.2 se realiza una breve revisión de la literatura relacionada con la estimación de la función de distribución y cuantiles, lo cual puede ayudar a entender las aportaciones realizadas en este campo, y la dificultad, en algunas ocasiones, de obtener las propiedades asintóticas de nuevos estimadores de un cuantil.

La estimación de cuantiles es crucial, entre otros aspectos, para poder entender la intensidad de la pobreza y medir la distribución de ingresos de la población de pobres. Además, como se comenta en Rueda y Muñoz (2009) o Singh et al. (2001), el uso de la función de distribución y de los cuantiles se hace indispensable al estudiar variables de gastos o ingresos, ya que estas variables poseen distribuciones asimétricas. En estos casos, medidas estadísticas como la media no son representativas, y el uso de la mediana es considerada una mejor medida de tendencia central que la media. La estimación de la mediana es un caso particular de estimación de cuantiles. En resumen, los cuantiles son necesarios para la obtención de numerosas medidas de pobreza. El umbral de pobreza es un ejemplo de medida económica basada en cuantiles. La razón o cociente entre dos cuantiles (Dickens y Manning 2004, EUROSTAT 2000, Rueda y Muñoz 2011, etc) es también otra medida de pobreza basada en cuantiles. En este caso, también se trata de una medida muy conocida y altamente usada en estudios de pobreza.

Volviendo al tema que nos ocupa, es decir, la estimación de cuantiles en la población de pobres, destacamos el estudio realizado en INE (2005b), en el cual se utilizan los deciles de la población de pobres para analizar la profundidad de la pobreza. En particular, el estudio de los deciles (o de los cuantiles a nivel general) tiene el objetivo de proporcionar una idea de la intensidad de la pobreza y de la desigualdad existente entre los pobres. Según

INE (2005b) y con datos de la Encuesta de Condiciones de Vida 2004, los deciles de ingresos de la población clasificada como pobre pueden consultarse en la Figura 1.5 (por comodidad, véase también este gráfico en la Figura 2.1). A partir de este estudio comentamos que se pueden derivar conclusiones relevantes para la población pobre, tal como el hecho de que un 20% de los pobres (aproximadamente un 4% de la población total) están situados por debajo de la mitad del umbral de pobreza, es decir, tienen unos ingresos por unidad de consumo inferiores a 3139.4 euros anuales, lo que equivale a 259.7 euros al mes. El resto de los pobres, el 80% restante, se reparte casi en su totalidad y de forma más o menos homogénea en la segunda mitad del intervalo (0 euros, 6278.7 euros), es decir, tienen unos ingresos por unidad de consumo superiores a 3139.4 euros anuales y por debajo de la línea de pobreza situada en 6278.7.

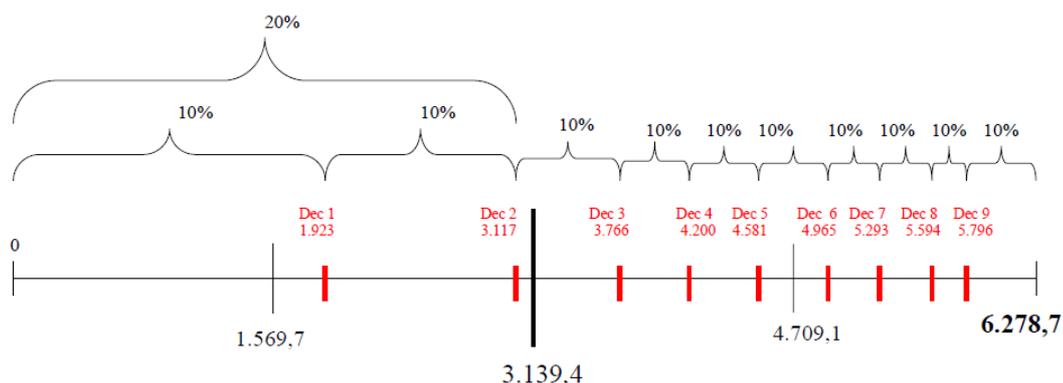


Figura 2.1. Deciles de los ingresos por unidad de consumo de la población clasificada como pobre (Figura también ilustrada en la Figura 1.5). Fuente: INE (2005b).

Tanto las medidas ya comentadas basadas en cuantiles, como este estudio de deciles de ingresos por unidad de consumo en la población pobre y realizado en INE (2005b), nos permite hacernos una idea de la importancia de los cuantiles en los estudios de pobreza, y en particular, en los estudios relacionados con la distribución y la intensidad de la pobreza. Por esta razón, el primer objetivo de esta tesis era realizar nuevas aportaciones sobre la

estimación de cuantiles en la población de pobres. El análisis de deciles, cuantiles, etc en la población de pobres tiene el objetivo de proporcionar una idea de la intensidad de la pobreza y la desigualdad existente entre los pobres.

Como se comentaba anteriormente, el problema de la estimación de un cuantil está directamente relacionado con el problema de la estimación de la función de distribución, y por tanto, la estimación de la función de distribución puede considerarse una cuestión tan importante como el problema de la estimación de un cuantil. De entre las numerosas aplicaciones de la función de distribución podemos destacar la obtención del umbral de pobreza, puesto que este indicador económico suele venir determinado como un porcentaje de la mediana de gastos o ingresos asociados al país, por lo que el uso de un estimador de la función de distribución con propiedades deseables se hace necesario para una óptima estimación de cuantiles o del propio umbral de pobreza.

En el estudio descriptivo de la Figura 2.1 tan sólo se conoce el valor de la estimación de los distintos deciles, pero se desconoce la precisión que tienen estas medidas así como si esta precisión puede estar seriamente afectada por algún factor como puede ser el orden del cuantil, la asimetría de la variable objeto de estudio, el diseño muestral utilizado, etc. En Álvarez et al. (2012) se investigan algunos de estos aspectos en el problema de la estimación de la función de distribución y los cuantiles en la población de pobres. En el mencionado trabajo se realizan estudios empíricos basados principalmente en poblaciones artificiales, lo cual resulta útil puesto que permite investigar los distintos parámetros bajo distintos escenarios posibles. Sin embargo, también puede resultar interesante el análisis de la estimación de estos parámetros en datos recientes sobre ingresos por unidad de consumo, y por tanto, éste es el principal objetivo de este capítulo, es decir, la idea de este capítulo es completar el estudio empírico de Álvarez et al. (2012) con las nuevas poblaciones descritas en el Apéndice I de este trabajo. En particular, analizamos la estimación de la

función de distribución y cuantiles en la población de pobres con los datos reales basados en la EU-SILC del año 2011 y extraídos de 8 países diferentes de la Unión Europea. Además, se realizan nuevos estudios de simulación con las poblaciones generadas a partir de la distribución Singh-Maddala. La precisión de estos estimadores en las nuevas poblaciones en estudio se medirá en términos de las medidas empíricas más conocidas y utilizadas, y que son el sesgo relativo y el error cuadrático medio relativo. Este estudio de simulación permitirá identificar nuevas situaciones en las que la precisión asociada a la estimación de un determinado cuantil en la población de pobres podría no ser apropiada, además de analizar la precisión de tales estimadores en nuevas poblaciones reales y en nuevas situaciones anteriormente no abordadas.

2.2. Antecedentes de estimadores de la función de distribución y cuantiles

La función de distribución es un parámetro que se ha estudiado de forma muy exhaustiva en las últimas décadas. Pueden existir numerosas razones para la realización de estas aportaciones, pero resaltamos el hecho de que la función de distribución es un parámetro fundamental en la determinación de muchas otras medidas estadísticas (Serfling 1980) como pueden ser los cuantiles, o bien por el hecho de tratarse de un parámetro más explicativo o que proporciona más información que otras medidas tal como una media o un total (Sedransk y Sedransk 1979).

Dada una población finita U con N individuos, la función de distribución de una variable de interés y , y evaluada en el argumento t , puede definirse como

$$F(t) = \frac{1}{N} \sum_{i=1}^N \delta(y_i \leq t),$$

donde y_i es el valor observado de la variable de interés y en el i -ésimo individuo de la población, y $\delta(\cdot)$ es la función indicadora que toma el valor 1 si su argumento es cierto, y esta función toma el valor 0 en caso contrario. En otras palabras, para un valor dado de t , la función de distribución $F(t)$ es la proporción de individuos en la población U para los cuales las correspondientes observaciones y_i son menores o igual que dicho argumento t .

Por su parte, el cuantil de orden α en la población U (el cual denotaremos como Y_α) se define como la inversa de la función de distribución anteriormente definida, es decir,

$$Y_\alpha = F^{-1}(\alpha) = \inf(t : F(t) \geq \alpha)$$

En el contexto del muestreo en poblaciones finitas se asume que los parámetros $F(t)$ y Y_α son desconocidos, pero puede estimarse a partir de una muestra s , con tamaño n , y seleccionada de la población U mediante un determinado diseño muestral con probabilidades de inclusión de primer orden dadas por π_i . Por su parte, las probabilidades de inclusión de segundo orden se denotarán como π_{ij} , para los individuos i y j de la población. Siguiendo la anterior definición de un cuantil, el estimador de un cuantil se obtiene, en general, invirtiendo el estimador de la función de distribución utilizado, es decir, dado el estimador de la función de distribución $\hat{F}(t)$, el correspondiente estimador de Y_α viene dado por

$$\hat{Y}_\alpha = \hat{F}^{-1}(\alpha) = \inf(t : \hat{F}(t) \geq \alpha)$$

Algunas de las referencias más importantes relacionadas con la estimación de la función de distribución y cuantiles son Chambers y Dunstan (1986), Kuk y Mak (1989), Rao et al. (1990), Francisco y Fuller (1991), Dorfman y Hall (1993), Mak y Kuk (1993), Silva y Skinner (1995), Chen y Wu (2002), Harm y Duchesne (2006), Dorfman (2009), Rueda y Muñoz (2009), Berger y Muñoz (2015), etc. Los estimadores propuestos por Chambers y Dunstan (1986) y Rao et al. (1990) son los clásicos estimadores de la función de distribución basados en información auxiliar. Silva y Skinner (1995) proponen un estimador de la función de distribución basado en estratificación a posteriori. Por su parte, Chen y Wu (2002) y Rueda y Muñoz (2009) hacen uso del reciente método de verosimilitud pseudo empírica (véase Chen y Sitter 1999) para la estimación de la función de distribución. Harm y Duchesne (2006) proponen estimadores de cuantiles mediante el método de calibración (véase Deville y Särndal 1992). Berger y Muñoz (2015) proponen un estimador de cuantiles basado en una transformación de las variables y un estimador eficiente de la media poblacional. Un resumen de los métodos de estimación de la función de distribución y cuantiles más importantes hasta el año 2009 puede consultarse en Dorfman (2009). A continuación se introducen los que consideramos los estimadores más importantes de la función de distribución y cuantiles existentes en la literatura.

2.2.1. Estimadores basados en el diseño

Los primeros estimadores de la función de distribución fueron propuestos por Horvitz y Thompson (1952) y Háyeek (1964). El estimador de Horvitz y Thompson viene dado por:

$$\hat{F}_{HT}(t) = \frac{1}{N} \sum_{i \in S} \frac{\delta(y_i \leq t)}{\pi_i} = \frac{1}{N} \sum_{i \in S} d_i \delta(y_i \leq t),$$

donde $d_i = \pi_i^{-1}$ son los denominados pesos básicos del diseño. El estimador de Horvitz y Thompson es insesgado, pero presenta el inconveniente de no verificar necesariamente una de las propiedades que ha de tener una verdadera función de distribución. En particular, no se verifica que $\hat{F}_{HT}(+\infty) = 1$. De hecho se verifica:

$$\hat{F}_{HT}(+\infty) = \frac{1}{N} \sum_{i \in S} d_i \neq 1$$

En general, y sólo si

$$d_i = \frac{N}{n}, \text{ para todo } i \in U,$$

el estimador $\hat{F}_{HT}(t)$ cumplirá todas las propiedades para ser una verdadera función de distribución. Por este motivo, es preferible utilizar el estimador de tipo Háyek para el problema de la estimación de la función de distribución, cuya expresión es la siguiente:

$$\hat{F}_H(t) = \frac{\sum_{i \in S} d_i \delta(y_i \leq t)}{\sum_{i \in S} d_i} = \frac{1}{\hat{N}} \sum_{i \in S} d_i \delta(y_i \leq t),$$

donde

$$\hat{N} = \sum_{i \in S} d_i.$$

Se puede comprobar fácilmente que este estimador sí verifica $\hat{F}_H(+\infty) = 1$, lo cual hace que sea un estimador basado en el diseño recomendable en el caso de evaluar la función de distribución en argumentos muy elevados. No obstante,

este estimador presenta el inconveniente de ser, en general, un estimador sesgado, aunque asintóticamente se ha demostrado que es insesgado.

En el caso de muestras seleccionadas mediante muestreo aleatorio simple sin reemplazamiento (MAS), se tiene que

$$d_i = \frac{N}{n}, \text{ para todo } i \in U,$$

y en esta situación ambos estimadores coinciden. Además, la expresión de estos estimadores bajo MAS es muy simple:

$$\hat{F}_{HT}(t) = \hat{F}_H(t) = \frac{1}{n} \sum_{i \in s} \delta(y_i \leq t).$$

2.2.2. Estimador de Chambers y Dunstan

Los estimadores de la función de distribución descritos en la sección anterior sólo hacen uso de la variable de interés, es decir, no hacen uso de otras variables en la etapa de estimación. A estas variables adicionales, las cuales deberían estar relacionadas con la variable de interés, se les denomina variables auxiliares, y su uso se debe a que pueden proporcionar estimadores más precisos. Es bastante común asumir que existe una variable auxiliar, la cual denotaremos como x , relacionada con la variable objeto de estudio. Esta variable auxiliar se puede utilizar en la etapa de estimación con el objetivo de mejorar las propiedades del correspondiente estimador de la función de distribución, tal como veremos a continuación.

Chambers y Dunstan (1986) propusieron el primer método para la estimación de la función de distribución en presencia de información auxiliar. A partir de este momento han sido numerosas las aportaciones realizadas en esta

línea de trabajo, puesto que el uso de esta información auxiliar puede mejorar considerablemente la precisión de los estimadores de la función de distribución.

El estimador de Chambers y Dunstan se basa en un enfoque predictivo, es decir, asume un modelo de superpoblación que relaciona la variable objeto de estudio con una o más variables auxiliares. Por supuesto, la eficiencia de este enfoque dependerá de la calidad del ajuste del modelo que se considere. Supongamos que el modelo de superpoblación es de la forma:

$$y_i = \beta x_i + v(x_i)u_i \quad (i = 1, \dots, N), \quad (2.1)$$

donde β es un parámetro desconocido, $v(x)$ es una función conocida de x , y u_i son variables aleatorias independientes e idénticamente distribuidas con media cero. El estimador de Chamber y Dunstan (1986) tiene la siguiente expresión:

$$\hat{F}_{CD}(t) = \frac{1}{N} \left[\sum_{i \in s} \delta(y_i \leq t) + \frac{1}{n} \sum_{j \in U-s} \sum_{i \in s} \delta \left(u_{ni} \leq \frac{t - b_n x_j}{v(x_j)} \right) \right]$$

donde

$$b_n = \left[\sum_{i \in s} \frac{x_i^2}{v^2(x_i)} \right]^{-1} \sum_{i \in s} \frac{y_i x_i}{v^2(x_i)}$$

y

$$u_{ni} = \frac{y_i - b_n x_i}{v(x_i)}.$$

Destacamos que este estimador de Chamber y Dunstan (1986) es asintóticamente insesgado bajo el modelo considerado, aunque tiene propiedades poco deseables bajo el diseño muestral considerado. Numerosos estudios también han indicado que el estimador de Chamber y Dunstan (1986) puede tener sesgos excesivamente altos en determinadas situaciones, aunque su eficiencia puede ser moderadamente mejor que otros estimadores en otras situaciones.

2.2.3. Estimadores de Rao, Kovar y Mantel

Asumiendo también información auxiliar en la etapa de estimación, Rao et al. (1990) propusieron los siguientes estimadores de tipo razón y diferencia para la función de distribución:

$$\hat{F}_r(t) = \frac{1}{N} \frac{\sum_{i \in s} d_i \delta(y_i \leq t)}{\sum_{i \in s} d_i \delta(\hat{R}x_i \leq t)} \sum_{i \in U} \delta(\hat{R}x_i \leq t) \quad (2.2)$$

y

$$\hat{F}_d(t) = \frac{1}{N} \left\{ \sum_{i \in w} d_i \delta(y_i \leq t) + \sum_{i \in U} \delta(\hat{R}x_i \leq t) - \sum_{i \in s} d_i \delta(\hat{R}x_i \leq t) \right\}, \quad (2.3)$$

donde

$$\hat{R} = \frac{\sum_{i \in s} d_i y_i}{\sum_{i \in s} d_i x_i}$$

es un estimador consistente bajo el diseño de la razón poblacional

$$R = \frac{Y}{X}$$

y basada en el modelo de superpoblación descrito en la expresión (2.1). X e Y denotan, respectivamente, los totales poblacionales de la variable auxiliar y de la variable de interés. Un tercer estimador propuesto en Rao et al. (1990) también está basado en el diseño. Sin embargo, y a diferencia de los anteriores estimadores (2.2) y (2.3), este estimador es asintóticamente insesgado tanto bajo el diseño como bajo el modelo de superpoblación considerado. Dicho estimador, asumiendo por ejemplo el modelo

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad (i = 1, \dots, N), \quad E_\epsilon[\epsilon_i] = 0 \quad V_\epsilon[\epsilon_i] = \sigma_\epsilon^2$$

donde ϵ_i son errores aleatorios independientes e idénticamente distribuidos, y donde α y β son parámetros poblacionales desconocidos, viene dado por:

$$\hat{F}_{dm}(t) = \frac{1}{N} \left[\sum_{i \in s} d_i \delta(y_i \leq t) + \sum_{j=1}^N \hat{G}_j - \sum_{k \in s} d_k \hat{G}_{kc} \right]$$

donde

$$\hat{G}_j = \sum_{i \in s} d_i \delta(\tilde{\alpha} + \tilde{\beta} x_j + \tilde{\epsilon}_i \leq t) / \sum_{i \in s} d_i ; \quad (2.4)$$

$$\hat{G}_{kc} = \sum_{i \in s} \frac{\pi_k}{\pi_{ki}} \delta(\tilde{\alpha} + \tilde{\beta} x_j + \tilde{\epsilon}_i \leq t) / \sum_{i \in s} \frac{\pi_k}{\pi_{ki}} ;$$

$$\tilde{\beta} = \frac{\sum_{i \in s} d_i (x_i - \hat{x}_H)(y_i - \hat{y}_H)}{\sum_{i \in s} d_i (x_i - \hat{x}_H)^2} ;$$

$$\tilde{\alpha} = \hat{y}_H - \tilde{\beta}\hat{x}_H;$$

$$\tilde{\epsilon}_i = y_i - \tilde{\alpha} - \tilde{\beta}x_i;$$

$$\hat{y}_H = \frac{\sum_{i \in s} d_i y_i}{\sum_{i \in s} d_i};$$

y

$$\hat{x}_H = \frac{\sum_{i \in s} d_i x_i}{\sum_{i \in s} d_i}.$$

El estimador $\hat{F}_{dm}(t)$ fue motivado como un estimador de tipo diferencia, y asumiendo que las cantidades:

$$G_j = \frac{1}{N} \sum_{i=1}^N \delta(\alpha + \beta x_j + \epsilon_i \leq t)$$

son desconocidas, pero reemplazadas por sus estimaciones dadas en la expresión (2.4). Rao et al. (1990) demostraron que este estimador es consistente bajo el diseño.

2.2.4. Estimador de Silva y Skinner

Silva y Skinner (1995) propusieron el siguiente estimador basado en muestreo estratificado a posteriori o post-estratificación

$$\hat{F}_{PS}(t) = \frac{1}{N} \sum_{g=1}^G \frac{N_g}{\hat{N}_g} \sum_{i \in s} d_i \delta(y_i \leq t) \delta(i \in U_g)$$

donde U_1, \dots, U_G son los G estratos de la población formados a posteriori, N_g es el tamaño del estrato U_g , y

$$\hat{N}_g = \sum_{i \in s_g} d_i,$$

con $g = 1, \dots, G$.

2.2.5. Algunos estimadores de cuantiles

Como se comentaba con anterioridad, cualquier estimador de la función de distribución se podría utilizar para estimar cuantiles, y para ello tan sólo es necesario invertir el estimador de la función de distribución, es decir, el estimador de Y_α viene dado por

$$\hat{Y}_\alpha = \hat{F}^{-1}(\alpha) = \inf(t : \hat{F}(t) \geq \alpha)$$

donde $\hat{F}(t)$ denota a un determinado estimador de la función de distribución. No obstante, algunos estimadores de la función de distribución no cumplen todas las propiedades requeridas en una función de distribución, por lo que en estas situaciones es necesario el uso de alguna transformación en la función de distribución para que se verifiquen tales propiedades. Por ejemplo, el estimador $\hat{F}_{dm}(t)$ es el que tiene la mejor precisión en la mayoría de las ocasiones de entre sus estimadores rivales, pero presenta el inconveniente que no siempre es una función no decreciente monótona, la cual es una propiedad que debe satisfacer una verdadera función de distribución. Por este motivo, para el problema de la estimación de cuantiles, Rao et al. (1990) propusieron el siguiente estimador:

$$\hat{F}_{dm}^*(t) = \max \left\{ \tilde{F}_{dm}(y_{(i)}) : y_{(i)} \leq t \right\}$$

donde $y_{(i)}$ denota el estadístico ordenado de la muestra $\{y_i, i \in s\}$ y $\tilde{F}_{dm}(y_{(i)})$ se define a través de la siguiente fórmula recursiva

$$\tilde{F}_{dm}(y_{(i)}) = \max\{\tilde{F}_{dm}(y_{(i-1)}), \hat{F}_{dm}(y_{(i)})\}$$

donde $\tilde{F}_{dm}(y_{(1)}) = \min\{\hat{F}_{dm}(y_{(1)})\}$. De esta forma, el estimador de cuantiles propuesto por Rao et al. (1990) está dado por

$$\hat{Y}_{\alpha;dm} = \inf(t : \hat{F}_{dm}^*(t) \geq \alpha)$$

En el caso de que el cuantil de la variable auxiliar sea conocido a nivel poblacional, es decir, X_α sea conocido, Rao et al. (1990) también propusieron estimadores de cuantiles basados en los métodos de razón y diferencia, pero aplicados sobre cuantiles, es decir, el estimador de tipo razón de Y_α y basado en cuantiles viene dado por

$$\hat{Y}_{\alpha;r} = \frac{\hat{Y}_{\alpha;H}}{\hat{X}_{\alpha;H}} X_\alpha,$$

donde $\hat{X}_{\alpha;H}$ y $\hat{Y}_{\alpha;H}$ son los estimadores de tipo Hájek de X_α y Y_α , respectivamente, es decir,

$$\hat{Y}_{\alpha;H} = \hat{F}_H^{-1}(\alpha) = \inf(t : \hat{F}_H(t) \geq \alpha).$$

De forma similar se define el estimador de tipo Hájek de X_α . Por su parte, el estimador de tipo diferencia propuesto por Rao et al. (1990) viene dado por

$$\hat{Y}_{\alpha;d} = \hat{Y}_{\alpha;H} + \hat{R}(X_\alpha - \hat{X}_{\alpha;H})$$

Existen numerosos estimadores de cuantiles en la literatura, pero destacamos también el estimador propuesto por Harms y Duchesne (2006), debido a su flexibilidad para incluir la información auxiliar y la precisión que ha mostrado tener en distintos estudios de simulación. Este estimador está basado en el método de calibración y viene dado por

$$\hat{Y}_{cal;\alpha} = \hat{F}_{cal}^{-1}(\alpha) = \inf(t : \hat{F}_{cal}(t) \geq \alpha),$$

donde

$$\hat{F}_{cal}(t) = \left(\sum_{i \in S} w_i \right)^{-1} \sum_{i \in S} w_i H(t, y_i)$$

es el estimador ponderado de la función de distribución $F(t)$ y basado en el método de calibración. Los pesos de calibración w_i se obtienen minimizando con respecto a la variable $\mathbf{v} = (v_1, \dots, v_n)'$ la función $\mathbf{w} = \arg \min_{\mathbf{v}} D(\mathbf{v}, \mathbf{d})$, y sujeto a las restricciones de calibración

$$\sum_{i \in S} v_i = N$$

y

$$\hat{X}_{cal;\alpha} = X_{\alpha},$$

donde $\mathbf{d} = (d_1, \dots, d_n)'$, y $\hat{X}_{cal;\alpha}$ es el estimador de Harms y Duchesne basado en la variable auxiliar. La función $H(\cdot, \cdot)$ se define como

$$H(t, y_i) = \begin{cases} 1 & \text{si } y_i \leq L(t) \\ \frac{t - L(t)}{U(t) - L(t)} & \text{si } y_i = U(t) \\ 0 & \text{si } y_i > U(t) \end{cases}$$

donde

$$L(t) = \max \{ \{y_i, i \in s | y_i \leq t\} \cup \{-\infty\} \}$$

Y

$$U(t) = \min \{ \{y_i, i \in s | y_i > t\} \cup \{\infty\} \}.$$

Cuando se considera la función de distancia cuadrática

$$D(\mathbf{v}, \mathbf{d}) = \sum_{i \in s} \frac{(v_i - d_i)^2}{d_i q_i},$$

Harms y Duchesne (2006) obtienen la siguiente expresión analítica para los pesos de calibración

$$w_i = d_i (1 + q_i \mathbf{a}_i' \lambda_s), \quad i \in s$$

donde q_i determina la importancia de la unidad $i \in s$, y

$$\lambda_s = \left(\sum_{i \in s} d_i q_i \mathbf{a}_i \mathbf{a}_i' \right)^{-1} \left(\mathbf{T} \mathbf{a} - \sum_{i \in s} d_i \mathbf{a}_i \right),$$

con $\mathbf{T} \mathbf{a} = (N, \alpha)$, $\mathbf{a}_i = (1, \mathbf{a}_i)$ y $\mathbf{a}_i = N^{-1} H(X_\alpha, x_i)$.

2.3. Función de distribución y cuantiles en la población de pobres

2.3.1 Notación y definición de parámetros

En primer lugar, antes de abordar el problema de la estimación de cuantiles en la población de pobres, vamos a recordar la definición formal del umbral de pobreza que consideramos en esta tesis doctoral, la cual nos permitirá también introducir la notación general que seguiremos a lo largo de este trabajo. Como ya indicamos anteriormente, consideramos una población finita U con N individuos en la cual se desea estudiar una variable de interés y . En este trabajo supondremos que la variable y son los ingresos familiares por unidad de consumo. Esta variable es utilizada comúnmente para la determinación del umbral de pobreza, el cual delimita la población que se considera pobre de la población que no es considerada como tal. En concreto, el umbral de pobreza relativo suele obtenerse como un porcentaje de un determinado cuantil de los ingresos por unidad de consumo. El cuantil poblacional de orden β de la variable y se ha definido anteriormente como

$$Y_{\beta} = \inf \{t : F(t) \geq \beta\},$$

donde $F(t)$ es la función de distribución de la variable y , evaluada en el argumento específico t , y la cual también se ha definido en la sección anterior. Numerosos organismos oficiales de estadística definen la línea o umbral de pobreza relativa (L) como un porcentaje α del cuantil de orden β , es decir,

$$L = \alpha Y_{\beta}.$$

Capítulo 2. Función de distribución y cuantiles en la población de pobres

Por ejemplo, el Instituto Nacional de Estadística o EUROSTAT consideran $\alpha = 0.6$ y $\beta = 0.5$ para la determinación del umbral de pobreza, es decir, consideran el 60% de la mediana de ingresos por unidad de consumo.

Una vez determinado el umbral de pobreza, se considera que un individuo es pobre o pertenece a la población de pobres si su ingreso por unidad de consumo es menor o igual que el mencionado umbral de pobreza, es decir, $y_i \leq L$. Esta clasificación nos permite definir los valores de la variable y para la población de pobres como

$$y_i^* = \{y_i : y_i \leq L\} \quad i = 1, \dots, N$$

Supongamos que el número de valores que verifican esta condición a nivel poblacional es N^* , es decir, N^* denotará el número total de pobres en la población U .

Uno de los objetivos de este capítulo es plantear y analizar la estimación de la función de distribución en la propia población de pobres, es decir, estimar el parámetro

$$F^*(t) = \frac{1}{N^*} \sum_{i=1}^{N^*} \delta(y_i^* \leq t).$$

A partir de la función de distribución anterior se puede plantear la estimación del cuantil de orden γ en la población de pobres, es decir, en este capítulo se analiza mediante nuevos estudios de simulación Monte Carlo el problema de la estimación del cuantil

$$Y_\gamma^* = \inf \{t : F^*(t) \geq \gamma\}.$$

Siguiendo Álvarez et al. (2012), en las Secciones 2.3.2 y 2.3.3 se describe la estimación de los parámetros poblacionales anteriormente descritos a partir de muestras aleatorias seleccionadas en la población U objeto de estudio. En la Sección 2.4 se presentan los resultados obtenidos en nuevos estudios de simulación. Estos estudios están basados en las poblaciones descritas en el Apéndice I.

2.3.2. Estimación bajo un diseño muestral general

En esta sección se describen estimadores para la función de distribución en la población de pobres, $F^*(t)$, y para el cuantil de orden γ , Y_γ^* , de la mencionada población de pobres. Con el fin de realizar esta estimación asumiremos que en la población U se extrae una muestra s , de tamaño n , y mediante un diseño muestral arbitrario con probabilidades de inclusión de primer orden dadas por π_i . Además, asumiremos que la línea de pobreza relativa es desconocida y por tanto también ha de ser estimada. El estimador estándar del umbral de pobreza L viene dado por

$$\hat{L} = \alpha \hat{Y}_\beta,$$

donde

$$\hat{Y}_\beta = \inf \{t : \hat{F}(t) \geq \beta\}$$

es el estimador estándar para Y_β , y $\hat{F}(t)$ denota un estimador para la función de distribución $F(t)$. Por ejemplo, $\hat{F}(t)$ puede ser cualquiera de los estimadores basados en el diseño de la función de distribución presentados en la Sección 2.2. Teniendo en cuenta las consideraciones anteriores, expresaremos los valores de

la variable de interés para los individuos de la muestra y considerados como pobres como

$$\hat{y}_i^* = \{y_i : y_i \leq \hat{L}\} \quad i = 1, \dots, n.$$

Por su parte, el número total de pobres en la muestra se denotará como n^* . Siguiendo la definición del estimador de tipo Horvitz-Thompson para la función de distribución (véase la Sección 2.2.1), el correspondiente estimador de tipo Horvitz-Thompson para $F^*(t)$ viene dado por

$$\hat{F}_{HT}^*(t) = \frac{1}{N^*} \sum_{i=1}^{n^*} d_i \delta(\hat{y}_i^* \leq t).$$

Tal como se indica en Álvarez et al. (2012), el estimador $\hat{F}_{HT}^*(t)$ no se puede obtener en la práctica, dado que el número de pobres en la población, lo cual se ha denotado como N^* , suele ser una cantidad desconocida. Por esta razón, Álvarez et al. (2012) proponen utilizar el estimador para $F^*(t)$ dado por la expresión

$$\tilde{F}^*(t) = \frac{1}{\tilde{N}^*} \sum_{i=1}^{n^*} d_i \delta(\hat{y}_i^* \leq t),$$

donde $\tilde{N}^* = n^* / f$ y $f = n/N$ es la fracción de muestreo. Por su parte, el estimador de tipo Hájek para $F^*(t)$ se puede definir directamente a partir de su definición, es decir,

$$\hat{F}_H^*(t) = \frac{1}{\hat{N}^*} \sum_{i=1}^{n^*} d_i \delta(\hat{y}_i^* \leq t),$$

donde $\hat{N}^* = \sum_{i=1}^{n^*} d_i$.

Respecto a la estimación de cuantiles en la población de pobres, el estimador de tipo Horvitz-Thompson tampoco se podría utilizar para estimar el parámetro Y_γ^* , por lo que en su lugar se puede definir a partir del estimador $\tilde{F}^*(t)$ anteriormente descrito, es decir,

$$\tilde{Y}_\gamma^* = \inf \left\{ t : \tilde{F}^*(t) \geq \gamma \right\}.$$

El estimador de tipo Hájek para Y_γ^* se define de forma similar, es decir,

$$\hat{Y}_\gamma^* = \inf \left\{ t : \hat{F}_H^*(t) \geq \gamma \right\}.$$

2.3.3. Estimación bajo muestreo aleatorio simple

En esta sección se describen los estimadores de la función de distribución y cuantiles para la población de pobres, pero en el caso de muestras seleccionadas mediante MAS. Bajo este diseño muestral, las probabilidades de inclusión de primer orden vienen dadas por $\pi_i = n/N$, y por tanto los pesos básicos del diseño están dados por $d_i = N/n$. Por otra parte, es fácil comprobar que el valor de \hat{N}^* anteriormente definido viene dado bajo MAS por:

$$\hat{N}^* = \sum_{i=1}^n d_i = \frac{N}{n} n^*,$$

y esto implica que el estimador de tipo Hájek para el parámetro $F^*(t)$ puede obtenerse a partir de la siguiente expresión más simple:

$$\hat{F}_H^*(t) = \frac{1}{n^*} \sum_{i=1}^{n^*} \delta(\hat{y}_i^* \leq t)$$

Respecto al estimador de tipo Horvitz-Thompson para $F^*(t)$ y en el caso de MAS, destacamos que tampoco podría obtenerse en la práctica, puesto que el número de pobres en la población, N^* , sigue siendo una cantidad desconocida. Por este motivo, se podría utilizar de nuevo el estimador $\tilde{F}^*(t)$ anteriormente comentado, y el cual coincide con el estimador de tipo Hájek $\hat{F}_H^*(t)$ en el caso de muestras seleccionadas bajo MAS, es decir, $\tilde{F}^*(t) = \hat{F}_H^*(t)$ bajo este diseño muestral (véase Álvarez et al. 2012). En consecuencia, los estimadores \hat{Y}_γ^* y \tilde{Y}_γ^* para el cuantil poblacional Y_γ^* y basados, respectivamente, en $\hat{F}_H^*(t)$ y $\tilde{F}^*(t)$ también coinciden en el caso de muestras seleccionadas bajo MAS.

2.4. Nuevos estudios de simulación Monte Carlo

En esta sección se llevan a cabo nuevos estudios de simulación de Monte Carlo para comparar numéricamente los estimadores de la función de distribución y de cuantiles para la población de pobres descritos en este capítulo (véase también Álvarez et al. 2012). Como aportación de este trabajo destacamos que tales estudios se han realizado para las 8 poblaciones reales extraídas de la EU-SILC del año 2011, y correspondientes a los 8 países diferentes. Además, estos estudios de simulación están basados en las poblaciones obtenidas a partir de la distribución Singh-Maddala. Todas las poblaciones, reales y simuladas, están descritas en el Apéndice I de este trabajo.

La precisión de los estimadores para los distintos parámetros ($F^*(t)$ y Y_γ^*) se evaluará de forma empírica mediante el Sesgo Relativo (SR) y el Error Cuadrático Medio Relativo (ECMR), los cuales se definen a partir de las siguientes expresiones en el caso del estimador \hat{Y}_γ^* para el cuantil Y_γ^* en la población de pobres:

$$SR(\hat{Y}_\gamma^*) = \frac{E[\hat{Y}_\gamma^*] - Y_\gamma^*}{Y_\gamma^*}; \quad ECMR(\hat{Y}_\gamma^*) = \frac{\sqrt{ECM[\hat{Y}_\gamma^*]}}{Y_\gamma^*}.$$

donde $E[\hat{Y}_\gamma^*]$ y $ECM[\hat{Y}_\gamma^*]$ denotan, respectivamente, el valor esperado y el error cuadrático medio del estimador \hat{Y}_γ^* , es decir,

$$E[\hat{Y}_\gamma^*] = \frac{1}{D} \sum_{d=1}^D \hat{Y}_{\gamma;d}^*; \quad ECM[\hat{Y}_\gamma^*] = \frac{1}{D} \sum_{d=1}^D (\hat{Y}_{\gamma;d}^* - Y_\gamma^*)^2.$$

donde $D = 10000$ es el número total de muestras que se han utilizado en los distintos estudios de simulación. Por su parte, $\hat{Y}_{\gamma;d}^*$ denota el estimador \hat{Y}_γ^* cuando éste se obtiene en la d -ésima simulación o experimento. Estas medidas se definen de forma simulada cuando se evalúan los estimadores de la función de distribución $F^*(t)$. Los resultados obtenidos de este estudio de simulación con muestras seleccionadas bajo MAS pueden consultarse desde la Tabla 2.1 hasta la Tabla 2.11.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-0.7	0.6	11.2	25.8
	0.10	-0.2	0.2	7.7	17.8
	0.20	0.2	0.3	4.9	11.8
0.3	0.05	-0.1	0.5	5.0	14.1
	0.10	-0.1	0.3	3.7	9.6
	0.20	-0.1	0.2	2.5	6.4
0.5	0.05	0.3	0.4	3.1	10.1
	0.10	0.2	0.3	2.0	7.0
	0.20	0.2	0.2	1.2	4.6
0.7	0.05	0.1	0.4	3.3	7.8
	0.10	0.1	0.3	2.3	5.5
	0.20	0.1	0.2	1.6	3.6
0.9	0.05	-0.4	0.2	3.2	6.1
	0.10	-0.4	0.3	2.3	4.3
	0.20	-0.4	0.2	1.6	3.0

Tabla 2.1. Valores de SR y ECMR de los estimadores tipo Hájek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Bulgaria. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	1.4	-0.3	9.0	28.6
	0.10	1.4	0.0	5.9	19.7
	0.20	1.1	-0.4	3.9	13.0
0.3	0.05	0.4	-0.2	4.0	15.1
	0.10	0.4	-0.2	2.8	10.4
	0.20	0.4	-0.3	1.9	6.8
0.5	0.05	-0.4	0.1	2.9	10.5
	0.10	-0.2	-0.1	1.9	7.1
	0.20	-0.2	-0.3	1.2	4.7
0.7	0.05	-0.2	0.0	2.4	7.7
	0.10	-0.2	-0.1	1.7	5.2
	0.20	-0.1	-0.2	1.2	3.4
0.9	0.05	0.1	0.0	1.8	5.5
	0.10	0.1	0.0	1.2	3.7
	0.20	0.1	-0.2	0.8	2.3

Tabla 2.2. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Eslovenia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-2.5	0.4	11.3	21.6
	0.10	-1.8	-0.3	8.0	14.9
	0.20	-1.4	-0.4	5.5	10.1
0.3	0.05	0.6	0.4	4.9	12.0
	0.10	0.6	-0.2	3.7	8.4
	0.20	0.6	-0.3	2.7	5.7
0.5	0.05	0.0	0.2	3.0	8.9
	0.10	0.2	-0.2	2.0	6.2
	0.20	0.2	-0.3	1.4	4.3
0.7	0.05	-0.4	0.2	2.5	7.2
	0.10	-0.2	-0.2	1.7	5.1
	0.20	-0.1	-0.3	1.2	3.6
0.9	0.05	-0.1	0.1	2.2	5.9
	0.10	0.0	-0.2	1.7	4.4
	0.20	0.1	-0.3	1.2	3.2

Tabla 2.3. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Polonia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	0.0	-1.3	19.6	36.7
	0.10	0.6	-0.6	13.4	25.5
	0.20	1.0	-0.5	9.1	16.6
0.3	0.05	-0.8	-0.6	5.6	20.5
	0.10	-0.6	-0.5	3.8	14.0
	0.20	-0.3	-0.6	2.4	9.1
0.5	0.05	-0.5	-0.4	4.1	15.1
	0.10	-0.3	-0.5	2.7	10.4
	0.20	-0.1	-0.6	1.8	6.8
0.7	0.05	0.0	-0.4	3.4	12.2
	0.10	0.1	-0.5	2.3	8.3
	0.20	0.1	-0.5	1.5	5.5
0.9	0.05	0.4	-1.1	3.4	9.4
	0.10	0.4	-0.6	2.4	6.7
	0.20	0.4	-0.6	1.6	4.6

Tabla 2.4. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Bélgica. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-0.7	0.9	26.7	32.3
	0.10	-1.0	0.8	18.3	22.0
	0.20	-1.0	0.3	12.3	14.8
0.3	0.05	-0.9	1.2	7.6	17.8
	0.10	-0.6	0.5	5.2	12.1
	0.20	-0.4	0.4	3.6	8.2
0.5	0.05	-0.4	1.3	4.7	13.3
	0.10	-0.1	0.7	3.1	8.9
	0.20	0.0	0.4	2.1	6.0
0.7	0.05	-0.8	1.3	4.0	10.5
	0.10	-0.6	0.6	2.8	7.2
	0.20	-0.5	0.4	1.9	4.9
0.9	0.05	-0.4	0.6	2.9	7.5
	0.10	-0.2	0.5	2.0	5.7
	0.20	-0.1	0.5	1.3	4.1

Tabla 2.5. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Reino Unido. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	11.9	0.1	78.1	17.5
	0.10	4.5	0.1	53.6	12.2
	0.20	-1.1	0.2	32.3	8.0
0.3	0.05	0.2	0.3	7.1	9.6
	0.10	0.2	0.0	4.9	6.6
	0.20	0.0	0.2	3.3	4.4
0.5	0.05	-0.3	0.3	3.9	7.0
	0.10	-0.2	0.1	2.6	4.7
	0.20	-0.3	0.2	1.8	3.1
0.7	0.05	-0.1	0.3	3.0	5.4
	0.10	0.0	0.1	2.1	3.6
	0.20	-0.1	0.2	1.4	2.3
0.9	0.05	0.0	0.2	2.0	4.3
	0.10	0.1	0.1	1.3	2.8
	0.20	0.0	0.2	0.9	1.8

Tabla 2.6. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en España. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	1.3	-0.8	29.2	18.5
	0.10	2.2	-0.9	20.9	12.8
	0.20	2.2	-0.6	14.0	8.4
0.3	0.05	0.5	-0.9	5.6	9.9
	0.10	0.5	-0.7	3.9	6.9
	0.20	0.5	-0.6	2.6	4.6
0.5	0.05	0.1	-0.7	2.8	7.2
	0.10	0.2	-0.7	1.9	5.0
	0.20	0.1	-0.6	1.3	3.4
0.7	0.05	0.3	-0.7	2.0	5.5
	0.10	0.3	-0.7	1.4	3.9
	0.20	0.3	-0.6	0.9	2.7
0.9	0.05	0.3	-0.7	1.8	4.4
	0.10	0.3	-0.8	1.2	3.1
	0.20	0.2	-0.6	0.8	2.2

Tabla 2.7. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Italia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-1.1	1.3	27.8	34.3
	0.10	-1.3	1.2	19.1	23.5
	0.20	-1.2	1.0	12.9	15.7
0.3	0.05	-0.5	1.4	8.5	18.3
	0.10	-0.3	1.3	5.9	12.6
	0.20	0.0	1.0	3.8	8.4
0.5	0.05	-1.1	1.3	5.9	12.9
	0.10	-0.8	1.2	4.1	8.9
	0.20	-0.6	0.9	2.7	6.0
0.7	0.05	-0.6	1.2	4.4	9.9
	0.10	-0.5	1.3	3.1	6.8
	0.20	-0.4	0.9	2.2	4.6
0.9	0.05	-0.9	1.0	3.3	6.9
	0.10	-0.8	1.2	2.3	5.1
	0.20	-0.6	1.0	1.6	3.7

Tabla 2.8. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Lituania. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	0.3	1.5	8.5	47.8
	0.10	0.4	1.0	5.7	32.5
	0.20	0.4	0.3	3.8	21.5
0.3	0.05	0.1	1.6	4.6	26.6
	0.10	0.2	0.6	3.1	17.9
	0.20	0.2	0.4	2.1	12.0
0.5	0.05	-0.7	1.5	3.7	19.8
	0.10	-0.4	0.5	2.5	13.4
	0.20	-0.3	0.4	1.6	8.8
0.7	0.05	-0.2	1.4	2.8	15.9
	0.10	0.0	0.5	1.8	10.9
	0.20	0.0	0.4	1.2	7.0
0.9	0.05	-0.2	-0.3	2.5	11.0
	0.10	-0.1	0.1	1.7	8.3
	0.20	-0.1	0.3	1.1	5.8

Tabla 2.9. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 11.5$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-0.4	0.2	7.6	48.4
	0.10	-0.4	0.5	5.1	32.7
	0.20	-0.2	0.1	3.5	22.0
0.3	0.05	-0.4	0.7	4.2	27.0
	0.10	-0.3	0.5	2.9	18.2
	0.20	-0.2	0.0	2.0	12.4
0.5	0.05	-0.3	0.8	3.4	19.6
	0.10	-0.1	0.3	2.4	13.6
	0.20	0.0	-0.1	1.7	9.2
0.7	0.05	-0.1	0.8	3.2	15.7
	0.10	0.0	0.4	2.2	10.8
	0.20	0.1	0.0	1.6	7.3
0.9	0.05	-0.1	-0.5	2.7	11.1
	0.10	0.0	0.0	1.9	8.3
	0.20	0.1	-0.1	1.3	6.1

Tabla 2.10. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 8.1$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

γ	f	SR (%)		ECMR (%)	
		\hat{Y}_γ^*	$\hat{F}_H^*(t)$	\hat{Y}_γ^*	$\hat{F}_H^*(t)$
0.1	0.05	-0.9	1.3	10.7	57.3
	0.10	0.3	-0.2	5.6	38.7
	0.20	0.2	0.3	3.4	25.4
0.3	0.05	-1.1	0.9	5.3	30.6
	0.10	-0.9	0.7	3.6	20.7
	0.20	-0.6	0.3	2.3	13.5
0.5	0.05	-0.6	1.4	3.5	21.3
	0.10	-0.3	0.7	2.3	14.4
	0.20	-0.1	0.4	1.5	9.6
0.7	0.05	-0.4	1.6	2.6	16.1
	0.10	-0.2	0.6	1.8	10.8
	0.20	-0.1	0.3	1.2	7.1
0.9	0.05	-0.3	0.1	2.0	10.0
	0.10	-0.1	0.2	1.3	7.5
	0.20	0.0	0.3	0.8	5.3

Tabla 2.11. Valores de SR y ECMR de los estimadores tipo Háyek para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 4.0$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante muestreo aleatorio simple.

