



Actualización del Modelo de Simulación del Sistema de Pensiones MSSP-OLG: ii) extensión de la heterogeneidad de ingresos y bases contributivas

Alfonso R. Sánchez Martín

Documento de Trabajo 2023/04

junio de 2023

fedea

Las opiniones recogidas en este documento son las de sus autores y no coinciden necesariamente con las de Fedea.

Actualización del *Modelo de Simulación del Sistema de Pensiones* MSSP-OLG: ii) extensión de la heterogeneidad de ingresos y bases contributivas

Alfonso R. Sánchez Martín ¹

Junio de 2023

Resumen

Este trabajo presenta los cambios introducidos en el modelo MSSP-OLG con el objeto de lograr una mejor reproducción de la heterogeneidad observada en los ingresos por pensiones y en las bases de cotización en España. Se motiva la importancia de ambos cambios para conseguir una buena reproducción de las sendas de gasto en pensiones y para explorar algunas de las reformas recientes del Sistema, que afectan especialmente a los segmentos más altos de la distribución de ingresos.

¹ Este trabajo forma parte de un proyecto de investigación que ha sido financiado en parte por la Fundación Ramón Areces.

Índice

1. Introducción	3
2. ¿Por qué expandir la heterogeneidad en cotizaciones/pagos de pensiones?	4
3. Ampliación de la heterogeneidad de ingresos de pensiones	7
3.1. Distribución de ingresos de pensiones: evidencia	7
3.1.1. Dispersión de los ingresos agregados de pensiones en la sección cruzada	7
3.1.2. Análisis desagregado de la sección cruzada: MCVL	9
3.1.3. Dinámica de la dispersión de ingresos	12
3.2. Modelización de la heterogeneidad de ingresos intra-agente	15
3.2.1. Estrategia de modelización	15
3.2.2. Implementación: formas funcionales y propiedades resultantes	15
3.2.3. Calibración en la simulación de pensiones	19
4. Ampliación de la heterogeneidad en bases de cotización	23
4.1. Dispersión de las bases de cotización: evidencia empírica	23
4.2. Modelización de la heterogeneidad de las bases contributivas	30
4.3. Resultados de calibración de la simulación de pensiones	32
Anexo	
A. Heterogeneidad en ingresos de pensiones	36
A.1. Datos de ingresos de pensiones en el Informe Económico-financiero de la Seguridad Social	37
A.2. Datos de pensiones en la MCVL-2017	38
A.3. Distribución de Pareto Truncada aplicada a pensiones	39
B. Heterogeneidad en bases de cotización	41
B.1. Elaboración de la base de datos de bases de cotización en la MCVL	41
B.2. Comportamiento de las bases contributivas en la MCVL-2017	42

1. Introducción

Este trabajo es el segundo de una serie de documentos que conforman el proyecto de actualización del modelo MSSP-OLG, utilizado en diversos trabajos previos para proyectar el gasto en pensiones contributivas en España. El objetivo de esta actualización no es otro que adecuar el modelo para estudiar los cambios introducidos en las últimas reformas de pensiones de 2021 y 2023.

Tras un primer documento introductorio, el presente trabajo aborda uno de los cambios importantes en la arquitectura del modelo: la introducción de una mayor diversidad en las bases contributivas e ingresos de pensiones de los agentes del modelo. Este cambio es necesario para mejorar la proyección de gasto en presencia de una política discrecional restrictiva respecto de la pensión máxima y para estudiar las medidas de la reforma de 2023, enfocadas en buena parte a segmentos de la población en el extremo de la distribución de ingresos.

El documento se organiza como sigue. Arranca en la Sección 2 con la motivación de la necesidad de mejorar el reflejo de la heterogeneidad existente en las bases contributivas e ingresos por pensiones para abordar la pregunta que motiva el proyecto. Sin esta mejora no podríamos hacer una evaluación precisa de algunos cambios legislativos recientes que afectan a la actualización anual de los topes legales de pensiones y cotizaciones e incluyen impuestos específicos a los trabajadores de ingresos elevados. El núcleo básico del artículo se encuentra en la discusión de la nueva modelización de la dispersión de los ingresos por pensiones (Sección 3) y de las bases contributivas (Sección 4). En ambos casos se revisa la evidencia empírica, se propone un modelo teórico y se discute su calibración en el contexto del modelo de equilibrio. Los Apéndices A y B proporcionan información complementaria a la presentada en el texto principal.

2. ¿Por qué aumentar la heterogeneidad en cotizaciones e ingresos por pensiones?

Los sistemas de pensiones actuales son complejos y es, en cierto modo, inevitable que se haya priorizado el estudio de alguno de sus elementos (por su visibilidad o importancia cuantitativa) mientras que otros aspectos han quedado más en un segundo plano. La normativa sobre los topes legales de las bases contributivas y la cuantía de las pensiones estarían en esta segunda categoría. Durante mucho tiempo hemos tendido a prestar más atención a las consecuencias de cambios en la fórmula de cálculo de las pensiones o a la normativa de jubilación que a la política seguida con respecto a estos topes (con la posible excepción del tratamiento de la pensión mínima, al que la introducción del IRP en 2013 otorgó cierto protagonismo). Una dificultad asociada con estos mecanismos radica en que, al actualizarse de modo regular (típicamente en la ley de presupuestos), es preciso estudiarlos en términos de *estrategias* o *reglas contingentes* de actualización, en lugar de como cambios puntuales. El “espacio de alternativas” es, por tanto, más abierto que el que se ofrece ante, por ejemplo, un

cambio en la fórmula de pensiones. Dos fenómenos recientes, sin embargo, han modificado este estado de cosas:

- La toma de conciencia de la importancia cuantitativa de la pensión máxima sobre la trayectoria futura del gasto en pensiones.

En su última evaluación de los determinantes del gasto futuro de pensiones, el *Ministerio de Economía* (representante español en el Grupo de Envejecimiento de la Comisión Europea) enfatiza específicamente el papel de los topes máximos a la cuantía de las pensiones. En un contexto de crecimiento salarial real, un tope que se revaloriza a una tasa menor puede tener un impacto cuantitativo importante en la proyección del gasto futuro del Sistema. Esto es, efectivamente, lo que se encuentra en las proyecciones de gasto para la economía española entre 2020 y 2070. En la *Country Fiche* que acompaña al informe para nuestro país (European Commission (2020)) encontramos un reconocimiento explícito de la importancia de esta omisión en los trabajos previos y una cuantificación precisa del impacto en el gasto asociado a este proceso:

“A methodological improvement compared with the model used for the AR18 projections refers to the proportion of new old-age pensions capped by the ceiling on new pensions (maximum pension). In the AR18 this proportion remained fixed through the projection at approximately 10%. This assumption could be inconsistent with the internal dynamics of the projections and was changed for the AR21.

More precisely, in the AR21 projections the proportion of capped, new, old age pensions is calculated endogenously in the model. This proportion results from the interplay between the growth rate of wages and contributory bases on the one hand (and especially the growth rate of the cap on contributory bases) and the growth rate of the cap on new, old-age pensions. In the model, the former grows with nominal wages while the latter is updated tracking CPI. As a result, a growing number of new pensioners see their old-age pensions capped at the maximum level. This methodological change reduces public expenditure in pensions as a share of GDP in 0.5 pp at its peak.”

Ministerio-Economía (2021), pag. 49

- La inclusión de cambios legislativos dirigidos específicamente a trabajadores con bases contributivas altas en la reforma de 2023.

El estudio del impacto de los topes legales en las bases contributivas también ha sido comparativamente menor en los trabajos previos. La reforma de 2023, que incluye cambios legislativos que afectan a tramos específicos de la distribución de bases contributivas, cambia apreciablemente las cosas.² Así, el RDL 2/2023 (BOE 2023) introduce dos medidas directamente relacionadas con la incidencia del máximo en las bases contributivas: el proceso de “destope” de cotizaciones (la especificación de una tasa de crecimiento de la base máxima ampliamente superior a la de la pensión máxima, con la consiguiente conversión de las nuevas cotizaciones en un impuesto implícito) y la inclusión de un impuesto explícito “de solidaridad” aplicado al tramo de ingresos salariales que excede al máximo contributivo.

² Un segundo suceso reciente que ha puesto el foco en los topes contributivos (en este caso el mínimo legal) es la extraordinaria subida, entre 2018 y 2019, en las bases mínimas de cotización de todos los grupos de asalariados.

De La Fuente (2023) es un primer análisis del impacto de ambos cambios legislativos, centrado en la capacidad recaudatoria de las nuevas medidas. El modelo de equilibrio permite mejorar en cierta medida este cálculo financiero al incorporar algunos de los márgenes de respuesta de los individuos frente a este aumento de la fiscalidad, así como los efectos de segundo orden vía precios.³ La aportación del modelo de equilibrio es aún más importante porque hace posible un cálculo de bienestar que permite evaluar el impacto intra- e inter- generacional de estas medidas. Lógicamente, el modelo tiene que incluir una representación razonablemente realista de la distribución de las bases contributivas para ser capaz de medir el impacto recaudatorio de la medida y sus consecuencias redistributivas.

Dimensiones de mejora del modelo antiguo

Las versiones previas de nuestro modelo de simulación (revisadas en detalle en Sánchez Martín (2023a)) han prestado mucha atención al comportamiento de las pensiones mínimas, reproduciendo su tasa de incidencia y su impacto en el gasto agregado como parte de una correcta modelización del IRP.⁴ Por contra, hemos dedicado una menor atención a la pensión máxima y a los topes contributivos, limitándonos a constatar que su comportamiento cualitativo era el esperable.⁵ En esta actualización del modelo hacemos el esfuerzo de reproducir adecuadamente las magnitudes cuantitativas asociadas a ambos topes, reconociendo que los errores cometidos al reproducir la incidencia de los mismos pueden llevar a: (i) una sobreestimación apreciable del gasto futuro en pensiones y (ii) una infraestimación de la capacidad recaudatoria de la Seguridad Social.

En principio, el modelo debería encontrarse bien equipado para este esfuerzo gracias al elevado número de hogares de distintos tipos presentes en él. Para los jubilados, el modelo incluye en cada año de calendario 40 agentes distintos por cohorte de pertenencia (entre 60 y 100 años) y, para cada uno de ellos, 4 posibles niveles educativos y 10 posibles edades de jubilación (expresión de diferencias en su preferencia entre ocio y consumo). Además, se reconocen pensiones distintas para cada uno de los dos cónyuges de estos *hogares representativos*. En total, la simulación produce un total de 3.200 niveles diferentes de pensiones de jubilación en cada año. Para los agentes en edad de trabajar las diferencias en educación, sexo y cohorte de pertenencia implican la existencia de 400 niveles distintos de ingresos laborales en cada año de calendario.⁶

³ Reflejamos los cambios en la edad de jubilación y en el ahorro, lo que puede entenderse como una primera aproximación. Por supuesto, la respuesta en el mundo real puede producirse a través de otras dimensiones no incluidas. Sólo un modelo específicamente diseñado para analizar esta pregunta podría proporcionar una respuesta enteramente satisfactoria a esta pregunta.

⁴ El valor agregado de los complementos de mínimos tiene la consideración de “gasto” no contributivo y es, por tanto, objeto de una transferencia específica a la Seguridad Social. Esta transferencia es parte de los ingresos considerados en el IRP, de modo que su calibración cuantitativa detallada es parte importante de la modelización del IRP.

⁵ Ajustando la pensión máxima con el IPC, se producen un aumento en su tasa de incidencia, más acusado cuanto mayor es el crecimiento proyectado en el nivel futuro de las pensiones. Los topes contributivos, por su parte, tienden a tener una incidencia bastante estable cuando los límites legales se actualizan conforme al aumento anual en la productividad del trabajo

⁶ El modelo incluye muchos tipos de hogares pero, ciertamente, no tenemos un hogar para *cada* agente económicamente diferente en la población de referencia. Es en este sentido que denominamos a los hogares del modelo “*agentes representativos*”. Por ejemplo, entre los hogares nacidos un cierto año y cuyos cónyuges tienen educación superior y una baja preferencia por el ocio (es decir, que prefieren una jubilación tardía) hay, en el mundo real, una gran variedad de

Estos niveles de heterogeneidad se han revelado bastante adecuados para reproducir las colas inferiores de la distribución de ingresos salariales y de pensiones, pero no así con las colas superiores de ambas variables. Como vemos en la Sección 3.1, la distribución de los ingresos de pensiones tiene un alto grado de dispersión, incluyendo un porcentaje apreciable de pensiones “topadas” en el máximo legal. Un resultado similar, aunque menos acusado, se produce respecto del tope máximo en las bases contributivas (Sección 4.1). De hecho, la heterogeneidad proporcionada por los *Agentes Representativos* del modelo no ha sido suficiente para reproducir cuantitativamente estos estadísticos. Por ello, la Sección 3.2 propone desarrollar un módulo específico que incorpora una modelización en forma “reducida” de la heterogeneidad de ingresos de pensiones de los agentes del modelo. Con las bases contributivas seguimos una estrategia similar, que revisamos con detalle en la Sección 4.2. Comprobamos, en ambos casos, la capacidad de esta nueva modelización para generar una amplia heterogeneidad de ingresos. Finalmente, el éxito alcanzado al incorporar estas especificaciones en el modelo de simulación de pensiones se discute en las Secciones 3.2.3 (ingresos de pensiones) y 4.3 (bases contributivas).

3. Ampliación de la heterogeneidad de ingresos de pensiones

3.1. Distribución de los ingresos por pensiones: evidencia

El análisis de la dispersión de los ingresos procedentes de pensiones de jubilación debe contemplar dos perspectivas interrelacionadas. Evidentemente, debemos comenzar estudiando la evidencia disponible en las secciones cruzadas recientes, de forma que el modelo pueda reproducir la realidad al comienzo de la simulación. Al mismo tiempo, los aspectos dinámicos de la dispersión constituyen una segunda dimensión más compleja y difícil de explorar. Responden a tendencias a más largo plazo tanto en el ciclo vital de las personas como en los procesos de serie temporal del entorno económico. Como es de esperar, la evidencia sobre la primera dimensión es más extensa y completa que sobre la segunda.

3.1.1. Dispersión de los ingresos agregados de pensiones en la sección cruzada

La información sobre la dispersión de ingresos de pensiones está, lógicamente, contenida en el archivo de prestaciones de la Seguridad Social (tanto en su dimensión estática como dinámica). Pero el contenido del mismo no es de acceso público para los investigadores, de modo que hemos de recurrir a diversas fuentes de información:

- Los estadísticos resumen recogidos en diversas publicaciones de la Seguridad Social y las series cronológicas en la página eSTADISS-Estadísticas del INSS.⁷
- La información en la Muestra Continua de Vidas Laborales (MCVL).

experiencias en términos de historiales de participación laboral e ingresos. Esta variedad llevaría a múltiples pensiones en el instante de la jubilación. El modelo hace colapsar esa heterogeneidad en un único perfil laboral/pensión: su promedio entre los perfiles de las personas que comparten educación/cohorte/jubilación. Este uso del término “agente representativo” no es el habitual en macroeconomía (donde se reserva al modelo con un único agente y, típicamente, vida infinita), pero entendemos que resulta adecuado en este contexto.

⁷ Accesible desde la página <https://sede.seg-social.gob.es>.

- Estudios de simulación calibrados que incluyan los determinantes de la evolución dinámica de los ingresos: tendencias salariales, cambios en la esperanza de vida, cambios en las normas de pensiones y evolución de los topes legales

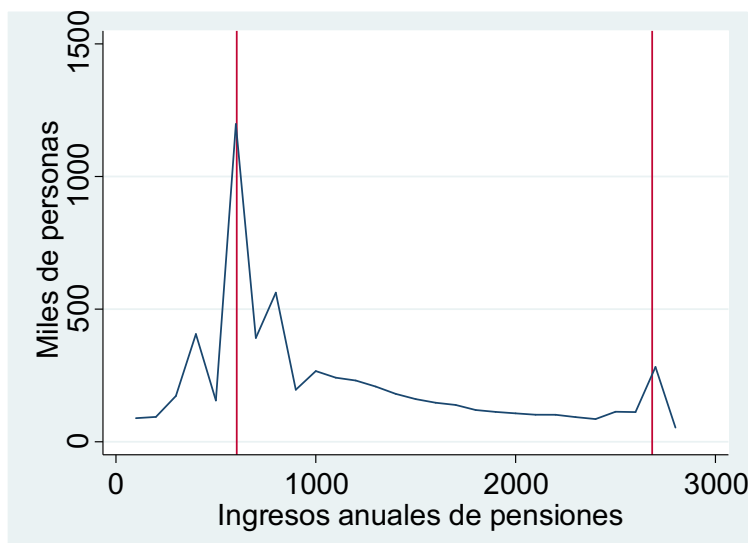


Figura 1: Distribución de los ingresos de pensiones en la sección cruzada de 2020. Fuente: estadística de pensiones por tramos en la base de datos eSTADISS-Estadísticas del INSS.

Con la primera fuente de información es posible construir algunas distribuciones agregadas de ingresos de pensiones. Así, la base eSTADISS-Estadísticas del INSS permite descargar la serie del número total de pensiones de la Seguridad Social por tramos de cuantía. Usando esta información, el Gráfico 1 representa la distribución del *stock* total de pensiones en 2020. La cuantía del pico observado en la cola derecha de la distribución permite medir la incidencia de la pensión máxima en ese año: el 4,6% del stock total de pensiones de la Seguridad Social.

Para calcular este estadístico para otros tipos de pensiones (en particular, para las pensiones de jubilación) hay que acudir a otras fuentes de datos. Una posibilidad es acudir al resumen de estadísticos incluidos en el Boletín Económico-Financiero que acompaña a la ley de presupuestos (el último, Seguridad-Social (2021)). El Cuadro IV.30 de la página 218 proporciona el número de pensiones en vigor por tramos de cuantía para cada uno de los cinco tipos de pensiones del Sistema.⁸ A partir del Cuadro se puede reconstruir la distribución del *stock* de pensiones de jubilación, pero es preciso tratar los datos originales. Siguiendo el procedimiento descrito en el Apéndice A.1 encontramos que la incidencia de la pensión máxima en el stock de pensiones de jubilación se estimaría entorno al 5,1% del total.⁹ Por otra parte, el

⁸ Los Cuadros IV.32 y IV.33 proporcionan información adicional sobre el desglose de la pensión por conceptos: inicial/complementos/revalorización.

⁹ Se reporta un 0,8% de pensiones cuyo valor excede a la máxima en vigor en el instante de medida (agosto de 2020).

informe del AWG para España (en Ministerio-Economía (2021)) proporciona una estimación de la incidencia de la pensión máxima entre las altas de pensiones en 2020 del 8,5%.

Medir la incidencia de la pensión mínima con los datos del Informe Económico-Financiero (o con los datos de eSTADISS) es más difícil, ya que hay varios valores distintos del complemento dependiendo de las características familiares y la edad de jubilación. En el Gráfico 1 se observa un pico de extraordinaria importancia (20,3% de la muestra) en la categoría de entre 600 y 700 euros mensuales, que corresponde al complemento en presencia de un cónyuge no dependiente en caso de jubilación antes de los 65 años. Hay un segundo pico en el tramo entre 800 y 900 euros donde se incluyen los complementos de mínimos con cónyuge a cargo, pero que también debe incluir pensiones sin complementos. Los ingresos por debajo de 400 euros al mes corresponden a pensiones sin garantía de mínimo (por concurrir con otras rentas) y pensiones de convenios internacionales de las que sólo se paga una parte en España. Calcular el total de pensiones afectadas por complementos es difícil con estos datos. Afortunadamente la Seguridad Social nos proporciona una estimación específica del porcentaje de pensiones en vigor con complemento de mínimos (por clase de pensión) en la misma publicación (Cuadro IV.44). Para 2020, esta cifra es del 23,53% en el conjunto de pensiones y 21,85% entre las pensiones de jubilación. La tendencia de ambos porcentajes en los últimos años ha sido decreciente.

3.1.2. Análisis desagregado de la sección cruzada: MCVL

La extracción aleatoria del archivo de la Seguridad Social que representa la Muestra Continua de Vidas Laborales (MCVL) nos permite estimar estadísticos poblacionales a un nivel mucho más desagregado (tanto para el stock de pensiones como para las altas en un período determinado). Esta capacidad se revela importante en la medida en que las distribuciones condicionadas resultantes son, en muchos casos, muy distintas de las agregadas.

La pauta resumen de la distribución de los ingresos de las altas de pensiones de jubilación se muestra en el Gráfico 2.¹⁰ Un 8.7% de las pensiones están “topadas” en el máximo legal, mientras que el porcentaje en el mínimo es del 15.3% (columnas de *Totales* de los Cuadros 1 y 2 para los topes inferior y superior respectivamente). Estos estadísticos cambian profundamente al considerar diversos subgrupos poblacionales. En particular, al condicionar en los observables que constituyen las *variables de estado* que definen los distintos *agentes representativos* del modelo de simulación: el nivel de educación y la edad de jubilación (y el sexo dentro de cada agente).¹¹ Como muestran los histogramas en el Gráfico 3 y los estadísticos en los Cuadros 1, 2 y 3 hay una variedad importante en los niveles medios de ingresos y en las pautas de distribución de los mismos:

¹⁰ Corresponde a las altas durante 2016, construidas utilizando la ola 2017 de la MCVL.

¹¹ Los *agentes representativos* del modelo también difieren por su cohorte de pertenencia pero, al centrarnos en las altas de jubilación en un año determinado, ésta queda determinada por la edad de jubilación.



Figura 2: Distribución de los ingresos de pensiones de las altas de jubilación en la MCVL de 2016

- Los ingresos medios (Cuadro 3) son mayores para los trabajadores de educación más elevada y para los varones (salvo en el grupo con nivel educativo elevado, en que no hay diferencia). Llama la atención que la salida más temprana a la jubilación se asocia con trabajadores de pensiones elevadas.
- La incidencia de las pensiones máximas y mínimas se sigue de modo bastante natural de las pautas observadas en los valores medios: los topes inferiores son vinculantes con más frecuencia entre mujeres y trabajadores de ingresos bajos (Cuadro 1), mientras que los superiores son más frecuentes si el nivel educativo es alto y, en otro caso, entre varones (Cuadro 2). La edad de jubilación también es un predictor muy significativo, aunque las pautas aquí son más complejas. Para la pensión mínima la relación es monótona (mayor incidencia en jubilaciones más tardías). En coherencia con una distribución de ingresos más concentrada, las mujeres muestran menos incidencia de las pensiones máximas conforme la edad de jubilación es más alta. Los hombres, por contra, tienen más incidencia de la máxima en salidas tardías, evidenciando una distribución más dispersa
- El grado de dispersión varía de modo complejo con las variables observables (ie, no responde a reglas sencillas como, por ejemplo, ser una función creciente de los ingresos). Esto es visible de modo bastante claro en los histogramas del Gráfico 3. Así, mientras que la distribución de ingresos de las mujeres está más concentrada que la de los varones, se observa lo contrario respecto de la educación (mayor dispersión con menor nivel educativo). Respecto de la jubilación, hay claramente más dispersión en las salidas anteriores a los 65 años.

Complementando este análisis, en el Cuadro 10 del Apéndice A.2 se realiza una partición algo más detallada de la muestra (con dobles condicionamientos). De este Cuadro extraemos algunos casos extremos que nos sirven, en la Sección 3.2.2, para testar el modelo de esta heterogeneidad que desarrollamos en la Sección 3.2.

	Sexo		
	Varón	Mujer	Total
	Incidencia%	Incidencia%	Incidencia%
Nivel educativo	0.081	0.293	0.165
Medio	0.069	0.137	0.097
Alto	0.074	0.132	0.092
Total	0.079	0.267	0.153
Jubilación Antes	0.058	0.215	0.115
65	0.194	0.392	0.314
Después 65	0.288	0.504	0.399
Total	0.079	0.266	0.153

Cuadro 1: Incidencia de la pensión mínima entre las altas de jubilación en la MCVL de 2016 por sexo, educación y edad de jubilación

	Sexo		
	Varón	Mujer	Total
	Incidencia%	Incidencia%	Incidencia%
Nivel educativo	0.082	0.035	0.063
Medio	0.198	0.124	0.167
Alto	0.284	0.253	0.275
Total	0.108	0.055	0.087
Jubilación Antes	0.106	0.062	0.090
65	0.088	0.031	0.054
Después 65	0.163	0.035	0.097
Total	0.108	0.055	0.087

Cuadro 2: Incidencia de la pensión máxima entre las altas de jubilación en la MCVL de 2016 por sexo, educación y edad de jubilación

	Sexo		
	Varón	Mujer	Total
	Pensión media	Pensión media	Pensión media
Nivel educativo	1341	1018	1213
Medio	1763	1578	1686
Alto	1890	1886	1888
Total	1424	1124	1306
Jubilación Antes	1463	1214	1374
65	1057	842	927
Después 65	1192	810	996
Total	1424	1124	1306

Cuadro 3: Pensión mensual media de las altas de jubilación en la MCVL de 2016 por sexo, educación y edad de jubilación

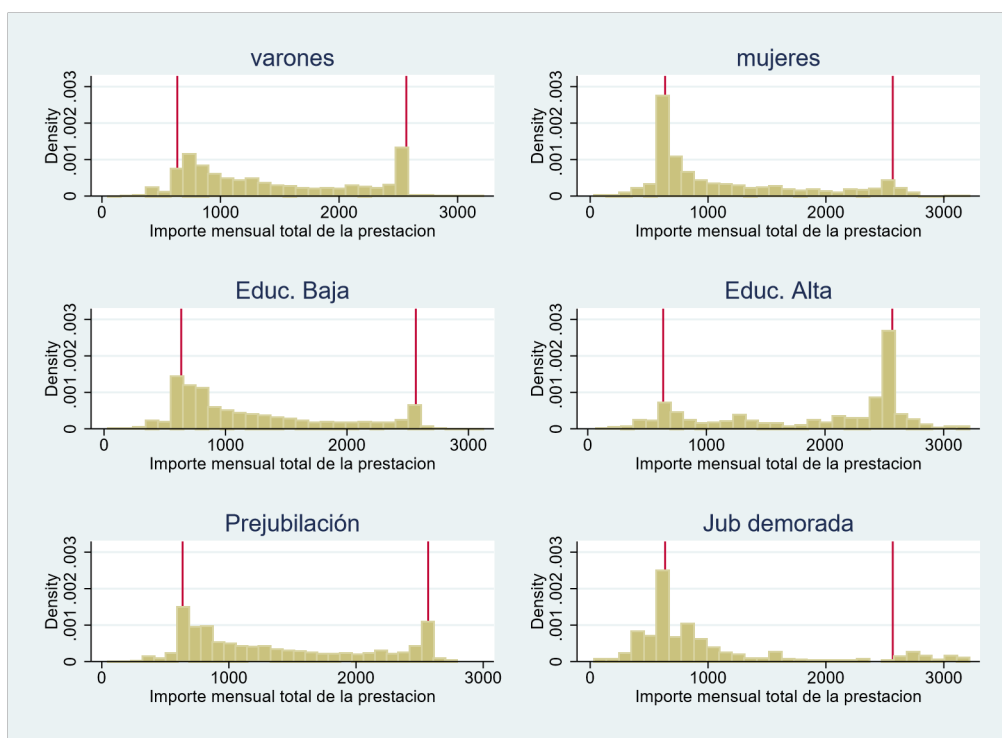


Figura 3: Distribución de los ingresos de pensiones de las Altas de jubilación en la MCVL de 2016: histogramas por sexo, educación y edad de jubilación.

3.1.3. Dinámica de la dispersión de ingresos

El paso del tiempo modifica las pautas observables de dispersión de ingresos. Este carácter dinámico de la distribución de pensiones se manifiesta en las dos perspectivas desde las que se pueden aproximar los fenómenos temporales:

1. Perspectiva de serie temporal.

Con el paso de tiempo de calendario la composición de las nuevas altas (en las dimensiones estudiadas en la sección previa) cambia en respuesta a procesos económicos, demográficos y legislativos. Entre los primeros destacan los cambios en la productividad, la distribución sectorial de la economía y sus ventajas comparativas, y los cambios en el mercado de trabajo. Estos factores determinan las bases reguladoras sobre las que se calculan las pensiones. Los procesos demográficos como el envejecimiento y la feminización de la población laboral también tienen un impacto evidente en la composición del colectivo de jubilados y, por tanto, en estadísticos como la incidencia de los topes de pensiones. Finalmente, cambios en la legislación de pensiones tienen consecuencias tanto directas (cambios en la fórmula de pensiones) como indirectas a través de los comportamientos de jubilación. Estos cambios en la composición de las altas se trasladan progresivamente en forma de cambios en la composición del stock total de pensiones.

2. Perspectiva de ciclo vital

El paso del tiempo implica un inexorable cambio en la edad de los individuos, con implicaciones en su salud y tasas de supervivencia. El avance en el ciclo vital no tiene lugar de modo homogéneo en la población, ya que los patrones de supervivencia varían con las características demográficas (sexo) y económicas (ingresos) de las personas. En consecuencia, la distribución de ingresos de pensiones de un cierto grupo poblacional cambia en el tiempo incluso en ausencia de otros cambios económicos. Por ejemplo, un agente representativo de una cierta cohorte y nivel de educación verá cambiar la incidencia de las pensiones máximas debido a la mayor supervivencia de los individuos de mayores ingresos del grupo. La edad produce, en resumen, cambios sistemáticos en la composición de los miembros de cada grupo muestral. Esto hace que la distribución de ingresos intra-grupo no sea invariable en el tiempo.