Los resultados empíricos obtenidos para MAS y con datos extraídos de la EU-SILC del año 2011 pueden observarse desde la Tabla 2.1 hasta la Tabla 2.8, en cada caso, para distintos países de la Unión Europea. En general, podemos observar que los sesgos relativos del estimador de la función de distribución $F^*(t)$ toman valores razonables, con valores por debajo del 1%, en valor absoluto. Tan sólo en los resultados derivados de Bélgica, Reino Unido y Lituania se obtienen sesgos para el estimador de este parámetro ligeramente por encima del 1%, y en la mayor parte de los casos esto ocurre cuando la fracción de muestreo es pequeña ($f = 5\%$). La situación es similar en el caso de la estimación de diversos cuantiles en la población de pobres, aunque en este caso se pueden obtener sesgos relativos algo mayores, y en general, este hecho ocurre cuando el orden del cuantil es muy pequeño ($\gamma = 0.1$). Por ejemplo, en el caso de datos de Polonia, el sesgo se sitúa en el -2.5% cuando $\gamma = 0.1$ y $f = 5\%$, mientras que bajo estas condiciones y en el caso de datos de España, el sesgo alcanza el 11.9%. Los datos procedentes de Italia y Lituania también tienen sesgos ligeramente por encima del 1% cuando el orden del cuantil a estimar es $\gamma = 0.1$.

Analizamos ahora los sesgos obtenidos en las poblaciones basadas en la distribución Singh-Maddala (Tablas 2.9, 2.10 y 2.11). Podemos observar que los estimadores para ambos parámetros (la función de distribución y los cuantiles en la población de pobres) tienen sesgos relativos con propiedades deseables, puesto que la mayoría están por debajo del 1%, en términos absolutos, y en ningún caso se supera el 2%, también hablando en términos absolutos.

A continuación nos centraremos también en las tablas anteriormente comentadas para analizar la eficiencia de los estimadores de los dos parámetros en el caso de muestras seleccionadas mediante MAS. Recordamos que la eficiencia se medirá en términos de la medida empírica *ECMR*. Respecto a la estimación de la función de distribución en la población de pobres, podemos observar que el estimador de este parámetro se vuelve más eficiente a medida

que aumenta el orden de cuantil (γ), o bien aumenta la fracción de muestreo (f). Esto ocurre en todas las poblaciones analizadas en este capítulo y en el caso del MAS que estamos actualmente analizando. Los valores de *ECMR* en las distintas poblaciones basadas en la distribución Singh-Maddala indican que los distintos grados de asimetría considerados en la variable de interés no tienen un impacto importante en la eficiencia de este estimador de la función de distribución, puesto que tales valores de *ECMR* son muy similares en las distintas poblaciones basadas en la distribución Singh-Maddala. Respecto a la estimación de cuantiles en la población de pobres, y como es también de esperar, el estimador \hat{Y}_γ^* es más eficiente a medida que aumenta la fracción de muestreo. Este estimador de cuantiles también es más eficiente a medida que aumenta el orden del cuantil, pero la ganancia en eficiencia es bastante menor en comparación con el problema de la estimación de la función de distribución a partir de órdenes superiores o iguales a 0.5, es decir, $\gamma \geq 0.5$. Al igual que en el caso de la estimación de la función de distribución, las distintas asimetrías de las variables de las poblaciones basadas en la distribución Singh-Maddala no tienen un impacto relevante en la eficiencia del estimador \hat{Y}_γ^* . Un estudio complementario sobre el impacto de la asimetría en los estimadores que estamos analizando puede consultarse en Álvarez et al. (2012).

Todos los resultados y comentarios anteriores están basados en muestras seleccionadas mediante MAS. A continuación, desde la Tabla 2.12 hasta la Tabla 2.22, pueden observarse los resultados obtenidos con muestras seleccionadas mediante el método de Midzuno, el cual se trata de un diseño muestral con probabilidades desiguales (véase Midzuno 1952, Singh 2003).

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-0.7	-0.8	0.3	0.5	11.3	11.3	25.9	25.9
	0.10	-0.2	-0.3	0.2	0.3	7.5	7.6	17.6	17.6
	0.20	0.2	0.1	0.3	0.3	4.9	4.9	11.7	11.7
0.3	0.05	-0.1	-0.1	0.3	0.4	5.1	5.1	14.2	14.2
	0.10	-0.1	-0.1	0.2	0.2	3.6	3.6	9.5	9.5
	0.20	-0.1	-0.1	0.2	0.2	2.5	2.5	6.4	6.4
0.5	0.05	0.3	0.3	0.2	0.3	3.1	3.1	10.1	10.1
	0.10	0.2	0.2	0.2	0.3	2.0	2.0	6.9	6.9
	0.20	0.2	0.2	0.2	0.2	1.2	1.2	4.6	4.6
0.7	0.05	0.1	0.1	0.2	0.3	3.3	3.3	7.9	8.0
	0.10	0.0	0.0	0.3	0.3	2.3	2.3	5.3	5.3
	0.20	0.1	0.0	0.2	0.2	1.6	1.6	3.6	3.6
0.9	0.05	-0.4	-0.4	0.0	0.2	3.2	3.2	6.2	6.2
	0.10	-0.4	-0.4	0.2	0.3	2.2	2.2	4.3	4.3
	0.20	-0.4	-0.4	0.2	0.2	1.5	1.5	2.9	2.9

Tabla 2.12. Valores de SR y ECMR de los estimadores tipo Hájek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Bulgaria. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	1.2	1.2	0.3	0.4	9.0	9.0	28.3	28.4
	0.10	1.1	1.1	0.3	0.3	5.9	5.9	19.7	19.7
	0.20	0.9	0.9	0.2	0.2	3.8	3.8	13.1	13.1
0.3	0.05	0.3	0.3	-0.1	0.0	3.9	3.9	15.1	15.1
	0.10	0.2	0.2	0.1	0.1	2.8	2.8	10.3	10.3
	0.20	0.3	0.3	-0.1	-0.1	1.9	1.9	6.9	6.9
0.5	0.05	-0.4	-0.4	0.1	0.2	2.9	2.9	10.6	10.6
	0.10	-0.3	-0.3	-0.1	0.0	1.9	1.9	7.2	7.2
	0.20	-0.2	-0.2	-0.1	-0.1	1.2	1.2	4.7	4.7
0.7	0.05	-0.3	-0.3	0.0	0.1	2.4	2.4	7.7	7.7
	0.10	-0.2	-0.2	0.0	0.0	1.7	1.7	5.2	5.2
	0.20	-0.2	-0.2	-0.1	-0.1	1.2	1.2	3.4	3.4
0.9	0.05	0.1	0.1	0.0	0.1	1.8	1.8	5.5	5.5
	0.10	0.1	0.1	0.0	0.0	1.2	1.2	3.8	3.8
	0.20	0.1	0.1	-0.1	-0.1	0.8	0.8	2.4	2.4

Tabla 2.13. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Eslovenia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-2.2	-2.3	-0.3	-0.2	11.2	11.3	21.6	21.6
	0.10	-1.7	-1.8	-0.4	-0.3	7.9	7.9	14.8	14.8
	0.20	-1.4	-1.4	-0.3	-0.2	5.5	5.5	10.0	10.0
0.3	0.05	0.7	0.7	-0.2	-0.1	5.0	5.0	12.1	12.1
	0.10	0.6	0.6	-0.2	-0.2	3.7	3.7	8.4	8.4
	0.20	0.5	0.5	-0.2	-0.2	2.6	2.6	5.7	5.7
0.5	0.05	0.1	0.1	0.0	0.0	3.0	3.0	8.8	8.8
	0.10	0.2	0.2	-0.2	-0.2	2.1	2.1	6.3	6.3
	0.20	0.2	0.2	-0.2	-0.2	1.4	1.4	4.3	4.3
0.7	0.05	-0.3	-0.3	0.0	0.0	2.5	2.5	7.2	7.2
	0.10	-0.2	-0.2	-0.3	-0.3	1.7	1.7	5.2	5.2
	0.20	-0.1	-0.1	-0.3	-0.2	1.2	1.2	3.6	3.6
0.9	0.05	0.0	0.0	0.0	0.0	2.2	2.2	5.9	5.9
	0.10	0.1	0.1	-0.3	-0.3	1.6	1.6	4.4	4.4
	0.20	0.1	0.1	-0.3	-0.3	1.2	1.2	3.1	3.1

Tabla 2.14. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Polonia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-0.1	-0.4	-0.4	-0.2	19.8	20.0	37.4	37.5
	0.10	0.7	0.4	-0.3	-0.2	13.3	13.3	25.1	25.1
	0.20	1.2	1.1	-0.8	-0.7	9.0	9.1	16.7	16.7
0.3	0.05	-0.7	-0.8	-0.8	-0.6	5.6	5.6	20.6	20.6
	0.10	-0.7	-0.7	-0.4	-0.3	3.8	3.8	14.1	14.1
	0.20	-0.4	-0.4	-0.6	-0.6	2.4	2.4	9.2	9.2
0.5	0.05	-0.2	-0.4	-0.9	-0.7	4.1	4.1	15.3	15.4
	0.10	-0.2	-0.3	-0.5	-0.4	2.7	2.8	10.4	10.4
	0.20	0.0	0.0	-0.7	-0.7	1.7	1.7	6.8	6.8
0.7	0.05	0.1	0.1	-0.7	-0.5	3.4	3.4	12.3	12.3
	0.10	0.1	0.0	-0.4	-0.3	2.3	2.3	8.3	8.3
	0.20	0.1	0.1	-0.7	-0.7	1.5	1.5	5.5	5.5
0.9	0.05	0.5	0.5	-1.4	-1.2	3.5	3.5	9.5	9.5
	0.10	0.3	0.3	-0.6	-0.5	2.4	2.4	6.8	6.8
	0.20	0.4	0.4	-0.7	-0.7	1.6	1.6	4.5	4.5

Tabla 2.15. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Bélgica. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-0.8	-0.9	1.4	1.5	26.5	26.5	32.1	32.1
	0.10	-1.1	-1.1	0.8	0.9	18.4	18.4	22.1	22.1
	0.20	-1.1	-1.1	0.4	0.4	12.3	12.3	14.8	14.8
0.3	0.05	-0.9	-0.9	1.2	1.4	7.6	7.6	17.8	17.8
	0.10	-0.6	-0.6	0.7	0.7	5.2	5.2	12.1	12.1
	0.20	-0.5	-0.5	0.4	0.5	3.6	3.6	8.2	8.2
0.5	0.05	-0.3	-0.4	1.1	1.3	4.7	4.7	13.2	13.3
	0.10	-0.1	-0.1	0.6	0.7	3.2	3.2	9.0	9.0
	0.20	0.0	0.0	0.4	0.5	2.1	2.1	6.0	6.0
0.7	0.05	-0.7	-0.8	1.0	1.2	4.0	4.0	10.5	10.6
	0.10	-0.5	-0.6	0.5	0.6	2.8	2.8	7.3	7.3
	0.20	-0.5	-0.5	0.4	0.5	1.9	1.9	4.9	4.9
0.9	0.05	-0.3	-0.3	0.4	0.5	3.0	3.0	7.6	7.6
	0.10	-0.1	-0.1	0.4	0.5	2.0	2.0	5.7	5.8
	0.20	-0.1	-0.1	0.4	0.5	1.3	1.3	4.1	4.1

Tabla 2.16. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Reino Unido. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	12.7	11.6	0.1	0.2	78.1	77.8	17.5	17.6
	0.10	5.0	4.4	0.1	0.1	53.4	53.2	12.2	12.2
	0.20	0.1	-0.2	0.0	0.1	33.4	33.4	8.2	8.2
0.3	0.05	0.1	0.1	0.2	0.3	7.1	7.1	9.6	9.6
	0.10	0.1	0.0	0.2	0.2	4.9	4.9	6.6	6.6
	0.20	0.1	0.1	0.1	0.1	3.3	3.3	4.3	4.3
0.5	0.05	-0.2	-0.3	0.2	0.3	3.9	3.9	7.0	7.0
	0.10	-0.3	-0.3	0.2	0.2	2.7	2.7	4.7	4.7
	0.20	-0.3	-0.3	0.1	0.1	1.7	1.7	3.0	3.0
0.7	0.05	-0.1	-0.1	0.1	0.2	3.0	3.0	5.4	5.4
	0.10	-0.1	-0.1	0.2	0.2	2.1	2.1	3.6	3.6
	0.20	0.0	0.0	0.1	0.1	1.4	1.4	2.3	2.3
0.9	0.05	0.0	0.0	0.2	0.3	2.0	2.0	4.3	4.3
	0.10	0.0	0.0	0.1	0.2	1.4	1.4	2.9	2.9
	0.20	0.0	0.0	0.2	0.2	0.9	0.9	1.8	1.8

Tabla 2.17. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en España. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	1.1	1.1	-0.8	-0.7	28.9	28.8	18.5	18.5
	0.10	1.7	1.7	-0.8	-0.8	20.9	20.9	12.7	12.7
	0.20	1.9	1.9	-0.5	-0.5	14.3	14.2	8.5	8.5
0.3	0.05	0.4	0.4	-0.8	-0.7	5.5	5.5	9.9	9.9
	0.10	0.5	0.5	-0.8	-0.7	3.9	3.9	6.9	6.9
	0.20	0.4	0.4	-0.5	-0.5	2.7	2.7	4.7	4.7
0.5	0.05	0.1	0.1	-0.7	-0.7	2.8	2.8	7.2	7.2
	0.10	0.2	0.2	-0.8	-0.7	1.9	1.9	5.0	5.0
	0.20	0.1	0.1	-0.6	-0.6	1.3	1.3	3.4	3.4
0.7	0.05	0.3	0.3	-0.7	-0.7	2.0	2.0	5.6	5.6
	0.10	0.3	0.3	-0.7	-0.7	1.3	1.3	3.9	3.9
	0.20	0.3	0.3	-0.6	-0.6	0.9	0.9	2.7	2.7
0.9	0.05	0.3	0.2	-0.7	-0.7	1.8	1.8	4.4	4.4
	0.10	0.3	0.3	-0.7	-0.7	1.2	1.2	3.1	3.1
	0.20	0.2	0.2	-0.6	-0.6	0.8	0.8	2.2	2.2

Tabla 2.18. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Italia. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-1.6	-1.6	1.6	1.8	28.0	28.0	33.9	34.0
	0.10	-2.1	-2.1	1.8	1.8	19.2	19.2	23.3	23.4
	0.20	-1.1	-1.1	0.8	0.8	12.8	12.8	15.5	15.5
0.3	0.05	-0.7	-0.7	1.4	1.6	8.7	8.7	18.3	18.4
	0.10	-0.5	-0.5	1.4	1.5	5.9	5.9	12.6	12.6
	0.20	0.0	0.0	0.8	0.9	3.9	3.9	8.5	8.5
0.5	0.05	-1.2	-1.2	1.4	1.6	5.9	5.9	12.9	12.9
	0.10	-0.9	-0.9	1.2	1.3	4.1	4.1	8.9	8.9
	0.20	-0.6	-0.6	0.9	0.9	2.7	2.7	6.0	6.0
0.7	0.05	-0.6	-0.7	1.3	1.5	4.4	4.4	9.7	9.8
	0.10	-0.5	-0.5	1.2	1.3	3.1	3.1	6.8	6.8
	0.20	-0.4	-0.4	0.9	1.0	2.2	2.2	4.6	4.7
0.9	0.05	-0.9	-0.9	0.9	1.1	3.3	3.3	6.9	6.9
	0.10	-0.8	-0.8	1.2	1.3	2.3	2.3	5.2	5.2
	0.20	-0.6	-0.6	0.9	1.0	1.6	1.6	3.6	3.6

Tabla 2.19. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo datos extraídos de la EU-SILC en Lituania. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	0.4	0.3	1.0	1.1	8.4	8.4	47.7	47.8
	0.10	0.4	0.4	0.7	0.7	5.6	5.6	32.0	32.0
	0.20	0.4	0.4	0.7	0.8	3.7	3.7	21.5	21.5
0.3	0.05	0.1	0.1	0.8	0.0	4.6	4.6	26.4	26.4
	0.10	0.1	0.1	0.6	0.6	3.1	3.1	18.1	18.2
	0.20	0.1	0.1	0.3	0.4	2.1	2.1	11.9	11.9
0.5	0.05	-0.5	-0.6	0.9	1.0	3.6	3.6	19.5	19.6
	0.10	-0.4	-0.4	0.7	0.8	2.5	2.5	13.5	13.6
	0.20	-0.3	-0.3	0.3	0.3	1.6	1.6	8.8	8.8
0.7	0.05	-0.1	-0.1	1.0	1.1	2.8	2.8	15.9	15.9
	0.10	-0.1	-0.1	0.6	0.6	1.9	1.9	11.0	11.0
	0.20	0.0	0.0	0.3	0.3	1.2	1.2	7.1	7.1
0.9	0.05	-0.2	-0.2	-0.6	-0.5	2.5	2.5	11.2	11.2
	0.10	-0.1	-0.1	0.1	0.1	1.7	1.7	8.3	8.3
	0.20	-0.1	-0.1	0.2	0.2	1.1	1.1	5.8	5.8

Tabla 2.20. Valores de SR y ECMR de los estimadores tipo Hájek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 11.5$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-0.4	-0.6	1.1	1.2	7.4	7.5	48.2	48.2
	0.10	-0.3	-0.3	0.2	0.2	5.1	5.1	32.9	32.9
	0.20	-0.2	-0.2	0.1	0.2	3.4	3.4	21.7	21.8
0.3	0.05	-0.5	-0.5	0.7	0.9	4.2	4.2	26.8	26.8
	0.10	-0.3	-0.3	0.0	0.1	2.9	2.9	18.3	18.3
	0.20	-0.2	-0.2	0.0	0.0	2.0	2.0	12.1	12.1
0.5	0.05	-0.2	-0.2	0.7	0.8	3.4	3.4	19.8	19.9
	0.10	-0.1	-0.1	0.2	0.2	2.4	2.4	13.7	13.7
	0.20	0.0	0.0	-0.1	0.0	1.6	1.6	9.0	9.0
0.7	0.05	-0.1	-0.1	0.5	0.6	3.2	3.2	15.6	15.7
	0.10	0.0	0.0	0.0	0.1	2.3	2.3	10.9	10.9
	0.20	0.1	0.0	0.0	0.0	1.6	1.6	7.4	7.4
0.9	0.05	-0.1	-0.1	-0.8	-0.7	2.7	2.7	11.1	11.1
	0.10	0.1	0.1	-0.3	-0.2	1.9	1.9	8.5	8.5
	0.20	0.1	0.1	-0.1	0.0	1.3	1.3	6.1	6.1

Tabla 2.21. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 8.1$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

γ	f	SR (%)				ECMR (%)			
		\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$	\hat{Y}_γ^*	\tilde{Y}_γ^*	$\hat{F}_H^*(t)$	$\tilde{F}^*(t)$
0.1	0.05	-0.8	-1.1	1.1	1.2	10.8	10.9	56.8	56.9
	0.10	0.2	0.2	0.3	0.4	5.6	5.6	38.9	38.9
	0.20	0.2	0.2	0.1	0.1	3.4	3.4	25.7	25.7
0.3	0.05	-1.3	-1.4	1.4	1.5	5.4	5.4	30.2	30.3
	0.10	-0.9	-0.9	0.5	0.5	3.6	3.6	20.5	20.5
	0.20	-0.6	-0.6	0.1	0.1	2.3	2.3	13.6	13.6
0.5	0.05	-0.5	-0.6	1.2	1.3	3.5	3.5	21.3	21.3
	0.10	-0.2	-0.3	0.5	0.5	2.4	2.4	14.5	14.5
	0.20	0.0	-0.1	0.1	0.1	1.5	1.5	9.7	9.7
0.7	0.05	-0.4	-0.4	1.4	1.5	2.6	2.6	16.0	16.0
	0.10	-0.2	-0.2	0.5	0.6	1.8	1.8	10.9	10.9
	0.20	-0.1	-0.1	0.2	0.2	1.2	1.2	7.2	7.2
0.9	0.05	-0.3	-0.3	0.0	0.2	2.0	2.0	10.0	10.0
	0.10	-0.1	-0.1	0.2	0.2	1.3	1.3	7.6	7.6
	0.20	0.0	0.0	0.2	0.2	0.8	0.8	5.3	5.3

Tabla 2.22. Valores de SR y ECMR de los estimadores tipo Háyek y Horvitz-Thompson para Y_γ^* y $F^*(t)$, con $t = Y_\gamma^*$, y asumiendo la población basada en la distribución Singh-Maddala con coeficiente de correlación lineal $\rho = 0.9$ y coeficiente de asimetría $g = 4.0$. $f = n/N$ es la fracción muestral. Muestras seleccionadas mediante el método de Midzuno.

En la Sección 2.3 se adelantaba que para el caso de diseños muestrales con probabilidades desiguales, tal como el caso del método de Midzuno, se podrían plantear dos estimadores diferentes basados en el diseño, tanto para el caso de la estimación de la función de distribución en la población de pobres, como para el caso de la estimación de un determinado cuantil en la población de pobres. A continuación se analizan estos estimadores mediante nuevos estudios de simulación basados en algunas de las poblaciones del Apéndice I.

En primer lugar analizamos los sesgos de ambos estimadores y para todas las poblaciones de este estudio. Podemos observar que para muestras seleccionadas bajo el método de Midzuno también se obtienen resultados muy positivos con respecto al sesgo relativo. Tan sólo para el caso de Polonia cuando $\gamma = 0.1$ y $f = 5\%$ se obtienen sesgos por encima del 2%. Al igual que en el caso de muestras seleccionadas mediante MAS, los sesgos obtenidos en la población EU-SILC y con datos de España son muy elevados cuando $\gamma = 0.1$ y $f \leq 10\%$, aunque en el resto de las situaciones los sesgos si toman valores razonables.

Respecto a la eficiencia de los estimadores se pueden tener en consideración las conclusiones y comentarios realizados en el caso de muestras seleccionadas bajo MAS. Además, podemos observar que ambos tipos de estimadores proporcionan resultados similares en términos de eficiencia, por lo que la elección de uno u otro estimador para la estimación de los parámetros en estudios no implicaría, según este estudio, una diferencia significativa en cuanto a sus estimaciones.

Capítulo 3

Estimación de la proporción de pobres

3.1. Introducción

El problema de la estimación eficiente de medias y totales en el contexto del muestreo en poblaciones finitas y en presencia de información auxiliar ha sido altamente estudiado, y numerosas técnicas de estimación para tales parámetros han sido propuestas. Por ejemplo, los métodos de calibración (Deville y Sarndall 1992) y verosimilitud pseudo empírica (Chen y Qin 1993, Chen y Sitter 1999) surgieron inicialmente para este propósito. Incluso la estimación de otros parámetros más complejos, como funciones de distribución y cuantiles, han tenido una investigación relevante por medio de artículos como Chambers y Dunstan (1986), Kuk y Mak (1989), Rao et al. (1990), Francisco y Fuller (1991), Dorfman y Hall (1993), Chen y Wu (2002), Mak y Kuk (1993), Silva y Skinner (1995), Chen y Wu (2002), Harms y Duchesne (2006), Dorfman (2009), etc, tal como se pudo constatar en la breve revisión realizada en el Capítulo 2.

Los estimadores más conocidos para la estimación de la media poblacional y basados en información auxiliar son los conocidos como estimadores de tipo razón y regresión (véase, por ejemplo, Cochran 1977, Gordon 1983, Chaudhuri y Vos 1988, Hedayat y Sinha 1991, Särndal et al. 1992, Fernández-García y Mayor-Gallego 1994, Mukhopadhyay 2000, Singh 2003, Chaudhuri y Stenger 2005, etc). A continuación se introducen tales estimadores con el fin de poder analizar las

Capítulo 3. Estimación de la proporción de pobres

similitudes y diferencias existentes en comparación con los estimadores de tipo razón y regresión propuestos en este capítulo.

Existen numerosos métodos de estimación que tienen el objetivo de mejorar la estimación de la variable de interés, y la mayoría de ellos están basados en el uso de información auxiliar relacionada con la variable de interés en la etapa de estimación. La mayoría de encuestas oficiales sobre ingresos y condiciones de vida contienen variables adicionales relacionadas con la variable de interés. Tales variables pueden ser utilizadas en la etapa de estimación como variables auxiliares para mejorar la estimación de parámetros poblacionales tal como distintos indicadores de pobreza. Éste es el objetivo de este capítulo, es decir, el uso de información auxiliar para obtener estimadores con mejores propiedades para la proporción poblacional de pobres o proporción poblacional de individuos que viven por debajo del umbral oficial de pobreza. Por ejemplo, en Molina y Rao (2010) se utilizan variables auxiliares para mejorar la estimación de distintos indicadores de pobreza en el contexto de estimaciones en áreas pequeñas o subpoblaciones. Destacamos que al igual que se hace en este trabajo, el estudio llevado a cabo por Molina y Rao (2010) también está basado en datos reales extraídos de la Encuesta Europea sobre Ingresos y Condiciones de Vida (EU-SILC). Por simplicidad, en este capítulo, se asume el uso de una única variable auxiliar (la cual estamos denominando x) relacionada con la variable de interés (la cual se seguirá denotando como y). La extensión de los métodos de estimación propuestos al caso de varias variables auxiliares es bastante simple a partir de las metodologías existentes basadas en más de una variable auxiliar (véase por ejemplo Särndal et al. 1992 p. 225).

Los métodos más conocidos, basados en técnicas que incluyen información auxiliar, para la estimación de la media poblacional

Capítulo 3. Estimación de la proporción de pobres

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$$

de la variable de interés y son los estimadores de tipo razón y regresión. Recordamos que N denotaba el número total de individuos en la población en estudio U . Asumiendo una muestra s , con tamaño n , seleccionada mediante MAS, el estimador de razón de la media poblacional \bar{Y} (Särndal et al. 1992, p.229) se define como

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}} \bar{X}$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i \in s} y_i$$

es el estimador de Horvitz-Thompson de \bar{Y} . Del mismo modo que se han definido las medias \bar{y} e \bar{Y} de la variable de interés se definen las medias \bar{x} y \bar{X} de la variable auxiliar y anteriormente utilizadas, es decir, para ello simplemente sustituimos y_i por x_i en las definiciones de \bar{y} e \bar{Y} , respectivamente.

El estimador de tipo regresión de la media \bar{Y} (Cassel et al. 1976, 1977; Särndal et al. 1992, p. 219) se define como

$$\bar{y}_{reg} = \bar{y} - \hat{\beta}_x (\bar{X} - \bar{x})$$

donde

$$\hat{\beta}_x = \left[\sum_{i \in s} (x_i - \bar{x})^2 \right]^{-1} \sum_{i \in s} (x_i - \bar{x})(y_i - \bar{y}).$$

Destacamos que las técnicas de estimación de tipo razón y regresión se basan en modelos lineales de superpoblación, sin y con término independiente o intercepto, respectivamente. Además, es importante destacar que para obtener estimadores más precisos y para que estas técnicas de tipo razón y regresión con una variable auxiliar tengan un buen comportamiento, es necesario que la variable de interés y la variable auxiliar tengan una relación lineal fuerte.

Existen otros métodos de estimación para la media y otros parámetros en la literatura. Por ejemplo, el método de calibración (Deville y Särndal 1992) y el método de verosimilitud pseudo empírica (Chen y Sitter 1999) son metodologías recientes de estimación basadas en información auxiliar.

3.2. Definición y estimación de la proporción de pobres

El objetivo de este capítulo es la estimación de la proporción poblacional de pobres o proporción poblacional de individuos que viven por debajo del umbral de pobreza L , la cual se define como

$$P_0 = \frac{N_p}{N},$$

donde N_p denota el número de individuos considerados como pobres en toda la población U . Por simplicidad y sin pérdida de generalidad, consideraremos que L es una línea de pobreza relativa. Destacamos que el número total de pobres puede obtenerse como

Capítulo 3. Estimación de la proporción de pobres

$$N_p = \sum_{i=1}^N A_i$$

donde A es una variable binaria o dicotómica que toma el valor 1 si un individuo es considerado pobre y toma el valor 0 en caso contrario, es decir, $A_i = \delta(y_i \leq L) = 1$ si el i -ésimo individuo es pobre y $A_i = 0$ en caso contrario. Recordamos que $\delta(\cdot)$ es la función indicadora que toma el valor 1 si su argumento es cierto, y esta función toma el valor 0 en caso contrario. Podemos observar que la proporción P_0 puede verse como la media aritmética de la variable binaria A , cuyos valores están dados por A_1, \dots, A_N .

En la práctica, la proporción de pobres poblacional P_0 es desconocida, ya que resulta imposible obtener esta medida de pobreza a nivel poblacional. Bajo este escenario, las agencias de estadística y otros organismos utilizan datos muestrales para obtener una estimación de este indicador de pobreza. Asumiendo MAS, el estimador estándar de P_0 viene dado por el estimador de tipo Horvitz-Thompson, cuya expresión es la siguiente:

$$\hat{p}_0 = \frac{n_p}{n}, \quad (3.1)$$

donde

$$n_p = \sum_{i=1}^n A_i$$

denota el número de pobres en la muestra s . Se puede observar que el estimador \hat{p}_0 también está basado en la variable binaria A , y esto implica que para obtener este estimador tan sólo es necesario conocer el número de individuos que en la

Capítulo 3. Estimación de la proporción de pobres

muestra s están clasificados como pobres. Bajo este escenario se podría utilizar tanto líneas de pobreza relativas como líneas de pobreza absolutas. Desatacamos que en este capítulo se asume que el umbral de pobreza L viene establecido por el organismo oficial correspondiente, es decir, L es una constante fijada a través de una cantidad oficial. Por ejemplo, EUROSTAT considera una línea de pobreza relativa L , donde L se determina como el 60% de la mediana de los ingresos por unidad de consumo. Para el caso de líneas de pobreza absolutas, la variable binaria A se define de una forma similar, es decir, $A_i = 1$ si el i -ésimo individuo es pobre según el criterio establecido por la línea de pobreza absoluta, y $A_i = 0$ en caso contrario.

Como se comentaba anteriormente, la proporción de pobres requiere la especificación de una línea de pobreza que divida la población en pobres y no pobres. Las líneas de pobreza más conocidas y utilizadas se pueden clasificar como absolutas y relativas. Las líneas de pobreza absolutas las suelen utilizar en las estadísticas oficiales de Estados Unidos y del Banco Mundial. Las líneas de pobreza relativas vienen especificadas, en general, como un porcentaje de la mediana de ingresos. Este método se utiliza, por ejemplo, en EUROSTAT. Estas dos metodologías (enfoques relativos y absolutos para el umbral de pobreza) para analizar el fenómeno de la pobreza presentan sus ventajas e inconvenientes, y su uso depende del objetivo a seguir en el análisis de la pobreza. Para un mayor detalle de estos temas puede consultarse Khandker (2005) pp. 43-52, Foster (1998), Madden (2000), Jolliffe (2001), etc. Resaltamos que el objetivo de este trabajo no es llevar a cabo una discusión o análisis entre el uso de medidas de pobreza relativa y absolutas, puesto que las aportaciones de este capítulo, por ejemplo, podrían aplicarse tanto en el caso de líneas de pobreza absolutas como en el caso de líneas de pobreza relativas. En particular, los métodos de estimación propuestos en este capítulo tan sólo requieren una variable auxiliar cuantitativa relacionada linealmente con la variable binaria de interés que

Capítulo 3. Estimación de la proporción de pobres

clasifica a los individuos como pobre y no pobres (véase, por ejemplo, la variable binaria A , definida anteriormente).

A continuación, se presentan algunos métodos existentes basados en información auxiliar para el problema de la estimación de una proporción poblacional.

En el caso de una variable auxiliar x dicotómica, el problema de la estimación de la proporción de pobres con información auxiliar puede resolverse fácilmente. Para ello, simplemente se pueden utilizar las técnicas de estimación para proporciones propuestas por Rueda et al. (2011a), Rueda et al. (2011b) y Muñoz et al. (2012). Sin embargo, en este trabajo asumimos que x es una variable cuantitativa. En efecto, la proporción de pobres suele obtenerse a partir de variables tal como el ingreso o el gasto de individuos o familias, las cuales son variables cuantitativas. En esta situación es muy probable que las variables auxiliares con una mayor relación lineal con la variable de interés sean también cuantitativas. Por tanto, los métodos propuestos por Rueda et al. (2011a), Rueda et al. (2011b) y Muñoz et al. (2012) no pueden aplicarse en esta situación donde la variable auxiliar es cuantitativa, puesto que tales metodologías asumen, exclusivamente, una variable auxiliar dicotómica.

En el caso de una variable auxiliar cuantitativa es posible utilizar el estimador de regresión logística (Lehtonen y Veijanen 1998) para el problema de la estimación de la proporción de pobres. Este estimador está basado en un modelo de regresión logística, y por tanto requiere la estimación del parámetro asociado a dicho modelo. Generalmente se utilizan algoritmos complejos como el de Newton-Raphson para estimar el parámetro asociado a este método de estimación (véase Duchesne 2003). Destacamos también, como aportación en este trabajo, que el análisis del comportamiento empírico del estimador de regresión

Capítulo 3. Estimación de la proporción de pobres

logística bajo diferentes escenarios relacionados con el problema de la estimación de la proporción de pobres es un tema que aún no había sido estudiado. En este capítulo se compara el comportamiento empírico del estimador de regresión logística con el comportamiento de los métodos de estimación propuestos para la proporción de pobres. El estimador de regresión logística para el problema de la estimación de la proporción de pobres y en presencia de una variable auxiliar cuantitativa x (Lehtonen y Veijanen 1998) viene dado por

$$\hat{p}_{0.Lgreg} = \hat{p}_0 + \bar{\mu} - \hat{\mu}, \quad (3.2)$$

donde

$$\bar{\mu} = \frac{1}{N} \sum_{i \in U} \hat{\mu}_i$$

y

$$\hat{\mu} = \frac{1}{n} \sum_{i \in s} \hat{\mu}_i.$$

Los valores

$$\hat{\mu}_i = \frac{\exp(x_i \hat{\theta})}{1 + \exp(x_i \hat{\theta})}, \quad i = 1, \dots, N,$$

son las predicciones basadas en el modelo de regresión logística

$$\Pr(A_i = 1) = \mu_i = \frac{\exp(x_i \theta)}{1 + \exp(x_i \theta)}$$

$$\Pr(A_i = 0) = 1 - \mu_i = 1 - \Pr(A_i = 1)$$

Capítulo 3. Estimación de la proporción de pobres

Para estimar el parámetro θ de este modelo se puede utilizar el método de máxima verosimilitud. La función de verosimilitud correspondiente viene dada por (véase también Särndal et al. 1992, p. 499, Lehtonen y Veijanen 1998, Duchesne 2003)

$$L(\theta) = \prod_{i \in U_1} \mu_i \prod_{i \in U_0} (1 - \mu_i),$$

donde $U_1 = \{i : i \in U \text{ y } A_i = 1\}$ y $U_0 = U - U_1$. Como puede verse en Duchesne (2003), puede utilizarse el algoritmo de Newton-Raphson para encontrar la solución a este problema de maximización.

El problema de la estimación de una proporción, y la proporción de pobres en particular, mediante técnicas basadas en información auxiliar que permitan una estimación más precisa no ha sido tan discutido como la estimación de parámetros tal como la media, total, función de distribución, cuantiles, etc. Este hecho supone que se podrían mejorar considerablemente los métodos de estimación existentes para la proporción de pobres. Destacamos que aunque la estimación de una proporción poblacional puede considerarse como un caso particular de estimación de la media poblacional, donde la variable de interés en lugar de ser cuantitativa es dicotómica, la estimación de una proporción poblacional requiere una investigación específica, puesto que en la estimación de una media poblacional no se plantean escenarios como que las estimaciones deben encontrarse dentro del intervalo $[0,1]$, el estimador se puede definir a partir de su complementario, etc. Algunos estimadores de proporciones basados en información auxiliar pueden consultarse en Rueda et al. (2011a), Rueda et al. (2011b) y Muñoz et al. (2012).

Capítulo 3. Estimación de la proporción de pobres

Cuando se desea investigar a fondo el problema de la estimación de un determinado parámetro, como por ejemplo una proporción, no es suficiente el estudio de técnicas de estimación eficientes para el problema de la estimación puntual. Al contrario, el estudio con datos obtenidos mediante una encuesta por muestreo puede plantear problemas adicionales que también resulta necesario conocerlos y tratarlos mediante las técnicas más oportunas. De este modo, adicionalmente al problema de la estimación puntual, parece razonable abordar problemas complementarios como el estudio de la varianza de los estimadores, y a partir de este estudio, analizar también la estimación mediante intervalos de confianza de la proporción de pobres.

La estimación de una proporción poblacional tiene importantes aplicaciones en el campo de la Economía, y en general en las Ciencias Sociales, puesto que algunos de los indicadores económicos o sociales están basados en variables dicotómicas o requieren el uso de una proporción para la determinación de los mismos. Por ejemplo, la información de numerosas variables recogidas en las Encuestas de Presupuestos Familiares, Encuestas de Condiciones de Vida (ECV), incluso en la propia EU-SILC, se puede recoger de una forma cualitativa a través de categorías, y el uso de proporciones se hace necesario en esta situación. Los últimos resultados publicados por EUROSTAT basados en las encuestas de EU-SILC de 2012 muestran que la proporción de población en riesgo de pobreza (o tasa de pobreza) es del 17% en la zona Euro (17 países) comparado con el 22,2% en España y el 16,1% en Alemania (véase <http://epp.eurostat.ec.europa.eu>).

La aplicación que analizaremos en este capítulo es la estimación de la proporción de individuos que viven por debajo del umbral de pobreza, también conocida como proporción de pobres. Berger y Skinner (2003), Medeiros (2006), Molina y Rao (2010), Rueda y Muñoz (2011), Crettaz y Sutter (2013), Navicke et

Capítulo 3. Estimación de la proporción de pobres

al. (2013), Giambona y Vasallo (2014), etc. son algunas de las numerosas referencias que analizan la proporción de pobres y otros indicadores de pobreza relacionados con este parámetro.

Una definición formal (aceptada por el INE, EUROSTAT, etc) de persona pobre es cuando tal individuo se encuentra en una situación de clara desventaja, económica y social, respecto al resto de personas de su entorno. El estudio de la pobreza y en especial del umbral de pobreza y de la proporción de pobres debería tener especial interés en la sociedad de un país, así como en las políticas sociales que pretenda llevar a cabo, sobre todo en un momento del tiempo en el que la lucha contra la pobreza es una prioridad en las principales agendas de muchos países y organizaciones. Por ejemplo, erradicar la extrema pobreza es el primero de los objetivos del programa “*Millennium Development Goals*”. Además, una de las iniciativas del programa “*Europe 2020 strategy*” es que al menos 20 millones de personas salgan de la pobreza y exclusión social (véase <http://ec.europa.eu/europe2020>). Por todo lo anterior y como se discute en Molina y Rao (2010), debido a que la proporción de pobres real es desconocida, pero estimada a partir de datos muestrales, es muy importante la disponibilidad de un buen procedimiento de estimación que utilice información más precisa relacionada con las condiciones de vida de la población, lo cual puede resultar útil para este importante objetivo de reducir la pobreza.

El objetivo de este capítulo es plantear y definir nuevos métodos de estimación para la proporción poblacional P_0 de individuos que viven por debajo del umbral de pobreza. Se persigue el objetivo de que estos nuevos procedimientos de estimación sean más precisos que los actuales. En la Sección 3.3 se definen los nuevos estimadores para la proporción de pobres o proporción de individuos que viven por debajo del umbral de pobreza. Las metodologías propuestas utilizan información auxiliar en la etapa de estimación con el fin de

Capítulo 3. Estimación de la proporción de pobres

obtener estimaciones más precisas. En esta Sección 3.3 son varias las aportaciones realizadas. En primer lugar (Sección 3.3.1), se describen las transformaciones propuestas y requeridas para poder obtener los estimadores propuestos. En la Secciones 3.3.2 y 3.3.3 se describen los estimadores de la proporción de pobres de tipo razón y regresión, y se analizan sus correspondientes propiedades teóricas. Para completar este estudio, se plantea la estimación de la proporción de pobres mediante intervalos de confianza y se realiza una extensión de este estudio al caso de muestras seleccionadas mediante un diseño muestral general.

En la Sección 3.4, tanto los estimadores puntuales como los intervalos de confianza, son evaluados empíricamente mediante estudios de simulación Monte Carlo. En estos estudios el objetivo es estimar la proporción de pobres con datos reales extraídos de la EU-SILC. Destacamos que con el fin de obtener resultados más consistentes, este estudio se ha realizado con datos de 8 países de la Unión Europea con distintas características, principalmente en lo que respecta a la proporción de pobres en tales países. Además del estudio de simulación basado en estos datos reales, el análisis del comportamiento de los estimadores propuestos se completa con poblaciones artificiales basadas en las distribuciones Gamma y Singh-Maddala. Destacamos que este estudio es similar al realizado en en Capítulo 2, pero en esta ocasión se están evaluando otros parámetros y procedimientos de estimación tal como los intervalos de confianza. En el Apéndice I de esta tesis se describen todas las poblaciones consideradas en los estudios de simulación de este capítulo. Además, y como también se ha mencionado con anterioridad, en este apéndice se realiza un estudio descriptivo completo y se analizan otros aspectos que pueden ser interesantes tal como el efecto de los outliers o datos anómalos en algunas medidas estadísticas y de pobreza.

3.3. Métodos de estimación propuestos

En esta sección se describen los métodos de estimación alternativos para la proporción de pobres, los cuales asumen que la variable auxiliar x es de carácter cuantitativo. El objetivo de tales métodos es transformar la variable cuantitativa x en una variable dicotómica, para posteriormente utilizar los métodos de razón y regresión descritos anteriormente en la Sección 3.1. Las transformaciones propuestas están descritas en la Sección 3.3.1.

3.3.1. Transformaciones propuestas

A partir de las expresiones (3.1) y (3.2) podemos observar que tales estimadores de P_0 están basados en la variable dicotómica A . Es muy probable que los modelos lineales de superpoblación de los estimadores de razón y regresión (véase la Sección 3.1) no sean válidos si la variable auxiliar x es una variable cuantitativa, puesto que la variable dependiente es la variable dicotómica A , mientras que la variable independiente es una variable cuantitativa. En estos casos se recomienda utilizar un estimador basado en un modelo de regresión logística, tal como el estimador $\hat{p}_{0.Lgreg}$ definido en la ecuación (3.2). Sin embargo, en este capítulo se propone transformar la variable cuantitativa x en una variable dicotómica, con el objetivo de que puedan utilizarse los estimadores de tipo razón y regresión para una proporción propuestos por Rueda et al. (2011a), Rueda et al. (2011b) y Muñoz et al. (2012).

A continuación se definen tres transformaciones para la variable auxiliar cuantitativa, las cuales nos permitirán obtener las variables dicotómicas B , C y D . Las transformaciones propuestas vienen dadas por

Capítulo 3. Estimación de la proporción de pobres

$$B_i = B_i(\kappa) = \delta(x_i \leq x_{(k)}) \quad (3.3)$$

$$C_i = C_i(\alpha, \beta) = \delta(x_i \leq \alpha X_{\beta}), \quad (3.4)$$

y

$$D_i = D_i(\gamma) = \delta(x_i \leq \gamma \bar{X}), \quad (3.5)$$

donde κ , β y γ son constantes, X_{β} y \bar{X} son el cuantil de orden β y la media poblacional de la variable x , respectivamente, y los valores $x_{(i)}$ denotan el estadístico ordenado de los valores x_i .

3.3.2. Estimadores de tipo razón

Asumiendo la variable dicotómica B definida por la ecuación (3.3), el estimador de tipo razón para la proporción de pobres P_0 se obtiene como

$$\hat{P}_{0,r\kappa} = \frac{\hat{P}_0}{\hat{P}_{0,B}} P_{0,B}, \quad (3.6)$$

donde \hat{p}_0 es el estimador de Horvitz-Thompson de la proporción de pobres definido en la ecuación (3.1),

$$P_{0,B} = \frac{1}{N} \sum_{i \in U} B_i$$

es la media poblacional de la variable dicotómica B , y

Capítulo 3. Estimación de la proporción de pobres

$$\hat{p}_{0.B} = \frac{1}{n} \sum_{\epsilon \in S} B_i$$

es el estimador de tipo Horvitz-Thompson de $P_{0.B}$.

Como se puede observar, el valor B_i depende de la constante κ . Se propone la utilización del valor óptimo de κ , con $\kappa = \{1, \dots, N\}$, que minimiza la varianza muestral de $\hat{p}_{0.r\kappa}$ y la cual viene dada por (véase Rueda et al. 2011b; Muñoz et al. 2012):

$$\hat{V}(\hat{p}_{0.r\kappa}) = \frac{1-f}{n-1} \left(\hat{p}_0 \hat{q}_0 + \hat{R}^2 \hat{p}_{0.B} \hat{q}_{0.B} - 2\hat{R} \hat{\phi} \sqrt{\hat{p}_0 \hat{q}_0 \hat{p}_{0.B} \hat{q}_{0.B}} \right), \quad (3.7)$$

donde $f = n/N$ es la fracción muestral, $\hat{q}_0 = 1 - \hat{p}_0$, $\hat{q}_{0.B} = 1 - \hat{p}_{0.B}$, $\hat{R} = \hat{p}_0 / \hat{p}_{0.B}$ y

$$\hat{\phi} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}}}$$

es el coeficiente V de Cramer basado en la siguiente clasificación de doble entrada entre los atributos A y B (véase también Rueda et al. 2011b):

	B	B^C	
A	n_{11}	n_{12}	$n_{1\bullet}$
A^C	n_{21}	n_{22}	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	n

A^C y B^C denotan, respectivamente, los atributos complementarios de A y B .

Capítulo 3. Estimación de la proporción de pobres

El problema de hallar una expresión analítica para el valor óptimo de κ no es una cuestión simple. En estas situaciones es común utilizar un método de búsqueda intensiva para la obtención del valor óptimo de un parámetro. Por ejemplo, esta metodología ha sido usada por Becsey et al. (1968) y Lerman (1980) para estimar modelos no lineales y parámetros relacionados con otro tipo de regresiones. De esta forma, se propone minimizar numéricamente la varianza estimada de la expresión (3.7) usando este método de búsqueda intensiva y utilizando los puntos $\kappa = \{1, \dots, N\}$, es decir, este método consiste en evaluar la expresión (3.7) en los diferentes puntos y seleccionar el valor de κ que minimiza la expresión (3.7).

Se puede observar que este método podría requerir de la evaluación de la varianza (3.7) en hasta N puntos. En este sentido, debemos tener en cuenta que el número de veces que habría que evaluar la varianza disminuye si la variable auxiliar x tiene valores repetidos, es decir, $x_{(i)} = x_{(j)}$ para cualquier $i \neq j$. Debido a esta posible evaluación excesiva de puntos, se propone utilizar la transformación B anteriormente definida cuando el tamaño de la población N es pequeño. En caso contrario, se recomienda el uso de los atributos C y D definidos en las expresiones (3.4) y (3.5), puesto que la evaluación de la varianza en la búsqueda intensiva es considerablemente menor en estos casos.

Es posible que la solución al problema de minimizar la varianza estimada definida en la expresión (3.7) no sea única, es decir, se obtengan J valores óptimos de κ . Sean $\{\kappa_1, \dots, \kappa_J\}$ los J valores óptimos de κ que minimizan la varianza muestral (3.7). Bajo este escenario se pueden utilizar varias soluciones diferentes. Por ejemplo, la mayoría de los métodos de estimación, tal como los estimadores de tipo razón y regresión, están diseñados para incorporar fácilmente varias variables auxiliares en la etapa de estimación, de forma que el

Capítulo 3. Estimación de la proporción de pobres

uso de tales variables auxiliares adicionales puede producir resultados más precisos. De esta forma, una posible solución es obtener estimadores de la proporción de pobres basados en las J variables dicotómicas derivadas de los J valores óptimos de κ . Esta solución es bastante compleja, especialmente para valores elevados de J , y de aquí que en este trabajo se proponga una solución más simple. En particular, se propone utilizar el valor óptimo κ , donde $\kappa \in \{\kappa_1, \dots, \kappa_J\}$, que minimiza el valor absoluto de la diferencia entre la proporción poblacional del atributo B con respecto a su correspondiente estimador estándar, es decir,

$$|P_{0,B} - \hat{P}_{0,B}| \quad (3.8)$$

Destacamos que la condición definida en la expresión (3.8) es bastante común en distintos métodos de estimación, tal como el método de calibración (véase por ejemplo Deville y Särndal 1992) y el método de verosimilitud pseudo empírica (véase por ejemplo Chen y Sitter 1999).

Teniendo en cuenta la variable dicotómica C , el correspondiente estimador de tipo razón se puede definir de manera similar, es decir,

$$\hat{P}_{0,r\alpha\beta} = \frac{\hat{P}_0}{\hat{P}_{0,C}} P_{0,C} \quad (3.9)$$

donde α y β son las constantes óptimas que minimizan la varianza muestral de $\hat{P}_{0,r\alpha\beta}$. La expresión de esta varianza viene dada por la expresión (3.7) pero sustituyendo la variable dicotómica B por la variable dicotómica C . La transformación C está justificada por la propia definición de la línea de pobreza relativa, la cual puede definirse como $L = \alpha^* Y_{\beta^*}$, donde por ejemplo EUROSTAT o

Capítulo 3. Estimación de la proporción de pobres

el INE considera $\alpha^* = 0.6$ y $\beta^* = 0.5$. Y_{β^*} denota el cuantil de orden β^* de la variable de interés y . Si las variables x e y tienen una relación lineal fuerte, es muy probable que los atributos A y C puedan tener una mayor relación a medida que los valores de α y β se aproximen, respectivamente, a α^* y β^* . Por esta razón, se propone que la búsqueda intensiva de los valores óptimos se centre en valores cercanos a α^* y β^* , es decir, se propone que el conjunto de los puntos para hallar los valores óptimos α y β se obtengan de entre los intervalos $(\alpha^* - \varepsilon, \alpha^* + \varepsilon)$ y $(\beta^* - \varepsilon, \beta^* + \varepsilon)$, y para un valor dado de ε . En comparación con el atributo B , este método puede requerir un menor número de puntos para la búsqueda intensiva. Este método para la obtención del atributo auxiliar también será analizado empíricamente mediante estudios de simulación Monte Carlo en la Sección 3.4.

La mayoría de los métodos de estimación basados en variables auxiliares están basados en la media poblacional de la variable auxiliar, y esto se debe a que en la práctica es bastante común el uso de este parámetro poblacional en la etapa de estimación. La transformación D , definida en la expresión (3.5), está justificada en este escenario, es decir, en los casos en los que la media poblacional sea conocida. El estimador de tipo razón basado en el atributo D está dado por:

$$\hat{p}_{0,r\gamma} = \frac{\hat{P}_0}{\hat{P}_{0,D}} P_{0,D}, \quad (3.10)$$

donde la constante γ minimiza la varianza muestral de $\hat{p}_{0,r\gamma}$. En esta situación, el conjunto de puntos para la búsqueda intensiva vienen dados por los valores de γ que se encuentran dentro del intervalo $(\gamma_{\min}, \gamma_{\max})$. Por ejemplo, podemos utilizar los valores $\gamma_{\min} = x_{(1)}/\bar{X}$ y $\gamma_{\max} = x_{(N)}/\bar{X}$, los cuales proporcionan el intervalo más amplio posible. Sería posible el uso de un intervalo más estrecho de modo que se

sigan obteniendo estimadores eficientes. Por ejemplo, se podría utilizar $\gamma_{\max} = 1$ en el caso que la variable auxiliar x tenga una distribución asimétrica positiva y $\beta^* = 0.5$. Destacamos que la hipótesis de una distribución asimétrica positiva es bastante factible que se satisfaga en el caso de variables económicas. Bajo este escenario, la elección $\gamma_{\max} = 1$ se debe al hecho de que $X_{0.5} < \bar{X}$ cuando la distribución es asimétrica positiva, y también debido al criterio de elección de puntos discutido previamente para la transformación C . Destacamos que en la Sección 3.4 también se analizará la elección $\gamma_{\max} = 1$ mediante estudios de simulación Monte Carlo.

3.3.3. Estimadores de tipo regresión

De forma similar podemos definir estimadores de tipo regresión para estimar la proporción de pobres P_0 . El estimador de tipo regresión para P_0 basado en la variable dicotómica B viene dado por

$$\hat{p}_{0.reg\kappa} = \hat{p}_0 + \hat{b}(P_{0.B} - \hat{p}_{0.B}), \quad (3.11)$$

donde

$$\hat{b} = \hat{\phi} \sqrt{\frac{\hat{p}_0 \hat{q}_0}{\hat{p}_{0.B} \hat{q}_{0.B}}}$$

y el valor óptimo de κ que minimiza la varianza

$$\hat{V}(\hat{p}_{0.reg\kappa}) = \frac{1-f}{n-1} \hat{p}_0 \hat{q}_0 (1-\hat{\phi}^2) \quad (3.12)$$

A partir de la ecuación (3.12) podemos observar que el parámetro κ sólo depende de $\hat{\phi}$, y de ahí que el problema de minimizar la varianza indicada en la ecuación (3.12) sea equivalente al problema de maximizar $\hat{\phi}$. Los estimadores $\hat{P}_{0.reg\alpha\beta}$ y $\hat{P}_{0.reg\gamma}$ basados, respectivamente, en los atributos C y D se pueden definir fácilmente a partir de la expresión (3.11). Para ello es suficiente con sustituir en (3.11) el atributo B por los atributos C y D .

3.3.4 Extensión a un diseño muestral general

Las secciones anteriores están basadas en una muestra s seleccionada mediante muestreo aleatorio simple sin reemplazamiento (MAS), el cual puede considerarse como el diseño muestral más conocido en la práctica. Los métodos de estimación propuestos para la proporción de pobres pueden extenderse fácilmente cuando la muestra s es seleccionada mediante un diseño muestral general. A continuación se describen las expresiones para los diversos estimadores de la proporción de pobres analizados en este capítulo cuando la muestra s es seleccionada mediante un diseño muestral general con pesos muestrales dados por $d_i = 1/\pi_i$, donde π_i es la probabilidad de inclusión de primer orden para el i -ésimo individuo poblacional. Las probabilidades de inclusión de segundo orden se están denotadas como π_{ij} (véase también, Särndal et al. 1992, p. 30)

El estimador de Horvitz-Thompson de la proporción de pobres P_0 viene dado por

Capítulo 3. Estimación de la proporción de pobres

$$\hat{p}_0 = \frac{1}{N} \sum_{i \in s} d_i A_i. \quad (3.13)$$

El estimador de tipo regresión logística de P_0 viene dado por

$$\hat{p}_{0.Lg\text{reg}} = \frac{1}{N} \left(\sum_{i \in U} \hat{\mu}_i + \sum_{i \in s} d_i (A_i - \hat{\mu}_i) \right). \quad (3.14)$$

Los estimadores de tipo razón y regresión propuestos pueden definirse fácilmente a partir de las definiciones descritas en las Secciones 3.3.2 y 3.3.3. El estimador de tipo razón $\hat{p}_{0.r\kappa}$ está basado en la variable dicotómica B y viene dado por la ecuación (3.6), donde el estimador $\hat{p}_{0.B}$ de $P_{0.B}$ viene dado por

$$\hat{p}_{0.B} = \frac{1}{N} \sum_{i \in s} d_i B_i.$$

El valor óptimo de κ también minimiza la varianza muestral de $\hat{p}_{0.r\kappa}$, la cual se define en este caso como (véase también Rueda et al. 2011b):

$$\hat{V}(\hat{p}_{0.r\kappa}) = \frac{1}{N^2} \sum_{i < j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{A_i - \hat{R}B_i}{\pi_i} - \frac{A_j - \hat{R}B_j}{\pi_j} \right)^2.$$

Los estimadores de tipo razón $\hat{p}_{0.r\alpha\beta}$ y $\hat{p}_{0.r\gamma}$ (basados, respectivamente, en los atributos C y D), se definen de forma similar en el caso de muestras seleccionadas mediante probabilidades desiguales. Es decir, los estimadores $\hat{p}_{0.r\alpha\beta}$ y $\hat{p}_{0.r\gamma}$ vienen dados por las ecuaciones (3.9) y (3.10), pero los estimadores básicos

Capítulo 3. Estimación de la proporción de pobres

$$\hat{p}_{0,C} = \frac{1}{N} \sum_{i \in s} d_i C_i$$

y

$$\hat{p}_{0,D} = \frac{1}{N} \sum_{i \in s} d_i D_i$$

son, respectivamente, los estimadores de tipo Horvitz-Thompson de p_0 basados en los atributos C y D .

De forma similar, el estimador de tipo regresión $\hat{p}_{0,reg\kappa}$ basado en la variable dicotómica B y usando diseños muestrales con probabilidades desiguales viene dada por la expresión (3.11), donde

$$\hat{b} = \frac{\sum_{i \in s} d_i (A_i - \hat{p}_0)(B_i - \hat{p}_{0,B})}{\sum_{i \in s} d_i (B_i - \hat{p}_{0,B})^2}.$$

El valor óptimo de κ es el valor que minimiza la varianza muestral

$$\hat{V}(\hat{p}_{0,reg\kappa}) = \frac{1}{N^2} \sum_{i < j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{A_i - \hat{b}B_i}{\pi_i} - \frac{A_j - \hat{b}B_j}{\pi_j} \right)^2.$$

La definición de los estimadores de tipo regresión y usando las variables dicotómicas C y D es bastante simple a partir de las expresiones ya descritas, y por esta razón, se han omitido en este trabajo tales definiciones.

3.4 Estudios de Simulación Monte Carlo

En esta sección se utiliza el método de simulación Monte Carlo para analizar empíricamente el comportamiento de los distintos estimadores de la proporción de pobres P_0 . En este estudio de simulación se utilizan varias poblaciones, con diferentes características, con el objetivo de analizar el efecto de los estimadores en diferentes situaciones. A continuación se describe, de forma muy breve, las poblaciones utilizadas en este estudio. Para un mayor detalle de estas poblaciones puede consultarse el Apéndice I.

En primer lugar se consideran poblaciones basadas en un conjunto de datos reales obtenidos de la EU-SILC en 2011, la cual contiene información de los ingresos, pobreza, inclusión social y condiciones de vida de las familias y los individuos en la Unión Europea. Se consideran datos muestrales de países con diferentes proporciones de pobres. Por ejemplo, los países considerados tienen una proporción de pobres que se encuentra a ambos lados del valor de la proporción de pobres media de toda la Unión Europea, fijada en el 16.9 % en 2011. Para ser más concretos, se han tomado datos muestrales de Bélgica, Eslovenia y Reino Unido, los cuales tienen tasas de pobreza por debajo de la media europea, y también se toman datos muestrales de países con tasas de pobreza por encima de la media europea, los cuales son Bulgaria, Italia, Lituania, Polonia y España. Se han considerado estas poblaciones porque puede resultar interesante analizar el comportamiento de los estimadores de la tasa de pobreza tomando datos de situaciones reales diferentes. Para nuestro estudio de simulación, las distintas muestras se han seleccionado mediante MAS y muestreo aleatorio estratificado, a partir de los datos de cada país, y los cuales se considerarán en este estudio como una población. Esta metodología puede consultarse, por ejemplo, en Chambers y Dunstan (1986), Rao et al. (1990), Harms y Duchesne (2006) y Muñoz et al. (2014). El objetivo que se persiguen en las

Capítulo 3. Estimación de la proporción de pobres

simulaciones basadas en datos reales es estimar la proporción de pobres utilizando la definición dada por EUROSTAT, es decir, consideraremos el 60 % de la mediana de ingresos por unidad de consumo para construir el umbral y pobreza. Además, consideraremos las contribuciones del impuesto sobre la renta como variable auxiliar.

Los estudios de simulación Monte Carlo también están basados en poblaciones simuladas y generadas de las distribuciones de probabilidad Gamma y Singh-Maddala, y las cuales son muy comunes en estudios de pobreza. Estas poblaciones también están descritas de forma más exhaustiva en el Apéndice I. Destacamos que se han utilizado las poblaciones basadas en la distribución Gamma para analizar el impacto de los estimadores cuando el valor del coeficiente de correlación lineal entre las variables x e y , ρ , oscila entre 0.5 y 0.9, y el valor de la proporción de pobres, P_0 , se encuentra entre 0.15 y 0.35. Las poblaciones basadas en la distribución Gamma consisten en $N=10,000$ valores aleatorios generados de una distribución Gamma con media igual a 10. Utilizamos la distribución $\gamma(10/2,2)$ para obtener el valor de $P_0 = 0.15$, la distribución $\gamma(10/4,4)$ para obtener el valor de $P_0 = 0.25$, y la distribución Gamma $\gamma(10/11,11)$ para obtener el valor de $P_0 = 0.35$. En este caso, el valor de P_0 también se obtiene a partir del umbral de pobreza definido por EUROSTAT, es decir, el 60 % de la media de la variable de interés y . Las variables auxiliares también han sido generadas de distribuciones Gamma, y los parámetros han sido seleccionados de forma que ρ se encuentre entre 0.5 y 0.9.

Por otra parte, la distribución Singh-Maddala ha sido utilizada para analizar el impacto de los estimadores cuando las distribuciones de las variables tienen un alto grado de asimetría. Estas poblaciones se obtienen utilizando $N=10,000$ valores aleatorios extraídos de una distribución Singh-Maddala. Los

Capítulo 3. Estimación de la proporción de pobres

parámetros de la distribución Singh-Maddala se han seleccionado de forma que el coeficiente de asimetría, g , toma los valores $g = \{4,8,12\}$, aproximadamente. Los métodos de estimación propuestos se comparan numéricamente con el estimador de Horvitz-Thompson de P_0 definido en la ecuación (3.13), y el estimador de tipo regresión logística definido en la ecuación (3.14). Los estimadores propuestos están basados en las transformaciones propuestas B , C y D descritas en la Sección 3.3.1. Para la transformación C se ha utilizado una búsqueda intensiva en los intervalos $(\alpha - \varepsilon, \alpha + \varepsilon)$ y $(\beta - \varepsilon, \beta + \varepsilon)$, con $\varepsilon = 0.1$ y paso 0.01 entre dos puntos consecutivos. Para la transformación D se ha propuesto una búsqueda dentro del intervalo $(\gamma_{\min}, \gamma_{\max})$, con $\gamma_{\min} = x_{(1)}/\bar{X}$, $\gamma_{\max} = 1$ y paso 0.01. Destacamos que se ha propuesto el uso de $\gamma_{\max} = 1$ puesto que las variables tienen distribuciones asimétricas positivas, y también porque esta elección requiere de un menor número puntos a evaluar en la búsqueda intensiva.

Se han utilizado varias medidas empíricas para evaluar y comparar el comportamiento de los estimadores discutidos en este capítulo para la proporción de pobres P_0 . En particular, se considera el Sesgo Relativo (SR) y el Error Cuadrático Medio empírico (ECMR), los cuales están definidos, para un determinado estimador \tilde{p}_0 de P_0 , como:

$$SR[\tilde{p}_0] = \frac{E[\tilde{p}_0] - P_0}{P_0}$$

y

$$ECMR[\tilde{p}_0] = \frac{\sqrt{ECM[\tilde{p}_0]}}{P_0},$$

Capítulo 3. Estimación de la proporción de pobres

donde $E[\cdot]$ y $ECM[\cdot]$ denotan, respectivamente, el valor esperado y el error cuadrático medio del estimador \tilde{p}_0 basados en $R = 1000$ simulaciones o réplicas del estudios de simulación, es decir,

$$E[\tilde{p}_0] = \frac{1}{R} \sum_{r=1}^R \tilde{p}_0(r)$$

$$MSE[\tilde{p}_0] = \frac{1}{R} \sum_{r=1}^R (\tilde{p}_0(r) - P_0)^2,$$

donde $\tilde{p}_0(r)$ es el valor del estimador \tilde{p}_0 obtenido en la r -ésima réplica. Puede observarse que estas medidas empíricas son las más utilizadas en este trabajo. Destacamos que tanto el SR como el $ECMR$ son posiblemente las medidas más utilizadas en la literatura para el problema de comparación numérica de la precisión de metodologías de estimación (véase Chen y Sitter 1999; Deville y Särndal 1992; Rao et al. 1990; Silva y Skinner 1995, etc). No obstante, se han utilizado otras medidas empíricas complementarias, tal como el sesgo absoluto empírico (SA) y la varianza empírica (VE) de los estimadores, las cuales vienen definidas como

$$SA[\tilde{p}_0] = \frac{1}{R} \sum_{r=1}^R \frac{|\tilde{p}_0(r) - P_0|}{P_0}$$

y

$$VE[\tilde{p}_0] = \frac{1}{R} \sum_{r=1}^R (\tilde{p}_0(r) - E[\tilde{p}_0])^2.$$

Capítulo 3. Estimación de la proporción de pobres

Finalmente, se evalúan los métodos de estimación propuestos utilizando la cobertura empírica (CE) de los intervalos de confianza definidos por la expresión

$$[L(r), U(r)],$$

donde los límites inferior y superior del intervalo anterior están definidos, respectivamente, como

$$L(r) = \tilde{p}_0(r) - 1.96\sqrt{\hat{V}(\tilde{p}_0(r))};$$

y

$$U(r) = \tilde{p}_0(r) + 1.96\sqrt{\hat{V}(\tilde{p}_0(r))}.$$

La cobertura empírica, al 95 % de nivel de confianza, se define como

$$CE[\tilde{p}_0] = \frac{1}{R} \sum_{r=1}^R \delta(L(r) \leq P_0 \leq U(r)).$$

Destacamos que la medida *SA* es utilizada por Rueda et al. (2007), mientras que las medidas *VE* y *CE* son utilizadas, por ejemplo, por Harms y Duchesne (2006).

Los resultados derivados de estos estudios de simulación pueden observarse en las siguientes Tablas (desde la Tabla 3.1 hasta la Tabla 3.16).

Capítulo 3. Estimación de la proporción de pobres

<i>SILC</i>	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
Italia	0.57	1	12.1	10.2	10.1	10.0	10.1	10.5	10.5	10.5
($\rho = 0.88$)	0.55	5	5.5	4.6	4.5	4.6	4.5	4.7	4.7	4.8
	0.54	10	3.7	3.2	3.1	3.2	3.2	3.3	3.2	3.3
	0.54	20	2.4	2.1	2.0	2.0	2.1	2.1	2.1	2.1
Eslovenia	0.5	1	20.0	18.8	18.4	18.3	18.4	18.5	18.5	18.4
($\rho = 0.83$)	0.49	5	9.1	7.8	7.8	7.8	7.8	7.8	7.8	7.8
	0.48	10	6.0	5.3	5.2	5.2	5.2	5.2	5.2	5.2
	0.48	20	4.0	3.5	3.5	3.5	3.5	3.6	3.6	3.6
Reino Unido	0.52	1	21.1	18.4	18.5	18.4	18.4	18.8	18.6	18.8
($\rho = 0.83$)	0.49	5	9.4	8.2	8.2	8.2	8.2	8.3	8.3	8.3
	0.49	10	6.4	5.6	5.6	5.6	5.6	5.6	5.9	5.6
	0.48	20	4.3	3.7	3.7	3.7	3.7	3.7	3.7	3.7
Polonia	0.47	1	14.6	13.0	13.1	13.0	13.1	13.4	13.2	13.4
($\rho = 0.82$)	0.43	5	6.2	5.6	5.7	5.6	5.7	5.8	5.8	5.8
	0.42	10	4.4	4.0	4.1	4.0	4.1	4.1	4.1	4.1
	0.42	20	2.8	2.6	2.6	2.6	2.6	2.6	2.6	2.6

Tabla 3.1. Valores de *ECMR* (%) para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

SILC	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{p}_{0.Lgreg}$	$\hat{p}_{0.reg\kappa}$	$\hat{p}_{0.reg\alpha\beta}$	$\hat{p}_{0.reg\gamma}$	$\hat{p}_{0.r\kappa}$	$\hat{p}_{0.r\alpha\beta}$	$\hat{p}_{0.r\gamma}$
Bulgaria	0.51	1	16.8	14.8	14.9	14.9	14.9	15.1	15.0	15.1
($\rho = 0.70$)	0.48	5	7.0	6.3	6.3	6.3	6.3	6.4	6.4	6.4
	0.47	10	5.2	4.6	4.6	4.6	4.6	4.6	4.6	4.6
	0.47	20	3.4	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Lituania	0.45	1	22.0	21.4	20.9	20.8	20.9	21.1	20.9	21.1
($\rho = 0.66$)	0.42	5	10.2	9.5	9.3	9.3	9.3	9.3	9.3	9.3
	0.41	10	6.7	6.4	6.1	6.1	6.1	6.1	6.1	6.1
	0.41	20	4.7	4.5	4.3	4.3	4.3	4.4	4.3	4.4
España	0.39	1	11.8	10.8	10.9	10.9	11.0	11.0	11.1	11.1
($\rho = 0.65$)	0.37	5	5.0	4.6	4.7	4.7	4.7	4.7	4.8	4.7
	0.36	10	3.6	3.4	3.3	3.4	3.3	3.4	3.4	3.4
	0.36	20	2.4	2.2	2.2	2.2	2.2	2.2	2.2	2.2
Bélgica	0.49	1	24.5	21.9	22.7	22.4	22.7	22.5	22.3	22.5
($\rho = 0.65$)	0.45	5	11.0	9.8	9.9	9.9	9.9	10.2	10.1	10.3
	0.43	10	7.4	6.9	6.8	6.9	6.8	7.1	7.1	7.1
	0.43	20	5.2	4.6	4.7	4.6	4.6	4.7	4.7	4.7

Tabla 3.2. Valores de ECMR (%) para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, España y Bélgica. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
Gamma	0.75	1	22.9	20.0	18.6	18.5	18.7	18.9	18.8	19.1
($\rho = 0.9$)	0.71	5	10.3	8.9	7.8	7.8	7.9	7.8	7.7	8.0
	0.70	10	7.2	6.2	5.3	5.3	5.4	5.3	5.3	5.4
	0.69	20	4.6	3.9	3.4	3.4	3.4	3.4	3.4	3.4
Gamma	0.54	1	23.5	20.3	21.3	21.1	21.2	21.2	21.0	21.3
($\rho = 0.7$)	0.49	5	10.7	9.3	9.3	9.3	9.3	9.3	9.3	9.4
	0.47	10	7.2	6.4	6.4	6.4	6.4	6.5	6.5	6.5
	0.47	20	4.8	4.3	4.4	6.4	4.4	4.4	4.4	4.4
Gamma	0.39	1	24.0	22.5	23.1	22.8	23.2	23.2	23.2	23.1
($\rho = 0.5$)	0.34	5	10.0	9.5	9.7	9.6	9.7	9.9	9.8	9.8
	0.32	10	7.1	6.5	6.7	6.7	6.7	6.7	6.7	6.7
	0.32	20	4.8	4.6	4.6	4.6	4.6	4.6	4.6	4.6

Tabla 3.3. Valores de ECMR (%) para los estimadores de $P_0 = 0.15$ y asumiendo las poblaciones basadas en la distribución Gamma. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , y $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
SM-g12	0.35	1	33.5	35.3	33.4	33.0	33.4	33.8	33.7	33.8
($\rho=0.9$)	0.26	5	14.2	15.3	14.1	14.0	14.1	14.0	14.1	14.0
	0.24	10	10.2	11.2	10.2	10.1	10.2	10.1	10.3	10.2
	0.23	20	6.4	7.0	6.2	6.2	6.2	6.2	6.4	6.3
SM-g8	0.34	1	34.5	37.1	34.4	33.9	34.3	34.5	34.2	34.4
($\rho=0.9$)	0.24	5	14.1	15.7	14.1	14.1	14.2	14.0	14.1	14.0
	0.23	10	10.4	11.4	10.2	10.1	10.2	10.2	10.3	10.3
	0.21	20	6.7	7.2	6.5	6.5	6.6	6.5	6.5	6.5
SM-g4	0.44	1	42.1	38.9	39.1	39.0	39.3	41.6	41.6	41.7
($\rho=0.9$)	0.40	5	18.1	16.6	17.3	17.0	17.3	17.2	17.2	17.2
	0.38	10	12.7	11.7	12.2	12.0	12.2	12.2	12.2	12.3
	0.36	20	8.1	7.3	7.6	7.6	7.6	7.7	7.6	7.6

Tabla 3.4. Valores de $ECMR(\%)$ para los estimadores de P_0 y asumiendo las poblaciones basadas en la distribución Singh-Maddala (SM). $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B , y g es el coeficiente de asimetría

Capítulo 3. Estimación de la proporción de pobres

<i>SILC</i>	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
Italia	0.57	1	1.80	1.53	1.52	1.52	1.53	1.60	1.60	1.60
($\rho = 0.88$)	0.55	5	0.83	0.71	0.70	0.70	0.70	0.72	0.72	0.72
	0.54	10	0.56	0.48	0.47	0.47	0.48	0.48	0.48	0.48
	0.54	20	0.37	0.31	0.30	0.31	0.31	0.31	0.31	0.31
Eslovenia	0.50	1	1.96	1.88	1.83	1.81	1.84	1.84	1.81	1.83
($\rho = 0.83$)	0.49	5	0.88	0.76	0.76	0.76	0.76	0.76	0.76	0.76
	0.48	10	0.58	0.51	0.50	0.50	0.50	0.50	0.50	0.50
	0.48	20	0.40	0.34	0.34	0.34	0.34	0.34	0.34	0.34
Reino Unido	0.52	1	2.91	2.51	2.55	2.51	2.55	2.60	2.54	2.61
($\rho = 0.83$)	0.49	5	1.30	1.14	1.14	1.15	1.14	1.16	1.17	1.16
	0.49	10	0.88	0.77	0.77	0.76	0.77	0.77	0.77	0.77
	0.48	20	0.59	0.51	0.51	0.51	0.51	0.52	0.51	0.51
Polonia	0.47	1	2.06	1.80	1.84	1.84	1.86	1.87	1.86	1.88
($\rho = 0.82$)	0.43	5	0.87	0.79	0.80	0.79	0.80	0.81	0.81	0.81
	0.42	10	0.62	0.56	0.57	0.56	0.57	0.57	0.57	0.57
	0.42	20	0.40	0.36	0.36	0.36	0.36	0.37	0.37	0.37

Tabla 3.5. Valores de $SA(\%)$ para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

<i>SILC</i>	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
Bulgaria	0.51	1	2.94	2.57	2.58	2.59	2.58	2.61	2.59	2.61
($\rho = 0.70$)	0.48	5	1.22	1.10	1.11	1.11	1.11	1.12	1.12	1.13
	0.47	10	0.92	0.80	0.81	0.80	0.81	0.81	0.81	0.81
	0.47	20	0.61	0.53	0.53	0.53	0.53	0.53	0.53	0.53
Lituania	0.45	1	3.36	3.22	3.16	3.15	3.16	3.18	3.14	3.17
($\rho = 0.66$)	0.42	5	1.52	1.41	1.38	1.39	1.38	1.39	1.39	1.38
	0.41	10	1.02	0.95	0.92	0.91	0.92	0.92	0.91	0.92
	0.41	20	0.70	0.66	0.63	0.64	0.64	0.64	0.64	0.65
España	0.39	1	1.92	1.74	1.78	1.76	1.78	1.78	1.79	1.79
($\rho = 0.65$)	0.37	5	0.81	0.75	0.77	0.77	0.77	0.76	0.78	0.76
	0.36	10	0.58	0.54	0.54	0.54	0.54	0.55	0.55	0.55
	0.36	20	0.38	0.35	0.36	0.36	0.36	0.37	0.37	0.37
Bélgica	0.49	1	2.83	2.56	2.66	2.63	2.66	2.65	2.61	2.66
($\rho = 0.65$)	0.45	5	1.27	1.15	1.14	1.15	1.14	1.20	1.18	1.21
	0.43	10	0.86	0.80	0.79	0.80	0.79	0.82	0.82	0.83
	0.43	20	0.61	0.54	0.54	0.54	0.54	0.55	0.55	0.55

Tabla 3.6. Valores de SA (%) para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, España y Bélgica. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{p}_{0.Lgreg}$	$\hat{p}_{0.reg\kappa}$	$\hat{p}_{0.reg\alpha\beta}$	$\hat{p}_{0.reg\gamma}$	$\hat{p}_{0.r\kappa}$	$\hat{p}_{0.r\alpha\beta}$	$\hat{p}_{0.r\gamma}$
Gamma	0.75	1	2.72	2.39	2.21	2.21	2.24	2.25	2.25	2.29
($\rho=0.9$)	0.71	5	1.24	1.06	0.95	0.94	0.94	0.94	0.94	96.00
	0.70	10	0.87	0.74	0.63	0.64	0.63	0.63	0.63	0.64
	0.69	20	0.55	0.47	0.41	0.40	0.41	0.41	0.41	0.41
Gamma	0.54	1	2.77	2.42	2.56	2.55	2.56	2.56	2.52	2.58
($\rho=0.7$)	0.49	5	1.28	1.12	1.13	1.13	1.14	1.14	1.14	1.14
	0.47	10	0.87	0.78	0.78	0.78	0.78	0.78	0.79	0.79
	0.47	20	0.58	0.52	0.53	0.53	0.53	0.54	0.54	0.54
Gamma	0.39	1	2.89	2.72	2.80	2.76	2.80	2.82	2.82	2.82
($\rho=0.5$)	0.34	5	1.20	1.14	1.16	1.16	1.17	1.17	1.23	1.17
	0.32	10	0.85	0.77	0.80	0.80	0.80	0.80	0.80	0.79
	0.32	20	0.58	0.55	0.55	0.55	0.55	0.55	0.55	0.55

Tabla 3.7. Valores de SA (%) para los estimadores de $P_0 = 0.15$ y asumiendo las poblaciones basadas en la distribución Gamma. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , y $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B.

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
SM-g12	0.35	1	2.27	2.37	2.26	2.24	2.27	2.28	2.24	2.27
($\rho = 0.9$)	0.26	5	0.96	1.03	0.95	0.94	0.95	0.94	0.96	0.95
	0.24	10	0.69	0.75	0.68	0.67	0.68	0.67	0.68	0.68
	0.23	20	0.43	0.47	0.42	0.42	0.42	0.42	0.43	0.42
SM-g8	0.34	1	2.22	2.42	2.22	2.19	2.21	2.23	2.20	2.22
($\rho = 0.9$)	0.24	5	0.89	1.00	0.89	0.89	0.90	0.89	0.90	0.89
	0.23	10	0.67	0.73	0.66	0.65	0.66	0.66	0.66	0.66
	0.21	20	0.43	0.46	0.42	0.42	0.42	0.42	0.42	0.42
SM-g4	0.44	1	1.95	1.79	1.82	1.83	1.84	1.86	1.85	1.86
($\rho = 0.9$)	0.40	5	0.84	0.77	0.79	0.78	0.80	0.79	0.79	0.79
	0.38	10	0.58	0.54	0.56	0.55	0.56	0.56	0.56	0.56
	0.36	20	0.37	0.33	0.34	0.34	0.35	0.35	0.35	0.35

Tabla 3.8. Valores de SA(%) para los estimadores de P_0 y asumiendo las poblaciones basadas en la distribución Singh-Maddala (SM). $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B , y g es el coeficiente de asimetría

Capítulo 3. Estimación de la proporción de pobres

<i>SILC</i>	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{p}_{0.Lgreg}$	$\hat{p}_{0.reg\kappa}$	$\hat{p}_{0.reg\alpha\beta}$	$\hat{p}_{0.reg\gamma}$	$\hat{p}_{0.r\kappa}$	$\hat{p}_{0.r\alpha\beta}$	$\hat{p}_{0.r\gamma}$
Italia	0.57	1	5.2	3.6	3.6	3.6	3.6	4.0	4.0	4.0
($\rho = 0.88$)	0.55	5	1.1	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	0.54	10	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4
	0.54	20	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Eslovenia	0.5	1	6.0	5.3	5.0	5.0	5.0	5.0	5.0	5.0
($\rho = 0.83$)	0.49	5	1.2	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	0.48	10	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4
	0.48	20	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Reino Unido	0.52	1	13.2	10.0	10.2	10.0	10.2	10.6	10.3	10.7
($\rho = 0.83$)	0.49	5	2.6	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	0.49	10	1.2	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	0.48	20	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Polonia	0.47	1	6.7	5.2	5.3	5.2	5.3	5.6	5.4	5.6
($\rho = 0.82$)	0.43	5	1.2	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.42	10	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.42	20	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2

Tabla 3.9. Valores de VE ($\times 10,000$) para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

SILC	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{p}_{0.Lgreg}$	$\hat{p}_{0.reg\kappa}$	$\hat{p}_{0.reg\alpha\beta}$	$\hat{p}_{0.reg\gamma}$	$\hat{p}_{0.r\kappa}$	$\hat{p}_{0.r\alpha\beta}$	$\hat{p}_{0.r\gamma}$
Bulgaria	0.51	1	13.7	10.6	10.9	10.9	10.8	11.1	10.9	11.1
($\rho = 0.70$)	0.48	5	2.4	1.8	2.0	2.0	2.0	2.0	2.0	2.0
	0.47	10	1.3	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.47	20	0.6	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Lituania	0.45	1	16.9	15.7	15.4	15.2	15.4	15.5	15.3	15.5
($\rho = 0.66$)	0.42	5	3.6	3.2	3.0	3.0	3.0	3.0	3.0	3.0
	0.41	10	1.6	1.4	1.3	1.3	1.3	1.3	1.3	1.3
	0.41	20	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7
España	0.39	1	5.7	4.8	4.9	4.9	4.9	5.0	5.0	5.0
($\rho = 0.65$)	0.37	5	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	0.36	10	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.36	20	0.2	0.2	2.0	0.2	0.2	0.2	0.2	0.2
Bélgica	0.49	1	12.8	10.2	11.1	10.8	11.1	10.8	10.6	10.8
($\rho = 0.65$)	0.45	5	2.6	2.1	2.1	2.1	2.1	2.2	2.1	2.3
	0.43	10	1.2	1.0	1.0	1.0	1.0	1.1	1.1	1.1
	0.43	20	0.6	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Tabla 3.10. Valores de VE ($\times 10,000$) para los estimadores de p_0 y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, España y Bélgica. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B.

Capítulo 3. Estimación de la proporción de pobres

<i>Población</i>	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{p}_{0.Lgreg}$	$\hat{p}_{0.reg\kappa}$	$\hat{p}_{0.reg\alpha\beta}$	$\hat{p}_{0.reg\gamma}$	$\hat{p}_{0.r\kappa}$	$\hat{p}_{0.r\alpha\beta}$	$\hat{p}_{0.r\gamma}$
Gamma	0.75	1	11.8	9.0	7.7	7.7	7.9	8.0	7.9	8.2
($\rho=0.9$)	0.71	5	2.4	1.8	1.4	1.4	1.4	1.4	1.4	1.4
	0.70	10	1.2	0.9	0.6	0.6	0.7	0.6	0.6	0.7
	0.69	20	0.5	0.4	0.3	0.3	0.3	0.3	0.3	0.3
Gamma	0.54	1	12.5	9.4	10.2	10.1	10.2	10.1	9.9	10.2
($\rho=0.7$)	0.49	5	2.6	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	0.47	10	1.2	0.9	0.9	0.9	0.9	0.9	0.9	0.9
	0.47	20	0.5	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Gamma	0.39	1	13.0	11.4	12.0	11.6	12.1	12.1	12.1	11.9
($\rho=0.5$)	0.34	5	2.3	2.1	2.1	2.1	2.1	2.2	2.2	2.2
	0.32	10	1.1	0.9	1.0	1.0	1.0	1.0	1.0	1.0
	0.32	20	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Tabla 3.11. Valores de $VE(\times 10,000)$ para los estimadores de $P_0=0.15$ y asumiendo las poblaciones basadas en la distribución Gamma. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , y $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B.

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$	$\hat{P}_{0.r\gamma}$
SM-g12	0.35	1	7.9	8.8	8.0	7.7	8.0	8.1	8.0	8.1
($\rho=0.9$)	0.26	5	1.4	1.7	1.4	1.4	1.4	1.4	1.4	1.4
	0.24	10	0.7	0.9	0.7	0.7	0.7	0.7	0.7	0.7
	0.23	20	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
SM-g8	0.34	1	7.5	8.8	7.6	7.4	7.5	7.6	7.5	7.6
($\rho=0.9$)	0.24	5	1.3	1.6	1.0	1.0	1.3	1.3	1.3	1.3
	0.23	10	0.7	0.8	0.7	0.7	0.7	0.7	0.7	0.7
	0.21	20	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
SM-g4	0.44	1	5.9	5.1	5.4	5.3	5.3	5.7	5.7	5.8
($\rho=0.9$)	0.40	5	1.1	0.9	1.0	1.0	1.0	1.0	1.0	1.0
	0.38	10	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
	0.36	20	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

Tabla 3.12. Valores de VE ($\times 10,000$) para los estimadores de P_0 y asumiendo las poblaciones basadas en la distribución Singh-Maddala (SM). $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B , y g es el coeficiente de asimetría.

Capítulo 3. Estimación de la proporción de pobres

SILC	$\hat{\rho}^*$	$f(\%)$	\hat{P}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$
Italia	0.57	1	95.0	93.9	94.6	93.8	94.3	94.3	94.0
($\rho = 0.88$)	0.55	5	94.5	94.0	94.5	94.7	94.9	94.9	94.8
	0.54	10	95.1	94.8	94.7	94.2	93.8	94.1	94.1
	0.54	20	95.6	95.7	96.0	95.5	95.5	95.4	95.3
Eslovenia	0.5	1	93.6	92.8	92.9	92.6	92.2	92.3	92.3
($\rho = 0.83$)	0.49	5	92.9	93.7	93.8	93.7	93.7	93.7	93.4
	0.48	10	94.5	96.0	95.7	95.8	95.7	95.6	95.6
	0.48	20	95.2	94.2	94.4	94.3	94.4	94.4	94.7
Reino Unido	0.52	1	95.5	93.0	93.4	93.6	93.2	93.2	92.9
($\rho = 0.83$)	0.49	5	94.7	94.7	94.9	94.7	94.8	95.0	95.4
	0.49	10	94.9	94.7	94.2	94.8	94.9	94.8	94.8
	0.48	20	94.7	95.6	95.3	95.5	95.4	95.2	95.7
Polonia	0.47	1	93.6	93.5	93.6	92.7	93.7	94.2	93.9
($\rho = 0.82$)	0.43	5	94.9	94.5	94.3	94.5	93.7	94.4	94.1
	0.42	10	95.2	94.1	94.7	94.0	95.0	4.5	95.2
	0.42	20	95.3	94.6	94.6	95.1	95.1	95.3	94.9

Tabla 3.13. Valores de CE (%) para los estimadores de P_0 y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B.

Capítulo 3. Estimación de la proporción de pobres

SILC	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$
Bulgaria	0.51	1	94.9	93.4	93.3	93.5	93.9	93.5	93.0
($\rho = 0.70$)	0.48	5	96.1	95.0	94.8	94.8	95.2	95.1	94.4
	0.47	10	93.3	93.4	93.5	93.8	93.7	93.7	93.9
	0.47	20	94.9	94.5	94.5	94.4	94.4	94.9	94.5
Lituania	0.45	1	96.2	93.1	93.5	92.9	92.9	93.5	92.8
($\rho = 0.66$)	0.42	5	94.1	94.1	94.0	94.2	94.9	93.8	94.2
	0.41	10	95.6	96.1	96.4	95.8	95.8	96.0	95.8
	0.41	20	93.7	94.0	94.1	93.9	94.0	4.3	94.1
España	0.39	1	94.3	94.2	94.6	94.0	94.0	94.9	94.1
($\rho = 0.65$)	0.37	5	95.8	94.8	94.8	95.1	95.0	95.0	95.1
	0.36	10	94.9	94.7	94.6	94.7	93.8	93.6	94.3
	0.36	20	94.6	94.9	95.0	95.0	95.9	95.2	95.7
Bélgica	0.49	1	92.4	92.8	93.8	92.8	92.8	94.4	93.0
($\rho = 0.65$)	0.45	5	94.7	94.7	95.0	94.7	94.5	94.9	94.6
	0.43	10	94.2	93.8	94.3	94.6	93.9	93.9	94.0
	0.43	20	94.7	94.1	94.7	94.2	93.8	93.9	94.0

Tabla 3.14. Valores de CE (%) para los estimadores de p_0 y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, España y Bélgica. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B .

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$
Gamma	0.75	1	93.3	86.7	86.9	86.9	86.9	86.8	86.7
($\rho = 0.9$)	0.71	5	91.9	92.1	93.8	92.0	92.3	92.4	92.2
	0.70	10	94.7	93.4	93.7	93.9	93.6	94.0	93.8
	0.69	20	95.8	95.6	95.2	95.1	95.3	95.1	95.8
Gamma	0.54	1	93.1	92.3	92.5	92.1	94.2	93.5	93.4
($\rho = 0.7$)	0.49	5	92.9	93.4	93.3	93.2	93.9	94.1	93.8
	0.47	10	95.2	95.0	94.8	94.6	95.3	95.1	94.8
	0.47	20	95.1	94.6	94.9	94.8	94.6	94.3	94.6
Gamma	0.39	1	93.7	91.7	92.7	92.3	92.1	92.6	92.2
($\rho = 0.5$)	0.34	5	94.5	95.1	94.7	94.9	94.5	94.3	94.7
	0.32	10	94.8	94.9	95.4	95.2	95.5	94.9	95.2
	0.32	20	94.5	93.7	93.9	93.4	94.0	94.1	94.1

Tabla 3.15. Valores de CE (%) para los estimadores de $P_0 = 0.15$ y asumiendo las poblaciones basadas en la distribución Gamma. $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , y $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B.

Capítulo 3. Estimación de la proporción de pobres

Población	$\hat{\rho}^*$	$f(\%)$	\hat{p}_0	$\hat{P}_{0.Lgreg}$	$\hat{P}_{0.reg\kappa}$	$\hat{P}_{0.reg\alpha\beta}$	$\hat{P}_{0.reg\gamma}$	$\hat{P}_{0.r\kappa}$	$\hat{P}_{0.r\alpha\beta}$
SM-g12	0.35	1	91.1	89.2	91.3	89.5	90.2	90.8	90.1
($\rho=0.9$)	0.26	5	94.8	94.2	94.4	94.2	94.6	95.5	94.8
	0.24	10	94.0	93.4	94.0	93.3	94.1	93.4	94.1
	0.23	20	95.4	95.9	96.0	95.9	95.0	95.5	95.0
SM-g8	0.34	1	89.9	90.0	91.4	89.6	90.1	91.4	90.3
($\rho=0.9$)	0.24	5	96.0	95.2	95.6	94.9	95.2	95.0	95.0
	0.23	10	95.0	94.9	95.1	94.8	94.8	95.0	94.9
	0.21	20	95.1	95.2	95.3	95.2	95.5	95.5	95.1
SM-g4	0.44	1	92.2	91.2	91.0	91.8	91.4	91.2	91.6
($\rho=0.9$)	0.40	5	93.9	93.0	94.1	93.1	93.9	93.4	93.5
	0.38	10	94.1	92.3	92.7	92.9	92.7	93.0	93.0
	0.36	20	95.0	94.7	95.0	94.4	94.8	94.9	95.4

Tabla 3.16. Valores de CE (%) para los estimadores de P_0 y asumiendo las poblaciones basadas en la distribución Singh-Maddala (SM). $f = n/N$ es la fracción muestral, ρ es el coeficiente de correlación lineal entre y y x , $\hat{\rho}^*$ es el coeficiente de correlación lineal empírico entre el atributo A y el atributo propuesto B , y g es el coeficiente de asimetría.

Capítulo 3. Estimación de la proporción de pobres

En primer lugar, destacamos que los sesgos relativos empíricos (*SR*) de los estimadores analizados en este capítulo tienen unos valores insignificantes, puesto que éstos son todos menores, en términos absolutos, del 1%. Por tanto, se puede decir que los estimadores tienen un buen comportamiento en términos de sesgo relativo, y por este motivo están omitidos en este capítulo tales valores de *SR*.