Concluimos de los párrafos anteriores que la proyección de la dinámica de las distribuciones de ingresos es una tarea compleja. Afortunadamente, los modelos de equilibrio general incorporan de modo explícito muchos de los procesos económicos y demográficos relevantes. Además, como ya hemos indicado, nuestro modelo incluye varias de las dimensiones de heterogeneidad claves para estos cambios dinámicos. Aun así, es indiscutible que no incluye toda la heterogeneidad relevante, especialmente por la invarianza de los procesos de supervivencia respecto del nivel de renta. Es preciso, por tanto, introducir modelos reducidos de la heterogeneidad intra-agente del modelo puedan cambiar flexiblemente en el tiempo.

La Sección 3.2 discute la modelización de esta heterogeneidad y su calibración. Aquí simplemente constatamos la necesidad de: (1) documentar la evolución de serie temporal de los indicadores básicos de la distribución; y (2) alcanzar una buena comprensión de la dependencia de las tasas de supervivencia respecto de las características económicas (especialmente el nivel de ingresos). Las fuentes de evidencia empírica para ambas tareas incluyen:

1. Estudios de serie temporal.

Como hemos indicado en la sección 3.1.1, la base de datos eSTADISS-Estadísticas del INSS permite construir una serie temporal de histogramas de ingresos de pensiones y de la incidencia de las pensiones máximas (y, aproximadamente, de las mínimas), pero sólo para el agregado total de pensiones.¹² Alternativamente, podrían construirse las series temporales agregando la información longitudinal de la MCVL (que se describe a continuación).

2. Estudios longitudinales.

La mejor (y prácticamente única) fuente para estudiar el cambio en el tiempo de los patrones desagregados de los ingresos de pensiones es la MCVL. No nos referimos aquí a la información retrospectiva que se incluye en cada extracción anual, sino a la secuencia completa de los archivos de prestaciones publicados desde 2004. Dado el mecanismo de muestreo de la MCVL, esta serie histórica permitiría análisis representativos de los estadísticos anuales asociados tanto al stock total como a las altas de pensiones. También

¹² La serie cronológica de la incidencia de las mínimas se publica en otros informes, incluyendo Seguridad-Social (2021).

permitiría analizar, para cohortes determinadas, la evolución de ciclo vital de sus variables de pensiones.

3. Estudios de la dependencia de la supervivencia del nivel de ingresos.
4. Estudios de simulación calibrados.

Un mecanismo indirecto de exploración de las propiedades de los procesos de ingresos utilizaría simulaciones informáticas con base en los agentes individuales o familiares. En este caso sólo se utiliza información empírica a nivel agregado, que se compara con las previsiones producidas al consolidar los comportamientos simulados de los agentes (en contextos artificiales que reproducen con detalle el entorno económico e institucional). Los modelos se utilizan como bancos de prueba en los que testar hipótesis sobre las conductas micro, que son validadas o rechazadas por la similitud de sus resultados agregados con los procesos observados. La metodología de *inferencia indirecta* que describimos en la Fase 2 de nuestra calibración (Sección 3.2.3) encaja en este tipo de estudios sustituyendo la información histórica por las proyecciones agregadas del Grupo de Envejecimiento de la CE para el intervalo 2020/2070.

Este tipo de técnicas podrían, en principio, utilizarse para extraer información del comportamiento en el ciclo vital a partir de una cantidad muy reducida de evidencia de sección cruzada: las diferencias entre la distribución del stock de pensiones y la observada entre las altas de jubilación.¹³

3.2. Modelización de la heterogeneidad de ingresos

Revisamos la modelización de la heterogeneidad de ingresos de pensiones de los agentes representativos del modelo (discutidos en el pie de página 6 de la sección 2) en tres pasos: (i) proponemos una estrategia genérica (Sección 3.2.1), (ii) revisamos su implementación con una distribución de Pareto truncada (Sección 3.2.2) y (iii) discutimos la calibración de esta distribución en nuestro modelo de equilibrio utilizando la evidencia empírica de la Sección previa (Sección 3.2.1).

3.2.1. Estrategia de modelización

Realizamos una partición de la población en el instante t en N subgrupos, indexados por $i \in \{1, \dots, N\}$. En lo que sigue omitimos el superíndice t para aliviar notación. Suponemos que la pensión inicial de jubilación (B) de los agentes en cada uno de los N grupos poblacionales se distribuye conforme a una *función de distribución* $F_i(B)$. La heterogeneidad poblacional se manifiesta, cuando se produce el alta de jubilación¹⁴.

Hacemos operativo el modelo postulando una forma funcional paramétrica común $F(\theta)$ para la distribución de probabilidad de la pensión inicial de todos los grupos. Las diferencias observadas entre grupos de agentes se generan haciendo variar los valores paramétricos θ en función de las características definitorias de cada grupo. A nivel operativo, $F(\theta_i)$ se concreta en

¹³ La distribución del stock es fruto tanto de los procesos de ciclo vital como de la evolución histórica de serie temporal de las variables económica e institucionales. En teoría, sería posible separar la contribución de ambos factores combinando, vía simulación, la información observable sobre los procesos históricos (topes de pensiones, normativa de pensiones e ingresos laborales) con las pautas de ciclo vital que se desean probar.

¹⁴ El resto de pensiones del sistema (viudedad, incapacidad permanente, etc) se vincula a la modelización de la pensión de jubilación

un pequeño número de estadísticos, χ_i . En nuestro trabajo, χ_i incluye la esperanza matemática de la pensión y las tasas de incidencia de las pensiones mínima y máxima en el instante considerado.¹⁵

$$\chi_i = \{\mu_i, Im_i, IM_i\}$$

Si se dispone de una muestra de pensiones suficientemente grande, procederíamos a la calibración/estimación de los parámetros $\{\theta_i\}$ de cada grupo resolviendo un problema de mínima distancia:

$$\hat{\theta}_i = \arg \min || \tilde{\chi}_i - \chi(F(\theta_i)) || \quad \forall i \in \{1, \dots, N\} \quad (1)$$

donde $\tilde{\chi}_i$ representa la realización muestral del vector de estadísticos en el grupo i y donde $||x||$ es típicamente la norma cuadrática.

Si el número de subgrupos poblacionales (agentes) es pequeño es perfectamente factible calcular el vector χ_i para $i \in \{1, \dots, N\}$ y estimar θ_i aplicando la ecuación (1) *en cada grupo* por el Método Generalizado de Momentos o GMM. En caso contrario, siempre es posible realizar este “matching” de modo algo más agregado de forma que la simulación reproduzca las diferencias observadas por características como la educación, el sexo y/o el grupo general de jubilación (pre/post 65).

En la sección siguiente discutimos la implementación de esta estrategia, empezando con la elección de la forma paramétrica seleccionada para F y el anclaje con la simulación de pensiones del modelo.

3.2.2. Implementación: formas funcionales y propiedades resultantes

La elección de forma funcional está condicionada por el modo en que se integra el modelo de distribución de ingresos en el conjunto del equilibrio de pensiones. Nuestra estrategia en este sentido se basa en tratar de replicar, para cada *agente representativo* del modelo¹⁶ (considerando a cada uno de los cónyuges por separado) los siguientes estadísticos:

1. Como media matemática de la pensión de jubilación de cada grupo, $\bar{\mu}_i$, se asigna la pensión generada por el modelo en la jubilación del agente (conforme a la fórmula de pensiones en vigor en cada momento, el historial laboral del agente, los topes legales, etc). De este modo, las distribuciones de gasto quedan firmemente ancladas en la previsión cuantitativa original del modelo de pensiones.
2. Para las tasas de incidencia agregada de las pensiones máximas y mínimas, $\{\overline{Im}_i, \overline{IM}_i\}$, asignamos sus contrapartidas muestrales calculadas en la MCVL (ver cuadros de la Sección 3.1.2).¹⁷

¹⁵ Se puede operar tanto con la media de $F(\theta)$ como con la media condicionada de las observaciones *interiores* entre los topes legales. Ambos procedimientos son equivalentes para caracterizar la distribución de modo único.

¹⁶ Es decir, para cada cohorte de jubilados en el año t y, dentro de ella, para cada nivel de educación y cada edad de jubilación.

¹⁷ Pese al elevado tamaño muestral de la MCVL, no es posible disponer de un número suficiente de observaciones para estimar estos estadísticos por separado para cada agente. En su lugar, replicamos los estadísticos muestrales de un

Forma funcional: distribución de Pareto truncada

Como primera opción en la elección de forma funcional para $F(b)$ consideramos la distribución Log-normal $LN(\mu, \sigma)$, una de las más utilizadas y mejor conocidas para representar distribuciones de ingresos. Lamentablemente, su comportamiento en la simulación se ha revelado demasiado inflexible para nuestros propósitos. En general, al ajustar en cada subgrupo el valor de μ_i conforme al nivel generado por el modelo no hemos sido capaces de reproducir la incidencia simultánea de los dos topes de pensiones variando el parámetro restante de la distribución (σ_i). Así, cuando se representaba con éxito la incidencia de la pensión mínima en la cola izquierda de la distribución obteníamos una incidencia de la pensión máxima en la cola izquierda sistemáticamente inferior a la observada. Hemos encontrado, en resumen, que con dos parámetros libres no se consigue replicar satisfactoriamente los 3 estadísticos objetivo del proceso.

Nuestro siguiente intento se ha centrado en leyes de distribución con tres parámetros. Entre las diversas opciones, hemos trabajado con la familia de *distribuciones de Pareto*, que ha sido utilizada con éxito para reproducir variables con “colas pesadas” (en las que los valores más extremos tienen un peso poblacional superior al que tendrían con una distribución normal o lognormal).¹⁸ Finalmente, hemos decidido implementar la distribución más sencilla de esta familia, pero bajo el supuesto de que el soporte se encuentra truncado por arriba y por abajo. La ley estadística de *Pareto Truncada* se caracteriza como sigue:

- Parametrizada con 3 elementos: α , L y H .

El parámetro α determina la forma general de la distribución y toma valores positivos. L (“low”) y H (“high”) determinan los extremos inferior y superior de la variable representada (pensión inicial), de modo que el soporte de la misma es $x \in [L, H]$. En general no hay restricciones de signo sobre L o H , pero en nuestra aplicación sus valores están acotados por los topes legales: la pensión mínima Bm y la máxima BM :

$$x \in [Bm, BM] \Rightarrow L \leq Bm \quad H \geq BM$$

modo más agregado, de forma que se respetan las diferencias empíricas observadas a nivel de las características observables más importantes: educación, sexo y grupo general de jubilación.

¹⁸ Por ejemplo, distribuciones de riqueza o de rentas que exceden un cierto límite externo (página 118 de Mood et al. (1974)). Una revisión extensiva de los modelos estadísticos utilizados para representar la distribución de rentas se encuentra en Kleiber and Kotz (2003).

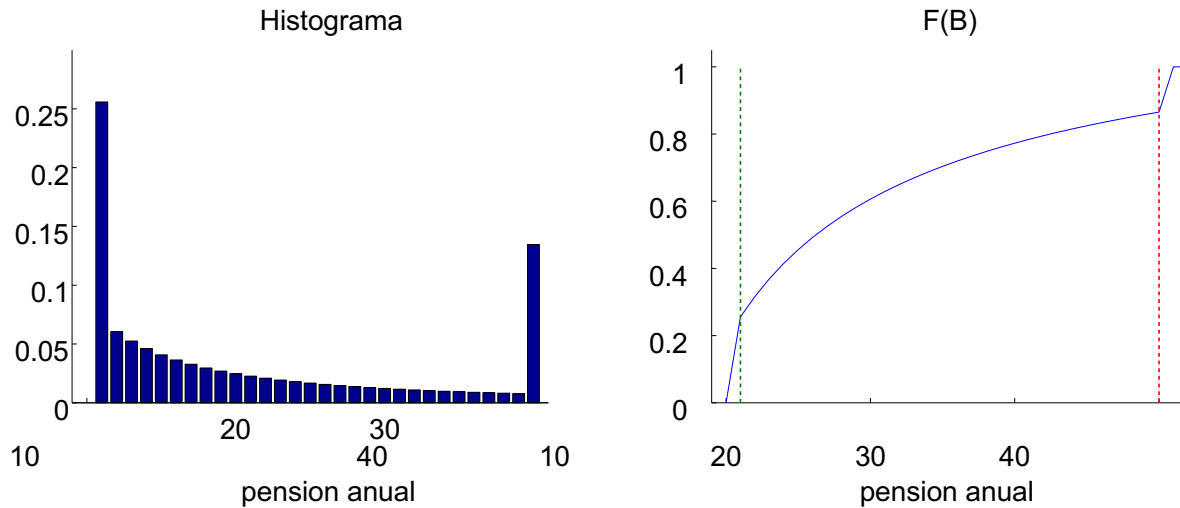


Figura 4: Histograma y función de distribución de Pareto truncada con parámetros $\alpha=0.7$, $L = 8$, $H = 70$ y topes legales $Bm = 11$, $BM = 40$ (miles de euros anuales) similares a los utilizados para algunos agentes del modelo en la simulación del año 2020.

Si no se cumplieren las desigualdades anteriores, existirían niveles de pensiones observables a los que el modelo atribuiría probabilidad nula. Con estas restricciones siempre se generan “picos” con valores razonables en los topes legales.¹⁹

- La densidad de probabilidad en esta distribución es $f(x) = \alpha L^\alpha x^{-\alpha-1} / (1 - (L/H)^\alpha)$, aunque resulta más operativo trabajar con su *función de distribución*:

$$F(x) = \frac{1 - (L/x)^\alpha}{1 - (L/H)^\alpha} \quad (2)$$

La probabilidad acumulada es una función potencial de la distancia (en ratio) entre el extremo inferior L y el valor x considerado, estandarizada por la distancia en ratio entre los dos límites de la distribución.

- La media poblacional de esta distribución tiene fórmula analítica cerrada:²⁰

$$E[x] = \frac{L^\alpha}{cte} \left(\frac{\alpha}{\alpha - 1} \right) \left(\frac{1}{L^{\alpha-1}} - \frac{1}{H^{\alpha-1}} \right) \quad \text{con} \quad cte = 1 - (L/H)^\alpha \quad (3)$$

El Gráfico 4 muestra un histograma y la función de distribución de la regla de Pareto para un caso típico de los valores paramétricos utilizados en el modelo de simulación en 2020. Para entender mejor las propiedades de $F(\cdot)$ en función de los parámetros implementados, el Gráfico 11 en el Apéndice A.3 muestra el comportamiento de esta ley en diversas combinaciones de parámetros (representando los histogramas generados y el valor del vector de estadísticos χ asociado).

¹⁹ La pensión observada sólo toma valores en $[Bm, BM]$, de forma que puede decirse que los límites $[L, H]$ del soporte de que $F(B)$ son límites *latentes* o *subyacentes* respecto de la variable observada.