A continuación analizaremos la eficiencia de los estimadores de la proporción de pobres, cuyos resultados pueden observarse en las Tablas 3.1, 3.2, 3.3 y 3.4. En primer lugar puede observarse que los valores de *ECMR* de los estimadores de regresión propuestos son más pequeños que los valores de *ECMR* de los estimadores de tipo razón, lo cual implica que el método de regresión para la estimación de la proporción de pobres es más eficiente que el método de tipo razón, y las diferencias empíricas entre estos métodos de estimación y para las poblaciones en estudio pueden observarse en las tablas presentadas. También podemos destacar que las transformaciones propuestas generalmente dan resultados similares en términos de eficiencia. Los estimadores propuestos son más eficientes que el estimador de Horvitz-Thompson de \hat{p}_0 para las poblaciones estudiadas. Como se esperaba, los estimadores propuestos son más precisos a medida que la fracción muestral f aumenta. Los valores de ρ y $\hat{\rho}^*$ tienen un impacto importante en la eficiencia de los estimadores propuestos. A partir de la Tabla 3.3, por ejemplo, puede observarse que los estimadores propuestos son más eficientes que el estimador de tipo regresión logística cuando ambos valores de ρ y $\hat{\rho}^*$ son elevados (véase, por ejemplo, el caso $\rho = 0.9$ en la Tabla 3.3). Cuando ρ no es muy elevado, los estimadores propuestos también pueden ser más eficientes que el estimador de tipo regresión logístico (véase, por ejemplo, los resultados derivados de Italia, Eslovenia y Lituania en las Tablas 3.1 y 3.2). En los casos restantes, los estimadores propuestos y el estimador de tipo regresión logística tienen un comportamiento similar. Finalmente, analizamos el

Capítulo 3. Estimación de la proporción de pobres

comportamiento de los estimadores cuando la distribución de las variables tienen un alto grado de asimetría (véase la Tabla 3.4). Se observa que los estimadores propuestos son ligeramente más precisos que el estimador estándar cuando $g = \{8,12\}$. Sin embargo, el estimador de regresión logística es menos eficiente que el estimador estándar en esta situación.

A continuación se analiza el comportamiento empírico de los distintos estimadores utilizando las medidas empíricas SA y VE . Tales resultados pueden verse en las Tablas 3.5, 3.6, ..., 3.12. Las conclusiones son similares a las obtenidas a partir de los valores de $ECMR$. Por ejemplo, los valores de SA muestran que las estimaciones obtenidas a partir de los métodos de regresión propuestos, en general, están más próximas a la proporción real de pobres que las estimaciones obtenidas a partir del estimador de regresión logística, especialmente cuando el valor ρ es elevado.

Finalmente, analizamos las coberturas empíricas de los intervalos de confianza basados en los estimadores propuestos y construidos con un nivel de confianza del 95% (Tablas 3.13, 3.14, 3.15 y 3.16). Debemos tener en cuenta que la situación esperada es que tales coberturas estén próximas al 95%, el nivel de confianza fijado. A partir de los resultados obtenidos en estas tablas, podemos comprobar que los estimadores de tipo razón y regresión tienen un buen comportamiento, con coberturas próximas al 95 % requerido. Cuando la fracción de muestreo f es pequeña (en torno al 1%), podemos observar que los estimadores de la varianza de los estimadores propuestos sufren de una pequeña subestimación, puesto que en este caso las coberturas están en torno al 90%, (véase, por ejemplo, la el caso $\rho = 0.9$ de la Tabla 3.15). Se puede comprobar que este problema se soluciona a medida que incrementa la fracción muestral f .

Capítulo 3. Estimación de la proporción de pobres

Destacamos, por ejemplo, que los resultados anteriores de distribución Gamma están basados en el caso $P_0 = 0.15$. También se han realizado estudios de simulación basados en otras situaciones. Por ejemplo, hemos considerado poblaciones basadas en la distribución Gamma y con valores de $P_0 = \{0.25, 0.35\}$. Además, los estudios de simulación se han repetido con muestras seleccionadas mediante muestreo aleatorio estratificado, en lugar del MAS utilizado en los resultados mostrados en este capítulo. Las conclusiones derivadas de todos estos estudios de simulación complementarios han sido similares, y por este motivo tales resultados no están incluidos.

En resumen, la proporción de pobres es un indicador de pobreza importante y muy utilizado en comparaciones de pobreza entre países. La proporción de pobres real es desconocida en la práctica, por lo que se utilizan datos muestrales de encuestas oficiales para obtener una estimación de este indicador de pobreza. Los métodos de estimación de la proporción de pobres existentes incluyen el estimador de tipo Horvitz-Thompson y el estimador de regresión logística. Desde una perspectiva política, Molina y Rao (2010) indican que el uso de métodos de estimación con propiedades deseables y relacionados con las condiciones de vida pueden ser una herramienta útil para organizaciones destinadas a abordar el problema de la pobreza.

En este capítulo se ha definido métodos alternativos de estimación de la proporción de pobres. Los estimadores propuestos consisten en la utilización de los estimadores clásicos de tipo razón y regresión después de transformar la variable auxiliar en una variable dicotómica. Las transformaciones propuestas están basadas en el criterio de mínima varianza. Destacamos que por simplicidad, sólo se ha considerado una variable auxiliar en la etapa de estimación. Sin embargo, de las transformaciones propuestas se pueden obtener J variables auxiliares dicotómicas, donde $J > 1$, y por tanto, se podrían obtener

Capítulo 3. Estimación de la proporción de pobres

mejores resultados en la fase de estimación. En otras palabras, los métodos de estimación propuestos pueden generalizarse fácilmente al caso de numerosas variables auxiliares, puesto que los métodos de tipo razón y regresión pueden adaptarse a esta situación (véase Särndal et al. 1992, p. 219).

Por otra parte, la proporción de pobres discutida en este capítulo es una medida unidimensional expresada como una proporción (véase, por ejemplo, la ecuación 3.1). La pobreza puede verse como un fenómeno multidimensional, y la utilización de índices más complejos como el Índice de Desarrollo Humano (IDH) puede ayudar a entender mejor el problema del fenómeno de la pobreza. El método de estimación propuesto puede ser fácilmente aplicado a cualquier medida que pueda expresarse como una proporción. La metodología de estimación propuesta también requiere variables auxiliares cuantitativas relacionadas con la variable de interés. En este sentido, el método de estimación desarrollado podría ser utilizado para estimar cada componente del IDH (expectativas de vida, educación e ingresos), siempre que éstos puedan ser expresados como proporciones, y asumiendo que los datos muestrales contengan variables adicionales relacionadas con las componentes. Estos temas que acabamos de mencionar podrían ser líneas de trabajo importantes para una investigación futura (véase también el Capítulo 5).

En este capítulo se han llevado a cabo los estudios de simulación Monte Carlo para evaluar el comportamiento de los distintos estimadores propuestos de la proporción de pobres en términos de varias medidas empíricas. También se han utilizado estos estudios de simulación Monte Carlo para comparar los estimadores propuestos con los estimadores existentes, tal como el estimador de regresión logística. En todos los casos, se ha analizado el efecto de los estimadores bajo diferentes situaciones. En particular, los estimadores de la proporción de pobres se han analizado para varios valores de P_0 (proporción de

Capítulo 3. Estimación de la proporción de pobres

pobres), f (fracción muestral), ρ (coeficiente de correlación lineal entre x e y), $\hat{\rho}^*$ (coeficiente de correlación lineal empírico entre los atributos A y B) y g (coeficiente de asimetría).

Por una parte, los estudios de simulación están basados en datos muestrales extraídos de las encuestas de 2011 de EU-SILC, utilizando datos de varios países. Por otra parte, se ha completado este estudio analizando escenarios alternativos obtenidos a partir de poblaciones simuladas basadas en distribuciones probabilísticas relacionadas con estudios de pobreza.

En primer lugar, a partir de los resultados obtenidos se ha observado que los estimadores tienen un buen comportamiento en términos de sesgo relativo. En segundo lugar, se ha observado que se obtienen resultados similares a partir de las transformaciones propuestas. Para el caso de poblaciones con coeficientes de correlación lineales elevados se han observado que los estimadores propuestos son los métodos de estimación más precisos. En otro caso, para poblaciones con un coeficiente de correlación lineal moderado, los estimadores propuestos y el estimador de regresión logística tienen un comportamiento similar. Los estimadores propuestos tienen un buen comportamiento cuando las distribuciones de las variables tienen un alto grado de asimetría, mientras que el estimador de regresión logística puede comportarse peor que el estimador estándar en esta situación. Por último, destacamos que también se ha analizado el comportamiento empírico de los intervalos de confianza basados en los estimadores propuestos y contruidos con un nivel de confianza del 95%. Tales intervalos de confianza tienen un comportamiento aceptable al tener éstos una cobertura empírica en torno al requerido 95%.

En conclusión, los resultados derivados de los diversos estudios de simulación indican que los estimadores de regresión propuestos pueden ser un

Capítulo 3. Estimación de la proporción de pobres

método de estimación alternativo para el problema de la estimación de la proporción de pobres. El comportamiento empírico de estos estimadores es similar al estimador de regresión logística cuando el coeficiente de correlación lineal no es demasiado elevado. Sin embargo, los estimadores propuestos son más precisos cuando el coeficiente de correlación es elevado. También se obtienen resultados deseables en la presencia de variables con un alto grado de asimetría.

Capítulo 4

Estimación de los índices de brecha de pobreza y pobreza severa

4.1. Introducción

Como se ha venido discutiendo a lo largo de la tesis, la medición de la pobreza a través de indicadores es uno de los aspectos más importantes para abarcar el conocimiento del bienestar social y tratar el problema de la reducción de la pobreza. La medición de la pobreza es un compromiso clave en la agenda política de numerosas organizaciones mundiales. Esta tarea es incluso más importante en la economía actual, en la que se observan incrementos considerables de la pobreza y la desigualdad tanto en países desarrollados como en países en vías de desarrollo. También se ha comentado que son numerosos los estudios de pobreza dirigidos a poder entender este problema social de una mejor manera, y así poder realizar políticas más precisas que permitan una reducción evidente de la pobreza. En temas anteriores hemos analizado en profundidad temas como la estimación de cuantiles, la línea o umbral de pobreza, la proporción de pobres, etc. Algunas referencias destacadas sobre la medición de la pobreza son Sen (1976), Clark et al. (1981), Foster et al. (1984), Atkinson (1987), Haughton y Khandker (2009), etc. En este capítulo se realizan aportaciones sobre los índices de brecha de pobreza y pobreza severa, dos indicadores de pobreza que suelen ayudar a entender mejor la profundidad e intensidad de la pobreza, y que son el complemento perfecto a otros indicadores

tal como la proporción de pobres. Destacamos que ambas medidas están relacionadas con el grado de pobreza que sufren los individuos clasificados como pobres. Al igual que la proporción de pobres abordada en el Capítulo 3, los índices de brecha de pobreza y pobreza severa pertenecen a la familia de medidas de pobreza propuestas por Foster, Greer y Thorbecke (1984), las cuales se han denotado con el acrónimo FGT.

El índice de brecha de pobreza es un indicador que mide la diferencia media entre los ingresos de los individuos clasificados como pobres y la línea de pobreza, expresando dicha diferencia como un porcentaje de la línea de pobreza. Destacamos que este indicador es muy útil y suele utilizarse para realizar comparaciones entre distintas poblacionales, regiones, etc. Además, ofrece una interpretación económica muy interesante de la pobreza, ya que expresa el coste o la cantidad de recursos necesarios para que los pobres dejen de serlo, es decir, consigan unos ingresos que lleguen a la línea de pobreza. En la práctica no puede afirmarse que sea suficiente con realizar esa transferencia de recursos para solucionar el problema, pero sí permite hacernos una idea de la magnitud del problema. En algunos estudios se compara esta cantidad requerida con el PIB de un determinado país. El principal inconveniente de la brecha de pobreza es que es insensible a la distribución interna de la población de pobres, es decir, este indicador no está afectado ante cambios internos en la distribución, y además no especifica la desigualdad dentro de la población de pobres.

Por el contrario, el índice de pobreza severa sí tiene en cuenta la desigualdad entre los pobres, de forma que los países que reportan altos valores de esta medida suelen tener niveles altos de pobreza y desigualdad. El índice de pobreza severa está altamente relacionado con el índice de brecha de pobreza. En particular, la principal diferencia entre ambos indicadores es el hecho de que el

índice de pobreza severa añade un mayor peso en aquellos pobres que se encuentran más alejados del umbral de pobreza.

La mayoría de las medidas de pobreza son desconocidas en la práctica, a nivel poblacional, y por esta razón son estimadas utilizando datos procedentes de encuestas familiares o de hogares, lo cual conllevaría los correspondientes errores de muestreo. Además, es muy común hacer comparaciones entre países o entre diferentes instantes de tiempo. Por todo lo anterior, resulta imprescindible conocer el margen de error que surge en la fase de estimación, siendo conveniente para ello obtener y analizar las varianzas de los estimadores utilizados y la construcción de intervalos de confianza para las medidas de pobreza en estudio.

El problema de encontrar una expresión analítica para la varianza de estimadores de los índices de brecha de pobreza y pobreza severa no es una cuestión simple, y esto se debe principalmente por dos razones. En primer lugar, se trata de parámetros complejos, con expresiones que no son lineales respecto a las observaciones. En segundo lugar, este problema surge por el hecho de que los datos relacionados con ingresos suelen obtenerse a través de encuestas complejas, en las que la obtención de expresiones para las varianzas también es un tema con difícil solución. Por ejemplo, Zheng (2001) propuso un método de inferencia asintótica para una clase de medidas de pobreza, y también se obtuvieron expresiones para las varianzas asintóticas. Sin embargo, estos resultados derivados de Zheng (2001) están limitados a muestreo aleatorio simple, muestreo estratificado y muestreo por conglomerados. Además, las expresiones propuestas precisan de estimaciones adicionales de funciones, tal como la función de densidad poblacional (véase Zheng 2001).

En el caso de parámetros o encuestas complejas, pueden utilizarse métodos de aproximación alternativos para estimar la varianza de estimadores de un determinado parámetro. En este capítulo se propone estimar la varianza de los estimadores de los índices de brecha de pobreza y pobreza severa propuestos utilizando el método o técnica bootstrap, el cual consiste en generar un elevado número de muestras a partir de la denominada población bootstrap, y la variabilidad observada en tales muestras servirán de base para estudiar el comportamiento de determinados estadísticos tal como la varianza, objetivo que se persigue en ese capítulo. Destacamos que la población bootstrap se obtiene a partir de la muestra original mediante también diversos métodos. El método bootstrap es muy común en las investigaciones empíricas y en las ciencias sociales. Por ejemplo, la agencia nacional de estadística de Canadá realizó un estudio de simulación (véase Kovacevik et al. 1995) para comparar varios métodos de remuestreo con el objetivo de estimar la varianza de medidas de desigualdad de ingresos. Este estudio estaba basado en un conjunto de datos reales extraídos de encuestas canadienses de consumidores (Canadian Survey of Consumer Finance), y los resultados obtenidos indicaban que los estimadores de la varianza basados en el método bootstrap son los que poseían un menor sesgo relativo.

Para profundizar en los métodos de Bootstrap puede consultarse a Efron (1979; 1982), Stine (1989), Hall (1992), Efron y Tibshirani (1993), Mooney y Duval (1993), Shao y Tu (1995), Davison y Hinkley (1997), Chernick (1999), Biewen (2002), Moran (2006), Davidson y Flachaire (2007) y Thompson (2013).

El objetivo de este capítulo es la estimación de los índices de brecha de pobreza y pobreza severa. En primer lugar, se propone utilizar un estimador de tipo razón para la estimación de tales medidas de pobreza. En segundo lugar, se propone utilizar un método bootstrap para la estimación de la varianza del

estimador de tipo razón propuesto, y la cual nos servirá para la construcción de intervalos de confianza para los índices de brecha de pobreza y pobreza severa. En último lugar, se llevan a cabo estudios de simulación para investigar las propiedades empíricas de los estimadores e intervalos de confianza propuestos en este capítulo. Al igual que en capítulos anteriores, estos estudios de simulación estarán basados en las poblaciones descritas en el Apéndice I de esta tesis, y las cuales se podían clasificar en dos tipos de poblaciones: (i) las basadas en datos reales extraídos de varios países de la Encuesta Europea sobre ingresos y condiciones de vida (EU-SILC) del 2011, y (ii) las basadas en datos artificiales y generadas a partir de la distribución Gamma.

A continuación, en la Sección 4.2, se definen los índices de brecha de pobreza y pobreza severa. Tales indicadores de pobreza son desconocidos a nivel poblacional, es decir, requieren información de la variable de interés para todos los individuos de la población en estudio, la cual se asume que es desconocida. Por esta razón, en esta sección también se define el estimador estándar basado en pesos muestrales de los indicadores de pobreza analizados en este capítulo. En la Sección 4.3 se presentan los estimadores de tipo razón propuestos para los índices de brecha de pobreza y pobreza severa. En la Sección 4.4 se aplica el método bootstrap rescalado para la estimación de la varianza de los estimadores propuesto, y a continuación se construyen los correspondientes intervalos de confianza. En la Sección 4.5 se evalúan numéricamente los estimadores de los índices de brecha de pobreza y pobreza severa utilizando estudios de simulación Monte Carlo, los cuales están basados en las poblaciones reales y artificiales ya comentadas. Además, se analizan las propiedades empíricas de las estimaciones de la varianza y de los intervalos de confianza propuestos, obteniendo resultados deseables en todos los casos.

4.2. Los índices de brecha de pobreza y pobreza severa

Sea $U = \{1, \dots, N\}$ la población de interés, la cual contiene N individuos. Los índices de brecha de pobreza y pobreza severa también son desconocidos en la práctica, es decir, son desconocidos para todos los individuos de la población U , por lo que en esta situación también asumiremos que serán estimados a partir de datos muestrales. Tal como se indicaba en capítulos anteriores, la mayoría de las medidas de pobreza se obtienen a partir de la información recogida de una variable cuantitativa de bienestar, tal como los ingresos o gastos de individuos o familiares. Ésta es la variable denominada como variable de interés y denotada en todo este trabajo como y .

El procedimiento para obtener los indicadores de pobreza de este capítulo es similar al explicado en capítulos anteriores, y el cual, no obstante, se vuelve a describir. En primer lugar, utilizamos la variable de interés y para obtener la línea de pobreza relativa L , la cual clasifica a la población en dos grupos: pobre y no pobre. Un individuo era considerado pobre si $y_i \leq L$, donde y_i denota el valor de la variable de interés y para el i -ésimo individuo en la población U . Recordamos que la línea de pobreza es un indicador económico estándar fijado, en general, en una cantidad oficial, es decir, se asume que la línea de pobreza viene establecida por la autoridad correspondiente. Por ejemplo, EUROSTAT establece que L es el 60% de la mediana de ingresos por unidad de consumo, y proporciona valores oficiales de L para varios países y regiones.

Los índices de brecha de pobreza y pobreza severa se pueden definir a partir de la familia de medidas de pobreza de FGT, la cual se ha definido en varias ocasiones a lo largo de esta tesis y puede verse como la suma ponderada de las brechas individuales, en las que la intensidad de las propias brechas

constituyen pesos o ponderaciones (véase Foster et al. 1984, Blackwood y Lynch 1994, Zheng 2001, Khandker 2005, Haughton y Khandker 2009, Molina y Rao 2010, Frick et al. 2012, etc). Asumiendo que la variable y es conocida para todos los individuos de la población, esta familia de medidas FGT viene definida por

$$P_{\alpha} = \frac{1}{N} \sum_{i=1}^N \left(\frac{L - y_i}{L} \right)^{\alpha} \delta(y_i \leq L), \quad (4.1)$$

donde $\delta(\cdot)$ denotaba a la variable indicadora, es decir, toma el valor 1 si su argumento es cierto y $\delta(\cdot)=0$ en caso contrario. En otras palabras $\delta(y_i \leq L)=1$ si el individuo i es considerado pobre, y $\delta(y_i \leq L)=0$ en caso contrario.

El parámetro α de la ecuación (4.1) es una medida que representa la sensibilidad del índice a la pobreza. Cuanto mayor sea α , el índice FGT será más sensible a las carencias de los grupos más pobres de la población. Cuando el parámetro es $\alpha=0$, la medida de la ecuación (4.1), P_0 , es la proporción de pobres ya estudiada en el Capítulo 3. El índice de brecha de pobreza, P_1 , viene definido cuando $\alpha=1$, mientras que el índice de pobreza severa, P_2 , se define cuando $\alpha=2$. El índice va perdiendo interpretación a medida que el valor α se va alejando de 3. Por tanto, el objetivo de este capítulo es estimar el parámetro P_{α} cuando $\alpha = \{1,2\}$, aunque la metodología propuesta podría aplicarse a cualquier valor de α .

En la práctica, el parámetro P_{α} es desconocido, ya que P_{α} está basado en los valores poblacionales de y , y por esta razón, el parámetro P_{α} será estimado usando la información extraída de una muestra s con n individuos y seleccionada de la población U . Asumiremos que la muestra s es seleccionada mediante un diseño muestral general con probabilidades de inclusión dadas por

π_i (véase Särndal et al. 1992 pag. 30). Los pesos muestrales vienen dados por $d_i = \pi_i^{-1}$.

El estimador de tipo Hájek (Hájek 1964) de P_α viene dado por la expresión

$$\hat{p}_\alpha = \frac{1}{\hat{N}} \sum_{i \in s} d_i \left(\frac{L - y_i}{L} \right)^\alpha \delta(y_i \leq L) \quad (4.2)$$

donde

$$\hat{N} = \sum_{i \in s} d_i .$$

Los estimadores básicos de los índices de brecha de pobreza y pobreza severa se obtienen mediante \hat{p}_α cuando $\alpha = \{1, 2\}$, respectivamente. Destacamos que el uso de los pesos muestrales d_i en la ecuación (4.2) garantiza que \hat{p}_α sea un estimador aproximadamente insesgado de P_α .

El estimador \hat{p}_α de P_α tiene una expresión más simple para el caso de una muestra seleccionada mediante muestreo aleatorio simple sin reemplazamiento (MAS). En este caso, el estimador de tipo Hájek de P_α viene dado por

$$\hat{p}_\alpha = \frac{1}{n} \sum_{i \in s} \left(\frac{L - y_i}{L} \right)^\alpha \delta(y_i \leq L).$$

4.3. Estimador de tipo razón

En capítulos anteriores se comentaba que la mayoría de encuestas sociales contienen habitualmente variables auxiliares relacionadas con la variable de interés. La variable de interés en la EU-SILC es el ingreso por unidad de consumo, y una posible variable relacionada con el ingreso por unidad de consumo puede ser los impuestos sobre la renta. Tanto en la Sección 4.5 como en el Apéndice I podemos observar que existe una relación lineal importante entre dichas variables y para varios países de la Unión Europea.

Las variables auxiliares relacionadas con la variable de interés pueden ser utilizadas en la fase de estimación para mejorar la estimación de un determinado parámetro, la cual es una técnica muy común para obtener resultados deseables, tal como se ha podido comprobar en el Capítulo 3. Recordamos que existen muchos métodos de estimación basados en variables auxiliares. Por ejemplo, para el problema de estimar una media poblacional, los métodos conocidos son el estimador de tipo razón (véase Särndal et al. 1992 pag. 219) y el estimador de tipo regresión (véase Särndal et al. 1992 pag. 219). Recientemente también se han propuesto para el problema de la estimación de la media poblacional los métodos de calibración (Deville y Särndal 1992) y verosimilitud pseudo empírica (Chen y Sitter 1999). Para el problema de la estimación de la función de distribución y cuantiles, los métodos más conocidos basados en variables auxiliares están propuestos por Chambers y Dunstan (1986), Rao et al. (1990), Silva y Skinner (1995), Dorfman (2009), etc. Finalmente, Molina y Rao (2010) y Muñoz et al. (2014) propusieron estimadores de algunas medidas de pobreza basados en variables auxiliares.

Sea x una variable auxiliar relacionada con la variable de interés y , el estimador de tipo razón de P_α propuesto viene dado por

$$\hat{p}_{\alpha;r} = \frac{\hat{p}_\alpha}{\hat{P}_{\alpha;x}} P_{\alpha;x} \quad (4.3)$$

donde \hat{p}_α es el estimador de tipo Hájek de P_α definido en la ecuación (4.2), y

$$\hat{P}_{\alpha;x} = \frac{1}{\hat{N}} \sum_{i \in s} d_i \left(\frac{L - \hat{\beta}x_i}{L} \right)^\alpha \delta(\hat{\beta}x_i \leq L)$$

es el estimador de tipo Hájek de

$$P_{\alpha;x} = \frac{1}{N} \sum_{i=1}^N \left(\frac{L - \beta x_i}{L} \right)^\alpha \alpha(\beta x_i \leq L)$$

El estimador definido por la ecuación (4.3) está motivado por el hecho que $\hat{\beta}x_i$ es un predictor de y_i bajo el modelo

$$y_i = \beta x_i + \nu(x_i)u_i \quad i = (1, \dots, N),$$

donde

$$\hat{\beta} = \frac{\sum_{i \in s} d_i y_i x_i / \nu^2(x_i)}{\sum_{i \in s} d_i x_i^2 / \nu^2(x_i)}.$$

Destacamos que es habitual considerar que $\nu(x_i) = x_i^{1/2}$. Si este es el caso, el estimador $\hat{\beta}$ puede expresarse como

$$\hat{\beta} = \frac{\sum_{i \in s} d_i y_i}{\sum_{i \in s} d_i x_i}.$$

Recordamos que el estimador $\hat{p}_{\alpha,r}$ se aplica para la estimación de los índices de brecha de pobreza y pobreza severa, es decir, para los valores $\alpha = \{1,2\}$. Destacamos que cuando $\alpha = 0$, el parámetro P_α se reduce a

$$P_0 = F(L) = \frac{1}{N} \sum_{i=1}^N \delta(y_i \leq L),$$

es decir, P_0 es la función de distribución $F(\cdot)$ evaluada en el argumento L . En este caso, el estimador $\hat{p}_{\alpha,r}$, con $\alpha = 0$, es el estimador de tipo razón de la función de distribución definido por Rao et al. (1990).

Tal como se comenta en Rao et al. (1990) para el problema de la estimación de la función de distribución, el estimador $\hat{p}_{\alpha,r}$ tiene propiedades deseables cuando la variable de interés es aproximadamente proporcional a la variable auxiliar. En particular, $\hat{p}_{\alpha,r}$ puede tener una ganancia considerable en eficiencia en comparación con el estimador habitual \hat{p}_α . Además, la varianza de $\hat{p}_{\alpha,r}$ es cero cuando la variable de interés es proporcional a la variable auxiliar, y esto implica que $\hat{p}_{\alpha,r}$ se reduce a P_α en esta situación.

4.4. Estimación mediante intervalos de confianza

En la sección anterior analizamos la estimación puntual de los índices de brecha de pobreza y pobreza severa. Sin embargo, la estimación de la varianza y la estimación de parámetros mediante intervalos de confianza también son temas importantes en estudios de pobreza. Por ejemplo, la variabilidad muestral y los intervalos de confianza relacionados con numerosas estimaciones de medidas de pobreza puede ser un tema muy importante cuando se desean realizar comparaciones entre países, distintas regiones dentro de un país, o bien a lo largo del tiempo.

El problema de la obtención de expresiones analíticas para la varianza de los estimadores de parámetros complejos, tales como los índices de brecha de pobreza y pobreza severa, es un tema que puede tener difícil solución. En primer lugar, esto se debe a que estas medidas de pobreza son funciones no lineales de los valores observados y_i . Por otra parte, los datos relacionados con ingresos de personas o familias generalmente se obtienen a partir de encuestas basadas en diseños muestrales complejos, tal como muestreo por conglomerados, estratificados, en varias fases, etc., con probabilidades de inclusión desiguales. En esta situación, encontrar expresiones teóricas para la varianza de estimadores también puede ser una cuestión con difícil solución.

Como alternativa al problema anteriormente comentado, se pueden utilizar métodos de aproximación para estimar la varianza de los estimadores de los índices de brecha de pobreza y pobreza severa. Por ejemplo, Kovacevik et al. (1995) llevaron a cabo un estudio de simulación con el propósito de comparar el comportamiento empírico de varios métodos de remuestreo para el problema de la estimación de algunas medidas de desigualdad. Los métodos de remuestreo

como el jackknife (Wolter 2007 pag. 151) o el bootstrap (Stine 1989, Wolter 2007, pag. 194) son los métodos más conocidos para el problema de estimar la varianza de estimadores. De hecho, en la práctica tales metodologías son altamente utilizadas dada su sencillez y facilidad para ser obtenidas. Por ejemplo, estos métodos están incluidos en numerosos programas estadísticos a través de paquetes y herramientas que implementan varios métodos de remuestreo. Además, los métodos de remuestreo se pueden aplicar fácilmente a una alta clase de diseños muestrales. También tenemos que tener en cuenta que técnicas como la de linealización (Wolter 2007 pag. 226, Särndal et al. 1992 pag. 172, Shao y Wu 1989) o la replicación repetida equilibrada (Shao y Wu 1992, Wolter 2007 pag. 107) podrían también utilizarse. Sin embargo la aplicación de los métodos de linealización, por ejemplo, podrían tener problemas adicionales tal como el hecho de requerir la estimación de la función de densidad probabilística de la variable de interés. Por todas las razones anteriormente comentadas, los métodos de remuestreo se utilizan con bastante frecuencia en la práctica, y son muy comunes en las ciencias sociales (véase Stine 1989, Biewen 2002, Moran 2006, Davidson y Flachaire 2007, Thompson 2013).

Para el problema de la estimación de la varianza de estimadores de los índices de brecha de pobreza y pobreza severa proponemos la utilización del bootstrap rescalado propuesto por Rao et al. (1992). La ventaja de este método en comparación con otros métodos de remuestreo alternativos es su sencillez. Este método consiste en calcular un nuevo conjunto de pesos $d_{i(b)}^*$, con $i = \{1, \dots, N\}$ y $b = \{1 \dots B\}$, denominados pesos bootstrap. Los valores de $d_{i(b)}^*$ se obtienen después de un ajuste de los pesos muestrales originales d_i (véase Rao et al. 1992). El valor B indica el número de muestras de bootstrap. Para cada muestra bootstrap, los correspondientes pesos bootstrap son utilizados para obtener el estimador de la medida de pobreza P_α , es decir, simplemente reemplazamos d_i por $d_{i(b)}^*$ en las

ecuaciones (4.2) y (4.3). Por ejemplo, si consideramos el estimador habitual de \hat{p}_α definido en la ecuación (4.2), el correspondiente estimador de la varianza basado en el bootstrap rescalado viene dado por

$$\hat{V}_{boot}(\hat{p}_\alpha) = \frac{1}{B} \sum_{b=1}^B (\hat{p}_{\alpha(b)}^* - \hat{p}_\alpha)^2$$

donde

$$\hat{p}_{\alpha(b)}^* = \frac{1}{\hat{N}_{(b)}^*} \sum_{i \in S_b} d_{i(b)}^* \left(\frac{L - y_i}{L} \right)^\alpha \delta(y_i \leq L)$$

y

$$\hat{N}_{(b)}^* = \sum_{i \in S_b} d_{i(b)}^*.$$

De la misma manera se define el estimador de la varianza mediante el bootstrap rescalado para el estimador de tipo razón $\hat{p}_{\alpha,r}$ propuesto en este trabajo.

Los intervalos de confianza pueden derivarse fácilmente utilizando los estimadores de la varianza basados en el bootstrap rescalado anteriormente comentado. Por ejemplo, el intervalo de confianza relacionado con el estimador estándar \hat{p}_α , basado en la aproximación Normal y con nivel de confianza $1 - \gamma$ viene dado por

$$CI_N(\hat{p}_\alpha) = [L_{\alpha;N}, U_{\alpha;N}]$$

donde los extremos o límites inferior ($L_{\alpha;N}$) y superior ($U_{\alpha;N}$) de dicho intervalo se definen como

$$L_{\alpha;N} = \hat{p}_{\alpha} - z_{1-\gamma/2} \sqrt{\hat{V}_{boot}(\hat{p}_{\alpha})};$$

$$U_{\alpha;N} = \hat{p}_{\alpha} + z_{1-\gamma/2} \sqrt{\hat{V}_{boot}(\hat{p}_{\alpha})}$$

y donde $z_{1-\gamma/2}$ denota el cuantil de orden $1-\gamma/2$ de la distribución Normal estándar.

El método de percentiles de las muestras bootstrap (en inglés denominado como “*percentile bootstrap*”) también puede utilizarse para la construcción de intervalos de confianza para los índices de brecha de pobreza y pobreza severa analizados en este capítulo. En este caso, el intervalo de confianza viene dado por la expresión

$$CI_P(\hat{p}_{\alpha}) = [L_{\alpha;P}, U_{\alpha;P}]$$

donde los extremos inferior y superior de dicho intervalo se definen, respectivamente, como

$$L_{\alpha;P} = \hat{p}_{\alpha}^*[\gamma/2]$$

y

$$U_{\alpha;P} = \hat{p}_{\alpha}^*[1-\gamma/2],$$

donde $\hat{p}_\alpha^*[a]$ denota el cuantil de orden a de los valores bootstrap $\hat{p}_{\alpha(b)}^*$. Aunque no están presentados en este trabajo, destacamos que los intervalos de confianza basados en el estimador de tipo razón $\hat{p}_{\alpha,r}$ se definen de forma similar, sólo que habría que utilizar los correspondientes valores bootstrap basados en el estimador $\hat{p}_{\alpha,r}$.

4.5. Estudios de simulación Monte Carlo

A continuación analizamos empíricamente el comportamiento de los estimadores de los índices de brecha de pobreza y pobreza severa descritos en la Sección 4.3, así como el comportamiento de los estimadores de la varianza y los intervalos de confianza descritos en la Sección 4.4. Para ello, se realizan estudios de simulación Monte Carlo basados en las poblaciones con diferentes características descritas en capítulos anteriores y en el Apéndice I. Esta metodología nos permitirá estudiar el impacto de diferentes escenarios analizados sobre las distintas técnicas de estimación abordadas en este capítulo.

De forma muy breve recordamos los dos tipos de poblaciones utilizadas en este estudio de simulación, aunque el Apéndice I puede proporcionar mayor información en este sentido.

En primer lugar utilizaremos los conjuntos de datos reales extraídos de la EU-SILC del año 2011. En particular, se consideran los 8 conjuntos de datos muestrales obtenidos de la EU-SILC, y correspondientes a los siguientes países: Italia, Eslovenia, Reino Unido, Polonia, Bulgaria, Lituania, Bélgica y España. Al igual que en capítulos anteriores, los datos correspondientes de cada país serán considerados como una población, a partir de la cual se seleccionarán $R = 1000$

muestras para poder llevar a cabo las simulaciones Monte Carlo. Recordamos que esta metodología ha sido utilizada por Harms y Duchesne (2006) y Muñoz et al. (2014). Las variables de estudio para cada población son el ingreso por unidad de consumo (variable de interés), y las contribuciones por impuesto sobre la renta (variable auxiliar). El objetivo es estimar los índices de brecha de pobreza y pobreza severa para las distintas poblaciones reales y utilizando, para cada país, la línea de pobreza L definida por EUROSTAT. Cada uno de los países considerados posee diferentes características de acuerdo con sus correspondientes medidas de pobreza. Por ejemplo, la proporción de pobres oscila entre el 12.2% en Eslovenia y el 22.1% en Bulgaria. El índice de brecha de pobreza en Eslovenia es del 2.7% en contraste con el caso español, el cual se sitúa en el 8%. Por último, el índice de pobreza severa oscila entre el 1% de Eslovenia y el 5.7% de España. El valor del coeficiente de correlación lineal ρ entre la variable auxiliar y de interés puede tener también un impacto importante en el comportamiento de los estimadores basados en información auxiliar. Destacamos que los valores de ρ oscilan entre el 0.65 en España y el 0.88 observado en los datos procedentes de Bélgica e Italia.

En segundo lugar, nuestro estudio estará basado en poblaciones simuladas o artificiales basadas en la distribución Gamma, la cual es una distribución continua altamente utilizada en estudios de pobreza. En este caso, el objetivo principal es evaluar numéricamente el impacto sobre los diferentes métodos de estimación para diferentes valores de ρ . Además, analizaremos el comportamiento de los estimadores cuando los valores de P_α son pequeños y elevados, donde obviamente asumiremos que $\alpha = \{1, 2\}$. Recordamos que cada población contiene $N = 10000$ valores aleatorios generados de una distribución Gamma con media igual a 10. El valor de L utilizado es el 60% de la mediana de la variable de interés y . Para obtener un valor de la proporción de pobres (P_0)

igual al 15%, un valor del índice de brecha de pobreza (P_1) igual al 3.6%, y un valor del índice de pobreza severa (P_2) igual al 1.3% se utiliza una distribución $\gamma(10/2,2)$. La distribución $\gamma(10/4,4)$ se utiliza para obtener una población con los valores $P_0 = 25%$, $P_1 = 8.9%$ y $P_2 = 4.5%$, mientras que los valores $P_0 = 35%$, $P_1 = 19.6%$ y $P_2 = 14%$ se han obtenido utilizando la distribución $\gamma(10/11,11)$. Las variables auxiliares se han generado igualmente a través de la distribución Gamma, y de forma que los correspondientes parámetros se han seleccionado de forma que ρ oscila entre 0.5 y 0.9. Para un mayor detalle descriptivo de estas poblaciones puede consultarse el Apéndice I.

En este estudio de simulación se analiza el comportamiento empírico de los estimadores \hat{p}_α y $\hat{p}_{\alpha,r}$ de P_α , con $\alpha = \{1,2\}$, mediante el sesgo relativo (SR) y el error cuadrático medio relativo (ECMR), los cuales se definen para los estimadores anteriormente comentados como:

$$SR[\tilde{p}_\alpha] = \frac{E[\tilde{p}_\alpha] - P_\alpha}{P_\alpha}$$

y

$$ECMR[\tilde{p}_\alpha] = \frac{\sqrt{ECM[\tilde{p}_\alpha]}}{P_\alpha}$$

donde \tilde{p}_α denota a cualquiera de los estimadores \hat{p}_α y $\hat{p}_{\alpha,r}$, y

$$E[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R \tilde{p}_\alpha(i)$$

y

$$ECM[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R (\tilde{p}_\alpha(i) - P_\alpha)^2$$

son respectivamente la esperanza empírica y el error cuadrático medio empírico basados en $R=1000$ muestras, y donde $\tilde{p}_\alpha(i)$ denota el valor del estimador \tilde{p}_α para la i -ésima muestra, con $i = \{1, \dots, R\}$. Como se comentaba en capítulos anteriores, las medidas SR y $ECMR$ son posiblemente las medidas más utilizadas para comparar la precisión entre distintos estimadores. En este sentido, pueden consultarse distintas referencias relevantes que incluyen tales medidas, tal como Rao et al. (1990), Deville y Särndal (1992), Silva y Skinner (1995), etc. Siguiendo estas referencias, las muestras han sido seleccionadas usando MAS y muestreo estratificado. Las fracciones de muestreo utilizadas en este estudio son $f = \{0.01, 0.05, 0.1, 0.2\}$, donde $f = n/N$.

Adicionalmente, y con el objetivo de corroborar las conclusiones finales obtenidas con otras medidas empíricas alternativas, también se ha utilizado en este estudio el sesgo absoluto empírico (SA) y la varianza empírica (VE) de los estimadores, las cuales vienen definidas como

$$SA[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R \frac{|\tilde{p}_\alpha(i) - P_\alpha|}{P_\alpha}$$

y

$$VE[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R (\tilde{p}_\alpha(i) - E[\tilde{p}_\alpha])^2. \quad (4.4)$$

La medida empírica SA se ha utilizado en Rueda et al. (2007), mientras que la medida VE se ha usado, por ejemplo, en Harms y Duchesne (2006).

Con respecto al comportamiento empírico de los estimadores de la varianza de los estimadores de los índices de brecha de pobreza y pobreza severa, así como sus correspondientes intervalos de confianza, es bastante común utilizar, respectivamente, el sesgo relativo de los estimadores de la varianza (SR_V) y la cobertura empírica (CE) de los intervalos de confianza con un nivel de confianza del 95%. El sesgo relativo de los estimadores de la varianza se define como

$$SR_V[\hat{V}_{boot}(\tilde{p}_\alpha)] = \frac{E[\hat{V}_{boot}(\tilde{p}_\alpha)] - VE[\tilde{p}_\alpha]}{VE[\tilde{p}_\alpha]}$$

donde

$$E[\hat{V}_{boot}(\tilde{p}_\alpha)] = \frac{1}{R} \sum_{i=1}^R \hat{V}_{boot}(\tilde{p}_\alpha(i))$$

y la varianza empírica se ha definido en la ecuación (4.4). Destacamos que se han utilizado $B=1000$ muestras bootstrap para obtener la varianza de los estimadores $\hat{V}_{boot}(\tilde{p}_\alpha(i))$.

La cobertura empírica de los intervalos de confianza basados en la aproximación normal y en los percentiles de las muestras bootstrap se definen, respectivamente, como

$$CR_N[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R \delta(L_{\alpha;N}(i) \leq P_\alpha \leq U_{\alpha;N}(i))$$

y

$$CR_P[\tilde{p}_\alpha] = \frac{1}{R} \sum_{i=1}^R \delta(L_{\alpha;P}(i) \leq P_\alpha \leq U_{\alpha;P}(i)).$$

Recordamos que los intervalos de confianza se han construido con un nivel de confianza del 95%. Los valores de SR_V y cobertura empírica se han utilizado por Berger y Muñoz (2015) y Harms y Duchesne (2006).

A continuación analizamos el comportamiento empírico de los distintos estimadores de los índices de brecha de pobreza y pobreza severa, y para ello utilizaremos las distintas medidas empíricas descritas anteriormente.

En primer lugar podemos destacar que los estimadores \hat{p}_α y $\hat{p}_{\alpha,r}$ de estos indicadores de pobreza mostraron un buen comportamiento en términos de sesgo y para las distintas poblaciones analizadas en este trabajo, puesto que los correspondientes valores de sesgo relativo, SR , fueron, en general y en términos absolutos, menores del 1%. Esta observación puede consultarse por ejemplo en las Tablas 4.1 y 4.2. Tan sólo en el caso de fracciones de muestreo pequeñas se han observado valores de sesgo relativo ligeramente mayores del 1% en términos relativos (véase por ejemplo los valores de SR de Reino Unido y Polonia de la Tabla 4.5 y para el valor $f = 0.01$).

Las Tablas 4.1 y 4.2 presentan los resultados derivados de la estimación del índice de brecha de la pobreza con poblaciones basadas en los datos de la EU-SILC. Como se comentaba anteriormente, los dos estimadores que se analizan en este estudio muestran un buen comportamiento en términos de sesgo relativo. Se observa que el estimador tipo razón propuesto $\hat{p}_{\alpha,r}$ es más eficiente que el estimador \hat{p}_α , puesto que el estimador $\hat{p}_{\alpha,r}$ tiene asociados unos errores cuadráticos medios más pequeños. Como se esperaba, ambos estimadores de P_α son más eficientes a medida que aumentan las fracciones de muestreo. El valor de ρ para afectar de forma significativa al estimador $\hat{p}_{\alpha,r}$, puesto que dicho estimador de tipo razón puede llegar a ser menos eficiente que el estimador

estándar \hat{p}_α cuando ρ es pequeño (véase por ejemplo el caso $f = 0.05$ y con resultados derivados de España). Las medidas SA y SE de las Tablas 4.1 y 4.2 también avalan una mejor precisión del estimador propuesto en comparación con el estimador estándar, puesto que ambos casos estas medidas son más pequeñas para el estimador de tipo razón propuesto.

SILC	f (%)	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
Italia ($\rho = 0.88$)	1	0.3	0.0	0.80	0.72	10.5	8.2	16.6	14.7
	5	0.2	0.1	0.38	0.32	2.2	1.6	7.7	6.5
	10	0.0	0.0	0.26	0.22	1.1	0.8	5.4	4.5
	20	-0.2	-0.1	0.17	0.15	0.4	0.3	3.4	3.0
Eslovenia ($\rho = 0.83$)	1	-0.2	-0.1	0.57	0.52	5.2	4.3	26.3	24.0
	5	0.0	0.0	0.26	0.24	1.1	0.9	11.8	10.7
	10	0.3	0.3	0.17	0.15	0.5	0.4	7.8	7.1
	20	-0.2	-0.1	0.12	0.11	0.2	0.2	5.4	4.9
Reino Unido ($\rho = 0.83$)	1	-0.5	-0.1	1.22	1.13	23.2	20.1	30.0	28.0
	5	0.3	0.3	0.51	0.49	4.3	3.8	12.9	12.2
	10	-0.5	-0.4	0.36	0.34	2.1	1.8	9.0	8.3
	20	0.2	-0.2	0.24	0.23	0.9	0.8	6.1	5.6
Polonia ($\rho = 0.82$)	1	-1.0	0.7	0.73	0.66	8.5	6.9	19.2	17.3
	5	-0.2	-0.1	0.32	0.29	1.6	1.4	8.2	7.7
	10	0.0	0.1	0.22	0.20	0.7	0.6	5.6	5.1
	20	0.1	-0.1	0.15	0.13	0.3	0.3	3.9	3.5

Tabla 4.1. Valores de SR, SA, VE y ECMR de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

SILC	$f(\%)$	SR		SA		VE		ECMR	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
Bulgaria ($\rho = 0.70$)	1	0.2	0.6	1.10	0.99	19.1	15.7	21.7	19.7
	5	0.1	0.2	0.48	0.44	3.6	3.1	9.4	8.7
	10	-0.1	-0.1	0.32	0.31	1.6	1.5	6.3	6.0
	20	-0.2	-0.2	0.21	0.20	0.7	0.6	4.2	3.8
Lituania ($\rho = 0.66$)	1	-0.1	0.1	1.48	1.41	34.2	32.1	30.9	29.9
	5	-0.3	0.0	0.63	0.62	6.2	5.9	13.1	12.8
	10	0.0	0.1	0.43	0.42	3.0	2.8	9.1	8.8
	20	0.2	0.3	0.29	0.28	1.4	1.2	6.2	5.9
Bélgica ($\rho = 0.65$)	1	-0.3	-0.8	1.18	1.13	25.3	22.7	43.2	40.9
	5	-0.2	-0.2	0.51	0.50	4.1	3.8	17.5	16.8
	10	-0.3	-0.6	0.36	0.36	2.1	2.0	12.3	12.2
	20	-0.2	-0.2	0.25	0.23	0.9	0.9	8.4	7.9
España ($\rho = 0.65$)	1	0.4	0.6	1.05	1.05	17.8	17.5	16.6	16.4
	5	-0.3	-0.2	0.47	0.47	3.4	3.6	7.3	7.4
	10	-0.1	0.1	0.32	0.66	1.6	6.9	5.0	5.0
	20	0.0	0.1	0.21	0.41	0.7	2.8	3.3	3.3

Tabla 4.2. Valores de SR, SA, VE y ECMR de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, Bélgica y España. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

En las Tablas 4.3 y 4.4 se analiza el comportamiento de los estimadores de la varianza obtenidos a partir del bootstrap reescalado. Para ello se ha utilizado el sesgo relativo (SR_V) de tales estimadores de la varianza y la cobertura empírica de los intervalos de confianza propuestos en la Sección 4.4 (CE_N y CE_P). Podemos observar que los sesgos son generalmente menores del 10% cuando f es menor que el 5%. Cuando $f = 20\%$ pueden observarse valores elevados de SR_V . Esta es la situación que se esperaba, puesto que el estimador de la varianza de bootstrap reescalado puede comportarse mejor cuando f es pequeño. Además, este hecho no supone un gran problema debido al hecho de que valores de f mayores que el 20% no son comunes en la práctica. Por otra parte, los sesgos de los estimadores de la varianza del bootstrap reescalado no tienen un impacto importante en los intervalos de confianza. Podemos observar que los intervalos de confianza tienen coberturas empíricas deseables, puesto que los valores de CE_N y CE_P están cercanos al 95%, el nivel de confianza fijado en este estudio. En general, podemos observar que las coberturas empíricas son menores que el 95% requerido cuando $f = 1\%$, y son mayores del 95% cuando $f = 20\%$. En general, las coberturas empíricas de los intervalos de confianza basados en los estimadores de tipo razón propuestos son ligeramente mejores que las coberturas empíricas de los intervalos de confianza basados en el estimador estándar. Por último, y atendiendo de nuevo a las coberturas empíricas observadas, podemos resaltar que el método de percentiles de las muestras bootstrap para la construcción de intervalos de confianza se comporta ligeramente mejor que los intervalos de confianza basados en la aproximación normal.

SILC	f (%)	SR_V		CE_N		CE_P	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
Italia ($\rho = 0.88$)	1	8.3	-0.4	94.4	93.1	94.2	93.5
	5	1.5	3.5	95.8	95.7	95.4	95.4
	10	4.8	5.6	95.8	94.8	96.2	95.0
	20	26.5	22.7	96.8	97.0	96.8	97.0
Eslovenia ($\rho = 0.83$)	1	0.5	0.5	91.7	92.8	92.7	93.4
	5	-1.5	0.2	94.3	94.9	94.0	94.3
	10	14.2	15.5	96.7	96.4	96.4	96.3
	20	19.1	21.3	96.6	97.2	96.9	96.9
Reino Unido ($\rho = 0.83$)	1	1.2	4.3	90.0	91.5	91.0	92.2
	5	8.4	4.3	94.9	94.6	94.6	95.0
	10	11.1	12.0	95.2	95.3	95.0	95.4
	20	22.3	22.1	96.3	96.1	96.4	96.2
Polonia ($\rho = 0.82$)	1	0.7	3.3	92.9	93.6	93.3	93.7
	5	9.9	3.3	95.4	95.1	95.2	95.3
	10	18.7	19.2	96.8	97.2	97.0	97.3
	20	25.0	26.4	97.3	97.5	97.1	96.7

Tabla 4.3. Valores de SR_V , CE_N y CE_P de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

<i>SILC</i>	<i>f</i> (%)	<i>SR_V</i>		<i>CE_N</i>		<i>CE_P</i>	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
Bulgaria ($\rho = 0.70$)	1	-1.3	-6.1	93.1	94.5	93.4	94.6
	5	4.1	6.7	95.3	95.2	95.0	95.0
	10	17.2	10.4	95.9	96.2	95.9	96.2
	20	31.7	34.6	97.9	98.1	97.8	98.2
Lituania ($\rho = 0.66$)	1	-2.9	-2.1	92.3	92.0	93.6	93.1
	5	7.8	5.0	95.7	95.5	95.6	95.3
	10	12.1	10.9	95.2	95.4	95.0	95.3
	20	22.0	24.0	96.3	97.1	96.6	97.2
Bélgica ($\rho = 0.65$)	1	-6.5	-0.1	87.0	86.6	87.6	88.0
	5	12.7	14.6	92.7	92.5	92.8	93.4
	10	11.1	5.6	93.1	92.8	93.3	92.5
	20	22.5	26.5	95.9	95.2	96.7	95.9
España ($\rho = 0.65$)	1	3.2	5.9	94.4	94.7	94.3	95.6
	5	4.5	-0.5	94.4	94.5	94.0	94.5
	10	12.7	8.2	95.3	96.2	95.7	96.4
	20	24.8	27.7	96.9	97.5	96.9	96.9

Tabla 4.4. Valores de SR_V , CE_N y CE_P de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, Bélgica y España. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

Los resultados anteriores están referidos a la estimación del índice de brecha de pobreza. En las Tablas 4.5, 4.6, 4.7 y 4.8 se presentan los resultados obtenidos a partir de la estimación del índice de pobreza severa, y considerando de nuevo las poblaciones basadas en la EU-SILC. También puede observarse que el estimador $\hat{p}_{\alpha,r}$ es el más eficiente, excepto en el caso de España. Esto puede deberse al hecho que valores pequeños de ρ pueden tener un impacto en el comportamiento de este estimador de tipo razón, $\hat{p}_{\alpha,r}$. Desde la Tabla 4.9 hasta la Tabla 4.20 se pretende analizar empíricamente este problema utilizando distribuciones Gamma y varios valores de ρ .

En cuanto a los valores de SR_V , CE_N y CE_P , tanto las Tablas 4.3 y 4.4, comentadas con anterioridad, como las Tablas 4.7 y 4.8 dan resultados similares, es decir, los valores de SR_V son generalmente menores del 10% cuando la fracción de muestreo es igual o inferior al 5%, mientras que las coberturas empíricas tienen un peor comportamiento cuando el coeficiente de correlación lineal ρ es más pequeño (véase por ejemplo los resultados obtenidos de los casos de Bélgica y España).

Desde la Tabla 4.9 hasta la Tabla 4.14 puede consultarse el impacto que diferentes valores del coeficiente de correlación lineal, ρ , puede producir sobre el estimador propuesto del índice de brecha de pobreza. De forma similar, este estudio se repite para el índice de pobreza severa desde la Tabla 4.15 hasta 4.20. Este estudio se ha llevado a cabo a través de las poblaciones descritas en el Apéndice I y basadas en la distribución Gamma. De forma adicional, en las mencionadas tablas se puede observar el comportamiento empírico de los estimadores de P_α , con $\alpha = \{1,2\}$, cuando las poblaciones Gamma tienen valores pequeños y elevados de P_α , con $\alpha = \{0,1,2\}$.

SILC	f (%)	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
Italia ($\rho = 0.88$)	1	-0.4	-0.4	0.73	0.71	9.0	8.2	27.0	25.8
	5	0.1	0.2	0.35	0.33	2.0	1.8	12.6	11.9
	10	0.1	0.1	0.24	0.24	1.0	0.9	8.8	8.4
	20	-0.2	-0.1	0.16	0.15	0.4	0.4	5.7	5.4
Eslovenia ($\rho = 0.83$)	1	0.0	0.2	0.30	0.29	1.4	1.3	38.0	36.4
	5	-0.1	-0.2	0.13	0.12	0.3	0.2	16.5	15.7
	10	0.5	0.5	0.09	0.08	0.1	0.1	10.9	10.4
	20	-0.3	-0.2	0.06	0.06	0.0	0.0	7.4	7.0
Reino Unido ($\rho = 0.83$)	1	1.7	2.4	1.04	1.02	18.6	18.7	50.6	50.6
	5	0.4	0.4	0.44	0.43	3.2	3.1	20.8	20.5
	10	-0.5	-0.4	0.31	0.30	1.6	1.5	14.7	14.2
	20	-0.1	-0.1	0.21	0.21	0.7	0.7	9.9	9.7
Polonia ($\rho = 0.82$)	1	-1.3	-0.9	0.48	0.45	3.6	3.2	27.8	26.2
	5	-0.3	0.0	0.20	0.20	0.7	0.6	12.0	11.5
	10	0.1	0.2	0.14	0.13	0.3	0.3	8.0	7.5
	20	-0.2	-0.1	0.09	0.09	0.1	0.1	5.5	5.1

Tabla 4.5. Valores de SR, SA, VE y ECMR de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

<i>SILC</i>	<i>f</i> (%)	SR		SA		VE		ECMR	
		\hat{P}_α	$\hat{P}_{\alpha:r}$	\hat{P}_α	$\hat{P}_{\alpha:r}$	\hat{P}_α	$\hat{P}_{\alpha:r}$	\hat{P}_α	$\hat{P}_{\alpha:r}$
Bulgaria ($\rho = 0.70$)	1	0.2	0.5	0.67	0.64	7.3	6.6	30.8	29.4
	5	0.1	0.3	0.29	0.28	1.3	1.2	13.2	12.7
	10	-0.2	-0.2	0.20	0.20	0.6	0.6	9.1	9.0
	20	-0.2	-0.1	0.13	0.13	0.3	0.3	6.0	5.8
Lituania ($\rho = 0.66$)	1	-0.2	1.0	1.10	1.09	19.1	19.1	43.7	43.7
	5	-0.4	0.6	0.47	0.48	3.3	3.6	18.3	19.0
	10	0.0	0.2	0.32	0.33	1.7	1.8	12.9	13.2
	20	0.3	0.5	0.22	0.23	0.7	0.8	8.6	9.1
Bélgica ($\rho = 0.65$)	1	1.4	1.4	1.85	1.84	150.4	147.6	30.8	29.4
	5	-0.4	-0.1	1.13	1.14	21.7	22.0	13.2	12.7
	10	-2.0	-2.1	0.81	0.81	10.2	10.4	9.1	9.0
	20	-0.8	-0.8	0.58	0.57	4.8	14.1	6.0	5.8
España ($\rho = 0.65$)	1	1.4	2.5	1.40	1.46	50.0	52.2	38.9	39.8
	5	-0.4	-0.1	0.71	0.74	9.8	10.4	17.2	17.7
	10	-0.2	0.1	0.52	1.02	4.4	16.7	11.3	11.6
	20	-1	0.1	0.63	0.70	2.0	7.7	7.5	7.7

Tabla 4.6. Valores de SR, SA, VE y ECMR de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, Bélgica y España. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

<i>SILC</i>	<i>f</i> (%)	<i>SR_V</i>		<i>CE_N</i>		<i>CE_P</i>	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
Italia ($\rho = 0.88$)	1	5.2	0.4	90.8	91.3	91.7	92.3
	5	0.4	0.7	93.6	94.3	93.7	94.5
	10	2.3	1.7	94.9	94.3	94.2	93.9
	20	23.4	23.1	96.0	96.2	96.4	96.0
Eslovenia ($\rho = 0.83$)	1	-2.7	-3.0	86.8	87.6	88.7	88.4
	5	1.5	1.9	94.2	93.7	94.5	94.1
	10	17.1	17.9	96.6	97.2	96.5	97.2
	20	25.1	28.6	96.5	97.6	97.2	98.0
Reino Unido ($\rho = 0.83$)	1	7.3	9.6	82.0	82.5	84.5	84.0
	5	10.4	8.7	92.8	92.3	93.9	93.4
	10	11.9	13.1	94.5	94.5	94.5	95.5
	20	22.8	22.9	95.5	95.5	96.1	95.6
Polonia ($\rho = 0.82$)	1	0.5	3.3	89.6	91.4	90.8	92.7
	5	9.1	3.8	94.3	94.9	95.1	94.9
	10	23.1	21.7	97.3	97.2	97.3	97.2
	20	29.6	29.5	97.4	97.4	97.6	97.0

Tabla 4.7. Valores de SR_V , CE_N y CE_P de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo datos extraídos de la EU-SILC en Italia, Eslovenia, Reino Unido y Polonia. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

SILC	f (%)	SR_V		CE_N		CE_P	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
Bulgaria ($\rho = 0.70$)	1	-2.9	3.8	88.8	89.5	89.9	90.3
	5	7.5	9.0	94.3	94.8	94.4	95.0
	10	13.1	8.1	95.8	95.5	95.6	95.7
	20	31.8	31.5	97.4	97.4	97.6	97.0
Lituania ($\rho = 0.66$)	1	-3.7	-5.5	86.4	87.7	87.9	89.6
	5	9.4	1.3	94.0	94.4	94.4	94.8
	10	9.5	9.3	94.6	94.9	94.7	95.3
	20	23.2	20.3	96.6	96.8	96.5	96.6
Bélgica ($\rho = 0.65$)	1	-5.7	-1.7	85.7	86.8	88.2	89.1
	5	12.4	12.0	88.5	89.2	90.1	90.7
	10	11.4	9.9	91.8	92.4	92.6	93.0
	20	20.9	23.0	96.1	96.2	96.3	96.4
España ($\rho = 0.65$)	1	2.6	7.5	90.1	89.5	91.0	90.7
	5	-4.5	-4.2	91.5	90.1	91.9	92.0
	10	6.2	6.0	91.7	91.9	91.6	92.8
	20	21.6	22.2	92.2	92.7	92.3	92.9

Tabla 4.8. Valores de SR_V , CE_N y CE_P de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo datos extraídos de la EU-SILC en Bulgaria, Lituania, Bélgica y España. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

A partir de las Tablas 4.9, 4.10 y 4.11 podemos observar que el estimador de tipo razón del índice de brecha de pobreza es claramente más eficiente que el estimador estándar cuando los valores de ρ son elevados. Este estimador, $\hat{p}_{\alpha,r}$, es también más preciso que el estimador estándar, \hat{p}_{α} , cuando el coeficiente de correlación lineal toma los valores $\rho = \{0.5, 0.7\}$, pero podemos comprobar cómo la ganancia en eficiencia descende a medida que el valor de ρ también descende. En otras palabras, $\hat{p}_{\alpha,r}$ es más eficiente a medida que el valor de ρ incrementa, la cual era es la situación que esperábamos. Por otra parte, ambos estimadores \hat{p}_{α} y $\hat{p}_{\alpha,r}$ son más eficientes a medida que los valores de P_{α} incrementan. Por ejemplo, el valor de la medida empírica *ECMR* del estimador $\hat{p}_{\alpha,r}$ es 20% cuando $P_0 = 15\%$, $\rho = 0.9$ y $f = 1\%$ (Tabla 4.11). Esta misma medida se sitúa aproximadamente en la mitad del valor anteriormente comentado (9.8%) en la Tabla 4.9, es decir, cuando $P_0 = 35\%$ y se mantienen constantes los valores de ρ y f ($\rho = 0.9$ y $f = 1\%$).

Respecto a los sesgos relativos (*SR*) podemos volver a resaltar que ambos estimadores ofrecen resultados deseables, puesto que en ningún caso se supera el 1%. A partir de las Tablas 4.9, 4.10 y 4.11 también podemos observar que el estimador de tipo razón propuesto muestra mejores resultados en términos de las medidas *SA* y *VE*.

ρ	$f(\%)$	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha:r}$	\hat{p}_α	$\hat{p}_{\alpha:r}$	\hat{p}_α	$\hat{p}_{\alpha:r}$	\hat{p}_α	$\hat{p}_{\alpha:r}$
0.9	1	0.0	0.1	2.38	1.53	93.4	37.7	15.5	9.8
	5	-0.2	0.2	1.12	0.67	19.4	7.1	7.1	4.3
	10	0.1	0.1	0.76	0.45	9.0	3.1	4.8	2.8
	20	-0.2	-0.1	0.52	0.29	4.1	1.4	3.2	1.9
0.7	1	0.3	0.7	2.58	2.12	101.9	69.7	16.5	13.7
	5	-0.2	0.2	1.11	0.91	18.7	12.9	7.1	5.9
	10	0.1	0.0	0.76	0.61	9.1	5.7	4.9	3.9
	20	-0.1	0.1	0.50	0.42	4.0	2.7	3.3	2.7
0.5	1	0.3	0.7	2.55	2.53	104.3	99.0	16.5	16.1
	5	-0.1	-0.3	1.10	1.08	10.5	19.1	7.1	6.9
	10	0.1	0.3	0.77	0.75	9.1	8.7	4.9	4.7
	20	-0.1	-0.1	0.52	0.49	4.1	3.7	3.3	3.1

Tabla 4.9. Valores de SR, SA, VE y ECMR de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma.

$P_0 = 0.35$, $P_1 = 0.196$ y $P_2 = 0.14$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	f (%)	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	0.1	0.6	1.56	0.92	37.7	13.5	21.8	13.1
	5	-0.2	-0.2	0.67	0.41	7.1	2.7	9.5	5.8
	10	0.2	0.2	0.45	0.28	3.2	1.2	6.4	3.9
	20	0.2	0.1	0.30	0.18	1.5	5.3	4.3	2.6
0.7	1	0.5	1.1	1.51	1.32	36.0	26.8	21.6	18.7
	5	0.1	0.5	0.66	0.58	6.7	5.2	9.4	8.2
	10	-0.1	0.1	0.46	0.39	3.2	2.4	6.5	5.6
	20	0.0	0.1	0.30	0.26	1.5	1.0	4.4	3.7
0.5	1	0.1	-0.1	1.56	1.48	37.7	34.2	21.8	20.8
	5	0.7	0.8	0.66	0.64	6.8	6.4	9.3	9.1
	10	-0.3	-0.2	0.45	0.43	3.2	3.0	6.3	6.2
	20	0.0	0.0	0.30	0.30	1.5	1.4	4.3	4.2

Tabla 4.10. Valores de SR, SA, VE y ECMR de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.25$, $P_1 = 0.089$ y $P_2 = 0.045$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	f (%)	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	0.1	1.4	0.76	0.54	0.1	4.6	27.9	20.0
	5	-0.3	-0.1	0.35	0.24	1.9	0.9	12.9	9.0
	10	0.6	0.3	0.25	0.17	1.0	0.4	9.2	6.1
	20	0.0	0.1	0.16	0.11	0.4	0.2	6.0	4.1
0.7	1	-0.1	1.2	0.88	0.80	11.9	10.1	30.6	28.3
	5	0.4	0.4	0.38	0.34	2.3	1.9	13.5	12.1
	10	0.1	0.0	0.26	0.23	1.1	0.8	9.2	8.0
	20	0.0	0.2	0.17	0.15	0.5	0.4	6.1	5.3
0.5	1	-0.9	-0.2	0.85	0.87	11.3	12.0	30.2	29.7
	5	0.2	0.4	0.39	0.37	2.4	2.2	13.4	13.0
	10	0.2	0.4	0.27	0.26	1.1	1.1	9.4	9.1
	20	0.1	0.0	0.18	0.17	0.5	0.5	6.2	5.9

Tabla 4.11. Valores de SR, SA, VE y ECMR de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.15$, $P_1 = 0.036$ y $P_2 = 0.013$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

Para los problemas de la estimación de la varianza y la construcción de intervalos de confianza de los distintos estimadores del índice de brecha de pobreza, en las Tablas 4.12, 4.13 y 4.14 podemos analizar el comportamiento de las diferentes estimaciones en términos de las medidas SR_V , CE_N y CE_P . Los valores de SR_V son generalmente menores del 10% cuando la fracción de muestreo f es menor del 5%. Los intervalos de confianza propuestos tienen un buen comportamiento en términos cobertura empírica, especialmente cuando $f = \{0.05, 0.1\}$. Por último, podemos resaltar a partir de estos resultados derivados de la poblaciones Gamma que los valores de ρ no tienen un impacto claro en el comportamiento de los intervalos de confianza.

Pueden observarse conclusiones muy similares si comparamos las Tablas sobre la estimación del índice de brecha de pobreza (4.9, 4.10, 4.11, 4.12, 4.13 y 4.14) con las tablas que aportan los resultados sobre el índice de pobreza severa (4.15, 4.16, 4.17, 4.18, 4.19 y 4.20). No obstante, a continuación destacamos algunas pequeñas diferencias. Por ejemplo, el estimador $\hat{p}_{\alpha,r}$ puede ser menos eficiente que \hat{p}_α para valores pequeños de P_0 , ρ y f (véase en la Tabla 4.17 el caso $P_0 = 15\%$, $\rho = 0.5$ y $f = 1\%$). Además, los intervalos de confianza pueden tener un comportamiento bastante pobre cuando $P_0 = 15\%$ y $f = 1\%$ (véase la Tabla 4.20), puesto que las correspondientes coberturas empíricas pueden ser menores del 90%.

ρ	$f(\%)$	SR_V		CE_N		CE_P	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
0.9	1	7.9	-4.9	94.6	94.8	93.1	93.8
	5	3.9	0.0	95.3	95.2	95.2	95.3
	10	11.6	14.6	96.2	96.1	96.1	96.3
	20	23.1	26.2	96.6	96.4	96.1	96.2
0.7	1	-2.3	1.4	93.6	94.6	94.3	94.2
	5	6.1	6.4	96.0	96.1	96.0	96.0
	10	9.4	15.9	95.9	96.4	95.8	97.2
	20	25.4	27.0	97.2	96.9	96.9	96.8
0.5	1	-2.8	-3.1	94.7	94.5	94.3	94.5
	5	3.9	2.8	95.5	95.7	95.5	95.5
	10	11.7	7.8	96.9	96.4	97.1	96.0
	20	24.4	25.1	97.6	97.6	97.6	97.3

Tabla 4.12. Valores de SR_V , CE_N y CE_P de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.35$, $P_1 = 0.196$ y $P_2 = 0.14$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	f (%)	SR_V		CE_N		CE_P	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	-2.1	5.9	93.5	94.4	93.9	94.1
	5	2.9	1.6	95.5	94.2	95.3	94.7
	10	14.5	9.6	96.3	95.6	96.1	95.5
	20	26.3	25.7	97.6	97.2	97.6	97.2
0.7	1	4.1	6.7	94.4	94.4	94.8	95.2
	5	8.5	8.2	95.6	96.0	95.4	95.8
	10	15.6	17.1	97.0	96.4	97.0	96.5
	20	27.9	27.6	97.1	97.4	96.8	97.4
0.5	1	-0.9	7.5	93.4	94.8	94.2	95.2
	5	10.5	11.2	95.9	95.7	95.7	95.3
	10	17.4	16.3	96.3	96.2	96.0	95.7
	20	27.1	25.1	97.1	97.3	96.9	96.8

Tabla 4.13. Valores de SR_V , CE_N y CE_P de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.25$, $P_1 = 0.089$ y $P_2 = 0.045$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	$f(\%)$	SR_V		CE_N		CE_P	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	6.7	8.9	93.9	95.4	94.4	95.0
	5	11.8	3.3	95.6	95.3	95.5	95.0
	10	12.9	11.2	96.5	95.9	96.2	95.9
	20	29.8	24.0	97.8	96.5	97.4	96.3
0.7	1	1.0	3.0	92.4	92.7	92.8	93.8
	5	3.1	1.0	95.3	94.9	95.3	94.7
	10	11.0	12.5	95.7	96.1	95.8	95.9
	20	27.7	28.7	97.6	97.4	97.2	97.3
0.5	1	4.8	0.9	92.1	92.3	93.5	93.7
	5	3.4	6.2	94.4	95.1	94.6	94.6
	10	6.3	7.8	96.1	96.1	96.6	95.8
	20	20.9	27.1	96.5	96.6	96.2	96.9

Tabla 4.14. Valores de SR_V , CE_N y CE_P de estimadores del índice de brecha de pobreza (P_α con $\alpha = 1$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.15$, $P_1 = 0.036$ y $P_2 = 0.013$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	f (%)	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	0.0	0.2	1.99	1.34	64.2	28.7	18.2	12.1
	5	-0.2	0.2	0.94	0.58	13.7	5.3	8.4	5.2
	10	0.1	0.2	0.62	0.39	6.2	2.3	5.6	3.4
	20	-0.2	-0.1	0.43	0.26	2.8	1.1	3.8	2.4
0.7	1	0.1	0.7	2.16	1.78	71.3	49.8	19.5	16.3
	5	-0.2	0.2	0.94	0.77	13.4	9.3	8.4	7.1
	10	0.1	0.1	0.63	0.53	6.3	4.3	5.8	4.8
	20	-0.1	0.1	0.42	0.35	2.7	1.9	3.8	3.2
0.5	1	0.3	0.8	2.18	2.16	75.8	72.2	19.7	19.2
	5	-0.3	-0.5	0.92	0.90	13.5	12.5	8.3	8.0
	10	0.1	0.3	0.65	0.64	6.5	6.3	5.8	5.7
	20	-0.1	-0.1	0.43	0.41	2.9	2.6	3.8	3.7

Tabla 4.15. Valores de SR, SA, VE y ECMR de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.35$, $P_1 = 0.196$ y $P_2 = 0.140$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	$f(\%)$	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	0.4	1.0	1.01	0.64	15.8	6.3	28.1	17.7
	5	-0.3	-0.1	0.43	0.27	3.0	1.1	12.2	7.6
	10	0.3	0.3	0.29	0.19	1.3	0.6	8.2	5.3
	20	0.2	0.0	0.20	0.12	0.6	0.2	5.5	3.5
0.7	1	0.8	1.8	1.01	0.89	16.0	12.6	28.3	25.1
	5	0.3	0.8	0.43	0.39	2.9	2.4	12.1	10.9
	10	-0.1	0.1	0.30	0.27	1.4	1.1	8.4	7.4
	20	0.0	0.1	0.20	0.18	0.6	0.5	5.6	4.9
0.5	1	0.4	0.1	0.99	0.94	15.5	14.1	27.6	26.2
	5	0.8	0.9	0.43	0.42	2.9	2.8	12.0	11.6
	10	-0.5	-0.3	0.29	0.28	1.3	1.3	8.1	7.9
	20	-0.1	-0.1	0.20	0.19	0.6	0.6	5.5	5.4

Tabla 4.16. Valores de SR, SA, VE y ECMR de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.25$, $P_1 = 0.089$ y $P_2 = 0.045$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	$f(\%)$	SR		SA		VE		ECMR	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	0.8	2.8	0.36	0.27	2.2	1.1	38.6	27.9
	5	-0.5	-0.1	0.17	0.12	0.5	0.2	17.6	12.8
	10	1.0	0.4	0.12	0.08	0.2	0.1	12.6	8.8
	20	0.0	0.1	0.09	0.06	0.0	0.0	8.2	5.8
0.7	1	0.5	2.6	0.44	0.41	3.1	2.8	43.0	40.0
	5	0.5	0.6	0.20	0.17	0.6	0.5	18.7	16.7
	10	0.0	0.1	0.13	0.12	0.3	0.2	12.4	11.1
	20	0.7	0.5	0.09	0.08	0.1	0.0	8.4	7.5
0.5	1	-2.1	-0.5	0.42	0.44	2.8	3.1	39.4	41.5
	5	0.5	0.8	0.19	0.19	0.6	0.6	17.8	17.5
	10	0.4	0.7	0.13	0.13	0.3	0.3	12.5	12.4
	20	0.2	0.2	0.09	0.09	0.1	0.1	8.2	7.9

Tabla 4.17. Valores de SR, SA, VE y ECMR de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.15$, $P_1 = 0.036$ y $P_2 = 0.013$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	$f(\%)$	SR_V		CE_N		CE_P	
		\hat{p}_α	$\hat{p}_{\alpha,r}$	\hat{p}_α	$\hat{p}_{\alpha,r}$	\hat{p}_α	$\hat{p}_{\alpha,r}$
0.9	1	9.3	-5.8	94.5	93.4	94.8	93.8
	5	2.3	2.1	94.8	95.1	94.9	95.0
	10	13.2	16.4	95.8	97.3	95.8	97.3
	20	24.1	23.9	97.3	95.6	97.3	95.9
0.7	1	-2.1	2.7	93.5	93.9	94.8	95.0
	5	4.4	5.7	95.7	95.9	95.7	95.9
	10	11.4	15.1	96.3	96.9	96.3	96.8
	20	27.5	29.4	97.2	97.6	96.8	97.1
0.5	1	-6.2	-6.4	93.1	93.1	93.3	93.6
	5	5.1	4.5	95.7	95.9	95.5	95.6
	10	10.1	5.3	96.6	96.2	96.7	95.8
	20	23.6	24.2	97.5	97.3	97.4	96.8

Tabla 4.18. Valores de SR_V , CE_N y CE_P de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.35$, $P_1 = 0.196$ y $P_2 = 0.140$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	f (%)	SR_V		CE_N		CE_P	
		\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$	\hat{p}_α	$\hat{p}_{\alpha;r}$
0.9	1	-2.5	4.2	91.9	93.2	93.1	93.6
	5	3.2	8.2	95.0	96.0	95.4	96.1
	10	14.2	9.7	95.1	96.0	96.1	96.0
	20	27.8	24.0	97.6	97.0	97.6	97.0
0.7	1	0.3	7.6	92.6	93.6	93.5	94.1
	5	9.3	8.4	95.5	95.9	95.4	95.2
	10	11.7	13.2	96.0	96.0	95.9	95.5
	20	25.0	27.1	97.0	97.7	97.0	97.6
0.5	1	1.3	7.8	92.4	92.6	93.0	94.1
	5	8.8	9.8	94.8	95.4	95.0	95.4
	10	17.2	16.6	96.0	96.1	95.9	96.2
	20	24.8	23.6	97.2	96.5	97.1	96.5

Tabla 4.19. Valores de SR_V , CE_N y CE_P de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.25$, $P_1 = 0.089$ y $P_2 = 0.045$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

ρ	$f(\%)$	SR_V		CE_N		CE_P	
		\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$	\hat{P}_α	$\hat{P}_{\alpha;r}$
0.9	1	7.6	9.8	90.0	93.5	91.8	93.8
	5	9.7	1.3	93.8	95.6	94.9	95.7
	10	10.4	6.8	96.2	95.4	95.8	95.5
	20	28.9	22.5	97.4	96.7	96.9	96.4
0.7	1	2.1	5.5	86.8	89.7	88.3	92.2
	5	0.9	3.2	93.4	94.0	93.5	94.2
	10	14.3	14.9	95.6	96.2	95.8	96.1
	20	28.1	29.3	97.8	97.6	97.4	97.5
0.5	1	3.4	4.0	87.4	88.6	88.8	90.2
	5	6.4	11.0	93.7	94.7	93.7	94.1
	10	7.4	9.6	95.3	95.3	95.1	95.6
	20	24.0	29.8	96.7	97.9	96.6	97.8

Tabla 4.20. Valores de SR_V , CE_N y CE_P de estimadores del índice de pobreza severa (P_α con $\alpha = 2$) y asumiendo las poblaciones basadas en la distribución Gamma. $P_0 = 0.15$, $P_1 = 0.036$ y $P_2 = 0.013$. $f = n/N$ es la fracción muestral y ρ es el coeficiente de correlación lineal entre y y x .

Para finalizar destacamos que todos los estudios de simulación Monte Carlo de esta sección se han llevado a cabo utilizando MAS. No obstante, estos estudios se han repetido bajo muestreo aleatorio estratificado, siguiendo las recomendaciones de importantes referencias tal como Rao et al. (1990) y Silva y Skinner (1995). Las conclusiones derivadas de estos estudios de simulación son similares a las conclusiones ya descritas en esta sección. Por este motivo, los resultados de los estudios de simulación basados en muestreo aleatorio estratificado se han omitido.

Capítulo 5

Conclusiones y futuras líneas de investigación

5.1. Conclusiones

El objetivo principal de este último capítulo de esta tesis es realizar un breve resumen con las principales conclusiones que se derivan a partir de este trabajo. Además, se plantean algunas líneas de investigación relacionadas con las aportaciones realizadas en este trabajo, y las cuales se podrían llevar a cabo en futuras investigaciones.

La idea principal de este trabajo ha sido el planteamiento o definición de nuevos estimadores para algunos indicadores de pobreza. Respecto a los métodos de estimación planteados, el objetivo era la definición de nuevos estimadores con propiedades deseables, especialmente en comparación con las propiedades de los estimadores estándar existentes en la literatura. Son numerosos los indicadores o medidas de pobreza existentes. En este trabajo nos hemos centrado en el estudio y análisis de los siguientes indicadores de pobreza:

- Cuantiles en la población de pobres para analizar la situación de éstos respecto el umbral de pobreza. De forma adicional y dada la relación con los cuantiles, también se ha analizado la estimación de la función de distribución en dicha población de pobres.
- Proporción de individuos con ingresos medios por unidad de consumo por debajo del umbral de pobreza, o proporción de pobres.

- Índices de brecha de pobreza y pobreza severa.

En el Capítulo 1 de esta tesis se introduce de forma muy breve el problema de la medición de la pobreza. Para ello se realizan algunas observaciones sobre la medición de este fenómeno. A continuación se definen y comentan algunos indicadores de pobreza, lo cual también sirve de introducción para las aportaciones realizadas en los capítulos siguientes de este trabajo. Además, se aportan resultados recientes de pobreza y otros estudios de pobreza y desigualdad con el fin de analizar la situación actual en algunas regiones, y como un complemento de la breve introducción realizada sobre el fenómeno de la pobreza.

En el Capítulo 2 se define un cuantil poblacional, medida que depende de la función de distribución. Por esta razón, se aportan los antecedentes más relevantes relacionados con la estimación de la función de distribución y cuantiles. El objetivo principal de este capítulo es la estimación de cuantiles en la población de individuos clasificados como pobres. Estas medidas nos sirven para analizar la situación de los pobres respecto el umbral de pobreza que delimita a la población de pobres y al resto de individuos de la población. En este sentido, la Figura 1.5 es determinante para entender la aplicación que pueden tener los cuantiles en el problema de la situación de los pobres de una población. Destacamos que las primeras aportaciones en el proceso de realización de esta tesis doctoral se realizaron precisamente en este Capítulo 2, por lo que podría decirse que son la base del resto de este trabajo, o bien la formación inicial para la realización de las aportaciones más completas descritas en los Capítulos 3 y 4. En particular, en el Capítulo 2 se han completado con una serie de nuevas poblaciones reales y artificiales los estudios llevados a cabo por Álvarez et al. (2012). En sentido, resultaba interesante analizar los estimadores de la función de distribución y cuantiles en la población de pobres con datos actuales y reales obtenidos a partir de encuestas sobre ingresos y condiciones de vida, tal como es la EU-SILC. En los

nuevos estudios de simulación se ha constatado que los estimadores de los parámetros anteriormente comentados tienen, en general, sesgos relativos con valores razonables. Tan sólo se observan algunos valores altos de sesgos relativos para órdenes de cuantiles muy bajos y fracciones de muestreo también muy pequeñas. Respecto a la eficiencia de estos estimadores, podemos destacar que se han evaluado a través de la medida empírica ECMR (Error Cuadrático Medio Relativo). En las tablas proporcionadas en este capítulo se puede observar la evolución de la eficiencia en función de distintos aspectos, tal como el orden del cuantil y la fracción de muestreo ya comentados. Con ayuda de poblaciones creadas a partir de la distribución Singh-Maddala también se ha podido analizar la precisión de los estimadores para los distintos coeficientes de asimetría considerados en estas poblaciones. Adicionalmente, también se pueden realizar comparaciones entre dos esquemas de muestreo diferentes: muestreo aleatorio simple sin reemplazamiento y muestreo con probabilidades desiguales (utilizando el método de Midzuno).

A lo largo de este trabajo hemos constatado en varias ocasiones que la mayoría de las medidas de pobreza dependen o están basadas en la línea o umbral de pobreza, puesto que este indicador económico es el encargado de dividir la población en pobres y no pobres. Una vez realizada esta clasificación, es posible que la medida de pobreza más inmediata que habría que conocer sea la proporción de pobres.

El objetivo del Capítulo 3 es la estimación de este indicador de pobreza. En este capítulo se define formalmente la proporción de pobres, es decir, básicamente es el cociente entre el número de pobres en una población y el número total de individuos de la mencionada población. Además, se introducen los estimadores conocidos de este parámetro, que son el estimador estándar de Horvitz-Thompson y el estimador de regresión logística (Lehtonen y Veijanen 1998, Duchesne 2003). Rueda et al. (2011a), Rueda et al. (2011b) y Muñoz et al. (2012) también proponen estimadores de una proporción

poblacional, aunque estos métodos de estimación requieren el uso de variables auxiliares dicotómicas.

Los métodos de estimación propuestos en el Capítulo 3 asumen que la variable auxiliar relacionada con la variable principal es de carácter cuantitativo, y tales métodos consisten en transformar la mencionada variable auxiliar en una variable dicotómica, de tal forma que se puedan aplicar fácilmente los estimadores de tipo razón y regresión propuestos en la literatura. Se ha propuesto 3 transformaciones con características similares, las cuales dependen de unas determinadas constantes. Se ha propuesto determinar las constantes mediante un método de búsqueda intensivo y aplicando criterios de minimización de varianzas. Una vez determinadas las constantes, los estimadores propuestos surgen de forma simple, usando expresiones básicas de tipo razón y regresión. Destacamos que se analiza tanto el caso de muestras seleccionadas mediante muestreo aleatorio simple sin reemplazamiento como el caso de muestras seleccionadas mediante un diseño muestral general.

Desde una perspectiva más económica y/o política, Molina y Rao (2010) indican que el uso de métodos de estimación con propiedades deseables y relacionadas con condiciones de vida de individuos puede ser una herramienta útil para organizaciones y agencias destinadas a luchar contra la pobreza, por lo que los resultados obtenidos en este apartado pueden tener este tipo de aplicaciones.

Se han llevado a cabo estudios de simulación para analizar el comportamiento empírico de distintos estimadores de la proporción de pobres. Por una parte, se han utilizado datos reales extraídos de la encuesta EU-SILC del año 2011. Por otra parte y con el fin de analizar escenarios alternativos a los abordados en las poblaciones basadas en datos reales, se han utilizado poblaciones simuladas a partir de distribuciones probabilísticas continuas y

relacionadas o utilizadas en estudios de pobreza, tal como la distribución Gamma o la distribución Singh-Maddala.

En primer lugar, los resultados empíricos indican que todos los estimadores analizados en este trabajo para la estimación de la proporción de pobres tienen un buen comportamiento en términos de sesgo relativo. En segundo lugar, podemos destacar que las transformaciones propuestas proporcionan, en general, resultados similares para los estimadores propuestos. Tal como se esperaba, los estimadores propuestos son más precisos que otros estimadores de la proporción de pobres en el caso de poblaciones con un coeficiente de correlación lineal elevado. En este caso se ha utilizado como criterio de comparación el error cuadrático medio relativo. En el caso de correlaciones lineales más bajas, los estimadores propuestos y el estimador de regresión logística tienen un comportamiento similar. Los distintos estimadores también se han analizado en función de la asimetría de la población en estudio. En este sentido, destacamos que los estimadores para la proporción de pobres propuestos han mostrado un buen comportamiento en el caso de variables con una alta asimetría, mientras que el estimador de regresión logística puede tener un peor comportamiento en esta situación. El resto de medidas utilizadas en este capítulo para analizar el comportamiento de estimadores puntuales también avalan los buenos resultados comentados con anterioridad.

Respecto a la estimación de la proporción de pobres mediante intervalos de confianza, podemos destacar que los métodos de estimación propuestos también han obtenido en este estudio de simulación resultados deseables, puesto que las correspondientes coberturas empíricas de los intervalos de confianza están en torno al 95%, el nivel de confianza considerado para los intervalos de confianza.

En el Capítulo 4 se han realizado aportaciones similares a las llevadas a cabo en el Capítulo 3, pero en este caso los parámetros o medidas de pobreza

objeto de estimación son los índices de brecha de pobreza y pobreza severa. Después de una breve introducción, se han definido formalmente los índices de brecha de pobreza y pobreza severa en el contexto de poblaciones finitas. Es importante resaltar que tanto la proporción de pobres como los índices de brecha de pobreza y pobreza severa surgen como casos particulares de la familia de medidas de pobreza propuestas por Foster et al. (1984), y este hecho se constata en la Sección 4.2 de esta tesis. También se introduce en esta sección el estimador estándar tanto en el caso de muestras seleccionadas mediante muestreo aleatorio simple como en el caso de muestras seleccionadas mediante un diseño muestral general.

El estimador propuesto para los índices de brecha de pobreza y pobreza severa se describe en la Sección 4.3. En particular, se trata de un estimador de tipo razón basado en predicciones obtenidas a partir de un modelo de regresión lineal con una única variable explicativa y sin término independiente. Dicha variable explicativa puede ser una variable auxiliar relacionada con la variable de interés, la cual juega el papel de variable dependiente o endógena en el modelo de superpoblación anteriormente comentado.

Además, se presentan intervalos de confianza basados en estimaciones de la varianza obtenidas a partir del método bootstrap reescalado (véase Rao et al. 1992). Para la construcción de los intervalos de confianza se propone el uso de distintas aproximaciones y metodologías. En particular, se propone utilizar la aproximación Normal y el método de percentiles de las muestras bootstrap.

Todas las aportaciones realizadas en este trabajo están analizadas o evaluadas empíricamente mediante una serie de estudios de simulación Monte Carlo. Al igual que en capítulos anteriores, se utilizan las poblaciones reales y simuladas descritas en el Apéndice I, así como distintas medidas empíricas para la comparación de estimaciones puntuales y mediante intervalos de confianza. A partir de los resultados obtenidos en estos estudios se puede observar que los

distintos estimadores de los índices de brecha de pobreza y pobreza severa presentan sesgos relativos insignificantes, es decir, con valores inferiores a 1% en términos absolutos. Por otra parte, el estimador propuesto es, en general, más eficiente que el estimador estándar, especialmente en aquellas situaciones donde el coeficiente de correlación lineal entre la variable auxiliar y la de interés es más elevado. Evidentemente, ésta es la conclusión que esperábamos. El resto de medidas también avalan la mejor precisión que parece tener el estimador propuesto.

Respecto a la estimación mediante intervalos de confianza, también se han obtenido resultados satisfactorios, los cuales pueden medirse a través del sesgo relativo de las estimaciones de las varianzas y de las coberturas empíricas de los intervalos de confianza. Para fracciones de muestreo superiores al 20% se han observado sesgos relativos para la varianza más elevados, aunque merece la pena destacar que ésta no es la situación habitual en la práctica. El método de percentiles de las muestras bootstrap para la construcción de intervalos de confianza parece tener un mejor comportamiento que los intervalos de confianza basados en la aproximación normal.

Destacamos que también se ha incluido en los estudios de simulación Monte Carlo del Capítulo 4 el método denominado bootstrap-t o studentized bootstrap (véase Efron y Tibshirani 1993 p. 160, Davison y Hinkley 1997 p. 194, Moran 2006), y los resultados obtenidos fueron similares a los resultados derivados de los intervalos de confianza basados en el método de percentiles de las muestras bootstrap, y por esta razón tales resultados basados en el método bootstrap-t no se han incluido en el Capítulo 4.

5.2. Limitaciones y futuras líneas de investigación

El objetivo de esta sección es resaltar algunas de las limitaciones que se han encontrado a lo largo de este trabajo y las consecuentes líneas de investigación que pueden surgir a partir de las mismas, o bien a partir de las propias aportaciones realizadas en este trabajo.

En el Capítulo 2 se ha planteado la estimación de cuantiles en la población de pobres y la propia estimación de la función de distribución en dicha población de pobres. En ambos casos se han utilizado estimadores basados en el diseño muestral, y utilizando únicamente la información proporcionada por la variable de interés. Como se puede comprobar en los Capítulos 3 y 4 de esta tesis, el uso de otras variables auxiliares en la fase de estimación puede producir resultados satisfactorios, tal como el hecho de mejorar la precisión de los estimadores en estudio. Álvarez et al. (2012) analizan los problemas que surgen para incorporar información auxiliar en el contexto de la estimación de la función de distribución y cuantiles en la población de pobres, lo que supone una importante limitación es este campo de investigación. Por tanto, ésta sigue siendo una línea de investigación abierta que puede proporcionar estimaciones más precisas para los parámetros anteriormente comentados.

En el Capítulo 3 de este trabajo se plantea la estimación de la proporción de pobres. Los estimadores propuestos están basados en transformaciones de la variable auxiliar cuantitativa en variables dicotómicas. Estas transformaciones dependen de parámetros que se obtienen minimizando la varianza estimada de los estimadores propuestos. Es posible que la solución al problema de minimizar esta varianza estimada no sea única, es decir, se obtengan J valores óptimos de la constante κ , por ejemplo (véase las expresiones (3.3) y (3.7)). Sean $\{\kappa_1, \dots, \kappa_J\}$ los J valores óptimos de κ que minimizan la varianza muestral (3.7), por ejemplo. Bajo este escenario se pueden utilizar varias soluciones diferentes.

En este trabajo y por motivos de simplicidad, se ha propuesto el uso de una única variable dicotómica auxiliar, la cual es seleccionada entre las J soluciones mediante el criterio descrito por la ecuación (3.8). No obstante, resaltamos de nuevo que esta solución no es única, y otras posibles soluciones son también factibles. Estos argumentos indican que se podrían plantear nuevas líneas de investigación. En particular, es conocido que la mayoría de los métodos de estimación, tal como los estimadores de tipo razón y regresión, están diseñados para incorporar fácilmente varias variables auxiliares en la etapa de estimación, de forma que el uso de tales variables auxiliares adicionales puede producir resultados más precisos. En otras palabras, una solución alternativa podría ser obtener estimadores de la proporción de pobres basados en las J variables dicotómicas derivadas de los J valores óptimos de κ . Esta solución es bastante compleja, especialmente para valores elevados de J , y de aquí que en este trabajo se haya propuesto una solución más simple (el uso de una única variable auxiliar). En el caso de para valores elevados de J , tal como se mencionaba con anterioridad, también puede resultar interesante obtener una selección óptima de variables auxiliares, de entre las J disponibles. Por ejemplo, diversos estudios indican que la ganancia en eficiencia no es significativa cuando en la etapa de estimación se incluyen más de 3 o 4 variables auxiliares (véase por ejemplo Rueda et al. 2008). En resumen, puede merecer la pena investigar las ganancias en términos de eficiencia de los estimadores propuestos basados en varias variables auxiliares dicotómicas.