²⁰ Esta propiedad es importante, dada la necesidad de computar este valor en un gran número de casos

Dinámica de la distribución

Los argumentos presentados en la Sección 3.1.3 sobre el impacto de los procesos de serie temporal y de ciclo vital en la distribución de ingresos de pensiones nos llevan a concluir que $F(x)$ debe poder cambiar en el tiempo. Para conseguir este resultado proponemos una especificación paramétrica de los extremos “latentes” L y H , manteniendo constante el parámetro α que da forma a la distribución:

$$L(B, Bm) = \beta_L B^\gamma Bm^\eta \quad H(B, BM) = \beta_H B^\gamma BM^\eta \quad (4)$$

donde B es el nivel de pensión al que se desea anclar la distribución $F()$ y Bm y BM representan la pensión mínima y máxima respectivamente.²¹ Es decir, asociamos a los parámetros específicos de cada tope (β_L, β_H) una función Cobb-Douglas del nivel de la pensión y del tope considerado. Proponemos la especificación Cobb-Douglas por ser flexible y bien conocida en economía. La necesidad de que $F(x)$ cambie con B y con los topes legales se argumenta como sigue:

- La dependencia respecto de B es enteramente natural. Las pensiones medias crecen en el tiempo de modo que el supuesto más sencillo es que $F(B)$ se desplaza tendencialmente “hacia la derecha” con su primer momento. Esto justifica la presencia de B en (4) con derivada parcial $\partial L / \partial B > 0$. Su forma, además, tiende a cambiar, ya que los aumentos en el valor medio de la pensión suelen acompañarse de aumentos en la dispersión.²² Esta propiedad puede conseguirse de nuestra forma funcional seleccionado adecuadamente los valores de los parámetros.
- La dependencia respecto de los topes se racionaliza con un argumento de flexibilidad. Si $F(x)$ no dependiese de los topes, un aumento en (por ejemplo) la pensión mínima Bm que excediese a una subida de la pensión media llevaría a un aumento de su incidencia Im . Esto es natural, pero la importancia cuantitativa de este aumento puede variar. Si Bm no está en la función de cálculo de L entonces depende de la curvatura de F en la cola inferior y de la sensibilidad de L respecto de B . Esta situación puede ser adecuada, pero la formulación de (4) permite un grado de libertad adicional. El caso en el que L no depende de los topes corresponde a $\eta = 0$, pero es posible que un $\eta > 0$ represente mejor la experiencia histórica. En la formulación de los ejemplos de la Sección 3.2.3 se ha usado un valor mayor que cero pero reducido, $\eta \ll \gamma$.

En resumen, el modelo queda especificado al asignar valores al vector de parámetros $\bar{\theta} = \{\alpha, \beta_L, \beta_H, \gamma, \eta\}$. A continuación, revisamos la calibración de estos parámetros y mostramos la flexibilidad de esta especificación para reproducir formas extremas de la distribución.

²¹ B es típicamente el nivel de pensión generado por el modelo para cada *agente representativo*. Recordemos que todo este esfuerzo de modelización procede de intentar reflejar la heterogeneidad de situaciones que subyacen a cada agente del modelo.

²² Como hemos visto en la Sección 3.1.2, esto no ocurre en todos los casos, pero sí en la mayoría.

Caso	Educación	Sexo	Jubilación	EST		
				$\bar{\mu}_B$	$\bar{I}m$	$\bar{I}M$
1. Extremo inferior	Baja	Femenino	65	11.3	0.41	0.02
2. Extremo superior	Alta	Varón	pre65	28.0	0.02	0.26
3. Disperso (colas grandes)	Alta	Femenino	65	17.0	0.33	0.19
4. Concentrado (colas pequeñas)	Baja	Varón	pre65	19.3	0.06	0.08

Cuadro 4: Selección de casos ilustrativos de la diversidad de formas en la distribución de ingresos de pensiones. Valores muestrales estimados en la sección cruzada para 2016 de la MCVL2017. Estadísticos (EST): ingreso medio de pensiones ($\bar{\mu}_B$), incidencia de la pensión mínima, $\bar{I}m$ e incidencia de la pensión máxima, $\bar{I}M$.

3.2.3. Calibración en la simulación de pensiones

El modelo queda plenamente operativo una vez que se han especificado los valores de los parámetros discutidos en la sección previa. Esta *calibración* paramétrica se realiza en dos fases, de acuerdo a una partición del espacio de parámetros en dos grupos: $\bar{\theta}_1 = \{\alpha, \beta_L, \beta_H\}$ se asignan para reproducir propiedades de la sección cruzada, mientras que $\bar{\theta}_2 = \{\gamma, \eta\}$ se determinan conforme al comportamiento de las series temporales agregadas.

1. FASE 1. Reproducción de la heterogeneidad de sección cruzada

El estudio empírico en la Sección 3.1.2 ha revelado un notable grado de heterogeneidad en la distribución de los ingresos de las altas de jubilación, resumida en los estadísticos $\{\bar{I}m, \bar{I}M, \bar{\mu}_B\}$ para cada *agente representativo* del modelo. Esta Fase 1 trata de introducir esta heterogeneidad en el modelo de equilibrio. Para ello ajusta los parámetros $\bar{\theta}_1$ de la ley de Pareto conforme a las características de educación, sexo y edad de jubilación de los agentes del modelo. Típicamente, identificamos cada uno de los parámetros en $\bar{\theta}_1$ con uno de los estadísticos χ que resumen las propiedades de $F(B)$: β_L se asigna para aproximar la incidencia de la pensión mínima, $\bar{I}m$; β_H se asigna para aproximar la incidencia de la pensión máxima, $\bar{I}M$, y α permite anclar el valor medio de los ingresos truncados de pensiones al nivel de pensión generado por el modelo, B (el cual, en virtud del proceso de calibración general del modelo, se alinea con el valor observable $\bar{\mu}_B$).

Como ilustración de este procedimiento, el Cuadro 4 presenta cuatro posibles configuraciones de la distribución de ingresos $F(B)$. Son casos extraídos directamente de la evidencia empírica de la MCVL, como puede comprobarse en las tablas de doble entrada del Cuadro 10 del Apéndice A.2. Muestran los estadísticos completos para una partición de la muestra por educación, sexo y edad de jubilación (agrupada en 3 categorías). Hemos seleccionado casos que reflejan configuraciones extremas de la distribución de pensiones: *concentradas* en uno de los extremos (Casos 1 y 2), *dispersa* en todo el soporte (Caso 3, que presenta picos elevados de probabilidad en ambos extremos) y *concentrada* en los valores medios del soporte (Caso 4, que presenta valores muy bajos de probabilidad en ambos extremos). Junto a los estadísticos proporcionados en los cuadros anteriores, las configuraciones elegidas pueden visualizarse (aproximadamente) en los histogramas del Gráfico 3. Utilizando el procedimiento

descrito en el párrafo previo, buscamos los valores paramétricos de la ley de Pareto que son capaces de aproximar cada uno de estos Casos.

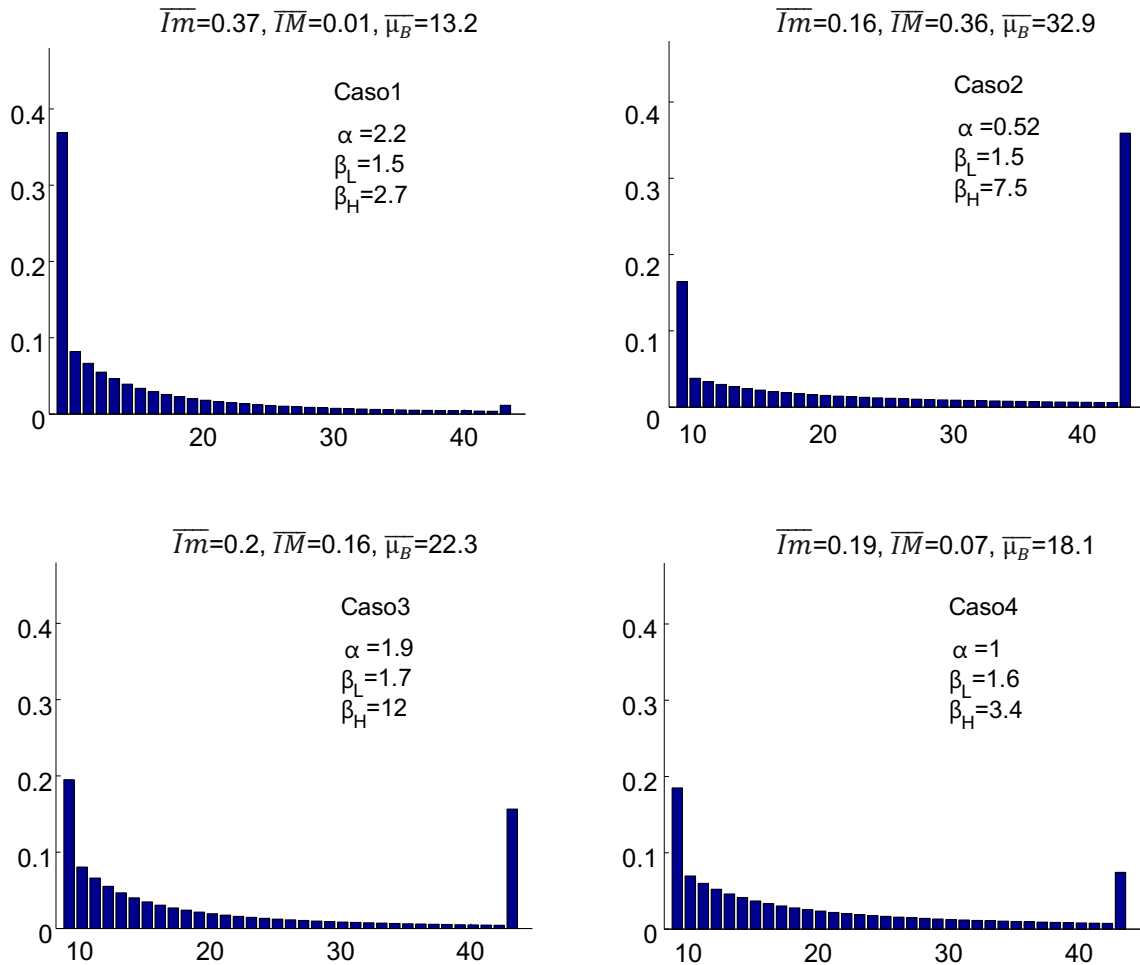


Figura 5: Reproducción con la distribución pareto-truncada de los casos ilustrativos de situaciones extremas del Cuadro 4

2. FASE 2. Reproducción del comportamiento agregado por Inferencia Indirecta

Los elementos en $\overline{\theta}_2$ determinan el comportamiento dinámico de las distribuciones intra-agente mediante la parametrización de su sensibilidad ante cambios en el nivel medio de la pensión y en los topes legales. Se calibran aplicando una técnica de *inferencia indirecta*, caracterizada por seleccionar aquellos valores de $\overline{\theta}_2$ que hacen que la simulación agregada del modelo de equilibrio reproduzca alguna serie temporal exógena que se toma como objetivo. En nuestro caso, elegimos la proyección del peso del gasto en pensiones sobre el PIB para España en la última entrega del Grupo de Envejecimiento de la CE (disponible en European-Commission (2020)). La razón para esta elección radica en que

ese trabajo es el ancla de calibración de nuestro ejercicio de evaluación de la contrarreforma de 2021.²³ Al contrario que con los parámetros en $\bar{\theta}_1$, γ y η se asignan de modo único sin desagregar para agentes de distintas características.

El procedimiento completo de asignación de parámetros itera recursivamente en $\bar{\theta}_2$. Operamos en 4 pasos:

Paso 1 se propone un valor para $\bar{\theta}_2$

Paso 2 Dado el paso 1, se asignan los valores en $\bar{\theta}_1$ para reflejar la heterogeneidad de sección cruzada.

Paso 3 Con un vector completo $\bar{\theta}$ se simula el modelo de equilibrio y se compara la previsión macroeconómica elegida (gasto en pensiones sobre PIB) con la serie objetiva.

Paso 4 Se itera el procedimiento hasta que la discrepancia entre la predicción del modelo y la serie objetivo se ha reducido a un nivel aceptable.

El Gráfico 6 muestra que este procedimiento de calibración tiene éxito, ya que nuestro modelo de pensiones ajusta de modo preciso la serie de gasto propuesta por el AWG.²⁴ El modelo también reproduce razonablemente bien los estadísticos resumen de la distribución de pensiones de la sección cruzada de 2020 (pensión media y tasas de incidencia de las pensiones máximas y mínimas).²⁵

²³ El objetivo de nuestro modelo es estudiar las últimas reformas de pensiones, pero nos exigimos como condición previa para su uso la alineación de las dos proyecciones (modelo y AWG) en el entorno previo a la reforma. La serie más importante en este sentido es la serie de gasto total en pensiones sobre PIB del Gráfico 6. El comportamiento de la pensión media y de las tasas de incidencia de los topes legales de pensiones que subyacen a la serie de gasto se muestran en el apéndice A.

²⁴ Por supuesto, este ajuste sólo se produce una vez alineados todos los procesos exógenos de nuestra simulación (series demográficas, productividad, mercado de trabajo y elementos discrecionales del sistema de pensiones) con los supuestos del AWG en Ministerio de Economía (2021).

²⁵ La incidencia de la pensión mínima en 2020 es del 21.3% por 21.2 en el modelo. Para la pensión máxima el modelo queda ligeramente por debajo de los datos observados: 8% frente a 8.5. Finalmente, la pensión media es de 14.2 miles de euros anuales (de 2019) por 13.9 en el modelo.

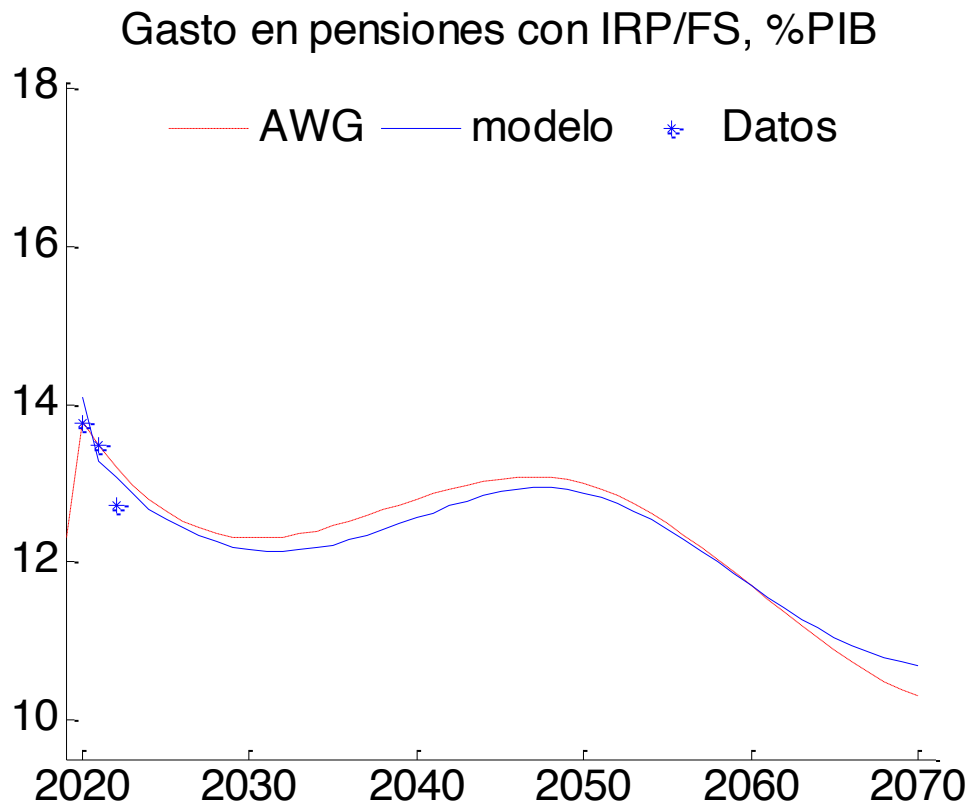


Figura 6: Serie temporal de gasto total en pensiones en ratio sobre el PIB del modelo con IRP/FS versus AWG2021

4. Ampliación de la heterogeneidad en bases de cotización

En esta sección presentamos las mejoras implementadas en la reproducción de la variabilidad en las bases de contribución en el modelo de simulación. La sección se estructura de modo similar a la anterior: revisamos la evidencia empírica (Sección 4.1), proponemos un modelo reducido de la evolución de la distribución (Sección 4.2) y lo aplicamos al modelo de simulación, calibrándolo conforme a la evidencia empírica antes elaborada (Sección 4.3).