Siguiendo la idea comentada en el párrafo anterior, destacamos que los datos muestrales recogidos en estudios de pobreza pueden contener varias variables auxiliares relacionadas con la variable de interés, las cuales se podrían utilizar, todas ellas o bien una parte de ellas, en la fase de estimación con el fin de mejorar la precisión de los estimadores. La extensión de los estimadores propuestos al caso de variables auxiliares es un tema simple a partir de la literatura existente. Por ejemplo, existen numerosas referencias en las que se extienden los estimadores de tipo razón al caso de múltiples variables

auxiliares, y tales metodologías se pueden aplicar sin excesiva dificultad en las aportaciones realizadas en este trabajo. Estas investigaciones pueden resultar interesantes y pueden producir importantes beneficios en términos de eficiencia.

Por otra parte, la proporción de pobres es una medida unidimensional que se puede expresar como una proporción (número de individuos clasificados como pobres entre número total de pobres). Sin embargo, la pobreza también suele verse como un fenómeno multidimensional, el cual se puede medir o interpretar mejor si se utilizan índices más complejos tal como el Índice de Desarrollo Humano. El método de estimación propuesto se puede aplicar fácilmente a cualquier medida que se pueda expresar como una proporción. En particular, tan sólo se requiere la disponibilidad de una variable auxiliar cuantitativa que esté relacionada con la variable de interés. Estos temas de investigación también pueden suponer una interesante línea de investigación para un futuro próximo.

Las aportaciones realizadas en los Capítulos 3 y 4 de este trabajo están basadas en un umbral de pobreza L fijo, y el cual se establece en función de la cantidad oficial proporcionada por los propios organismos de estadística. Una posible línea de investigación podría ser analizar el impacto en los distintos métodos de estimación cuando el umbral de pobreza L también debe obtenerse a partir de la propia variable de interés. En particular, puede resultar interesante analizar el impacto en el problema de la estimación de la varianza de los estimadores puntuales.

En el Capítulo 1 se realizó una breve revisión de algunas medidas de pobreza y desigualdad. En este trabajo se han realizado aportaciones en tan sólo algunas de estas medidas, en las más comunes y utilizadas. No obstante, puede resultar interesante realizar aportaciones en otras medidas de pobreza y desigualdad, tal como las descritas en el Capítulo 1. Adicionalmente, se podrían

realizar nuevas aportaciones en la familia de medidas de pobreza de Foster et al. (1984) cuando su parámetro α es mayor o igual que 3. Recordamos que en este trabajo se han discutido medidas de pobreza basados en valores de α menores de 3. En todos los casos, la idea a seguir puede ser la misma que la llevada a cabo a lo largo de este trabajo, es decir, empleando técnicas de estimación basadas en información auxiliar, investigando el problema de la estimación de la varianza, analizando el comportamiento empírico de todas las metodologías de estimación, etc. Esta línea de investigación también podría aportar numerosas e interesantes contribuciones.

Los organismos nacionales de estadística suelen obtener distintas medidas de pobreza tanto a nivel de toda la población en estudio como en diferentes subgrupos de la misma, tal como provincias, comunidades autónomas, etc. De esta forma, se pueden analizar los subgrupos de la población que son más vulnerables a la pobreza. Este tipo de estimaciones realizadas en subgrupos o subpoblaciones se denomina estimación en áreas pequeñas o en subpoblaciones. La estimación en áreas pequeñas es un tema con un interés relevante en el contexto del muestreo en poblaciones finitas. Algunas referencias importantes dentro de este campo pueden encontrarse en Purcell y Kish (1979), Särndal et al. (1992), Ghosh y Rao (1994), Estevao y Särndall (1999), Pfeffermann (2002), Rao (2003), Hidiroglou y Patak (2004), Zhang y Chambers (2004), etc. Como futuras líneas de investigación también se propone el estudio de nuevas metodologías de estimación de medidas de pobreza y desigualdad en áreas pequeñas.

En este trabajo se ha asumido en varias ocasiones un modelo de superpoblación lineal, con una única variable explicativa y sin término independiente. Las razones para la elección de este modelo son simples. En primer lugar, es el modelo de superpoblación asumido en la literatura más importante relacionada con las aportaciones realizadas (véase, por ejemplo, Chambers y Dunstan 1986 y Rao et al. 1990). En segundo lugar, la elección de

Capítulo 5. Conclusiones y futuras líneas de investigación

este modelo se debe a cuestiones de simplicidad. Es obvio que el hecho de buscar un modelo más complejo y que ajuste mejor a los datos puede proporcionar mejores resultados, pero éste es un tema complementario que no se ha perseguido en este trabajo, es decir, la idea ha sido utilizar un modelo simple, básico, y altamente utilizado en la literatura con el fin de poder realizar comparaciones bajo las mismas condiciones. No obstante, se puede incluir como una línea de investigación el problema de la búsqueda de un modelo óptimo, y analizar la ganancia en precisión que puede obtenerse a partir de modelos más complejos y en comparación con modelos más simples.

Referencias

- [1] **Álvarez, E.; Estudillo, M.D.; Castillo, S.** (2012). Estimación de la Función de Distribución y Cuantiles en la Población de Pobres. *Estudios de Economía Aplicada*, 30, pp. 1133-3197.
- [2] **Anand, S.** (1983). Inequality and Poverty in Malaysia: Measurement and Decomposition. *Oxford University Press*.
- [3] **Anand, S.; Sen, A.** (1992). Human Development Index: Methodology and Measurement. *Human Development Report Office Occasional Paper No. 12, UNDP, New York*.
- [4] **Atkinson, A.B.** (1987). On the Measurement of Poverty. *Econometrica*, 55, 4, pp. 749-764.
- [5] **Bárcena, E.; Cowell, F.** (2006). Static and Dynamic Poverty in Spain, 1993 - 2000. *Hacienda Pública Española*, 179, pp. 51-98.
- [6] **Becsey, J.; Berke, L.; Callan, J.R.** (1968). Nonlinear Least Squares Methods: A Direct Grid Search Approach. *Journal of Chemical Education*, 45, 728.
- [7] **Berger, Y.G.; Skinner, C.J.** (2003). Variance Estimation for a Low Income Proportion. *Journal of the Royal Statistical Society, Series B*, 52, pp. 457-468.
- [8] **Berger, Y.G.** (2008). A Note on the Asymptotic Equivalence of Jackknife and Linearization Variance Estimation for the Gini Coefficient. *Journal of Official Statistics*, 24, pp. 541-555.
- [9] **Berger, Y.G.; Torres O.** (2012). Empirical Likelihood Confidence Intervals: An Application to the EU-SILC Household Surveys. *Contributions to Statistics*, 2014, pp. 65-84
- [10] **Berger, Y.G.; Muñoz J.F.** (2015). On Estimating Quantiles Using Auxiliary Information. *Journal of Official Statistics*, 31, No. 1, pp. 101-119.
- [11] **Biewen, M.** (2002). Bootstrap Inference for Inequality, Mobility and Poverty Management. *Journal of Econometrics*, 108, pp. 317-342.
- [12] **Blackorby, C.; Donaldson, D.** (1980). Ethical Indices for the Measurement of Poverty. *Econometrica*, 48 (4), pp. 1053-1060.

Referencias

- [13] **Blackwood, D.L.; Lynch R.G.** (1994). The Measurement of Inequality and Poverty: A Policy Maker's Guide to the Literature. *World Development*, 22, pp. 567-578.
- [14] **Blyth, C.R.; Still, H.A.** (1983). Binomial Confidence Intervals. *Journal of the American Statistical Association*, 78, pp. 108-116.
- [15] **Carlson, M. D.** (1974). The 1972-73 Consumer Expenditure Survey. *Monthly Labor Review December*, 16-23.
- [16] **Cassel, C.M.; Särndal, C.E.; Wretman, J.H.** (1976). Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. *Biometrika*, 63, 615-620.
- [17] **Cassel, C.M.; Särndal, C.E.; Wretman, J.H.** (1977). Foundation of Inference in Survey Sampling. *John Wiley, New York*.
- [18] **Chakravarty, S.R.** (1983b). Ethically Flexible Measures of Poverty. *Canadian Journal of Economics*, 76, pp. 74-85.
- [19] **Chambers, R.L.; Dunstan, R.** (1986). Estimating Distribution Functions from Survey Data. *Biométrica*, 73, pp. 597-604.
- [20] **Chaudhuri, A.; Vos J.W.E.** (1988). Unified Theory and Strategies of Survey Sampling. *North-Holland*.
- [21] **Chaudhuri, A.; Stenger, H.** (2005). Survey Sampling: Theory and Methods. *Chapman & Hall/CRC, New York*.
- [22] **Chen, J.; Qin, J.** (1993). Empirical Likelihood Estimation for Finite Populations and the Effective Usage of Auxiliary Information. *Biometrika*, 80, pp. 107-116.
- [23] **Chen, J.; Sitter, R.R.** (1999). A Pseudo Empirical Likelihood Approach To the Effective Use of Auxiliary Information in Complex Surveys. *Statistica Sinica*, 9, pp. 385-406.
- [24] **Chen, J.; Wu, C.** (2002). Estimation of Distribution Function and Quantiles Using the Model-Calibrated Pseudo Empirical Likelihood Method. *Statistica Sinica*, 12, pp. 1223-1239.
- [25] **Chen, S.; Ravallion, M.** (2008). The Developing World is Poorer Than We Thought, But No less Successful in the Fight Against Poverty. *Policy Research Working Paper No. 4703, World Bank, Washington, DC*.
- [26] **Chernick, M.R.** (1999). Bootstrap Methods. Wiley Series in Probability and Statistics. *Wiley, New York*.

Referencias

- [27] **Citro, C.F.; Michael, R.T.** (1995). Measuring Poverty: A New Approach. *Canadian Journal of Economics*, 76, pp. 74-85.
- [28] **Clark, S.; Hemming, R.; Ulph, D.** (1981). On Indices for the Measurement of Poverty. *Economic Journal*, 91, pp. 515-526.
- [29] **Cochran, W.G.** (1997). Sampling Techniques (3rd edition). *John Wiley, New York*.
- [30] **Colosanto, D.; Kapteyn, A. y Van Der Gaag, J.** (1984). Two Subjects Definitions of Poverty: Results from the Wisconsin Basic Needs Studies. *Journal of Human Resources*, 19, pp. 127-137
- [31] **Crettaz, E.; Suter, C.** (2013). The Impact of Adaptive Preferences on Subjective Indicators: An Analysis of Poverty Indicators. *Social Indicators Research*, 114, pp. 139-152.
- [32] **Dalton, H.** (1920). The Measurement of the Inequality of Incomes. *Economic Journal*, 30, pp. 384-61.
- [33] **Davidson, A.C.; Hinkley, D.V.** (1997). Bootstrap Methods and their Application. *Cambridge, Cambridge University Press*.
- [34] **Davidson, R.; Flachaire, E.** (2007). Asymptotic and Bootstrap Inference for Inequality and Poverty Measures. *Journal of Econometrics*, 141, pp. 141-66.
- [35] **De Tocqueville, A.** (1997). Memoir on Pauperism, translated by S. Drescher. *Civitas, London*.
- [36] **De Vos, K; Garner, T.I.** (1991). An Evaluation of Subjetive Poverty Definitions: Comparing Resutls from the U.S. and the Netherlands. *The Behaviour of Income and Wealth*, 37, pp. 267-285.
- [37] **Deaton, A.** (1997). The Analysis of Household Surveys: A Microeconomic Approach to Development Policy. *Baltimore, MD: Johns Hopkins University Press for the Workd Bank*.
- [38] **Deville J.C.; Särndal, C.E.** (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87, pp. 376-382.
- [39] **Deville, J.C.** (1997). Estimation de la Variance du Coefficient de Gini Mesuré par Sondage (In French). *Actes des Journées de Methodologie Statistiques, Insee Méthodes*.

Referencias

- [40] **Dickens, R.; Manning, A.** (2004). Has the National Minimum Wage Reduced UK Wage Inequality?. *Journal of the Royal Statistical Society Series A*, 167, pp. 613-626.
- [41] **Domínguez, J.; Martín, A.M.** (2006). Medición de la Pobreza: Una Revisión de los Principales Indicadores. *Revista de Métodos Cuantitativos Para la Economía y la Empresa*, 2, pp. 27-66.
- [42] **Domínguez, J.; Núñez, J.J.** (2007). A Proposal of a Synthetic Indicator to Measure Poverty Intensity, With an Application to EU-15 Countries. *ECINEQ Working Paper*, 2007-81.
- [43] **Dorfman, A.H.; Hall, P.** (1993). Estimators of the Finite Population Distribution Function Using Nonparametric Regression. *The Annals of Statistics*, 21, No. 3, pp. 1952-1975.
- [44] **Dorfman, A.H.** (2009). Inference on Distribution Functions and Quantiles. *Handbook of Statistics 29B Sample Surveys: Inference and Analysis, D. Pfefermann and C.R. Rao, Eds, North-Holland*, 2009, pp. 371-395.
- [45] **Duchesne, P.** (2003). Estimation of a Proportion with Survey Data. *Journal of Statistics Education*, 11, pp. 1-24.
- [46] **Duclos, J.Y.; Araar, Ar.** (2006). Poverty and Equity: Measurement, Policy and Estimation with DAD
- [47] **Efron, B.** (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, pp. 1-26.
- [48] **Efron, B.** (1982). The Jackknife, the Bootstrap, and other resampling plans. *Conf. Board Math. Sci. Soc. Ind. Appl. Math*, 38, pp. 1-92.
- [49] **Efron, B.; Tibshirani, R.J.** (1993). An Introduction to the Bootstrap. *New York: Chapman & Hall*.
- [50] **Estevao, V.M.; Särndal, C.E.** (1999). The Use of Auxiliary Information in design-based estimation for domains. *Survey Methodology*, 25, pp. 213-221.
- [51] **Eurostat** (2000). Low-Wage Employees in EU Countries. In *Statistics in Focus: Population and Social Conditions. Luxembourg: Office for Official Publications of the EC*.
- [52] **Fernandez, A.** (1992). Los índices de pobreza FGT. Estimación de la distribución del ingreso en España. *Tesis Doctoral. Departamento de Estadística y Econometría. Universidad de Málaga*.

Referencias

- [53] **Fernández García, F.; Mayor Gallego, J.A.** (1994). Muestreo en poblaciones finitas. *Curso Básico. PPU, S.A.*
- [54] **Ferullo H.** (2006). El Concepto de la Pobreza de Amartya Sen. *Cultura Económica*, 66, pp. 10-16.
- [55] **Foster, J.E.; Greer, J.; Thorbecke, E.** (1984). A Class of Decomposable Poverty Measures. *Econometrica*, 81, pp. 761-766.
- [56] **Foster, J.E.** (1998). Absolute Versus Relative Poverty. *The American Economic Review*, 88, pp. 335-341.
- [57] **Foster, J.E.** (2007). A Class of Chronic Poverty Measures. *Working Paper*, 07-W01. Vanderbilt University.
- [58] **Francisco, C.A.; Fuller, W.A.** (1991). Quantile Estimation with a Complex Survey Design. *The Annals of Statistics*, 19 No. 1, pp 454-469.
- [59] **Frick, R.J.; Goebel, J.; Schechtman, E.; Wagner, G.G.; Yitzhaki, S.** (2006). Using Analysis of Gini (ANOGI) for Detecting Whether Two Sub-Samples Represent the Same Universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods and Research*, 34, pp. 427-468.
- [60] **Frick, R.J.; Grabcka, M.M.; Groh-Samberg, O.** (2012). Dealing with Incomplete Household Panel Data in Inequality Research. *Sociological Methods & Research*, 41, pp. 89-123.
- [61] **Friedman R.D.** (1965). Poverty: Definition and Perspective. *American Enterprise Institute for Public Policy Research. Washington, D.C.*
- [62] **Gallego, L.M.** (2009). Acercamiento al Problema Social de la Pobreza. De las Nociones de Pobreza a los Mecanismos Causales.
- [63] **Giambona, F.; Vassallo E.** (2014). Composite Indicator of Social Inclusion for European Countries. *Social Indicators Research*, 116, pp. 269-293.
- [64] **Giles, D.E.A.** (2004). Calculating a Standard Error for the Gini Coefficient: Some Further Results. *Oxford Bulletin of Economics and Statistics*, 66, pp. 425-433.
- [65] **Gini, C.** (1914). Sulla Misura della Concentrazione e della Variabilità dei Caratteri. *Atti del R. Istituto Veneto di Scienze Lettere ed Arti.*
- [66] **Glasser, G.J.** (1962). Variance Formulas for the Mean Difference and Coefficient of Concentration. *J. Amer. Stat. Assoc.*, 57, pp. 648-654.ç

Referencias

- [67] **Goedemé, T.** (2010). The Standard Error of Estimates Based on EU-SILC. An Exploration Through the Europe 2020 Poverty Indicators. Working Paper No. 10/09. University of Antwerp
- [68] **Gordon, L.** (1983). Successive Sampling in Finite Populations. *The Annals of Statistics*, 11, pp. 702-706.
- [69] **Gosh, M.; Rao, J.N.K.** (1994). Small Area Estimation: An Appraisal (with Discussion). *Statistical Science*, 9, 65-93.
- [70] **Gross, S.T.** (1980). Median Estimation in Sample Survey. *Proc. Surv. Res. Meth. Sect. Am. Stat. Assoc.*, pp. 181-184.
- [71] **Gradín, C.; Cantó, O.; Del Río, C.** (2008). Inequality, Poverty and Mobility: Choosing Income or Consumption as Welfare Indicators. *Investigaciones Económicas*, XXXII, 2 pp. 169-200.
- [72] **Hagenaars, A.; Van Praag, B.** (1982). A Syntesis of Poverty Line Definitions. *Leyden University. Center of Research in Public Economics*.
- [73] **Hagenaars, A.** (1986). The Perception of Poverty. *North Holland. Amsterdam*.
- [74] **Hall, P.** (1992). The Bootstrap and Edgeworth Expansion. *Springer, New York*.
- [75] **Harms, T.; Duchesne, P.** (2006). On Calibration Estimation for Quantiles. *Survey Methodology*, 32, pp. 37-52.
- [76] **Haughton, J.; Khandker, R.** (2009). Handbook on Poverty and Inequality. *The World Bank. Washington D.C.*
- [77] **Háyek, J.** (1964). Asymptotic Theory of Rejective Sampling With Varying Probabilities from a Finite Population. *Annals of mathematical Statistics*, 35 pp. 1491-1523.
- [78] **Hedayat, A.S.; y Sinha, K.S.** (1991). Design and Inference in Finite Population Sampling. *John Wiley, New York*.
- [79] **Hidiroglou, M.A., Patak, Z.** (2004). Domain Estimation Using Linear Regression. *Survey Methodology*, 30, pp. 67-78.
- [80] **Horvitz, D.G.; Thompson, D.J.** (1952). A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, 47 pp. 663-685.

Referencias

- [81] **INE** (2005). La Pobreza y su Medición. Presentación de Diversos Métodos de Obtención de Medidas de Pobreza. *Documento de Trabajo del Instituto Nacional de Estadística*.
- [82] **INE** (2005). Estudio Descriptivo de la Pobreza en España. Resultados Basados en la Encuesta de Condiciones de Vida 2004. *Documento de Trabajo del Instituto Nacional de Estadística*.
- [83] **Jolliffe, D.** (2001). Measuring Absolute and Relative Poverty. The Sensitivity of Estimated Household Consumption to Survey Design. *Journal of Economics and Social Measurement*, 27, 1-23.
- [84] **Kakwani, N.C.** (1980). Income Inequality and Poverty: Methods of Estimation and Policy Applications. *Oxford University Press*.
- [85] **Kahndker, S.** (2005). Introduction to Poverty Analysis. *World Bank Institute*.
- [86] **Kelly, A.C.** (1991). The Human Development Index: Handle with Care. *Popul. Dev. Rev.*, 17 (2), pp. 315-324.
- [87] **Kovacevic, M.S.; Yung, W.; Pandher, T.** (1995). Estimating the Sampling Variances of Measures of Income Inequality and Polarization – An Empirical Study. *Statistics Canada, Methodology Branch Working Paper*, HSMD-95-007E.
- [88] **Kovar, J.G.; J.N.K.; Wu, C.F.J.** (1988). Bootstrap and Other Methods to Measure Errors in Survey Estimates. *Canadian Journal of Statistics*, 16, pp. 25-45.
- [89] **Kuk, A.Y.C.** (1988). Estimation of Distribution Functions and Medians Under Sampling With Unequal Probabilities. *Biometrika*, 75, pp. 97-103.
- [90] **Kuk, A.Y.C.; Mak, T.K.** (1989). Median Estimation in the Presence of Auxiliary Information. *J.R. Stat. Soc. Ser.*, 51, pp. 261-269.
- [91] **Kuk, A.Y.C.; Mak, T.K.** (1994). A Functional Approach to Estimating Finite Population Distribution Functions. *Theory Meth.*, 23, pp. 883-896.
- [92] **Lehtonen, R.; Veijanen, A.** (1998). On Multinomial Logistic Generalized Regression Estimators. *Survey Methodology*, 24, pp. 51-55.
- [93] **Lelkes, O.; Medgyesi, M.; István György Tóth, I.G.; Ward, T.** (2009). Income Distribution and the Risk of Poverty in European Inequalities: Social Inclusion and Income Distribution in the European Union, Budapest. *TARKI Social Research Institute Inc.*

Referencias

- [94] **Lerman, P.M.** (1980). Fitting Segmented Regression Models by Grid Search. *Applied Statistics*, 29, pp. 77-84.
- [95] **Lerman, R.I.; Yitzhaki, S.** (1989). Improving the Accuracy of Estimates of Gini Coefficients. *Journal of Econometrics*, 42, pp. 43-47.
- [96] **Madden, D.** (2000). Relative or Absolute Poverty Lines: A New Approach. *Review of Income and Wealth*, 46, pp. 181-199.
- [97] **Mak, T.K.; Kuk, A.** (1993). A New Method for Estimating Finite-Population Quantiles Using Auxiliary Information. *Canadian Journal of Statistics*. 21, pp. 29-38.
- [98] **McGillivray, M.** (1991). The Human Development Index: Yet Another Redundant Composite Development Indicator? *World Dev.*, 19 (10), pp. 1461-1468.
- [99] **Medeiros, M.** (2006). The Rich and The Poor: The Construction of an Affluence Line from the Poverty Line. *Social Indicators Research*, 78, pp. 1-18.
- [100] **Midzuno, H.** (1952). A Method of Sample Selection Providing Unbiased Ratio Estimates. *Bull. Int. Stat. Inst.*, 3, pp. 133-140.
- [101] **Modarres, R.; Castwirth, J.L.** (2006). A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality. *Oxford Bulletin of Economics and Statistics*, 68, pp. 385-390.
- [102] **Molina, I.; Rao, J.N.K.** (2010). Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics*, 38, pp. 369-385.
- [103] **Moneo, C.; Adiego, M.** (2005). Pobreza y Pobreza Persistente en España, 1994-2001. *En Indicadores Sociales de España 2004*, INE. Madrid.
- [104] **Mooney, C.Z.; Duval, R.D.** (1993). Bootstrapping: A Nonparametric Approach to Statistical Inference, *Sage Publications, Beberly Hills, CA*.
- [105] **Moran, T.P.** (2006). Statistical Inference for Measures of Inequality with a Cross-National Bootstrap Application. *Sociological Methods and Research*, 34, pp. 296-333.
- [106] **Mukhopadhyay, P.** (2000). Topics in Survey Sampling. *Springer*.
- [107] **Muñoz, J.F.; Álvarez, E.; Arcos, A.; Rueda, M.M.; González, S.; Santiago, A.** (2012). Optimum Ratio Estimators for the Population Proportion. *International Journal of Computer Mathematics*, 89, 357-365.

Referencias

- [108] **Muñoz, J.F.; Álvarez, E., García-Fernandez, R.M.; Barroso, L.J.** (2014). Efficient Estimation of the Headcount Index. *Social Indicators Research*. DOI 10.1007 / s11205-014-0757-9.
- [109] **Navicke, J.; Rastrigina, O.; Sutherland, H.** (2013). Nowcasting Indicators of Poverty Risk in the European Union: A Microsimulation Approach. *Social Indicators Research*, doi: 10.1007/s11205-013-04918.
- [110] **Newcombe, R.G.** (1998). Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods. *Statistic in Medicine*, 17, pp. 857-872.
- [111] **Nuñez, J.J.** (2009). Estado Actual y nuevas Aproximaciones a la Medición de la Pobreza. *Estudios de Economía Aplicada*, 27, pp. 325-344.
- [112] **Nygaard, F.; Sandström, A.** (1985). The Estimation of the Gini and the Entropy Inequality Parameters in Finite Populations. *Journal of Official Statistics*, 1, pp. 399-412
- [113] **Ogwan, T.** (2000). A Convenient Method of Computing the Gini Index and its Standard Error. *Oxford Bulletin of Economics and Statistics*, 62, pp. 123-129.
- [114] **Ogwan, T.** (2004). Calculating a Standar Error for the Gini Coefficient: Some Further Results: Reply. *Oxford Bulletin of Economics and Statistics*, 66, pp. 435-437.
- [115] **Osberg, L.; Xu K.** (1999). Poverty Intensity: How Well Do Canadian Provinces Compare?. *Canadian Public Policy – Analyse de Politiques*, 25 (2): 199-221.
- [116] **Pérez-Mayo, J.; Fajardo, M.A.** (2003). Un Método Alternativo Para Identificar los Hogares en Situación de Pobreza. *Estadística Española*, 152, pp. 115-134.
- [117] **Pérez-Moreno, S.** (2009). El estudio de la Pobreza en España Desde una Óptica Económica: Medición y Políticas. *Estudios de Economía Aplicada*, 27, pp. 349-372.
- [118] **Pfefferman, D.** (2002). Small Area Estimation-New Developments and Directions. *International Statistical Review*, 70, pp. 125-143.
- [119] **Prescott, N; Menino, P.** (1997). A Poverty Profile of Cambodia. *Discussion Paper No. 373, World Bank, Washington, DC.*

Referencias

- [120] **Purcell, N.J.; Kish, L.** (1979). Estimation for Small Domains. *Biometrics*, 35, pp. 365-384.
- [121] **Rao, C.R.; Zhao, L.C.** (1992). Aproximation to the Distribution of M-Estimates in Linear Models by Randomly Weighted Bootstrap. *The Indian Journal of Statistics*, 54, Series A, Pt. 3, pp. 323-331.
- [122] **Rao, J.N.K.; Kovar, J.G.; Mantel, H.J.** (1990). On Estimating Distribution Function and Quantiles from Survey Data Using Auxiliary Information. *Biometrika*, 77, pp. 365-375.
- [123] **Rao, J.N.K.** (2003). Small Area Estimation. *John Wiley, New York*.
- [124] **Ravallion, M.** (1992). Poverty Comparisons. A Guide to Concepts and Methods. *Living Standards Measurements Surveys Working Paper No. 88*, World Bank, Washington, D.C.
- [125] **Ravallion, M.** (1998). Poverty Lines in Theory and Practice. *Living Standards Measurements Surveys Working Paper No. 133*, World Bank, Washington, D.C.
- [126] **Rowntree, S.** (1941). Poverty and Progress: A Second Survey of York. *London: Longmans, Greens*
- [127] **Rueda, M.M.; Martínez, S.; Martínez, H.; Arcos, A.** (2007). Estimation of the Distribution Function With Calibration Methods. *Journal of Statistical Planning and Inference*, 137, pp. 435-448.
- [128] **Rueda, M.M.; Muñoz, J.F.; Berger, Y.G.; Arcos, A.; Martínez, S.** (2007). Pseudo Empirical Likelihood Method in the Presence of Missing Data. *Metrika*, 65, pp. 349-367.
- [129] **Rueda, M.M.; Muñoz, J.F.** (2009). New Model-Assisted Estimators for the Distribution Function Using Pseudo Empirical Likelihood Method. *Statistica Neerlandica*.
- [130] **Rueda, M.M.; Muñoz, J.F.** (2011). Estimation of Poverty Measures With Auxiliary Information in Sample Surveys. *Quality and Quantity*, 45, pp. 687-700.
- [131] **Rueda, M.M.; Muñoz, J.F.; Arcos, A.; Álvarez, E.; Martínez, S.** (2011). Estimators and Confidence Intervals for the Proportion Using Binary Information with Application to Pharmaceutical Studies. *L Journal of Biopharmaceutical Statistics*, 21, pp. 526-554.

Referencias

- [132] **Rueda, M.M.; Muñoz, J.F.; Arcos, A.; Álvarez, E.; Martínez, S.** (2011b). Indirect Estimation of Proportions in Natural Resources Surveys. *Mathematics and Computers in Simulation*, 81, pp. 2317-2325.
- [133] **Sandström, A.** (1983). Estimating Income Inequality. Large Sample Inference in Finite Population. *Department of Statistics, University of Stockholm, Research Report*.
- [134] **Sandström, A.; Wretman, J.H.; Waldén, B.** (1985). Variance Estimators of the Gini Coefficient, Simple Random Sampling. *Metron*, 43, pp. 41-70.
- [135] **Sandström, A.; Wretman, J.H.; Waldén, B.** (1988). Variance Estimators of the Gini Coefficient – Probability Sampling. *Journal of Business and Economics Statistics*, 6, pp. 113-119.
- [136] **Särndal, C.E.; Swensson, B.; Wretman, J.** (1992). Model Assisted Survey Sampling. *Springer Verlag*.
- [137] **Sedransk, N.; Sedransk, J.** (1979). Distinguishing among Distributions Using Data from Complex Sample Designs. *Journal of the American Statistical Association*, 74, pp. 368.
- [138] **Sedransk, J.; Meyer, J.** (1978). Confidence Intervals for the Quantiles of a Finite Population: Simple Random and Stratified Simple Random Sampling. *J.R. Stat. Soc. Ser.*, 40, pp. 239-252.
- [139] **Sedransk, J.; Smith, P.J.** (1988). Inference for Finite Population Quantiles. In: Krishnaiah, P.R., Rao, C.R. (eds.) *Handbook on Statistics*, 6, pp. 267-289.
- [140] **Sen, A.** (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44 (2), pp. 219-231.
- [141] **Sen, A.** (1979). Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics?. *The Economic journal*, 89, pp. 537-538
- [142] **Sen, A.** (1987). Commodities and Capabilities. *Amsterdam: North-Holland*.
- [143] **Serfling, R.J.** (1980). Aproximation Theorems of Mathematical Statistics. *Wiley, New York*.
- [144] **Shao, J.; Tu, D.** (1995). The Jackknife and Bootstrap. *New York, Springer*.
- [145] **Shao, J.; Wu, C.F.J.** (1989). A General Theory for Jackknife Variance Estimation. *The Annals of Statistics*, 9, pp. 1187-1195.

Referencias

- [146] **Shao, J.; Wu, C.F.J.** (1992). Asymptotic Properties of the Balanced Repeated Replication Method for Sample Quantiles. *The Annals of Statistics*, 20, 1571-1593.
- [147] **Silber, J.** (2007). Measuring Poverty: Taking a Multidimensional Perspective. *E Hacienda Publica Española*, 182 3m, pp. 29-73.
- [148] **Silva, P.L.D.; Skinner, C.J.** (1995). Estimating Distribution Function With Auxiliary Information Using Poststratification. *Journal of Official Statistics*, 11, pp. 277-294.
- [149] **Singh, H.P.; Singh, S.; Kozak, M.** (2008). A Family of Estimators of finite-Population Distribution Functions Using Auxiliary Information. *Acta Appl. Math.*, 104, pp. 115-130.
- [150] **Singh, S.; Joarder, A.H.; Tracy, D.S.** (2001). Median Estimation Using Double Sampling. *Aust. NZ J. Stat.*, 43, pp. 33-46.
- [151] **Singh, S.** (2003). Advanced Sampling Theory With Applications: How Michael Selected Amy. *E Kluwer Academic Publishers, The Netherlands*.
- [152] **Singh, S.; Singh, H.P.; Upadhyaya, L.N.** (2007). Chain Ratio and Regression Type Estimators for Median Estimation in Survey Sampling. *Stat. Pap.*, 48, pp. 23-46.
- [153] **Singh, S.; Kim, J.M.; Jang, H.; Horn S.** (2011). Truncated Midzuno-Sen Sampling Schemes for Estimating Distribution Functions. *N Commun. Stat - Simul. C.*, 40, pp. 1096-1110.
- [154] **Stine, R.** (1989). An Introduction to Bootstrap Methods: Examples and Ideas. *Sociological Methods and Research*, 18, pp. 243-291.
- [155] **Thompson, B.S.** (2013). Empirical Likelihood-Based Inference for Poverty Measures with Relative Poverty Lines. *Econometric Reviews*, 32, pp. 513-523.
- [156] **Thon, D.** (1979). On Measuring Poverty. *Review of Income and Wealth*, 25, pp. 429-439.
- [157] **Townsend, P.** (1992). The International Analysis of Poverty. *Hemel Hempstead. Eng.: Harvester-Wheatsheaf*.
- [158] **Trabold-Nübler, H.** (1991). The Human Development Index - A New Development Indicator?. *Intereconomics*, 26 (5), pp. 236-243.

Referencias

- [159] **UNDP** (1997). Human Development Report 1997. *Oxford University Press, New York*
- [160] **Ureña, C.** (1999). Contraste entre Medidas Objetivas y Subjetivas de Pobreza. *Reunión del Grupo Río. Lisboa.*
- [161] **Van Kerm, P.** (2007). Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC. *IRISS Working Paper Series No. 2007-01: CEPS-Instead, 51p.*
- [162] **Vaughan, R.N.** (1987). Welfare Approaches to the Measurement of Poverty. *The Economic Journal, 97*, pp. 160-170.
- [163] **Vietnam, General Statistical Office** (2000). Vietnam Living Standards Survey 1997-1998. *Hanoi: Statistical Publishing House.*
- [164] **Vollset, S.E.** (1993). Confidence Intervals for a Binomial Proportion. *Statistics in Medicine, 12*, pp. 809-824.
- [165] **Wang, S.; Dorfman, A.H.** (1996). A New Estimator for the Finite Population Distribution Function. *Biometrika, 83*, pp. 639-652.
- [166] **Watts, H.** (1968). An Economic Definition of Poverty, en D. P. Moynihan. On Understanding Poverty. *Basic Books. Inc. New York*, pp. 316-329
- [167] **Wilson, E.B.** (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association, 22*, pp. 209-212.
- [168] **Wolter, K.M.** (2007). Introduction to Variance Estimation. *Second Edition. Springer. New York.*
- [169] **World Bank**, (2000). World Development Report 2000-2001: Attacking Poverty. *Washington, D.C.: World Bank.*
- [170] **Wu, C.** (2002). Optimal Calibration Estimators in Survey Sampling. *Working paper 2002-01. Department of Statistics and Actuarial Science, University of Waterloo.*
- [171] **Wu, C.** (2003). Weighted Empirical Likelihood Inference. *Working Paper 2003-01. Department of Statistics and Actuarial Science, University of Waterloo.*
- [172] **Yitzhaki, S.** (1991). Calculating Jackknife Variance Estimators for Parameters of the Gini Method. *Journal of Business and Economic Statistics, 9*, pp. 235-239.

Referencias

- [173] **Zhang, L.C.; Chambers, R.L.** (2004). Small Area Estimates for Cross-Classifications. *Journal of the Royal Statistical Society: Series B*, 66, pp. 479-496.
- [174] **Zheng, B.** (1993). An Axiomatic Characterization of the Watts Poverty Index. *Economics Letters*, 43 (1), pp. 81-86.
- [175] **Zheng, B.** (2001). Statistical Inference for Poverty Measure with Relative Poverty Lines. *Journal of Econometrics*, 101, 337-356.

Apéndice 1

Descripción de las poblaciones en estudio

El objetivo de este apéndice es proporcionar información adicional sobre las poblaciones consideradas en esta memoria para los distintos estudios de simulación Monte Carlo que se han llevado a cabo. A modo de resumen, recordamos que a lo largo de esta memoria se han utilizado las siguientes poblaciones:

- Poblaciones basadas en datos reales de la EU-SILC.
- Poblaciones artificiales basadas en la distribución Gamma.
- Poblaciones artificiales basadas en la distribución Singh-Maddala.

Destacamos que en el caso de los datos procedentes de la EU-SILC, tales datos muestrales se consideran como una población en los estudios de simulación Monte Carlo que se llevan a cabo en esta tesis. Es decir, a partir de esta “población”, con tamaño N , se extraen las distintas muestras con el fin de obtener las estimaciones mediante las distintas metodologías. Por tanto, cuando se habla en estos casos de un parámetro o una medida de pobreza a nivel poblacional, no nos referimos al valor de dicha medida a toda la población del país correspondiente, sino que nos referimos a los N datos que hemos considerado como población en nuestros estudios de simulación Monte Carlo. En otras palabras, las medidas de pobreza consideradas en este apéndice se refieren a las medidas de pobreza que se obtendrían en las distintas “poblaciones” basadas en datos muestrales reales que hemos considerado en los varios estudios de simulación de este trabajo. Destacamos que en el Capítulo 1 de esta tesis se han definido las distintas medidas de pobreza consideradas en

Apéndice I. Descripción de las poblaciones en estudio

este trabajo y se ha aportado numerosas representaciones gráficas con los datos de pobreza correspondientes a regiones y países de la Unión Europea y obtenidos a través de la EU-SILC.

Para cada población, se obtiene en este apéndice la siguiente información estadística y de carácter económico:

- ρ : El coeficiente de correlación lineal poblacional entre la variable principal y la variable auxiliar.
- N : El tamaño de la población, y a partir del cual se seleccionan las muestras para los estudios de simulación Monte Carlo.
- P_0 : La proporción de pobres a nivel poblacional. En el caso de poblaciones artificiales, esta proporción se ha obtenido siguiendo la metodología propuesta por EUROSTAT, es decir, P_0 es la proporción de individuos cuyos valores de la variable principal se encuentran por debajo del 60% de la mediana poblacional de dicha variable principal.
- P_1 : El índice de brecha de pobreza poblacional.
- P_2 : El índice de pobreza severa poblacional. Debemos tener en consideración para P_1 y P_2 el mismo argumento discutido para el parámetro P_0 .
- g : El coeficiente de asimetría de la variable principal.

Además, para cada población también se obtienen diagramas de dispersión e histogramas. A partir de estos gráficos se puede apreciar la presencia de outliers o datos anómalos. Tales datos pueden tener un efecto importante en las distintas medidas anteriormente comentadas, e incluso en los distintos gráficos. Por esta razón, tanto las medidas estadísticas y económicas anteriormente mencionadas como las representaciones gráficas se han obtenido después de eliminar los outliers. Con este proceso se persigue poder describir mejor las poblaciones en estudio. En algunos casos podemos observar que la eliminación de outliers tiene un impacto significativo.

Poblaciones basadas en datos reales de la EU-SILC

A continuación se describen mediante medidas estadísticas y económicas las poblaciones basadas en datos reales de la EU-SILC. Las medidas pueden consultarse en la Tabla A.1, mientras que los distintos diagramas de dispersión e histogramas derivados de estas poblaciones pueden consultarse en la Figura A.1 hasta la Figura A.21. En el caso de presencia de outliers se puede observar que se repiten tanto medidas como representaciones gráficas después de eliminar de la población los mencionados outliers. Por último, destacamos que la presencia de outliers afecta principalmente al coeficiente de asimetría, tal como puede observarse en la Tabla A.1. El coeficiente de correlación lineal sufre una ligera disminución después de eliminar los outliers. El resto de análisis descriptivos referidos a las poblaciones artificiales consideradas en este trabajo se describen justo a continuación de este epígrafe.

Apéndice I. Descripción de las poblaciones en estudio

País	ρ	N	P_0	P_1	P_2	g
Italia	0.88	27553	18.9	6.2	3.5	23.1
	<i>0.81</i>	<i>27550</i>	<i>18.9</i>	<i>6.2</i>	<i>3.5</i>	<i>5.3</i>
Eslovenia	0.83	17612	12.2	2.7	1.0	2.6
Reino Unido	0.83	10586	17.2	5.1	2.7	20.5
	<i>0.82</i>	<i>10585</i>	<i>17.2</i>	<i>5.1</i>	<i>2.7</i>	<i>6.3</i>
Polonia	0.82	22113	17.7	4.8	2.1	11.0
	<i>0.77</i>	<i>22107</i>	<i>17.7</i>	<i>4.8</i>	<i>2.1</i>	<i>4.5</i>
Bulgaria	0.70	12469	22.1	6.4	2.8	18.3
	<i>0.66</i>	<i>12465</i>	<i>22.1</i>	<i>6.4</i>	<i>2.8</i>	<i>2.7</i>
Lituania	0.66	8421	18.7	6.0	3.2	5.6
España	0.65	28210	20.3	8.0	5.7	1.9
Bélgica	0.65	9305	14.6	3.7	2.3	51.6
	<i>0.48</i>	<i>9304</i>	<i>14.6</i>	<i>3.7</i>	<i>2.3</i>	<i>8.2</i>

Tabla A.1. Valores de ρ (coeficiente de correlación lineal), N (tamaño de la población), P_0 (proporción de pobres), P_1 (índice de brecha de pobreza), P_2 (índice de pobreza severa) y g (coeficiente de asimetría) para distintos países de la EU-SILC. Los datos en cursiva indican que se han eliminado los outliers.

Apéndice I. Descripción de las poblaciones en estudio

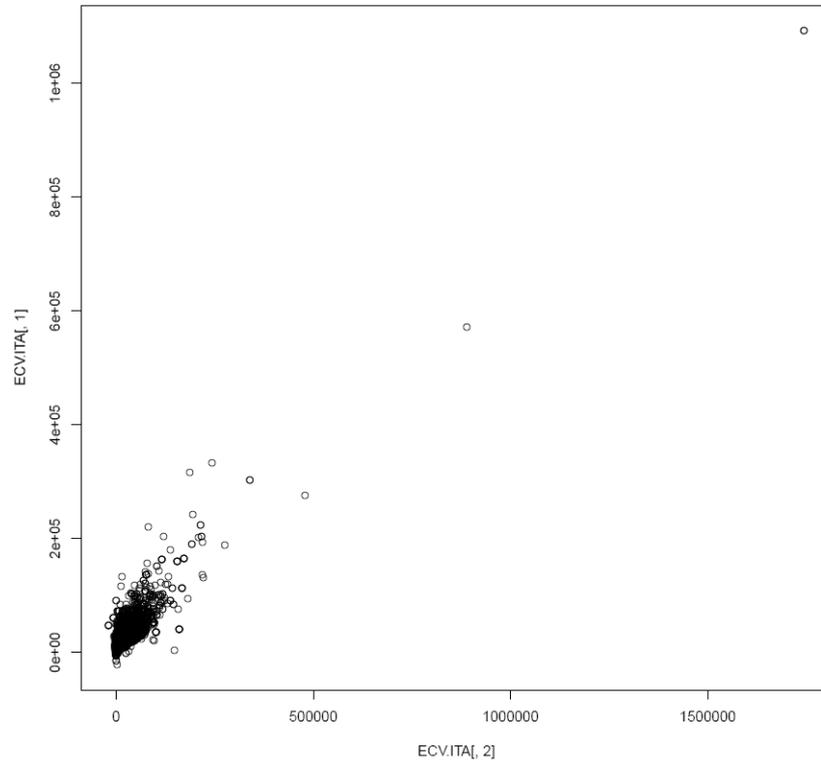


Figura A.1. Diagrama de dispersión de los datos de la EU-SILC en Italia.

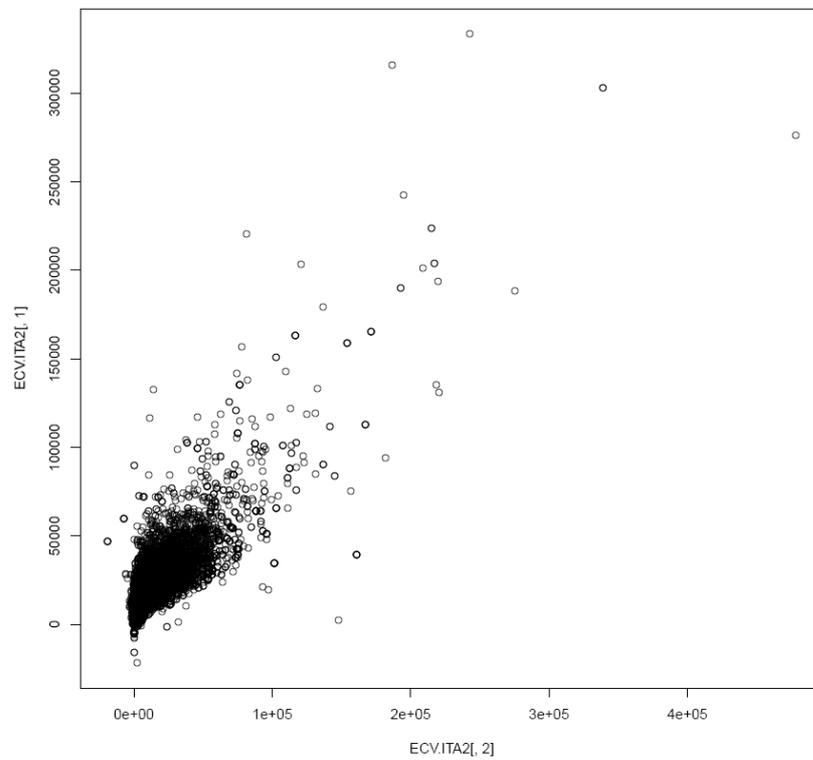


Figura A.2. Diagrama de dispersión de los datos de la EU-SILC en Italia y sin outliers.

Apéndice I. Descripción de las poblaciones en estudio

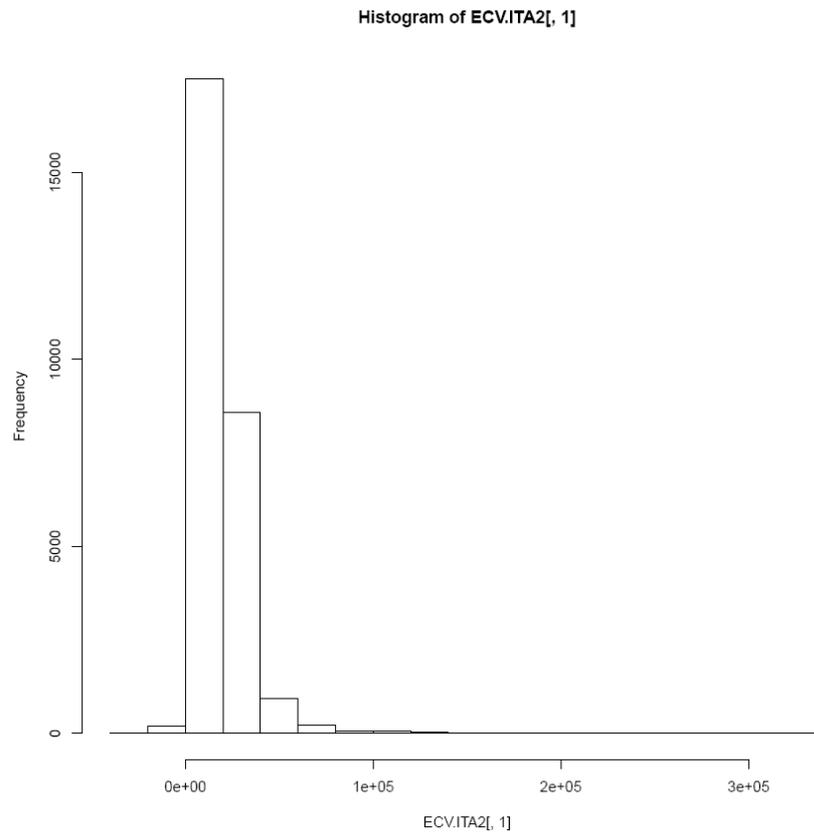


Figura A.3. Histograma de los datos de la EU-SILC en Italia y sin outliers.

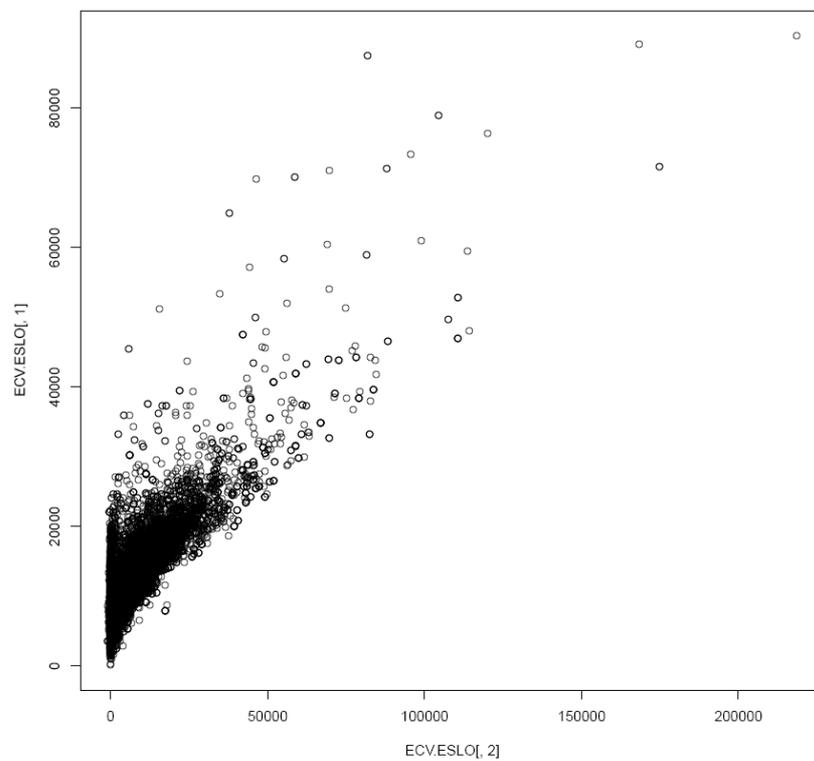


Figura A.4. Diagrama de dispersión de los datos de la EU-SILC en Eslovenia.

Apéndice I. Descripción de las poblaciones en estudio

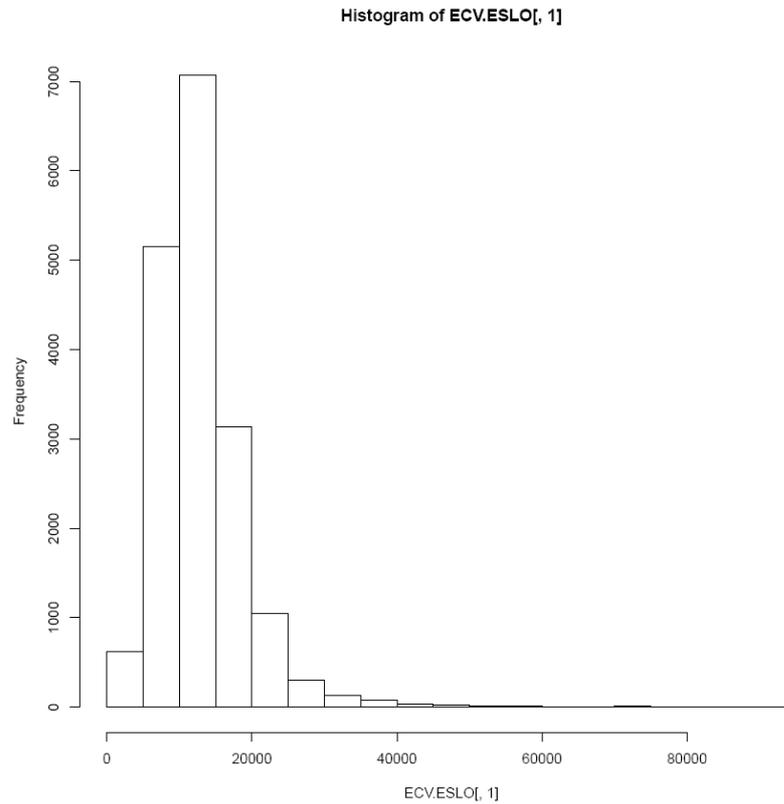


Figura A.5. Histograma de los datos de la EU-SILC en Eslovenia.

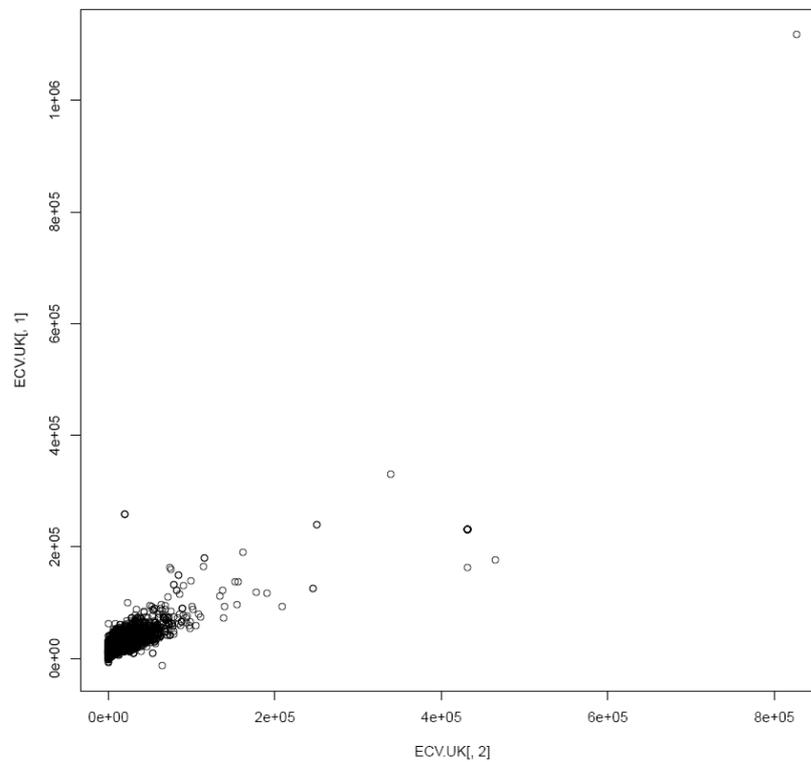


Figura A.6. Diagrama de dispersión de los datos de la EU-SILC en el Reino Unido.

Apéndice I. Descripción de las poblaciones en estudio

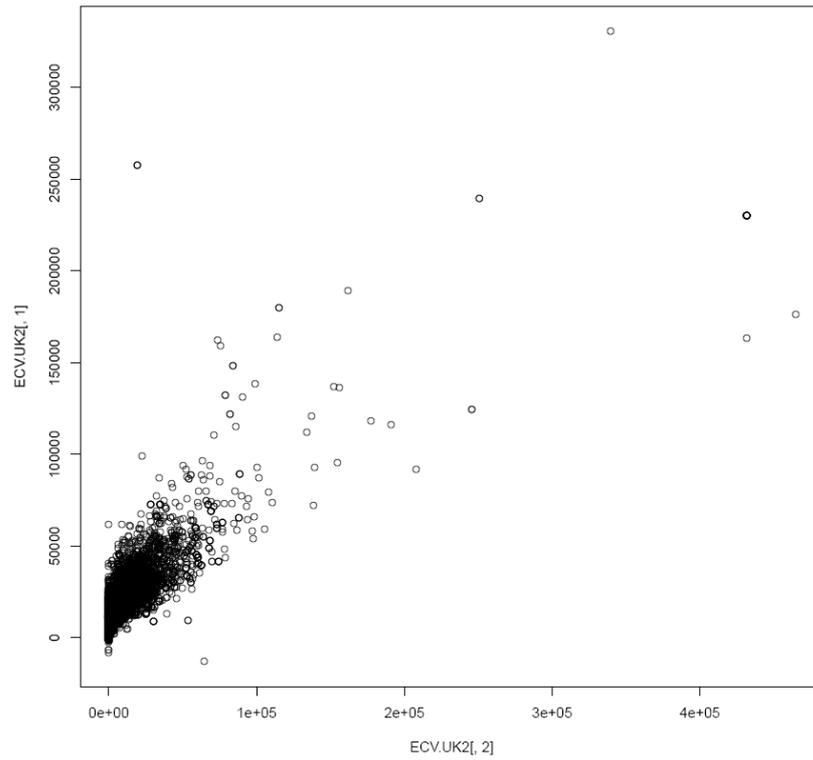


Figura A.7. Diagrama de dispersión de los datos de la EU-SILC en el Reino Unido y sin outliers.

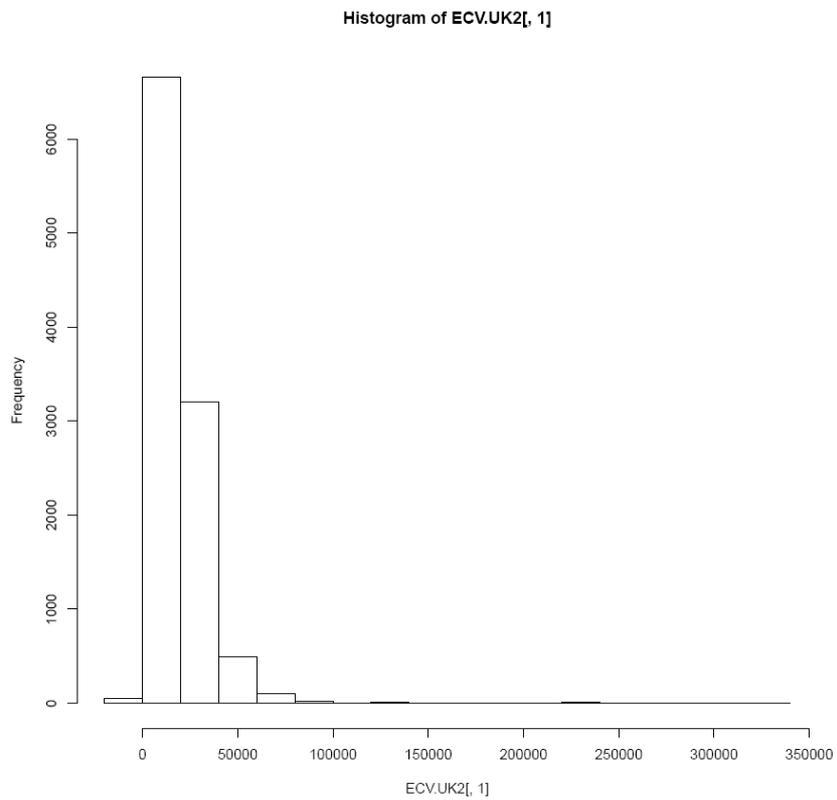


Figura A.8. Histograma de los datos de la EU-SILC en el Reino Unido y sin outliers.

Apéndice I. Descripción de las poblaciones en estudio

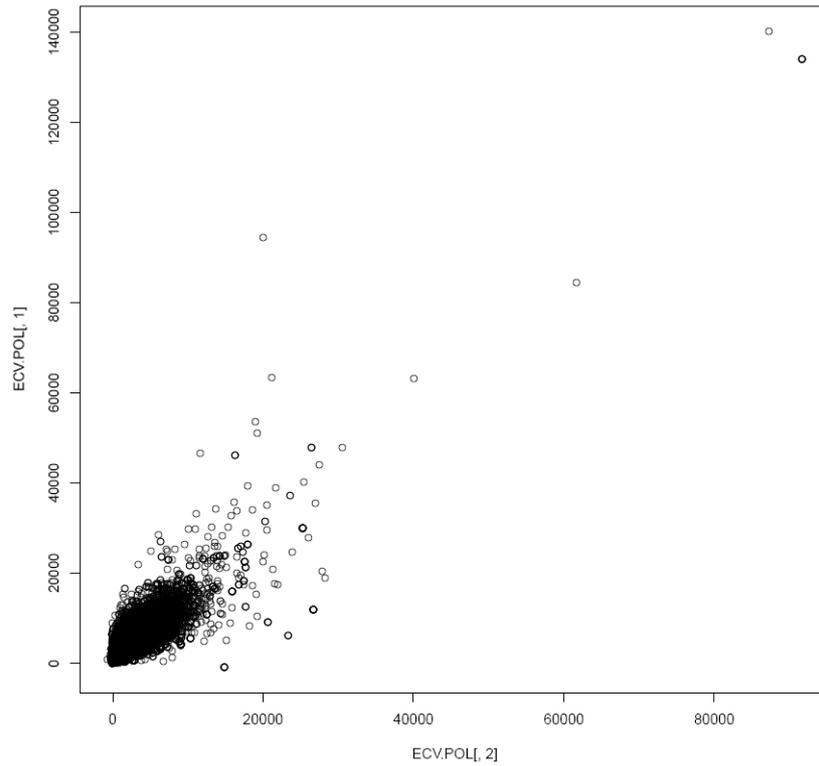


Figura A.9. Diagrama de dispersión de los datos de la EU-SILC en Polonia.

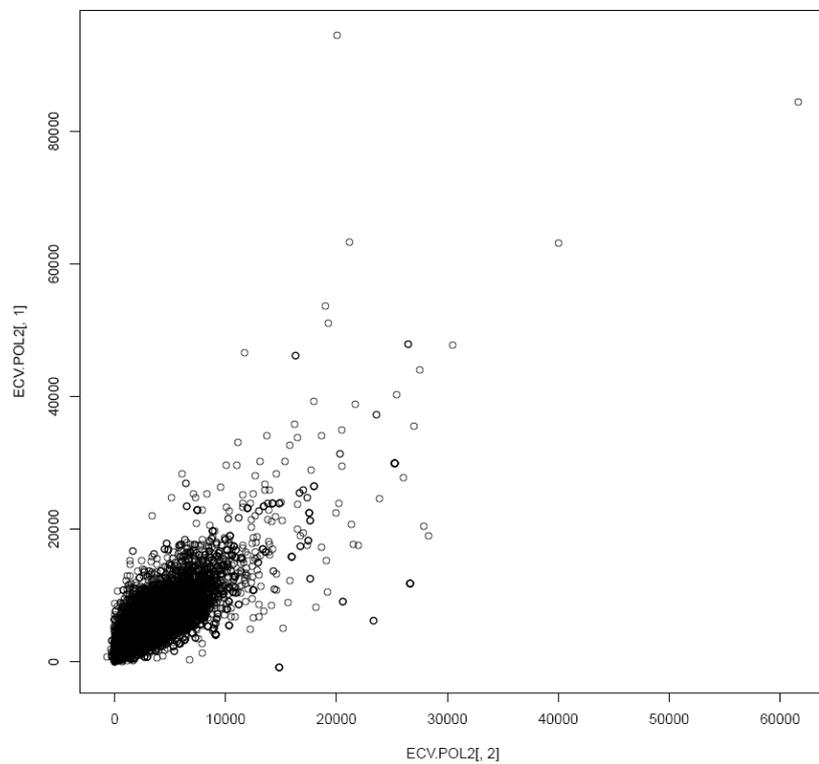


Figura A.10. Diagrama de dispersión de los datos de la EU-SILC en Polonia y sin outliers.

Apéndice I. Descripción de las poblaciones en estudio

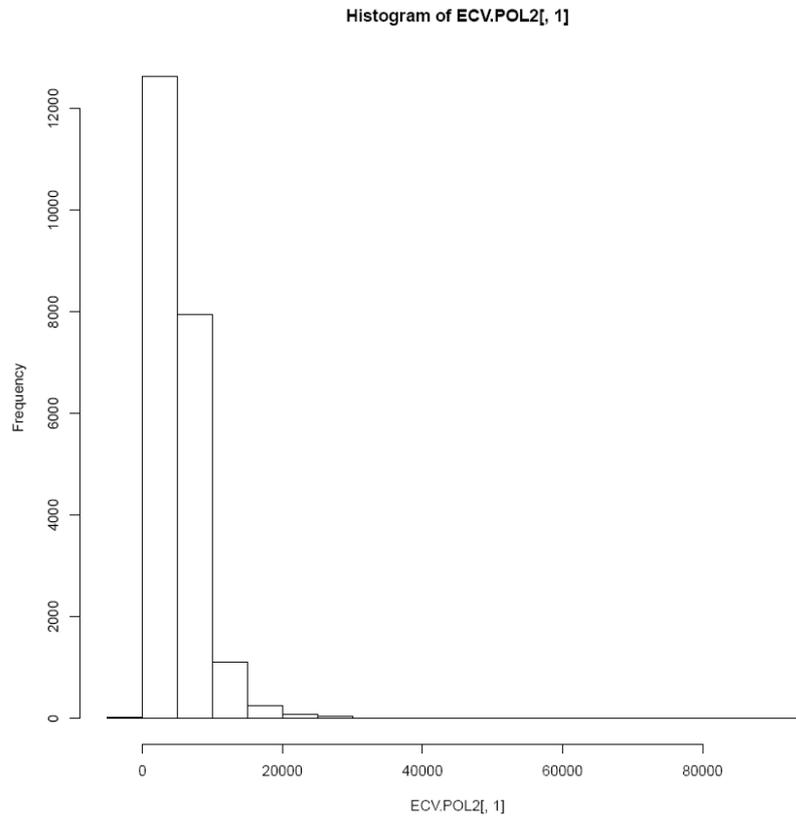


Figura A.11. Histograma de los datos de la EU-SILC en Polonia y sin outliers.

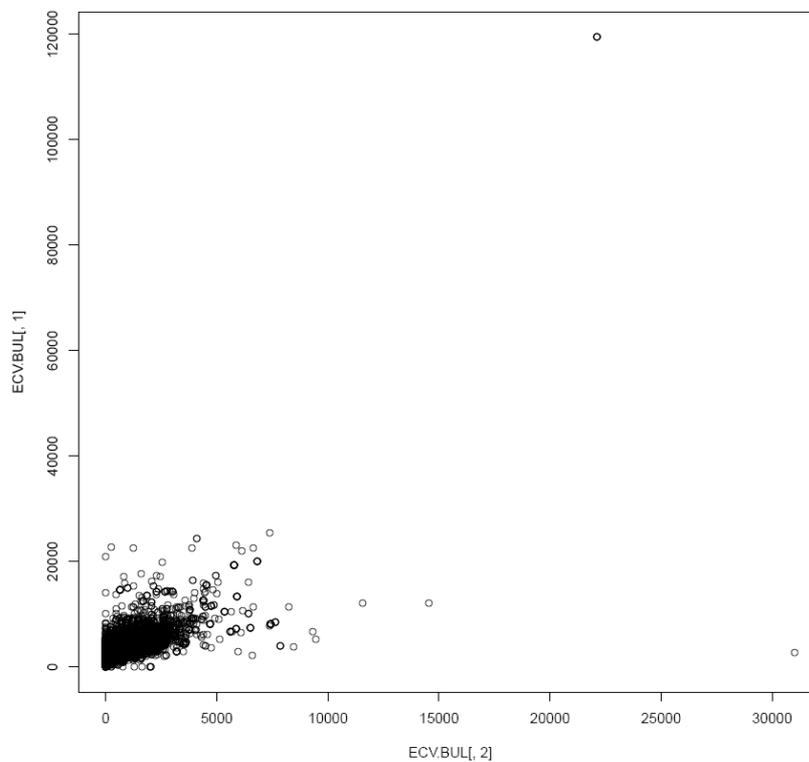


Figura A.12. Diagrama de dispersión de los datos de la EU-SILC en Bulgaria.

Apéndice I. Descripción de las poblaciones en estudio

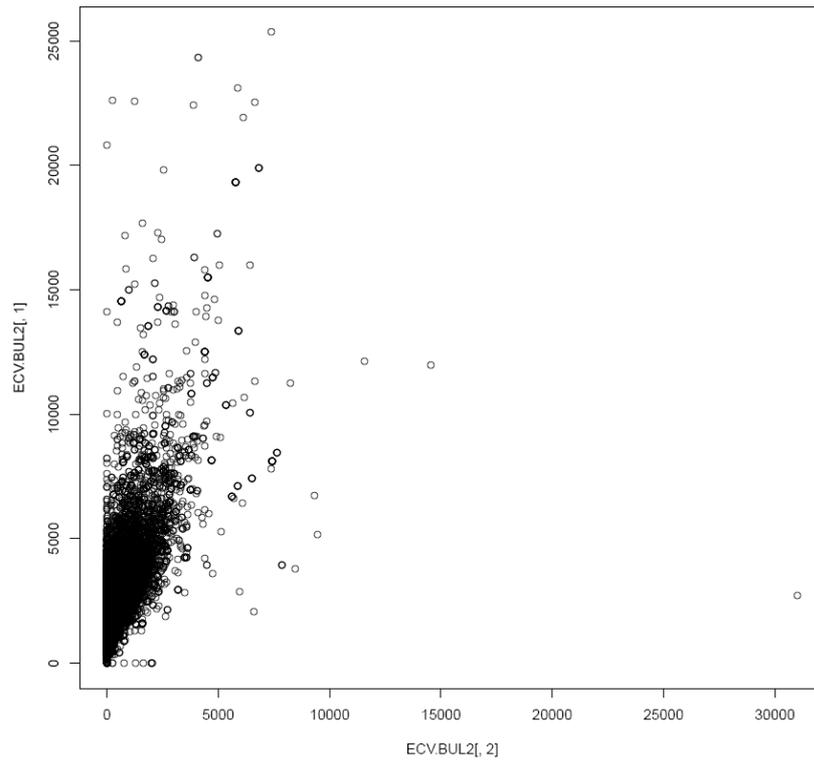


Figura A.13. Diagrama de dispersión de los datos de la EU-SILC en Bulgaria y sin outliers.

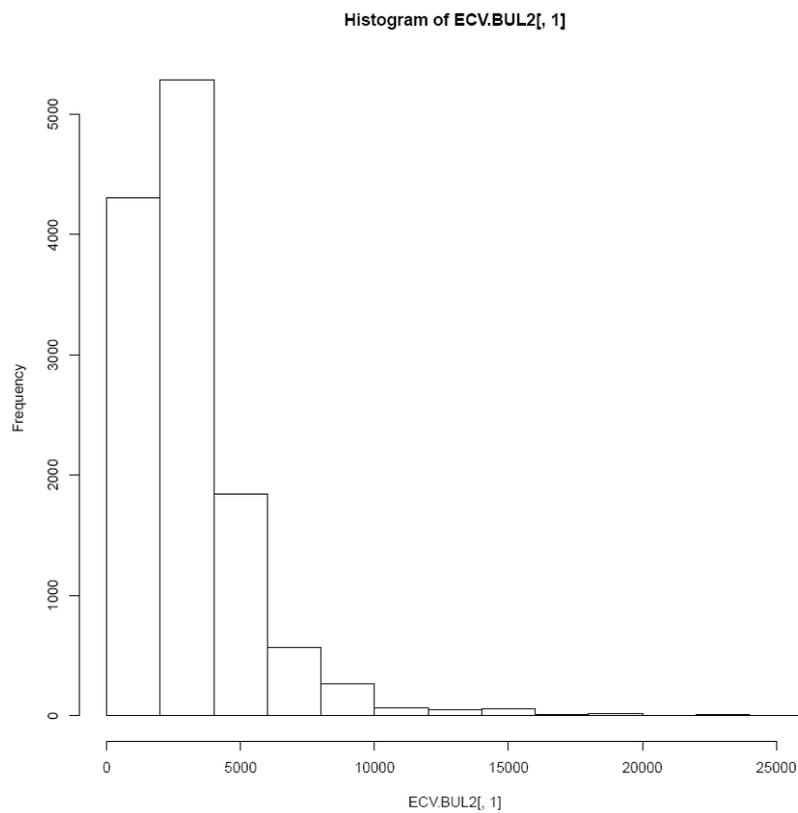


Figura A.14. Histograma de los datos de la EU-SILC en Bulgaria y sin outliers.

Apéndice I. Descripción de las poblaciones en estudio

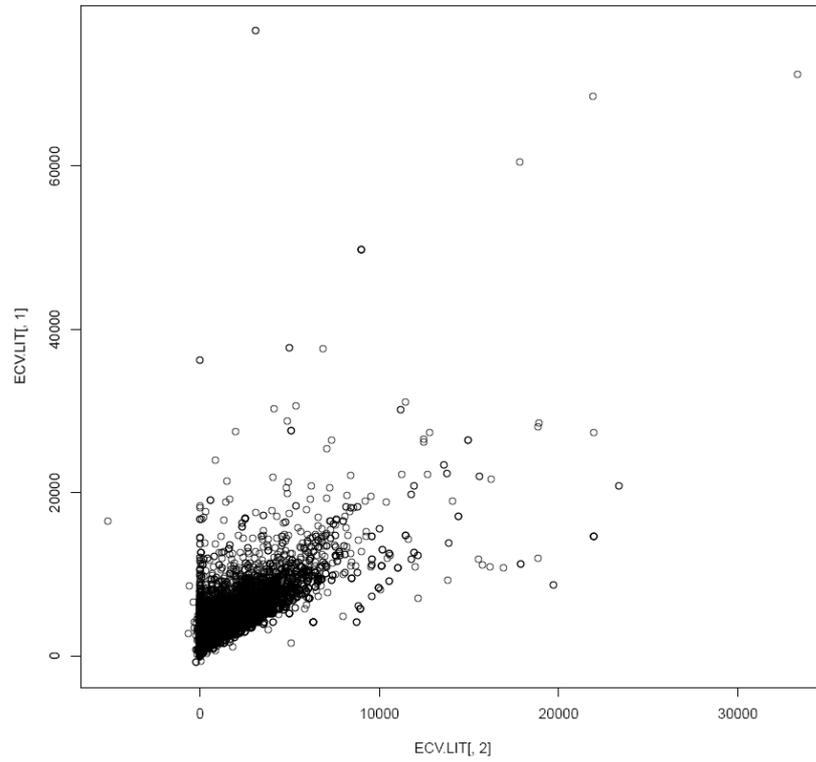


Figura A.15. Diagrama de dispersión de los datos de la EU-SILC en Lituania.

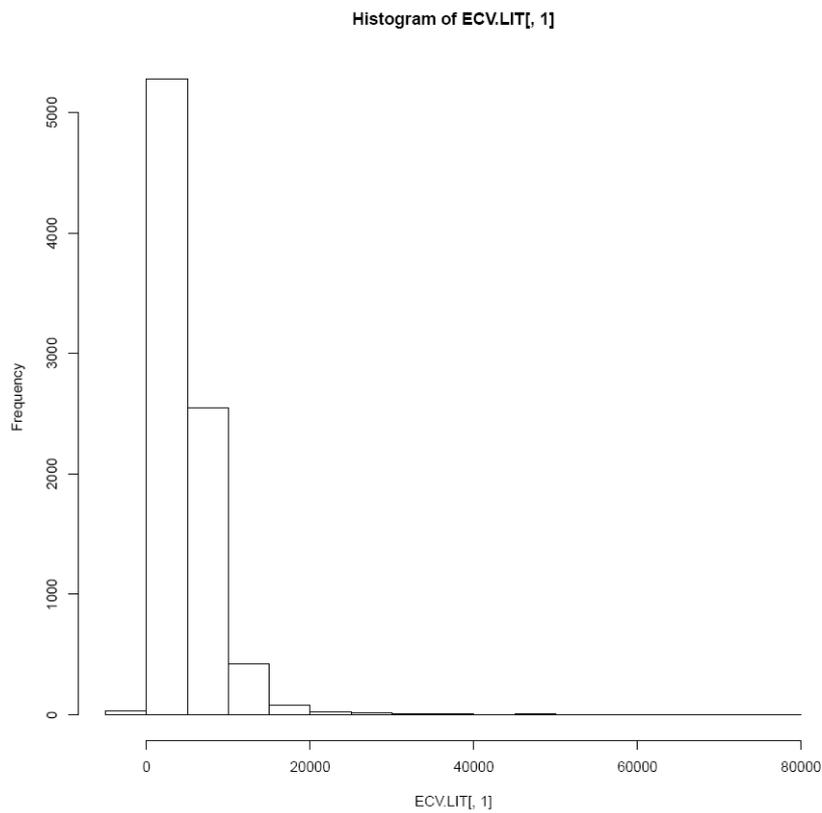


Figura A.16. Histograma de los datos de la EU-SILC en Lituania

Apéndice I. Descripción de las poblaciones en estudio

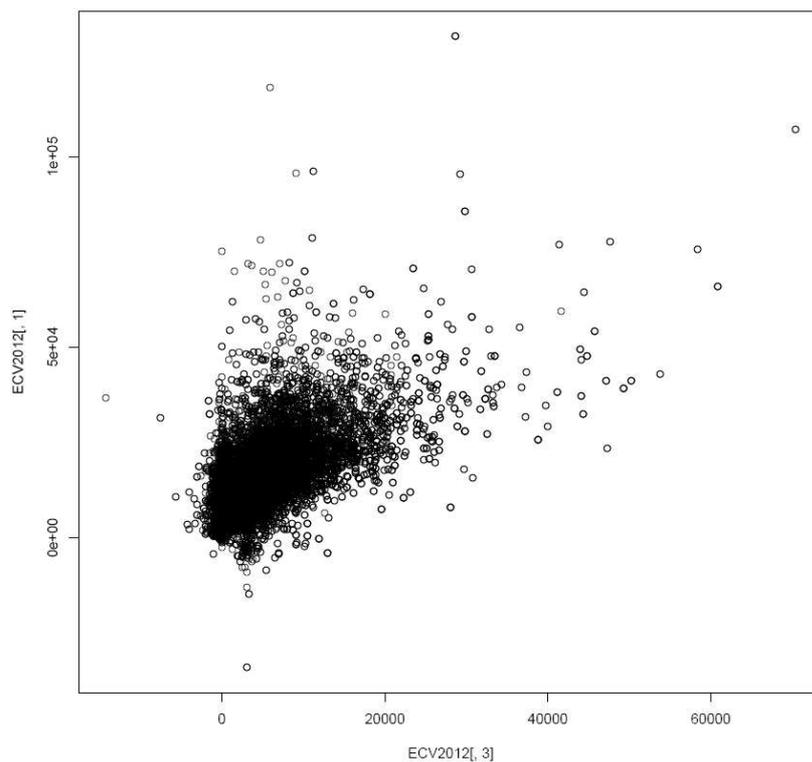


Figura A.17. Diagrama de dispersión de los datos de la EU-SILC en España.

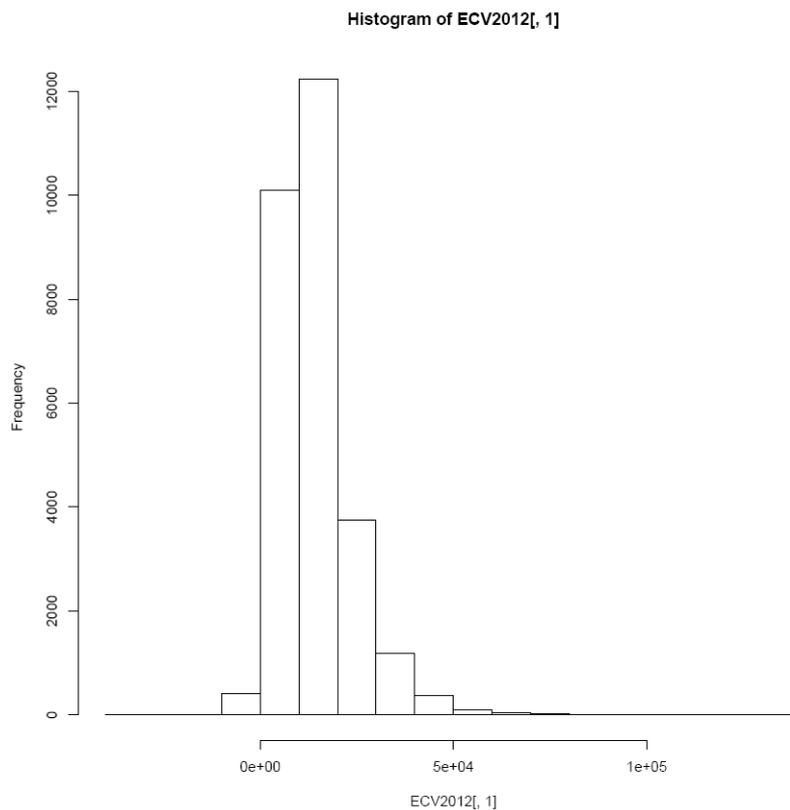


Figura A.18. Histograma de los datos de la EU-SILC en España.

Apéndice I. Descripción de las poblaciones en estudio

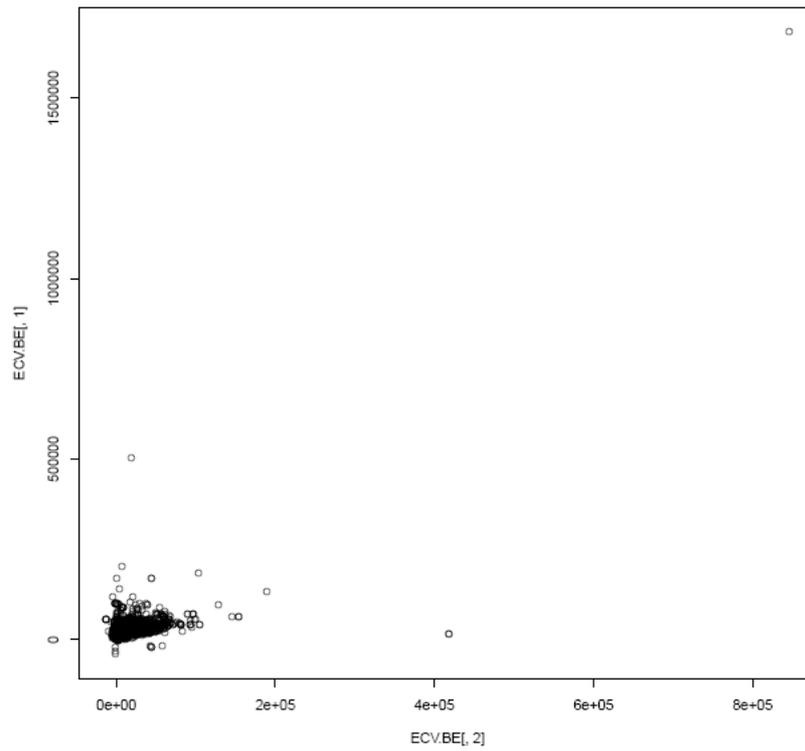


Figura A.19. Diagrama de dispersión de los datos de la EU-SILC en Bélgica.

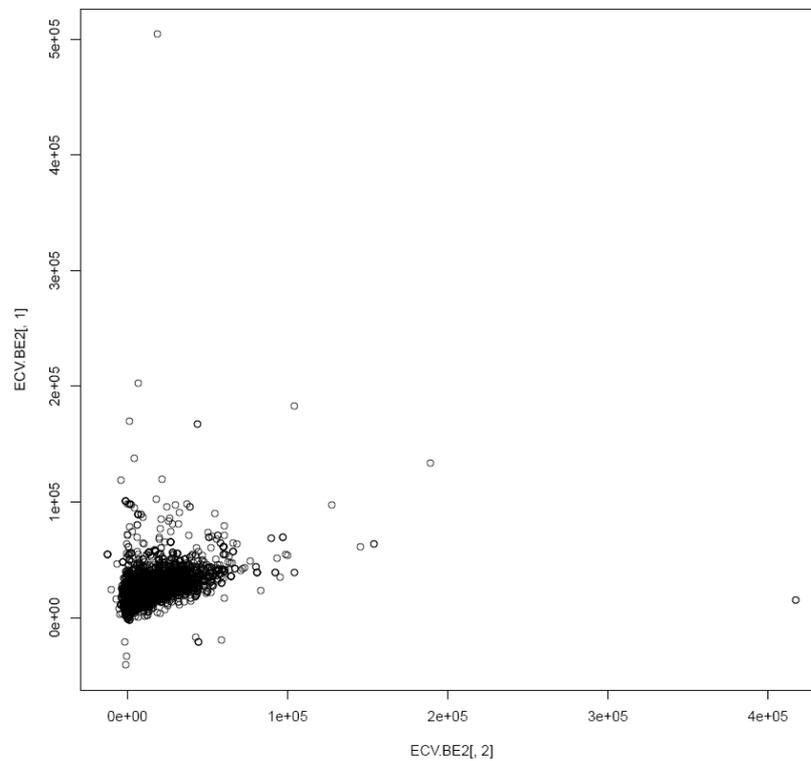


Figura A.20. Diagrama de dispersión de los datos de la EU-SILC en Bélgica y sin outliers.

Apéndice I. Descripción de las poblaciones en estudio

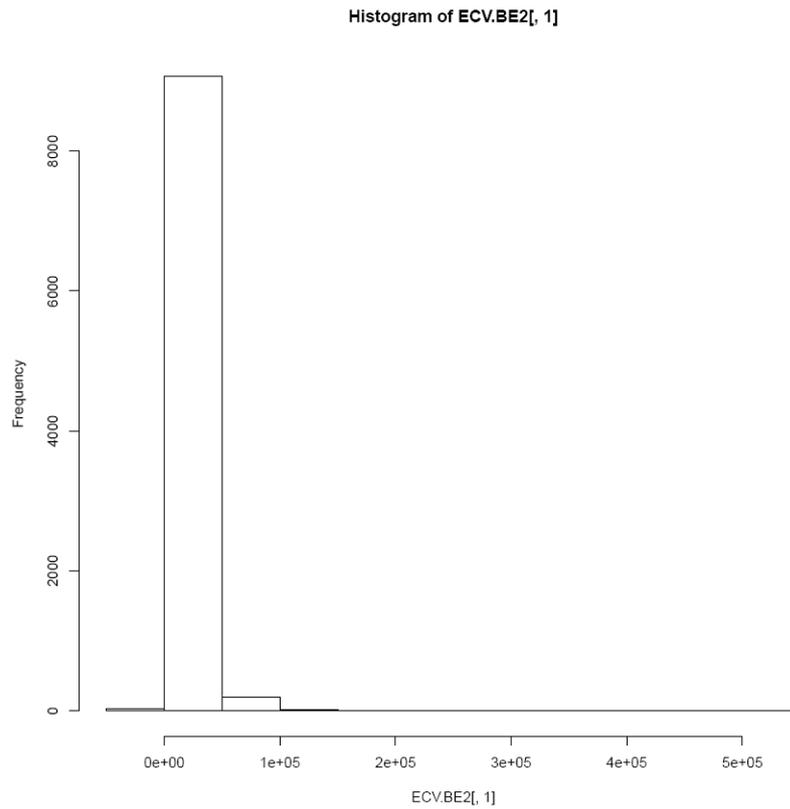


Figura A.21. Histograma de los datos de la EU-SILC en Bélgica y sin outliers.

Poblaciones artificiales basadas en la distribución Gamma

Apéndice I. Descripción de las poblaciones en estudio

P_0	ρ	P_1	P_2	g
15	0.5	3.6	1.3	0.86
	0.7	3.6	1.3	0.86
	0.9	3.5	1.2	0.93
25	0.5	8.9	4.5	1.22
	0.7	8.8	4.3	1.26
	0.9	8.9	4.5	1.25
35	0.5	19.6	14.0	2.12
	0.7	19.3	13.7	2.03
	0.9	19.7	14.0	2.20

Tabla A.2. Valores de ρ (coeficiente de correlación lineal), P_0 (proporción de pobres), P_1 (índice brecha de pobreza), P_2 (índice de pobreza severa) y g (coeficiente de asimetría) para distintas poblaciones artificiales basadas en la distribución Gamma. El tamaño de la población es $N = 10000$.

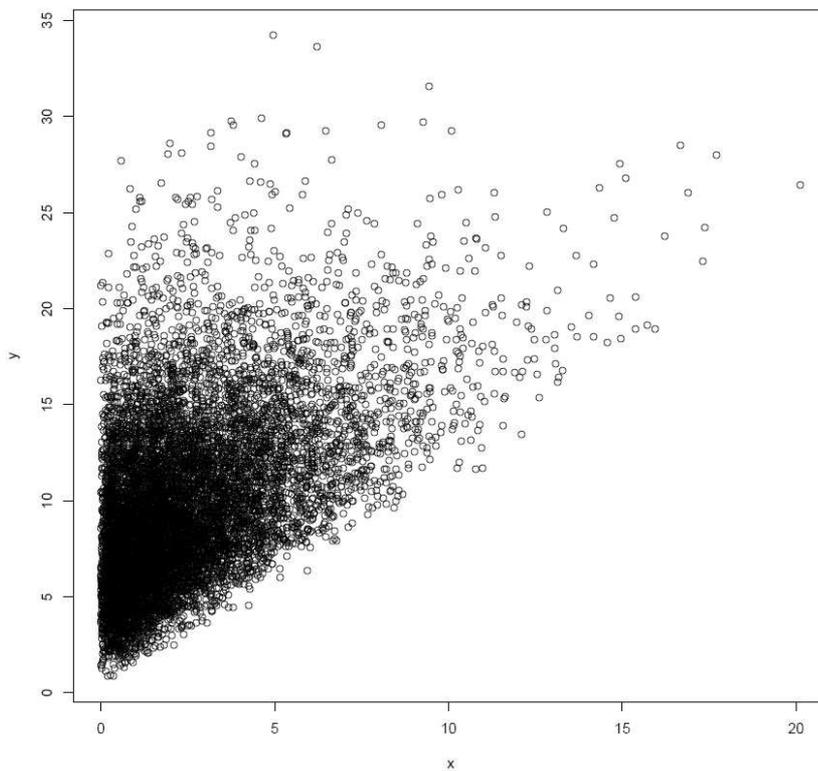


Figura A.22. Diagrama de dispersión de la población Gamma con $\rho = 0.5$ y $P = 0.15$.

Apéndice I. Descripción de las poblaciones en estudio

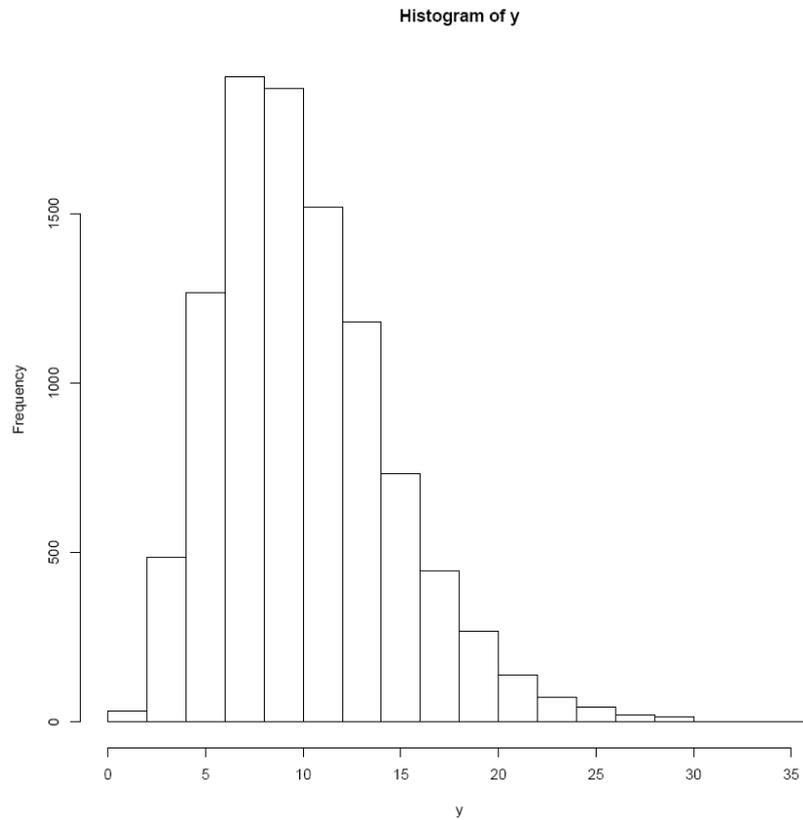


Figura A.23. Histograma de la población Gamma con $\rho = 0.5$ y $P_0 = 0.15$.