4.1. Dispersión de las bases de cotización: evidencia empírica

El primer paso en el análisis es revisar la evidencia empírica sobre la distribución de las bases de cotización en España. Como vemos a continuación, hay varias fuentes que proporcionan detalles útiles sobre este tema. Aunque, en última instancia, la pieza más valiosa para esta tarea va a ser la MCVL (como ya lo fue en el caso de los ingresos por pensiones). Nos va a permitir, en particular, caracterizar la heterogeneidad de la distribución de las bases entre los grupos de individuos que conforman los *agentes representativos* del modelo. Esta

información proporciona el soporte empírico necesario para la calibración del modelo de dispersión de bases incorporado a la simulación de pensiones.

1. La primera fuente de información a considerar son las estadísticas y publicaciones de la Seguridad Social. Lamentablemente, éstas proporcionan pocos datos sobre las distribuciones de las bases contributivas. Así, TGSS (2022) es una publicación mensual específica sobre bases de cotización y cotizantes, pero centrada en información de serie temporal sobre medias o agregados poblacionales. Aporta información útil sobre la variación de las bases medias con variables incluidas en el modelo (como sexo o grupo de edad), pero no resulta suficiente para la calibración del modelo de simulación. Una referencia clásica en toda búsqueda de datos del Sistema de Pensiones es el *Boletín económico-financiero* que acompaña a los presupuestos (por ejemplo, Seguridad-Social (2021) o Seguridad-Social (2023)). El capítulo 3 de esta publicación es una recopilación excelente del comportamiento, tanto reciente como histórico, de los agregados de cotizaciones (cuotas totales absolutas y en porcentaje del PIB y bases medias) y de las cifras de afiliación. Así, se revisa la evolución histórica de estas variables desde 2009 desglosando por tipo de contingencia (comunes, profesionales, desempleo, cese de actividad de autónomos) y régimen de Seguridad Social. También se incluye un comentario sobre los cambios más significativos en la evolución de los topes legales de cotización, como parte de una revisión detallada de las modificaciones legislativas ocurridas con respecto a publicaciones previas. El documento de 2021 es especialmente interesante para conocer los efectos económicos de la crisis del Covid y los programas específicos puestos en marcha en respuesta a esa contingencia. Lamentablemente, no se aporta la información más importante para este trabajo: indicadores de la dispersión de las bases contributivas del sistema, incluyendo la incidencia de los topes superiores e inferiores.
2. El historial de los topes legales aplicados a las bases contributivas ha sido objeto de estudio en varios trabajos, típicamente como parte de estudios más amplios sobre la evolución temporal del sistema de pensiones. El grupo español del proyecto de comparación internacional de incentivos de jubilación liderado por D. Wise para el NBER es un buen ejemplo de este tipo de trabajos. Boldrin et al. (2001) cubre el período anterior a 2001, mientras que García-Gómez et al. (2021) revisa los desarrollos más recientes.
3. En de la Fuente (2023) se utilizan datos fiscales de la recaudación del IRPF para aproximar (entre otros estadísticos) el porcentaje de bases contributivas afectadas por el truncamiento de la base máxima. El objetivo de este cálculo es estimar la capacidad recaudatoria del “destope” de las bases máximas y de la “cuota de solidaridad” que son parte de la *contrarreforma* de pensiones incluida en BOE (2023). Los datos de rentas del trabajo provienen de los microdatos de la Muestra Anual de Declarantes del IRPF 2019 del IEF-AEAT. Permiten reconstruir las colas de la distribución de ingresos laborales, pero hay algunos problemas para identificar exactamente las bases contributivas de la Seguridad Social a partir de ellos. Esta base de datos carece, además, de alguna información desagregada que sí está disponible para la Seguridad Social.
4. La Seguridad Social proporciona información directa sobre las bases de cotización en su MCVL. Esta información resulta muy adecuada para nuestros propósitos, ya que la muestra es grande y permite un análisis muy desagregado de la distribución de las bases. Existe una literatura creciente que utiliza estos datos para analizar cuestiones distributivas

y de desigualdad asociadas a los ingresos laborales. Por ejemplo, Bonhomme and Hospido (2017) estudia la evolución de la desigualdad de ingresos en el intervalo entre 1988 y 2010 con especial atención a su relación con el ciclo económico. En un trabajo más reciente, Arellano et al. (2022), se estudia la dinámica de los ingresos laborales entre 2005 y 2018, enfatizando el desigual reparto del riesgo en este tipo de ingresos entre individuos que varían de acuerdo con diversas características observables.

Esta base de datos es la elegida para el componente empírico de la calibración del modelo de pensiones. En la sección siguiente se describe con detalle el uso de estos datos (la MCVL de 2017 en concreto) para estudiar las propiedades de la distribución de las bases contributivas en la fase inicial de la simulación de pensiones.

Dispersión de las bases contributivas en la sección cruzada con la MCVL

En esta sección repasamos las propiedades de la distribución de las bases contributivas utilizando la MCVL de la Seguridad Social. Por compatibilidad con el modelo de simulación, la variable objeto de estudio corrige las observaciones correspondientes a contratos a tiempo parcial expresándolas en términos “equivalentes a tiempo completo”. Para ello, *elevamos* los datos observados conforme al “grado de parcialidad” del contrato recogido por la Seguridad Social. La técnica empleada se describe en el Apéndice B.1 (que revisa el procedimiento completo de elaboración de los datos originales). La MCVL del año 2017 se utiliza como base de datos fundamental para la calibración (utilizando las observaciones para el año precedente), recurriéndose ocasionalmente la información de la MCVL de 2021 para completar algunos aspectos de la calibración.²⁶

El histograma en la Figura 7 proporciona una visión de conjunto de la distribución de las bases de cotización *equivalentes*. El valor medio de las bases mensuales se estima en algo menos de 1.500 euros, mientras que la incidencia del máximo legal se estima en el 5,5% de la población y la de *alguna de los mínimos* se estima en el 30.1% del total poblacional. La existencia en el gráfico de numerosos “picos” muy marcados es reflejo de la fragmentación de la legislación aplicable en función de las circunstancias del trabajador.²⁷ Para medir la incidencia de los diversos topes inferiores en el conjunto de la población procedemos a calcular la incidencia de los topes específicos en cada subgrupo poblacional y la promediamos conforme a los pesos de cada subgrupo. El Gráfico 13 del Apéndice B.1 muestra la distribución de bases en los diversos subgrupos considerados: autónomos versus asalariados y dentro de

²⁶ El Apéndice B.1 también discute los problemas de representatividad de los resultados obtenidos con la MCVL de 2021, debido a la anómala situación de la economía durante la crisis del Covid de 2020. Las peculiares características de ese año desaconsejan su uso como año base de la calibración del modelo.

²⁷ Existen normativas diferentes para los trabajadores por cuenta propia (encuadrados en el Régimen Especial de Trabajadores Autónomos) frente los trabajadores por cuenta ajena (afiliados al Régimen General de la Seguridad Social). El régimen de autónomos se encuentra en transformación con la entrada en vigor del nuevo sistema de cotización (BOE, 2022) que reduce sustancialmente la libertad de elección de bases que caracterizó a este colectivo en el pasado. El sistema previo se analiza en Sánchez-Martín (2019), mientras que los detalles de la nueva configuración pueden estudiarse en la web de la Seguridad Social: <https://www.seg-social.es/wps/portal/wss/internet/HerramientasWeb/9d2fd4f1-ab0f-42a6-8d10-2e74b378ee24/32023cb>. Para asalariados, el RGSS establece bases mínimas de cotización diferentes según el “grupo contributivo” del trabajador: “1” (Ingenieros y licenciados), “2” (Ing. técnicos, peritos y ayudantes titulados) , “3” (Jefes administrativos y de taller) y “4 y superior” (Ayudantes, oficiales, subalternos y auxiliares administrativos). Los valores en vigor en un año concreto pueden encontrarse en la página web de la Seguridad Social. Por ejemplo, para el año en curso la dirección-web es: <https://www.seg-social.es/wps/portal/wss/internet/Trabajadores/CotizacionRecaudacionTrabajadores/36537#36538>.

estos, los 4 grupos diferenciables por el valor legal de su base mínima. Los valores obtenidos con este procedimiento pueden consultarse en las columnas de *Totales* de los Cuadros 5 y 6 (para la incidencia agregada de los topes inferior y superior respectivamente) y 7 para los valores medios de las bases en 2016.

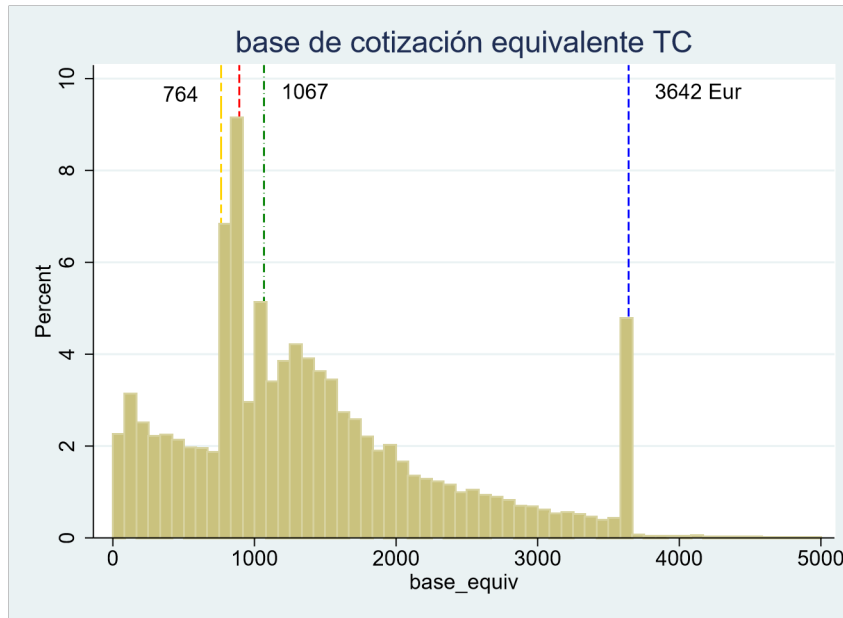


Figura 7: Histograma de la distribución de las bases de cotización mensuales para asalariados y autónomos en la MCVL de 2016. Se incluyen los valores vigentes de la base máxima (3642 Eur) y de las bases mínimas de cotización de autónomos (893 Eur) y asalariados de los grupos de tarifa 1 (1067 Eur) y 4 (764 Eur).

Para la calibración de nuestro modelo de simulación es muy conveniente explorar la variabilidad asociada a las variables que definen a los *agentes representativos del modelo*: el nivel educativo y la edad (junto con el sexo de los adultos cabeza del hogar).²⁸ De esa tarea se encarga el Gráfico 8 y los Cuadros 5 a 7 antes mencionados. Encontramos una variedad apreciable en las pautas de distribución de las bases contributivas:

Las bases medias son marcadamente mayores para los individuos de educación superior y, en menor medida, para los varones frente a las mujeres (Cuadro 7). Por grupo de edad se observa el perfil cóncavo característico de los procesos de ingresos durante el ciclo vital, alcanzándose el máximo entre

²⁸ Los *agentes representativos* del modelo también difieren por su edad de jubilación, pero esta variable no es relevante en términos de la cotización laboral de los agentes.

50 y 59 años para los varones y entre 40 y 49 para las mujeres.²⁹ Resulta reseñable la superioridad de las bases de cotización de las mujeres respecto de los varones en los grupos de edad más jóvenes.³⁰

La incidencia de los topes de pensiones correlaciona del modo esperable con las posiciones relativas de las bases medias: la probabilidad de estar en el mínimo legal es casi el doble para los trabajadores de menor nivel educativo frente a los de educación superior, y apreciablemente decreciente con la edad (Cuadro 5). Para el tope superior las pautas se invierten: la incidencia crece muy marcadamente con el nivel educativo y con la edad (Cuadro 6). Las diferencias por sexo son pequeñas en la incidencia de las bases mínimas y apreciables en el caso de la base máxima (aunque, de nuevo, se aprecia un cambio generacional en las pautas de las cohortes más jóvenes).

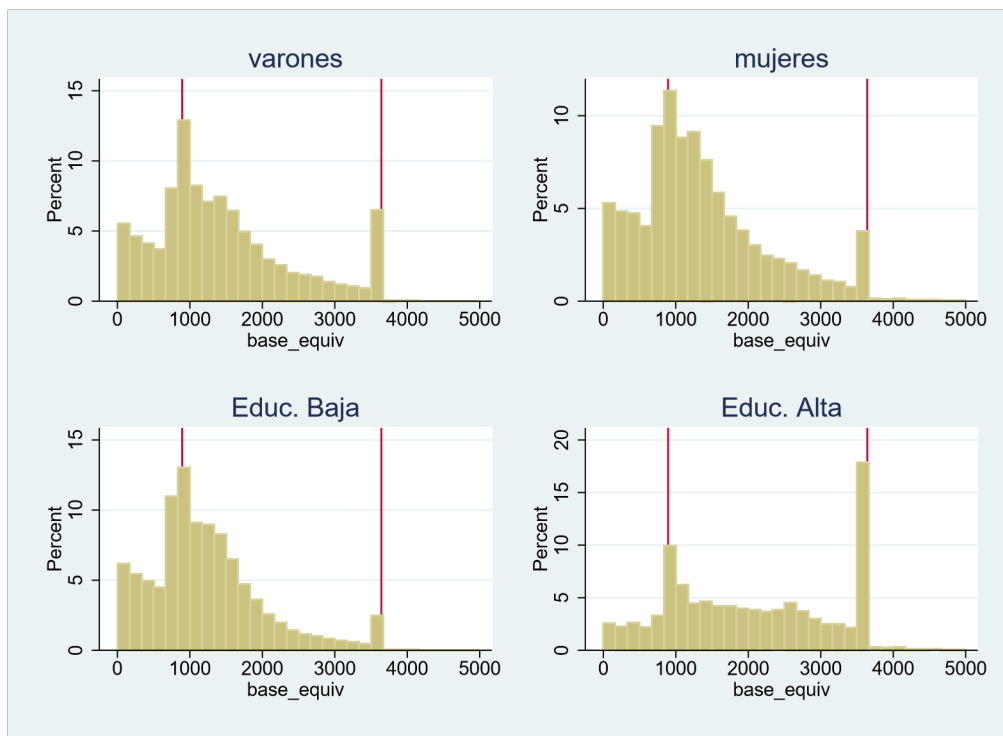


Figura 8: Distribución de las bases de cotización en la MCVL de 2016: histogramas por sexo y nivel educativo.

²⁹ Como es bien conocido, este resultado procede tanto de la variación de ingresos en el ciclo vital como de la evolución de serie temporal de los ingresos.

³⁰ Las mayores bases de las mujeres de las cohortes más jóvenes se acompañan de una apreciable superioridad en sus niveles educativos, así como en el peso de los grupos de tarifa asociados a mayor responsabilidad/cualificación. Estos resultados dependen del tratamiento del grado de parcialidad de los contratos.

	Sexo		
	Varón	Mujer	Total
	Incidencia%	Incidencia%	Incidencia%
Nivel educativo			
Bajo	0.338	0.348	0.343
Medio	0.257	0.247	0.252
Alto	0.183	0.192	0.188
Total	0.305	0.296	0.301
Grupo de edad			
menos de 20	0.588	0.518	0.561
entre 20 y 29	0.368	0.348	0.358
entre 30 y 39	0.304	0.277	0.291
entre 40 y 49	0.301	0.291	0.296
entre 50 y 59	0.283	0.293	0.288
60 y mas	0.252	0.278	0.264
Total	0.305	0.297	0.301

Cuadro 5: Incidencia de la base mínima de cotización en la MCVL de 2016 por sexo, educación y grupo de edad.

	Sexo		
	Varón	Mujer	Total
	Incidencia%	Incidencia%	Incidencia%
Nivel educativo			
Bajo	0.034	0.018	0.027
Medio	0.078	0.037	0.055
Alto	0.232	0.146	0.184
Total	0.065	0.044	0.055
Grupo de edad			
menos de 20	0.005	0.010	0.007
entre 20 y 29	0.011	0.018	0.015
entre 30 y 39	0.042	0.042	0.042
entre 40 y 49	0.073	0.053	0.063
entre 50 y 59	0.097	0.050	0.075
60 y mas	0.129	0.058	0.096
Total	0.065	0.044	0.055

Cuadro 6: Incidencia de la base máxima de cotización en la MCVL de 2016 por sexo, educación y grupo de edad.

Nivel educativo	Sexo		
	Varón	Mujer	Total
	base media	base media	base media
Bajo	15.8	14.5	15.3
Medio	20.0	19.1	19.5
Alto	26.7	27.2	27.0
Total	17.9	17.8	17.8
Grupo de Edad			
menos de 20	8.3	10.4	9.1
entre 20 y 29	13.6	14.7	14.1
entre 30 y 39	17.1	18.9	18.0
entre 40 y 49	18.7	19.0	18.9
entre 50 y 59	19.3	17.3	18.4
60 y mas	21.9	17.6	19.9
Total	17.9	17.8	17.8

Cuadro 7: Base media de cotización (equivalente anual en miles de euros) en la MCVL de 2016 por sexo, educación y grupo de edad

El lector interesado en una descripción más detallada de estas diferencias puede acudir a las Tablas 11 y 12 del Apéndice B. En ellas se muestra un doble condicionamiento simultáneo en grupo de edad y educación (para cada sexo), que utilizamos en la Sección 4.2 para construir ejemplos ilustrativos de las pautas más extremas observables en la distribución de las bases.

Dinámica de la dispersión de los ingresos laborales

La dinámica de la evolución de las bases contributivas es reflejo directo de la dinámica de los ingresos laborales salvo en los extremos de la distribución (que están sujetos a los topes institucionales y responden, por tanto, a cambios discrecionales en la política con respecto a la Seguridad Social). Los ingresos laborales muestran una doble dimensión de serie temporal/ciclo vital similar a la discutida en la Sección 3.1.3. La pauta del componente institucional es cambiante con el ciclo económico y con la aproximación a la problemática de las pensiones de los sucesivos gobiernos (como se muestra en las referencias citadas al comienzo de la Sección 4.1, siendo García-Gómez et al. (2021) de especial relevancia en este sentido).³¹

³¹ La sucesión de Boletines Económico-Financieros que acompañan a las propuestas de presupuestos de la Seguridad Social es, posiblemente, la fuente más detallada de los sucesivos cambios introducidos en la normativa de pensiones y cotizaciones.

De cara a la proyección del futuro de las cotizaciones serían, en principio, de aplicación los mismos enfoques discutidos en la sección 3.1.3. De entre ellos, el enfoque de *inferencia indirecta* parece la técnica más prometedora en nuestro caso. Los estudios basados en la reproducción de las pautas históricas no son muy atractivos en este caso por dos razones: (i) por la presencia de un marcado cambio estructural en el mercado de trabajo y (ii) por el condicionamiento de la evolución de esta variable a la (siempre cambiante) política de pensiones en España. Simplemente, no hay garantía de que las regularidades pasadas proporcionen una guía útil sobre la evolución futura de estas variables. Es más razonable tratar de condicionar la dinámica futura a especificaciones concretas de la política futura de pensiones. La *contra-reforma* de pensiones de 2023 es un ejemplo, al incluir a los topes de cotización como variables instrumentales de una política de pensiones específica. Las posibilidades de calibración del modelo de pensiones con inferencia indirecta es el objeto de estudio del último apartado (4.3) de esta Sección.

4.2. Modelización de la heterogeneidad de las bases contributivas

Las consideraciones teóricas respecto de la modelización de la distribución de bases contributivas son similares a las presentadas en la Sección 3.2 para las pensiones. Su aplicación al modelo de equilibrio de gasto en pensiones se realiza a través de la misma forma funcional: la distribución de Pareto Truncada desarrollada en la Sección 3.2.2. Todas las consideraciones teóricas presentadas entonces son de aplicación aquí en la modelización de las bases de cotización.

La capacidad de esta forma funcional para reflejar las regularidades empíricas se explora (como se hizo en la Sección 3.2.3) por medio de ejemplos. Así, el Cuadro 8 recoge los estadísticos básicos de la distribución empírica de las bases de 4 *agentes representativos* (por educación, cohorte y sexo). Son casos extraídos a partir de los estadísticos detallados de la MCVL presentados en las Tablas 11 y 12 del Apéndice B. Los casos seleccionados reflejan configuraciones extremas de la distribución de bases de cotización: los Casos 1 y 2 muestran distribuciones *concentradas* en cada uno de los dos extremos de la distribución; el Caso 3 muestra una distribución *dispersa* en todo el soporte con picos elevados de probabilidad en ambos extremos; finalmente, el Caso 4, presenta una distribución bastante *concentrada* en que la probabilidad de los valores extremos es relativamente reducida. Comparado con el Caso 4 de la sección de pensiones (Cuadro 5), es claro que las bases contributivas están más dispersas (en relación a los topes legales) de modo que las distribuciones muy concentradas no son muy representativas de la heterogeneidad real de las bases.

Caso	Educación	Sexo	Grupo Edad	EST		
				$\bar{\mu}_B$	$\bar{I}m$	$\bar{I}M$
1. Extremo inferior	Baja	Fem.	20/29	12.5	0.38	0.01
2. Extremo superior	Alta	Mas.	60 y más	32.2	0.15	0.39
3. Equilibrado (colas similares)	Alta	Mas.	30/39	25.6	0.19	0.16
4. Concentrado (colas reducidas)	media	Fem.	40/49	19.2	0.22	0.06

Cuadro 8: Selección de casos ilustrativos de la diversidad de formas en la distribución de bases de cotización. Valores muestrales estimados en la sección cruzada para 2016 de la MCVL2017. Estadísticos (EST): ingreso medio de pensiones ($\bar{\mu}_B$), incidencia de la pensión mínima, $\bar{I}m$, e incidencia de la pensión máxima, $\bar{I}M$.

Para replicar del mejor modo posible estos casos extremos seguimos el procedimiento descrito en la Sección 3.2.3, que particiona el espacio paramétrico en dos grupos y vincula propiedades de los datos con parámetros específicos de la distribución Pareto Truncada. El resultado de este procedimiento puede apreciarse en los gráficos recogidos en la Figura 9. La cabecera de cada gráfico reproduce los estadísticos resumen de cada configuración y los valores paramétricos que los generan. En líneas generales vemos que el modelo reproduce con éxito las configuraciones empíricas seleccionadas. El ajuste a los valores empíricos no es perfecto, pero si lo suficientemente bueno para aceptar su uso en el modelo de simulación de pensiones que, al centrarse en *agentes representativos*, tiene que lidiar con casos menos extremos que los recogidos en el Cuadro anterior. De este ejercicio concluimos que la forma funcional Pareto-truncada es lo suficientemente flexible para incorporar en el modelo de simulación los hechos estilizados básicos de la heterogeneidad en las bases contributivas. Para culminar el proceso de calibración nos falta un mecanismo de selección de los valores de los parámetros adecuada para la simulación de pensiones. Discutir el mecanismo que realiza esta asignación en nuestro caso es la tarea de la sección siguiente.

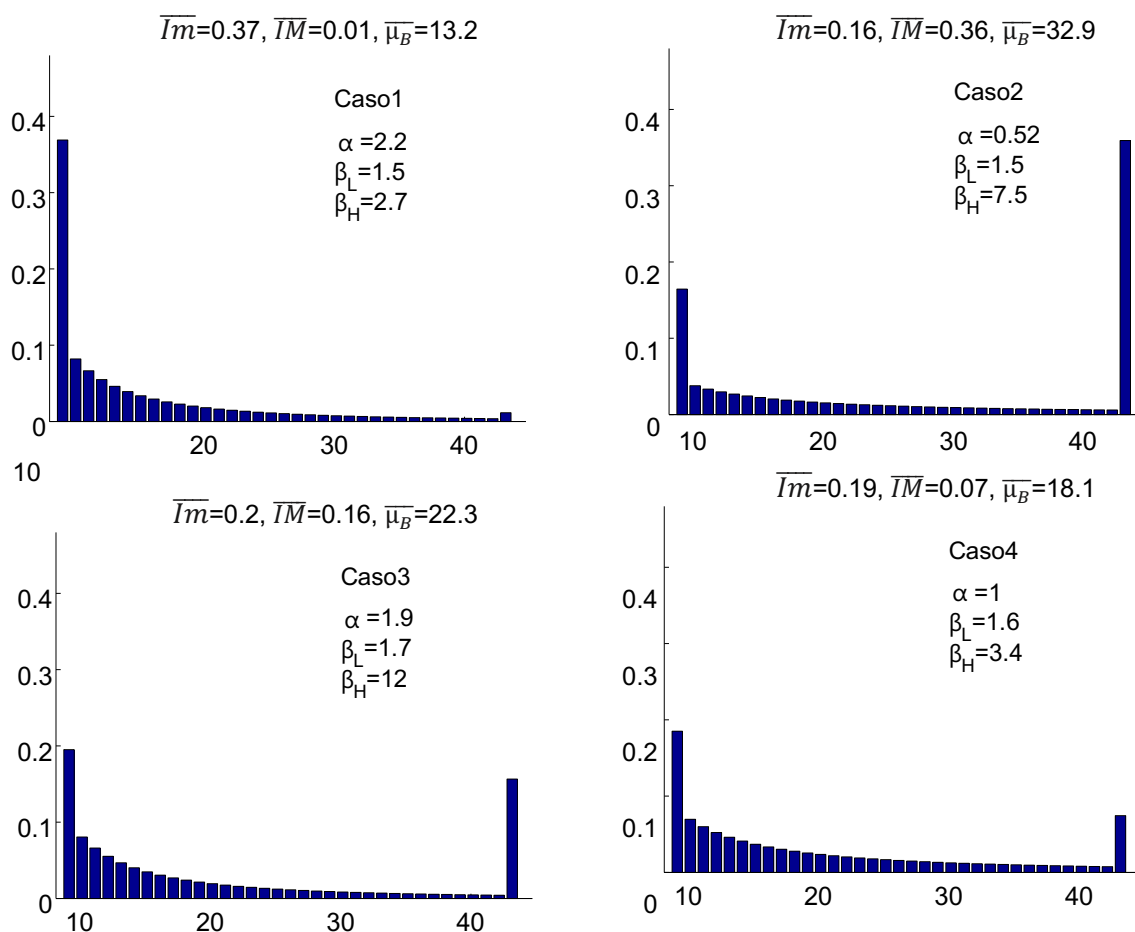


Figura 9: Ejemplos de calibración: resultados del modelo de Pareto Truncado aplicado a los cuatro *casos* empíricos ilustrativos de distribuciones extremas del Cuadro 8.

4.3. Resultados de calibración de la simulación de pensiones

La prueba de fuego para el modelo reducido de la distribución de bases de la sección previa es su comportamiento una vez incluido en el modelo de equilibrio del sistema de pensiones. En general, su programación en fortran-90 resulta sencilla y su integración con el resto del código informático no supone ninguna dificultad reseñable. La fase de *calibración* requiere más elaboración, ya que hay que asignar valores cuantitativos a varios parámetros y seleccionar los estadísticos empíricos adecuados como objetivos finales de la simulación. Para esta fase hemos adaptado el procedimiento en dos etapas descrito en la Sección 3.2.3 al calibrar las distribuciones de ingresos de pensiones. Las dos etapas se implementan como sigue:

- Realizamos la misma partición del espacio paramétrico en dos grupos $(\overline{\theta}_1, \overline{\theta}_2)$. El primero se enfoca a la reproducción de las condiciones iniciales y el segundo al comportamiento de largo plazo de la simulación.
- Los elementos del vector de parámetros $\overline{\theta}_1 = \{\alpha, \beta_L, \beta_H\}$ se asignan para reproducir los estadísticos descriptivos de la distribución de bases en la sección cruzada al inicio de la simulación: incidencia de los topes legales y valor medio de las bases. Tomamos como año de calibración el correspondiente a la evidencia empírica descrita en la Sección 4.1.³² Ajustamos los parámetros de los diversos *agentes representativos* del modelo (diferenciados por educación y sexo, pero no por cohorte de nacimiento) para reproducir las regularidades discutidas en la sección previa.
- Los elementos del segundo vector de parámetros, $\overline{\theta}_2 = \{\gamma, \eta\}$, son claves para moldear el comportamiento de la simulación en el largo plazo, es decir, en la senda de proyección. Lógicamente, no hay contrapartida empírica inmediata con la que hacer este ejercicio, de modo que es habitual recurrir a ejercicios de *Inferencia Indirecta*. Esta fue la solución empleada en el caso de los ingresos de pensiones. Las series “objetivo” más naturales en el caso de las bases contributivas son las series de incidencia de los topes contributivos en la proyección del informe para España del AWG, Ministerio-Economía (2021). Los supuestos utilizados en este ejercicio son, sin embargo, muy simplificados y dan lugar a una solución esencialmente trivial (tanto las bases medias como los topes crecen con los salarios nominales, haciendo que estas tasas sean constantes en el tiempo).³³ Es posible intentar reproducir esa constancia en nuestra simulación, pero hemos optado por una estrategia de calibración más sencilla: utilizamos los valores paramétricos aplicados para los ingresos de pensiones (cuyo comportamiento nos resulta bien conocido) y estudiamos hasta qué punto se aleja el resultado obtenido de la constancia en las tasas de incidencia. Como vemos a continuación, esta estrategia se comporta bien y evita una dificultad sobrevenida al modelo de simulación: la distorsión de las condiciones iniciales en 2020 por el shock del covid19. En conjunto, parece más adecuado dejar que el modelo recupere suavemente unas tasas de incidencia estables a lo largo de la simulación que imponerlas desde el inicio en lo que podría ser una condición de partida errónea.