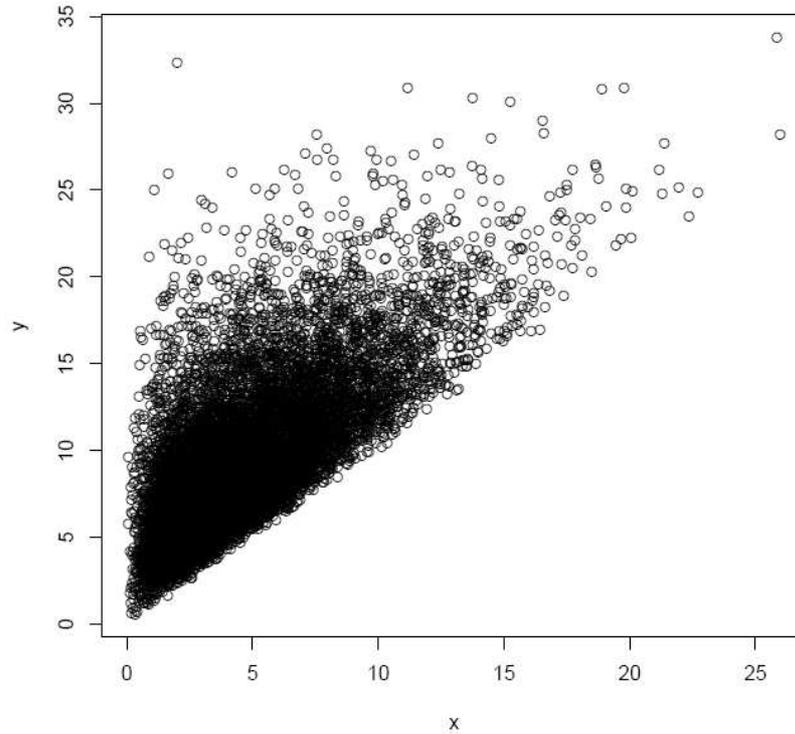


Figura A.24. Diagrama de dispersión de la población Gamma con $\rho = 0.7$ y $P_0 = 0.15$.

Apéndice I. Descripción de las poblaciones en estudio

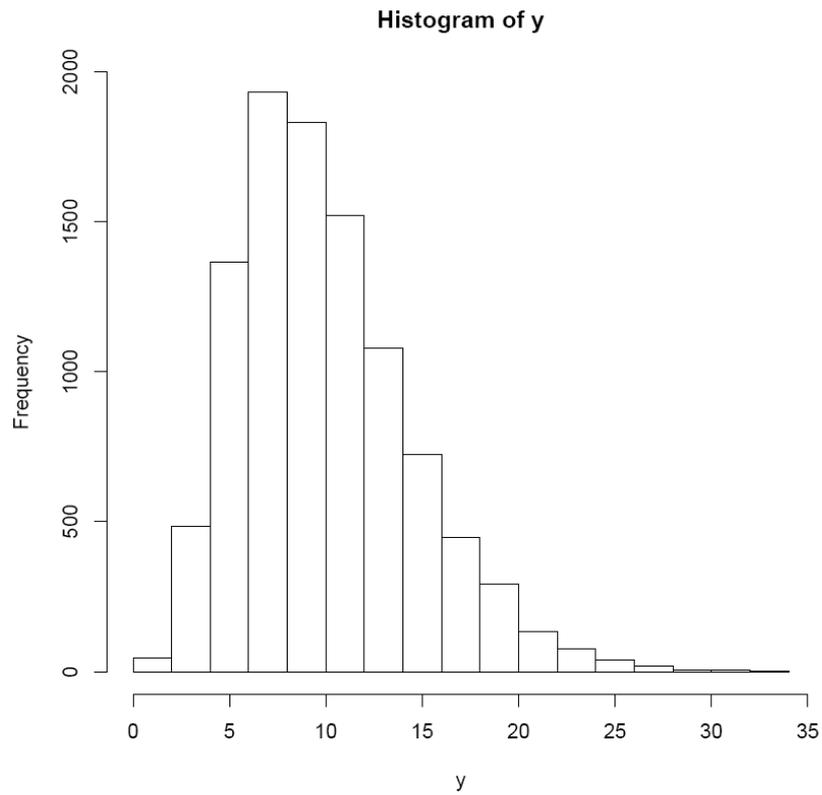


Figura A.25. Histograma de la población Gamma con $\rho = 0.7$ y $P_0 = 0.15$.

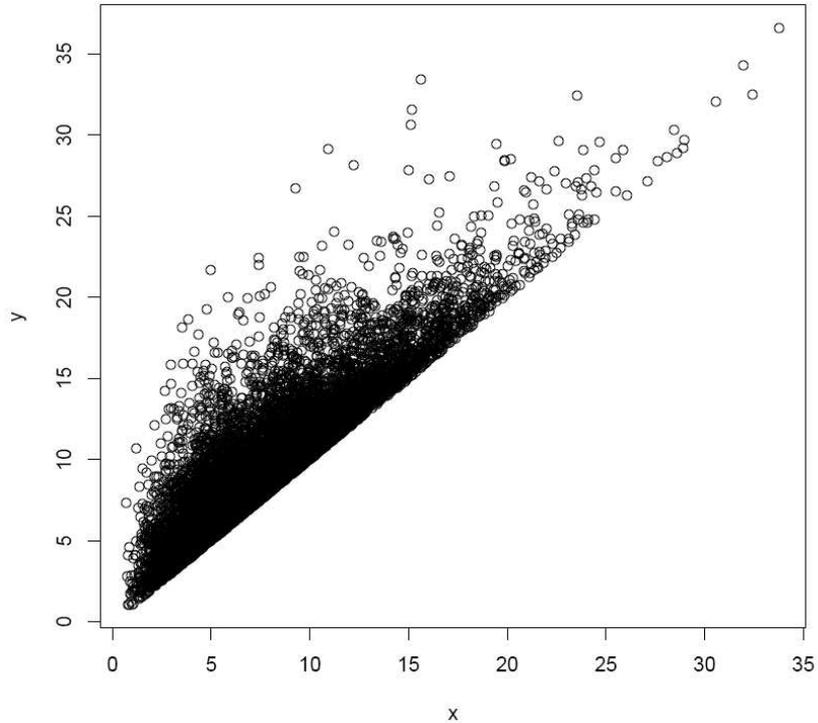


Figura A.26. Diagrama de dispersión de la población Gamma con $\rho = 0.9$ y $P_0 = 0.15$.

Apéndice I. Descripción de las poblaciones en estudio

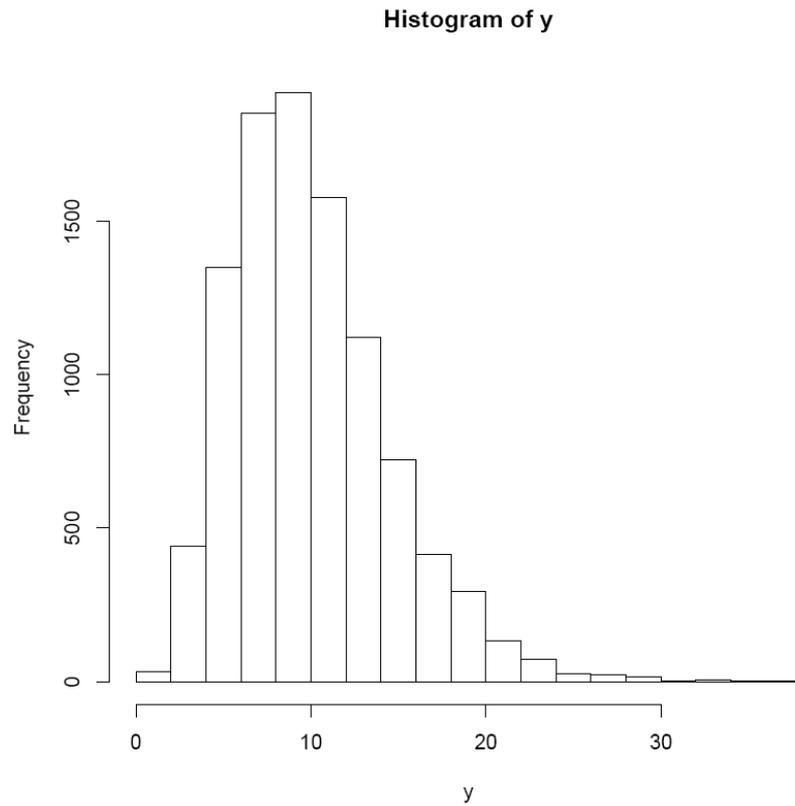


Figura A.27. Histograma de la población Gamma con $\rho = 0.9$ y $P_0 = 0.15$.

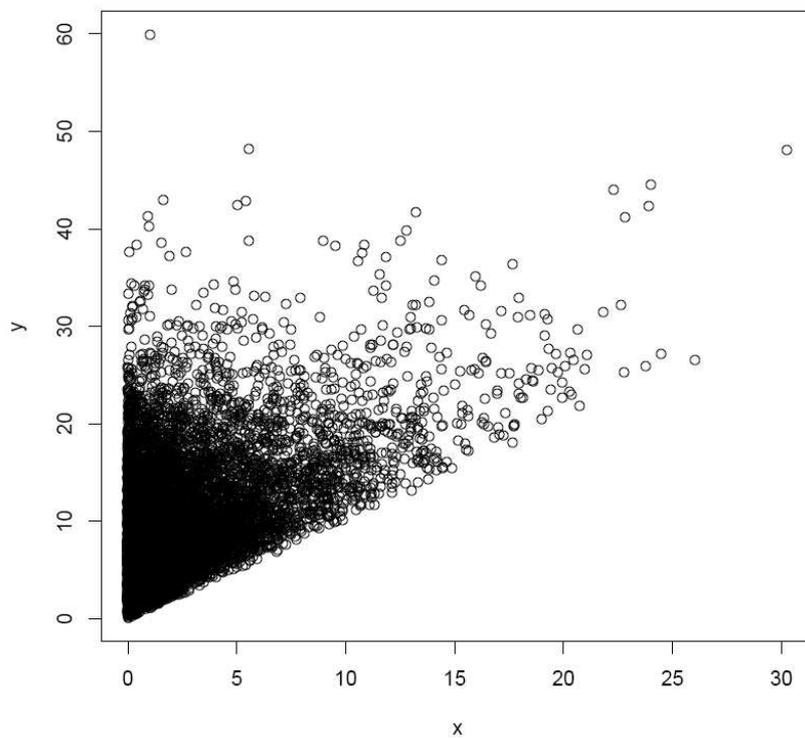


Figura A.28. Diagrama de dispersión de la población Gamma con $\rho = 0.5$ y $P_0 = 0.25$.

Apéndice I. Descripción de las poblaciones en estudio

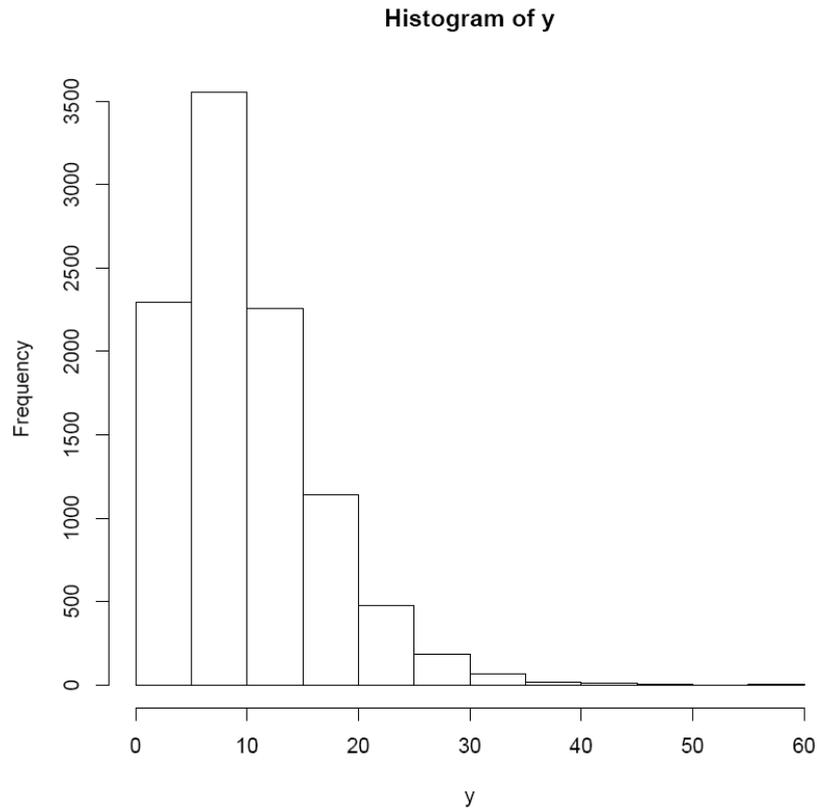


Figura A.29. Histograma de la población Gamma con $\rho = 0.5$ y $P_0 = 0.25$.

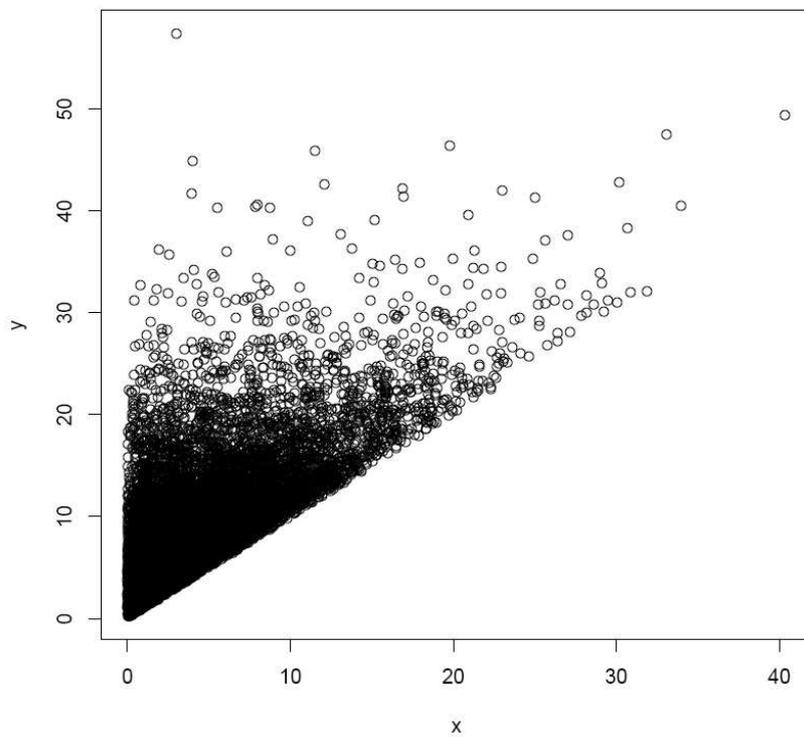


Figura A.30. Diagrama de dispersión de la población Gamma con $\rho = 0.7$ y $P_0 = 0.25$.

Apéndice I. Descripción de las poblaciones en estudio

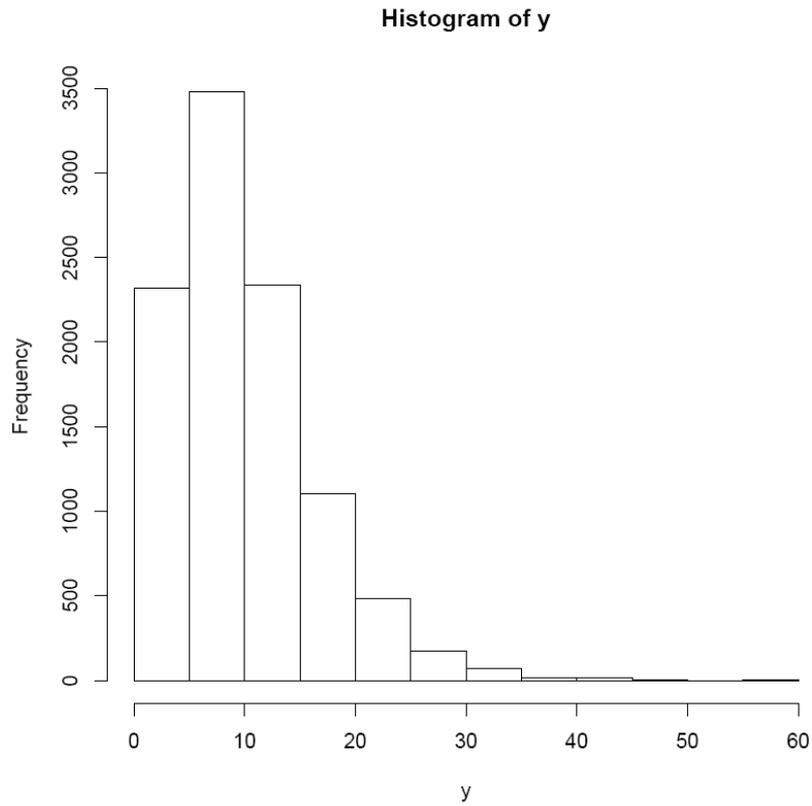


Figura A.31. Histograma de la población Gamma con $\rho = 0.7$ y $P_0 = 0.25$.

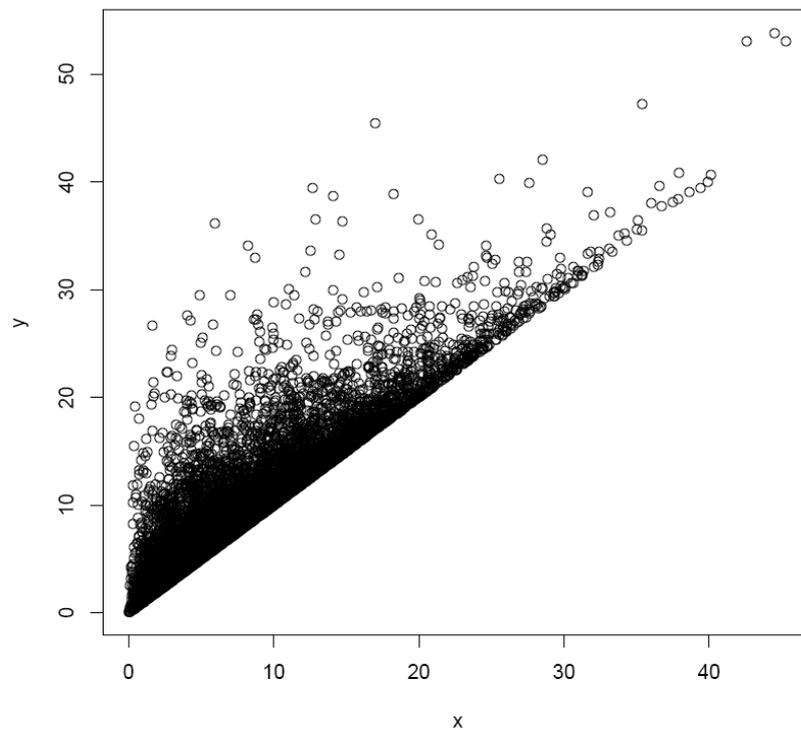


Figura A.32. Diagrama de dispersión de la población Gamma con $\rho = 0.9$ y $P_0 = 0.25$.

Apéndice I. Descripción de las poblaciones en estudio

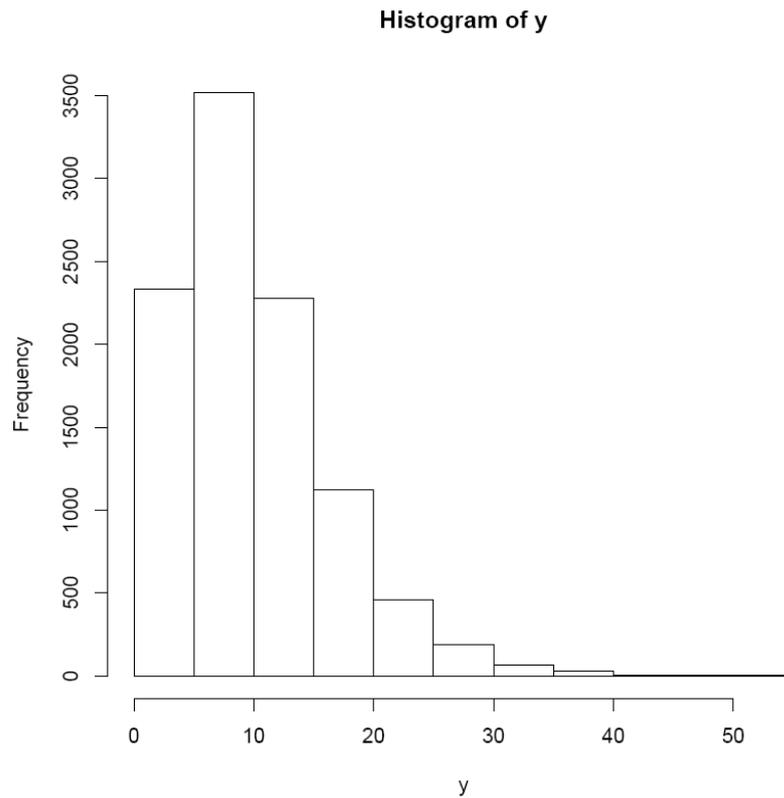


Figura A.33. Histograma de la población Gamma con $\rho = 0.9$ y $P_0 = 0.25$.

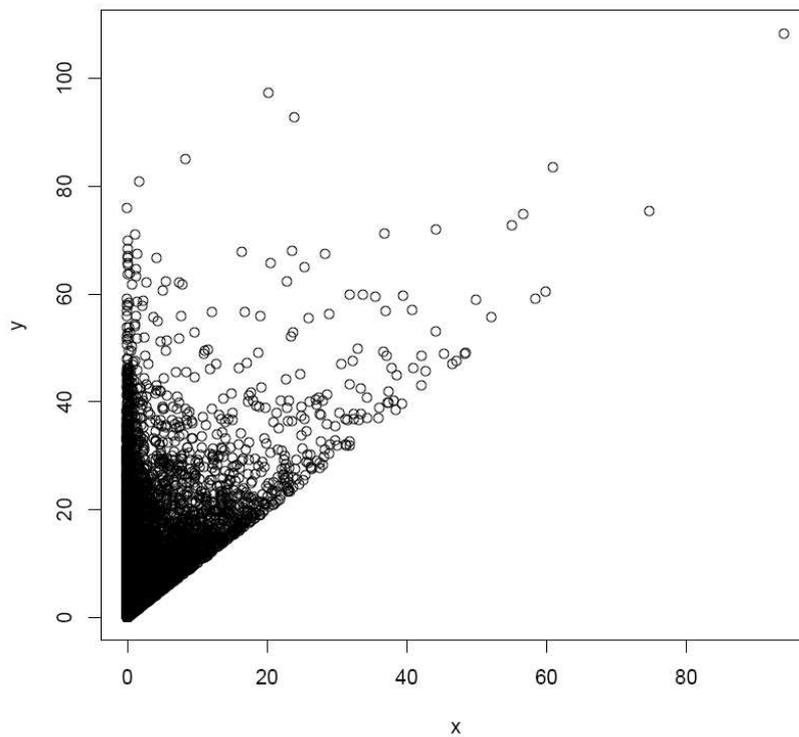


Figura A.34. Diagrama de dispersión de la población Gamma con $\rho = 0.5$ y $P_0 = 0.35$.

Apéndice I. Descripción de las poblaciones en estudio

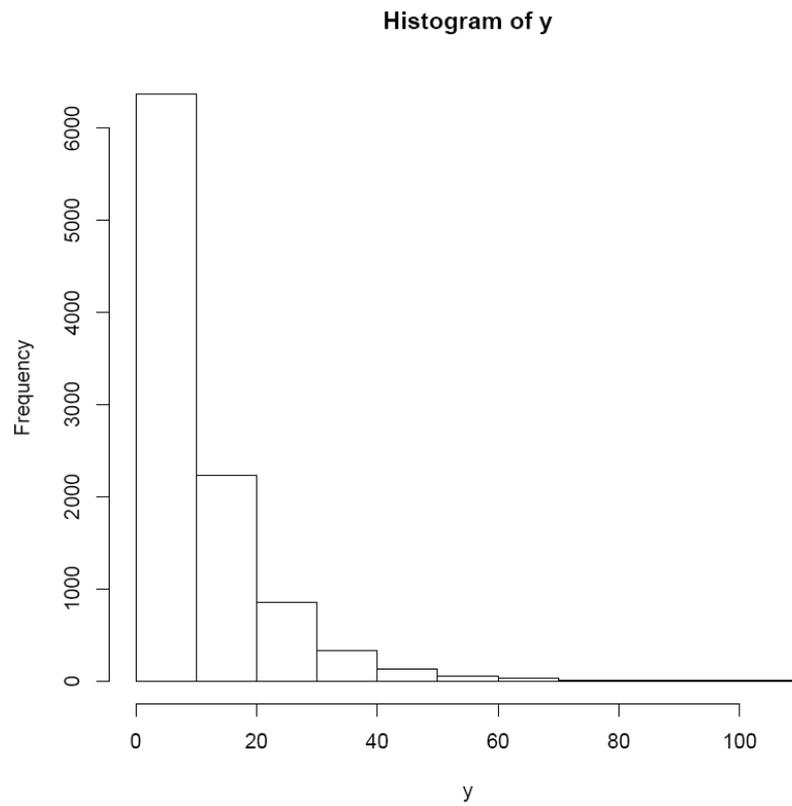


Figura A.35. Histograma de la población Gamma con $\rho = 0.5$ y $P_0 = 0.35$.

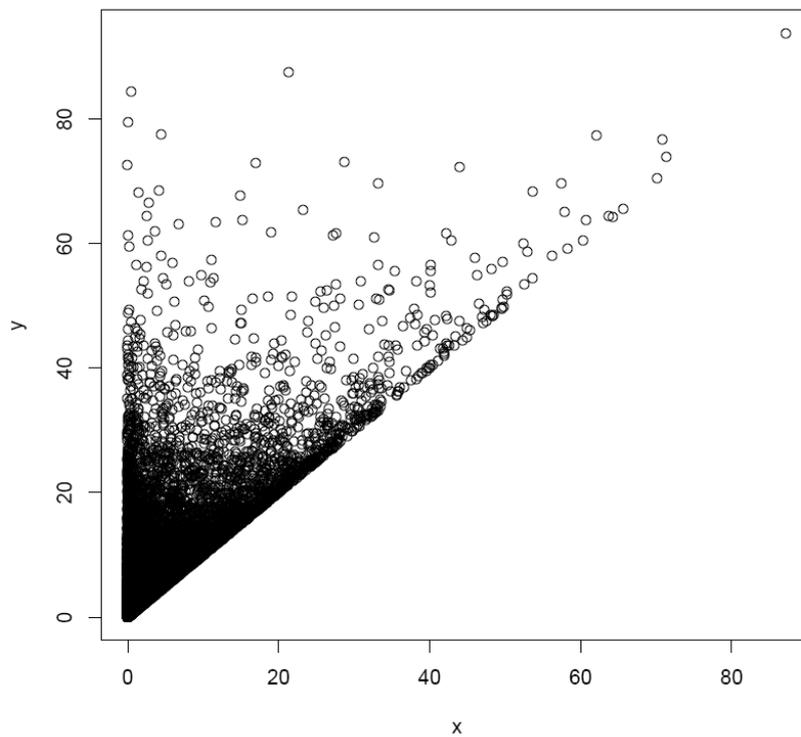


Figura A.36. Diagrama de dispersión de la población Gamma con $\rho = 0.7$ y $P_0 = 0.35$.

Apéndice I. Descripción de las poblaciones en estudio

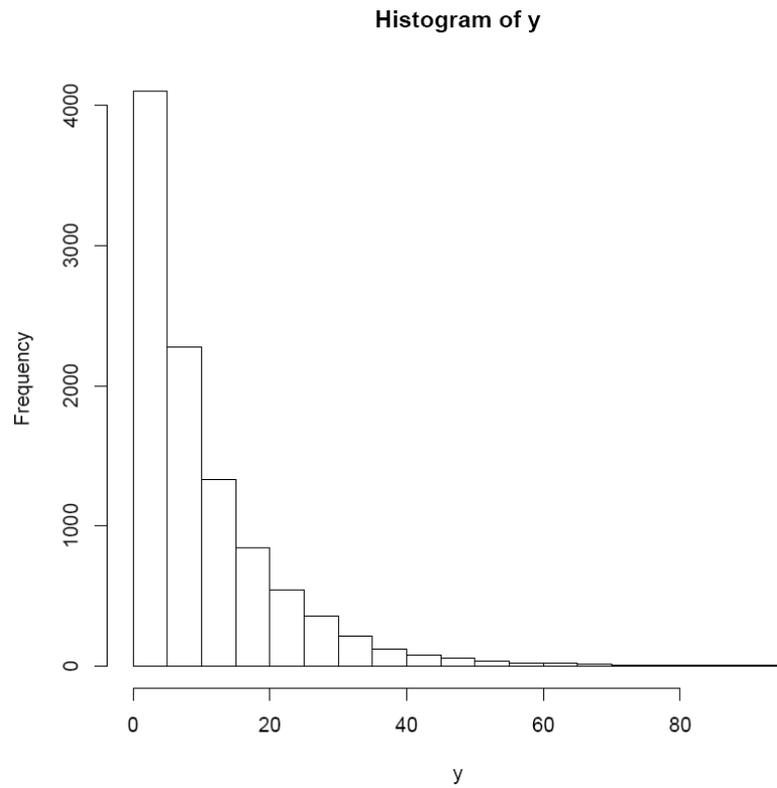


Figura A.37. Histograma de la población Gamma con $\rho = 0.7$ y $P_0 = 0.35$.

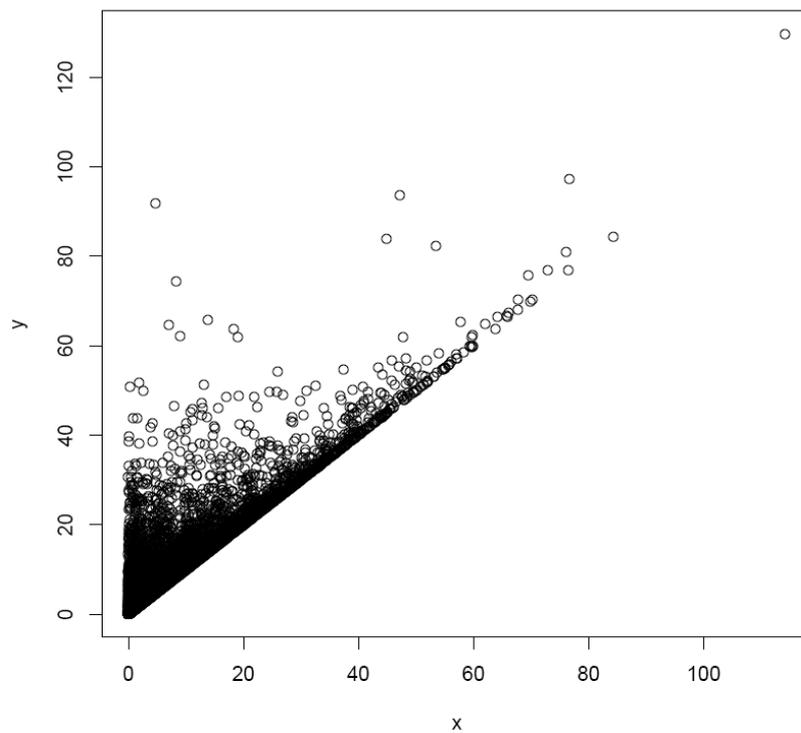


Figura A.38. Diagrama de dispersión de la población Gamma con $\rho = 0.9$ y $P_0 = 0.35$.

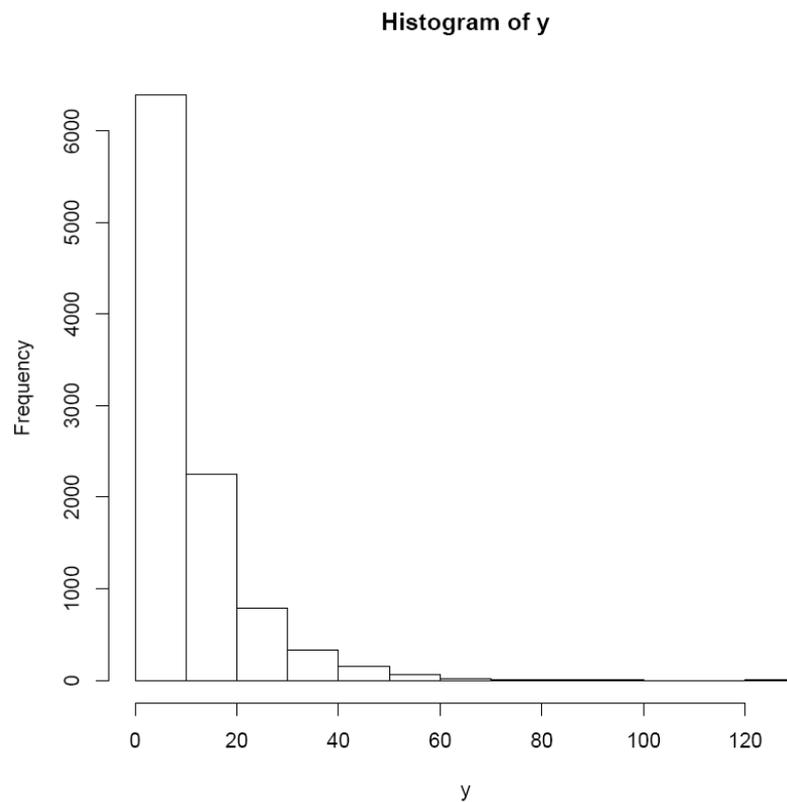


Figura A.39. Histograma de la población Gamma con $\rho = 0.9$ y $P_0 = 0.35$.

Poblaciones artificiales basadas en la distribución Singh-Maddala

g	P_0 (%)
4.0	6
8.1	8
11.5	8

Tabla A.3. Valores de P_0 (proporción de pobres) y g (coeficiente de asimetría) para distintas poblaciones artificiales basadas en la distribución Singh-Maddala. El tamaño de la población es $N = 10000$ y el coeficiente de correlación lineal es $\rho = 0.9$.

Apéndice I. Descripción de las poblaciones en estudio

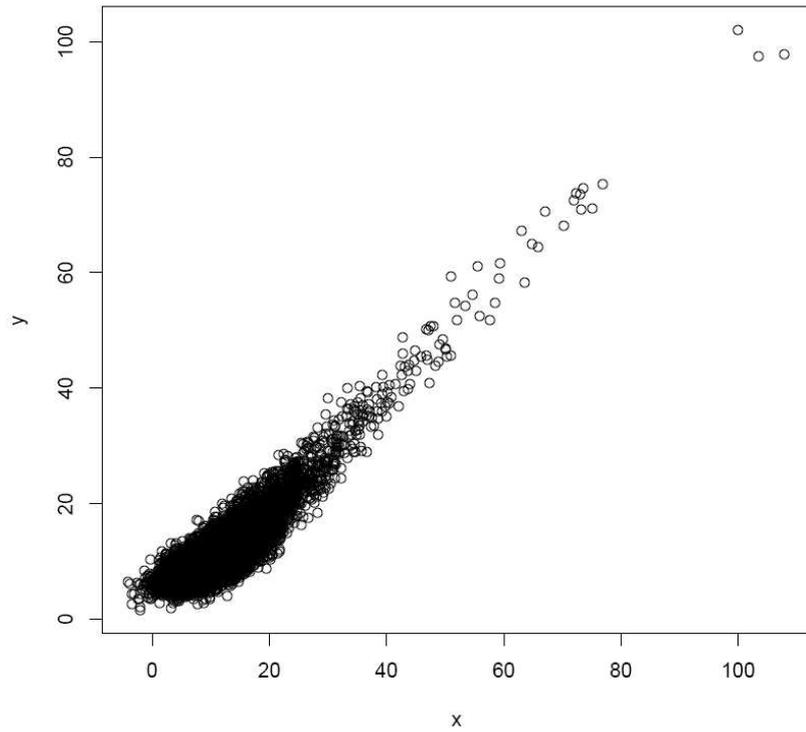


Figura A.40. Diagrama de dispersión de la población Singh-Maddala con $g = 4.0$.

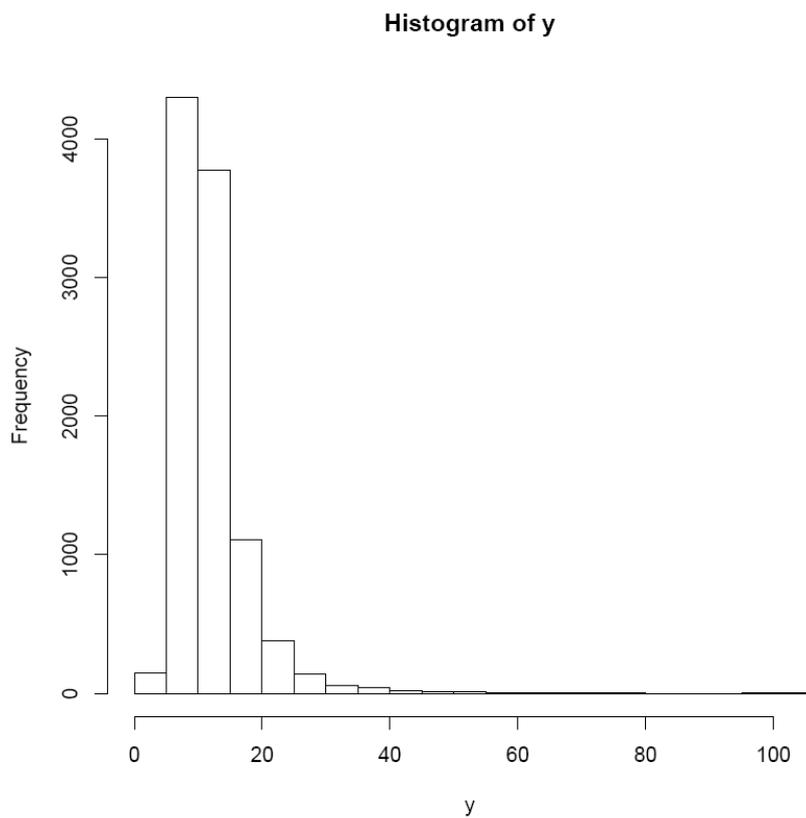


Figura A.41. Histograma de dispersión de la población Singh-Maddala con $g = 4.0$.

Apéndice I. Descripción de las poblaciones en estudio

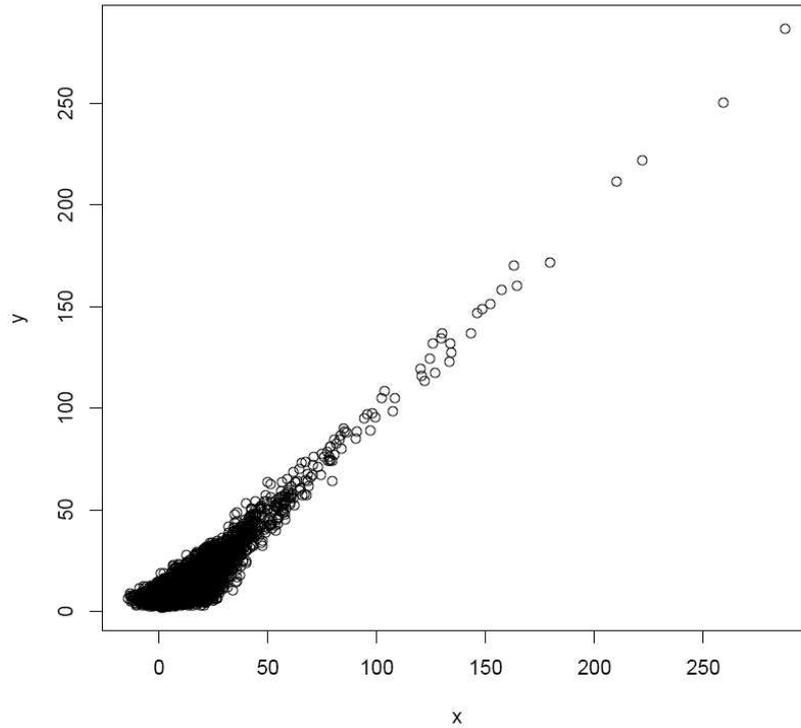


Figura A.42. Diagrama de dispersión de la población Singh-Maddala con $g = 8.1$.

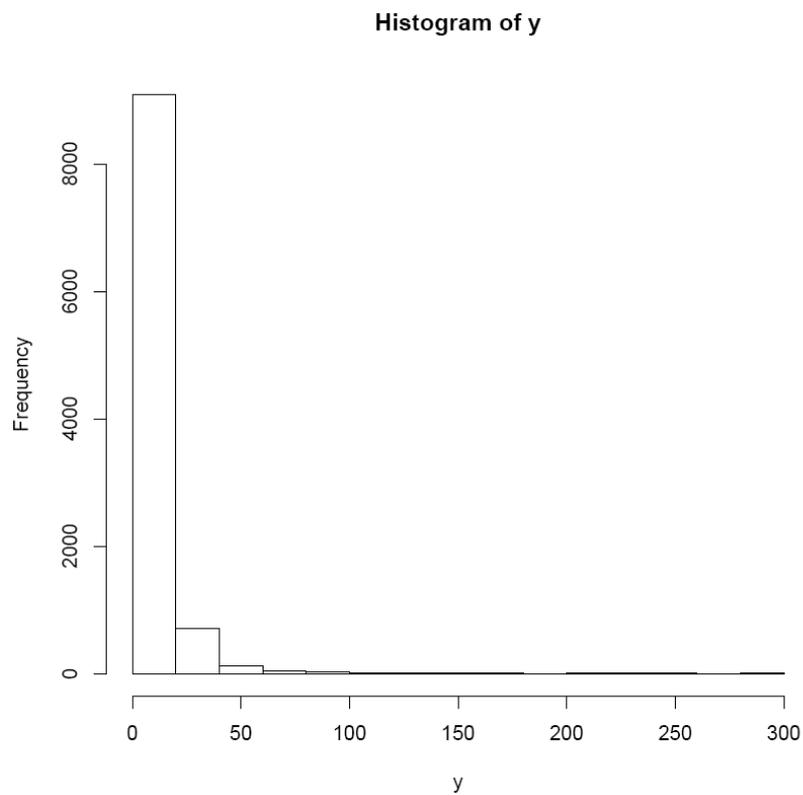


Figura A.43. Histograma de dispersión de la población Singh-Maddala con $g = 8.1$.

Apéndice I. Descripción de las poblaciones en estudio

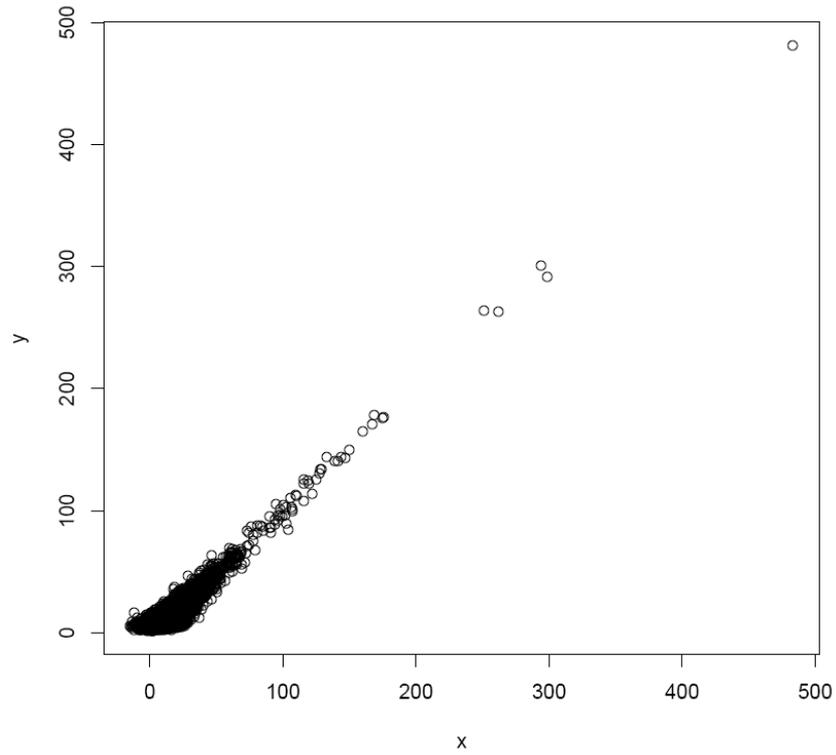


Figura A.44. Diagrama de dispersión de la población Singh-Maddala con $g = 11.5$.

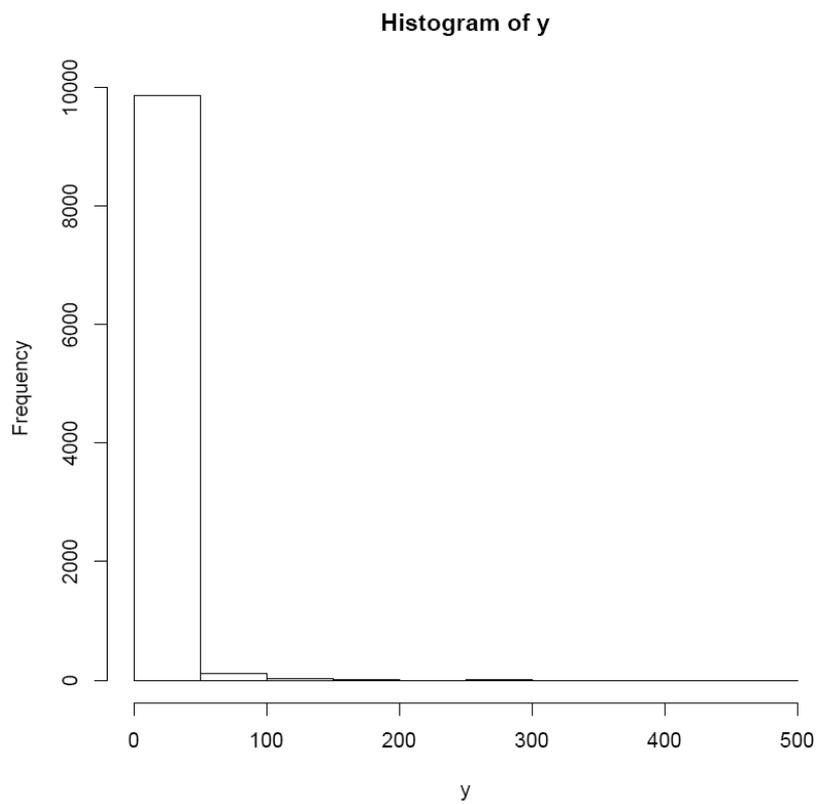


Figura A.45. Histograma de dispersión de la población Singh-Maddala con $g = 11.5$.