³² El apéndice B.1 revisa las razones por las que encontramos esos datos más adecuados que los obtenidos en la MCVL 2021.

³³ Página 12 del informe para España, Ministerio-Economía (2021)

Los resultados del proceso de calibración se muestran en el Cuadro 9 y en la Figura 10. En ellos observamos que el modelo tiene un comportamiento razonable en la fase “de prueba” previa a la proyección: reproduce con precisión las tasas de incidencia observadas en los datos (incluidas en los gráficos) y aproxima el valor de la base media en el año de calibración. El modelo genera un aumento apreciable en la incidencia de la base mínima (del 33 al 37%) en reflejo de la gran subida en el valor asignado al tope inferior de cotización introducida entre 2018 y 2019. Por ejemplo, para el grupo de cotización más numeroso (el correspondiente a las categorías 4 a 7) el aumento fue de 858 a 1050 euros al mes, reflejando una subida histórica del 22.3% en el SMI: la mayor subida observada en 4 décadas de funcionamiento del Sistema. Aparte, el modelo genera una nueva subida en la incidencia de las bases mínimas en 2020, esta vez asociada al shock covid. Los datos reflejan en parte esta subida, aunque el modelo la exagera al modelizar la caída de la producción en 2020 como una combinación de caídas en el empleo y en la productividad del trabajo (que en el modelo implica caídas en las rentas laborales). En el mundo real, varios programas *ad hoc* de protección de rentas amortiguaron este efecto. Por esta razón, evitamos tomar los estadísticos de 2020 como la referencia básica para la senda de proyección posterior.

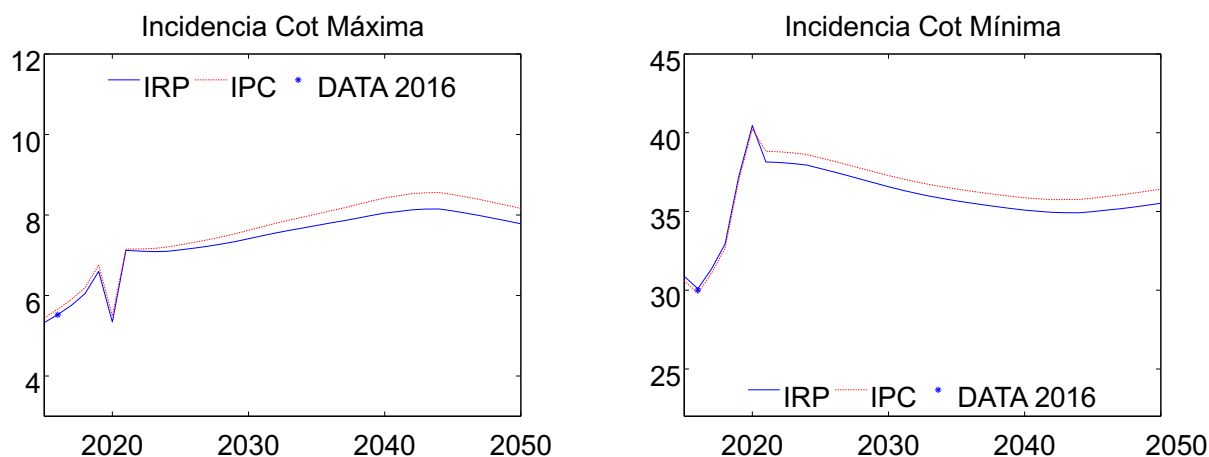


Figura 10: Proyección en el modelo de equilibrio de la incidencia de los topes superior (panel izquierdo) e inferior (panel derecho) en las bases de cotización. IRP= Economía que actualiza el stock de pensiones con el índice de Revalorización de Pensiones de 2013; IPC= Economía que actualiza el stock de pensiones con el IPC. Dato de 2016 obtenido en la MCVL 2017.

	\bar{m} %	\bar{M} %	base media de cotización
DATOS	30.0	5.5	19.6
Modelo	30.1	5.5	20.3

Cuadro 9: Estadísticos de calibración de la distribución de bases de cotización en el modelo de pensiones (calculado en el año 2016). \bar{m} =Incidencia de la base mínima; \bar{M} =Incidencia de la base máxima

Durante la fase de proyección que sigue a 2020 las tasas de incidencia de los topes se mueven muy suavemente en la dirección esperada, tanto si las pensiones se actualizan con el IRP como si se utiliza el IPC para ello. La incidencia de la base mínima se recupera del pico parcialmente “artificial” creado por el modelo en 2020 y se mantiene en una trayectoria muy suavemente decreciente que refleja los cambios subyacentes en la composición de la población (esencialmente el aumento en el nivel educativo medio de la fuerza laboral). Para la base máxima se observa un crecimiento muy suave. Es importante indicar que, de acuerdo con los supuestos de la simulación del AWG, el modelo actualiza los topes legales de las bases con el crecimiento salarial. No sorprende, por tanto, que no se produzca un aumento importante en la incidencia del tope máximo de contribución. Al mismo tiempo, las pensiones máximas sólo se mantienen constantes en términos reales, de modo que nuestra simulación incluye, de facto, al igual que el informe del AWG un “destope” bastante parecido al legislado en la reforma de pensiones de marzo de 2023.

Referencias

- Airef (2022, marzo). Documento técnico sobre el impacto de las reformas del sistema de pensiones entre 2021 Y 2023
- Arellano, M, S Bonhomme, M DeVera, L Hospido, and S Wei (2022, January). Income risk inequality: Evidence from Spanish administrative records. *Quantitative Economics* (13(4)), 1747–1801.
- BOE (2022, 27 de julio). Real decreto ley 13/2022 de 26 de julio por el que se establece un nuevo sistema de cotización para los trabajadores por cuenta propia o autónomos y se mejora la protección por cese de actividad”. Num 179, pag 107437 a 107500.
- BOE (2023, 16 de marzo). Real decreto ley 2/2023 de 16 de marzo de medidas urgentes para la ampliación de derechos de los pensionistas, la reducción de la brecha de género y el establecimiento de un nuevo marco de sostenibilidad del sistema público de pensiones”.
- Boldrin, M, S. Jiménez-Martín, and F. Peracchi (2001). Sistema de pensiones y mercado de trabajo en España. BBVA.
- Bonhomme, S. and L Hospido (2017). The cycle of earnings inequality in Spain: evidence from Spanish social security data. *The Economic Journal* (127(603)), 1244–1278.
- de la Fuente, Angel (2023, Marzo). Los efectos presupuestarios de la reforma de pensiones: un balance provisional. *Apuntes FEDEA 2003/06*.
- European-Commission (2020, May). The 2021 Ageing Report. economic and budgetary projections for the EU Member States (2019-2070). Institutional Paper (148).
- García-Gómez, P, S. García-Mandicó, Sergi Jiménez, and J. Vall-Castelló (2021). Social Security programs and Retirement around the world: Reform and Retirement incentives, Chapter 9, Trends in employment and Social Security Incentives in the Spanish pension system 1980-2016. NBER.
- Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and Actuarial Science*. J Willey and sons.
- Ministerio de Economía (2021). Ageing report 2021, Spain country fiche. Technical report.
- Mood, A, F Graybill, and D Boes (1974). *Introduction to the theory of statistics*. McGraw-Hill.
- Sánchez-Martín, Alfonso R (2019). Capacidad económica y pautas de cotización y formación de pensiones de los trabajadores autónomos en España: evidencia empírica en el intervalo 2008/2017. *Estudios sobre la Economía Española 2019/27*, FEDEA.
- Sánchez-Martín, A. R. (2023a, Abril). Actualización del modelo de Simulación del Sistema de Pensiones MSSP-OLG: i) una visión general. Documento de Trabajo dt2023-03, FEDEA.
- Sánchez-Martín, A. R. (2023b). Actualización del modelo MSSP-OLG: incorporación de “shocks” inesperados al entorno económico. Documento de Trabajo forthcoming, FEDEA.
- Seguridad-Social (2021). Proyecto de presupuestos ejercicio 2021. Informe económico y financiero. Technical report, Secretaría de Estado de la Seguridad Social.
- Seguridad-Social (2023). Proyecto de presupuestos ejercicio 2023. Informe económico y financiero. Technical report, Secretaría de Estado de la Seguridad Social.

TGSS (2022, Diciembre). Estadística de bases medias de cotización del régimen general. diciembre 2022.

<https://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/c9bd51e5-79bc-44f2-9618-14d6094eb585>.

A. Heterogeneidad en ingresos de pensiones

Esta sección proporciona información complementaria a la presentada en las Secciones 3.1 y 3.2 del texto principal:

- La Sección A.1 discute cómo utilizar la información en el Cuadro IV.30 del Boletín económico-financiero de la Seguridad Social de 2021 para estimar la incidencia de los topes legales de pensiones en un año determinado.
- El Cuadro 10 de la Sección A.2 explora con detalle los datos de ingresos de pensiones de las altas de jubilación en la MCVL 2017. Se explora un doble condicionamiento por dos de las características que definen a los *agentes representativos* del modelo: educación y edad de jubilación (para cada sexo). Se proporcionan los estadísticos que resumen la información de la distribución: media y tasas de incidencia de ambos topes. Esta información permite identificar agentes con distribuciones extremas, que van a servir como test de la flexibilidad del modelo reducido de la distribución introducido en la Sección 3.2.
- La Figura 11 explora rangos amplios del espacio paramétrico del modelo de distribución de ingresos de la Sección 3.2. Los gráficos muestran los estadísticos resumen en cada combinación de parámetros para una pensión inicial del modelo en el intervalo entre 10 y 50 mil euros anuales. El experimento es una ilustración gráfica del comportamiento del modelo y de la amplia variedad de soluciones que pueden conseguirse con el mismo.
- La Figura 12 muestra las series temporales de las pensiones medias y de la incidencia de las pensiones mínimas y máximas generadas por el modelo en los entornos con y sin IRP/FS. Las pautas mostradas son determinantes para el éxito del proceso de calibración de los parámetros de la distribución de Pareto-Truncada (Sección 3.2.3). En este trabajo estas series no se calibran directamente a sus homólogas del AWG. Esto es debido a que la proyección del valor real medio de las pensiones contributivas del AWG nos parece excesivamente baja (y, en consecuencia, el número de pensiones algo mayor que las generadas por el modelo). En este sentido nuestras simulaciones son más similares a las generadas por Airef (2023)

A.1. Datos de ingresos de pensiones en el Informe Económico-financiero de la Seguridad Social

La información original proporcionada en el Cuadro IV.30 del Boletín Económico-Financiero que acompaña a los Presupuestos (por ejemplo, Seguridad-Social (2021)), permite la estimación de la incidencia de los topes en el stock de pensiones de jubilación, pero sólo después de una cierta elaboración de la información. Específicamente, es preciso tratar los datos originales de dos maneras:

- Homogeneizando las categorías de ingresos en la cola izquierda de la distribución.
- Modelizando la cola de ingresos superiores, cuyos datos se proporcionan agregados.

Así, mientras que las categorías intermedias se agrupan en intervalos de 100 euros al mes, los ingresos a partir de 2000 euros se han agrupado en dos categorías (entre 2000 y 2683 y superiores a 2683). No es posible, en estas condiciones, determinar de modo preciso el porcentaje de pensiones para las que el tope superior es vinculante. De hecho, aunque el texto del Boletín indica el porcentaje de las pensiones que están en el límite superior (4.6%), la información del Gráfico sólo permite concluir que un 10.82% se encuentra por encima de 2000 euros. Es preciso modelizar la distribución en los grupos de cuantía intermedios entre 2000 y la pensión máxima para estimar la cifra deseada. Hemos comprobado que un procedimiento sencillo como la extrapolación lineal de la tasa de crecimiento de los grupos inmediatamente precedentes permite estimar con acierto la cifra exacta proporcionada en el texto (para las pensiones totales). Esto nos permite tener confianza para aplicar este método en *tipos* específicos de pensiones (jubilación, viudedad, etc).

A.2. Datos de pensiones en la MCVL-2017

Educ	Jubilación		
	Antes 65	65	Después 65
Bajo	19.33	13.27	14.98
	0.08	0.04	0.10
	0.06	0.19	0.31
Medio	25.52	18.43	15.86
	0.20	0.16	0.14
	0.05	0.25	0.23
Alto	28.03	21.43	23.27
	0.26	0.31	0.39
	0.02	0.19	0.23

Educ	Jubilación		
	Antes 65	65	Después 65
Bajo	15.33	11.27	10.32
	0.04	0.02	0.01
	0.24	0.41	0.54
Medio	23.58	15.04	13.87
	0.14	0.06	0.07
	0.11	0.24	0.31
Alto	28.46	16.97	22.04
	0.25	0.19	0.33
	0.09	0.33	0.24

Cuadro 10: Estadísticos de altas de pensiones en 2016 condicionados por nivel educativo, edad de jubilación y sexo (varones en el panel superior y mujeres en el panel inferior). En cada combinación se reproduce la pensión media en miles de euros anuales y las tasas de incidencia de la pensión máxima y mínima en tanto por uno. Los resultados se estiman con los microdatos de la MCVL 2017

A.3. Distribución de Pareto Truncada aplicada a pensiones

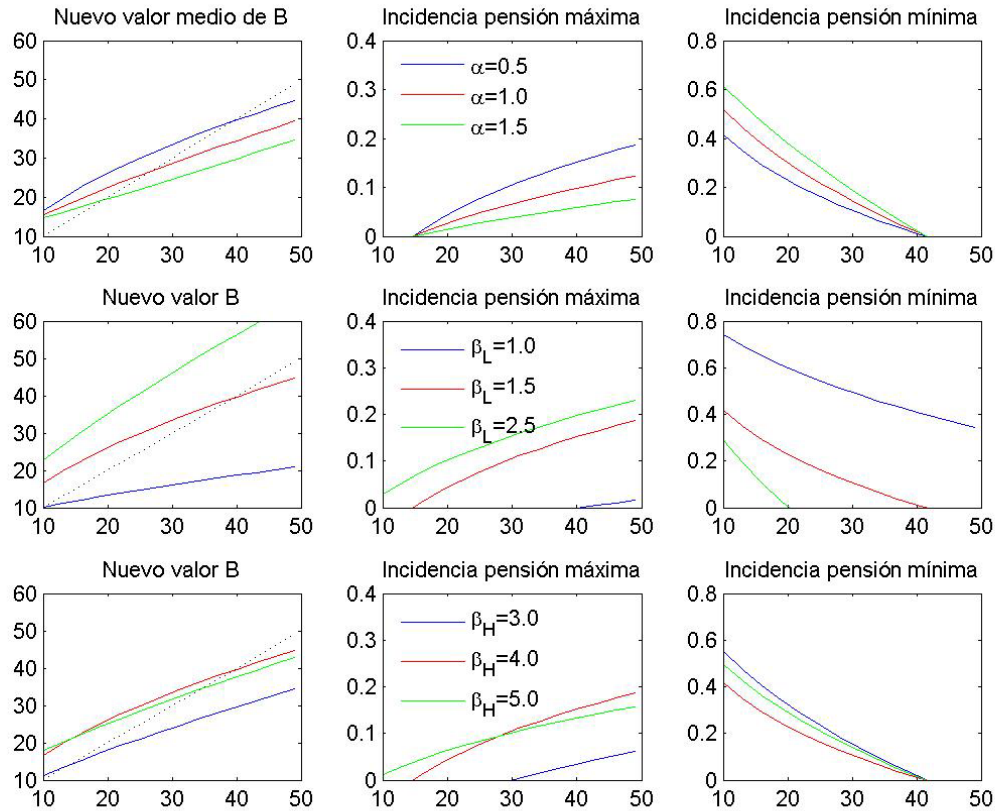


Figura 11: Simulación de la distribución Pareto Truncada para diversos valores de los parámetros α (fila 1), β_L (fila 2) y β_H (fila 3). Cada gráfico representa las soluciones obtenidas en el intervalo de valores de pensiones “ancla” (generados por el modelo de equilibrio) de entre 10 y 50 mil euros anuales. El nuevo valor medio combina los valores de los topes legales con sus probabilidades teóricas y con el valor medio del tramo *entre* los topes proporcionado por el modelo Pareto truncado.

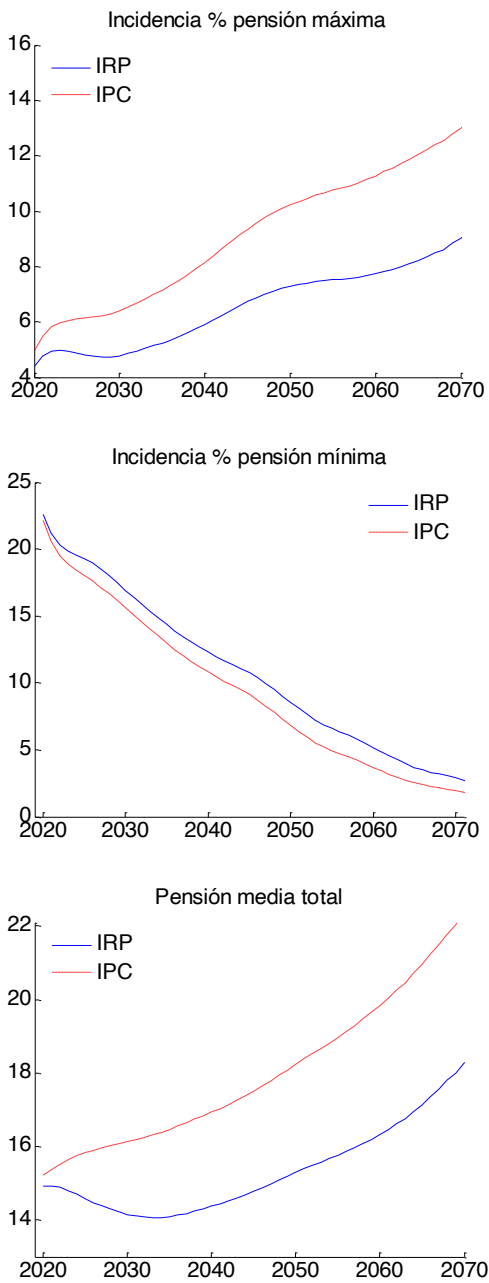


Figura 12: Serie temporal de la incidencia de los topes superior e inferior de pensiones (paneles superior e intermedio) y serie temporal de la pensión contributiva media (panel inferior). Economía que actualiza la pensión con IRP versus economía que actualiza las pensiones con el IPC

B. Heterogeneidad en bases de cotización

Esta sección completa la información presentada en las Secciones 4.1 y 4.2 del texto principal. Aporta detalles de la construcción de la muestra de bases contributivas y revisa algunos resultados construidos con ella:

- Elaboración de la base de datos de cotización en la MCVL y selección del año de calibración en la Sección B.1.
- Repaso de las regularidades empíricas en la Sección B.2.

B.1. Elaboración de la base de datos de bases de cotización en la MCVL

Debido a la estructura de los archivos de la MCVL, construir la sección cruzada de bases contributivas es más complejo que construir la base de datos de ingresos de pensiones. La siguiente secuencia de pasos resume como hemos procedido en este trabajo:

- Comenzamos utilizando los *archivos de afiliación*, que aportan la vida laboral completa de los individuos en la muestra. Para cada período de trabajo se dispone de variables como el régimen de Seguridad Social, tipo de contrato, grado de parcialidad, grupo de tarifa y fechas de inicio y final. En esta aplicación seleccionamos todos los contratos en vigor en diciembre de 2016.
- Las bases de cotización se almacenan en la serie de 13 *archivos de cotización* (12 para asalariados y el decimotercero para autónomos). En este trabajo utilizamos los valores mensuales correspondientes al mes de diciembre de 2016 para cada individuo de la muestra.
- Finalmente obtenemos las variables demográficas básicas de los individuos (sexo, educación, nacionalidad, edad) del *archivo de personal*.
Las tres bases de datos anteriores se fusionan conforme al identificador de individuo y de empresa para construir la muestra de trabajo en esta aplicación.
- En principio, la muestra incluye contratos a tiempo completo y contratos a tiempo parcial. Para determinar la activación de los topes de cotización en éstos últimos procedemos a hacer una “elevación” de las bases brutas conforme al grado de parcialidad de los contratos. De esta forma, la comparación con las topes legales (correspondientes según las características del individuo) se realiza sobre valores equivalentes a tiempo completo. Esta corrección es importante porque la incidencia de la parcialidad en la muestra está lejos de ser trivial: un 23% de la muestra tiene contratos con algún grado de parcialidad, y entre ellos el grado medio de parcialidad es del 49%.

La base de datos de 2020

El procedimiento apuntado en la sección previa puede aplicarse a la MCVL de cualquier año con ajustes inmediatos. Para este trabajo hemos usado la MCVL de 2017 por dos razones. En primer lugar, por continuidad respecto a la calibración de las versiones antiguas del modelo, que utilizaron esta misma base de datos. En segundo lugar, porque la versión más actualizada de los mismos de la que disponíamos al inicio del proyecto (la MCVL-2021) genera un conjunto de resultados cuya interpretación resulta bastante difícil. El problema se encuentra en que los datos para los que esta MCVL es representativa (2020) son muy peculiares por la contingencia única de la crisis del covid19. Los datos de bases contributivas presentan discontinuidades que, al contrario de lo que ocurre en 2016, no son siempre coincidentes con los diversos topes legales en vigor en ese año. Esto puede deberse a la existencia de numerosos programas específicos de sostenimiento del empleo y de las rentas laborales en ese año. En conjunto, los datos de 2016 se presentan mucho más “limpios” para ser usados como año base en la calibración del modelo.

B.2. Comportamiento de las bases contributivas en la MCVL-2017

Los datos originales tienen más riqueza descriptiva de la que, desafortunadamente, podemos reproducir en el modelo. Por ejemplo, el modelo no distingue explícitamente entre autónomos y asalariados, que tienen una distribución de sus bases contributivas muy diferente.³⁴ El Gráfico 13 presenta la distribución de las bases contributivas condicionadas al tipo de trabajador (autónomo o asalariado) y, para éstos últimos, muestra las notables diferencias que se observan por grupo de cotización. Las bases mayores se observan en el Grupo 1, que corresponde a las ocupaciones de mayor cualificación profesional. La incidencia de la base máxima en este grupo se aproxima al 40% de la muestra. Esta cifra desciende marcadamente al considerar los siguientes grupos contributivos, caracterizados por una cualificación profesional cada vez menor. Al tiempo, la incidencia de la base mínima es cada vez mayor (es importante notar las diferentes escalas utilizadas en el eje vertical de los histogramas). Los autónomos son, sin embargo, el colectivo con una mayor incidencia de la base mínima. Este hecho no es reflejo de sus ingresos por actividades económicas, ya que, en ese año, estos trabajadores podían elegir su base de cotización (con ciertas restricciones).

Los Cuadros 11 y 12 muestran los estadísticos resumen de la distribución de las bases condicionadas en dos de las variables que definen a los *agentes representativos* del modelo: educación y edad (en grupos decenales), separando hombres y mujeres.

³⁴ Un análisis empírico detallado de la afiliación y distribución de bases contributivas de los autónomos se encuentra en Sánchez-Martín 2019

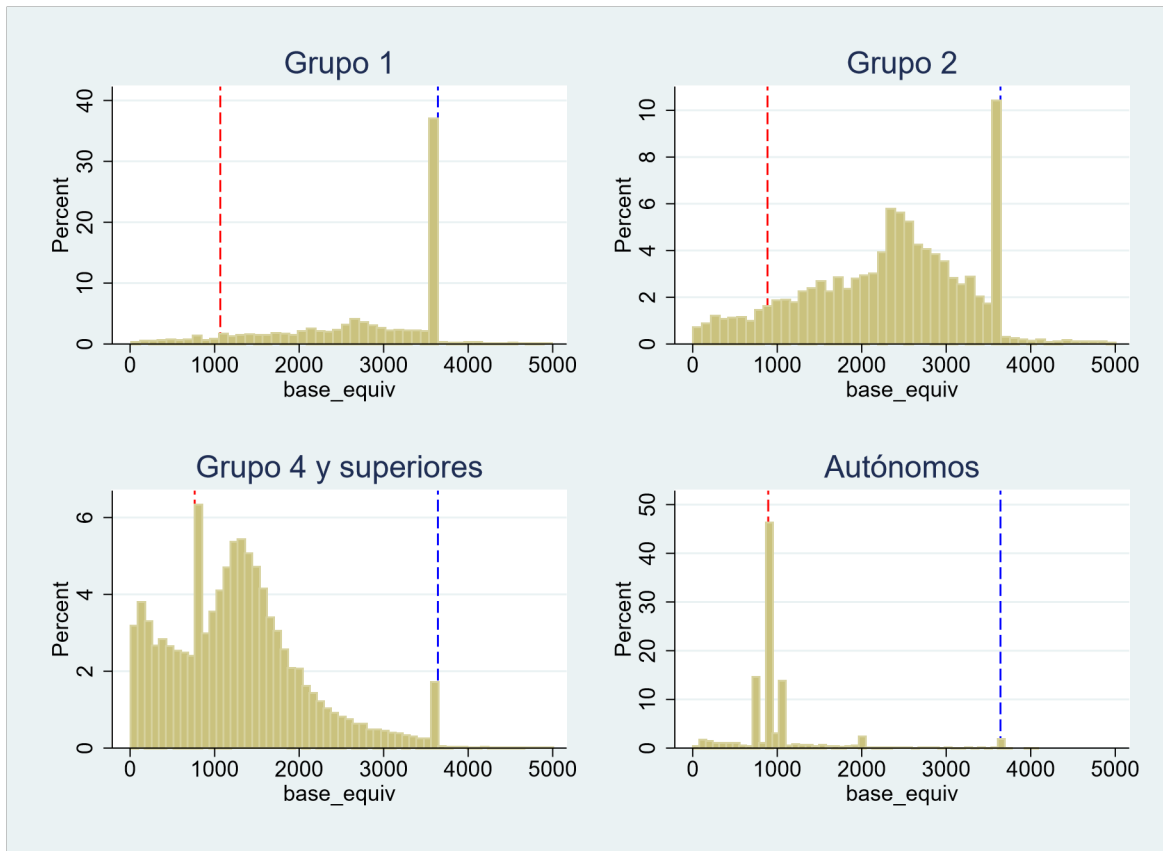


Figura 13: Distribución de las bases de cotización en el año 2016: estimación en la MCVL para autónomos y asalariados (separados por grupo de cotización).

Grupo edad	Bajo	Medio	Alto
menos de 20	6.37	8.38	
	0.58	0.53	0.0
	0.01	0.00	0.0
entre 20 y 29	12.29	13.02	17.55
	0.35	0.34	0.20
	0.01	0.01	0.04
entre 30 y 39	15.01	18.23	26.88
	0.27	0.20	0.13
	0.02	0.05	0.18
entre 40 y 49	15.97	20.73	33.33
	0.26	0.18	0.11
	0.03	0.10	0.32
entre 50 y 59	16.49	24.23	41.88
	0.24	0.13	0.06
	0.06	0.17	0.46
60 y mas	16.23	24.95	43.50
	0.20	0.09	0.12
	0.10	0.22	0.49

Cuadro 11: Estadísticos del nivel de las bases contributivas condicionados al nivel educativo y al grupo de edad para varones. En cada combinación se reproduce la base media de cotización (miles de euros anuales) y las incidencias de la base máxima y la base mínima en tanto por uno. Resultados estimados en la MCVL 2017

Grupo edad	Bajo	Medio	Alto
menos de 20	9.17	6.31	0
	0.50	0.57	0
	0.01	0.02	0
entre 20 y 29	11.97	12.82	16.90
	0.34	0.35	0.24
	0.01	0.02	0.04
entre 30 y 39	14.03	17.25	23.91
	0.26	0.18	0.14
	0.02	0.03	0.13
entre 40 y 49	14.52	18.58	30.06
	0.26	0.15	0.09
	0.02	0.04	0.22
entre 50 y 59	14.93	20.89	33.61
	0.22	0.13	0.06
	0.03	0.07	0.29
60 y mas	15.09	24.96	36.02
	0.15	0.08	0.07
	0.04	0.11	0.35

Cuadro 12: Estadísticos del nivel de las bases contributivas condicionados al nivel educativo y al grupo de edad para mujeres. En cada combinación se reproduce la base media de cotización (miles de euros anuales) y las incidencias de la base máxima y la base mínima en tanto por uno. Resultados estimados en la MCVL 2017