PROGRAMA DE DOCTORADO INTERUNIVERSITARIO EN ECONOMÍA: INSTRUMENTOS DEL ANÁLISIS ECONÓMICO.

TESIS DOCTORAL

APORTACIONES A LA ECONOMÍA DE LA INNOVACIÓN Y DEL CAMBIO TECNOLÓGICO: EL PROCESO INVENTIVO DESDE LA CIENCIA DE REDES

AUTOR

JOSÉ ALEJANDRO MENDOZA MEJÍA

DIRECTORES

FAUSTINO PRIETO MENDOZA

JOSÉ MARÍA SARABIA ALEGRÍA

UNIVERSIDAD DE CANTABRIA

UNIVERSIDAD DEL PAÍS VASCO | UNIVERSIDAD DE OVIEDO

Escuela de Doctorado de la Universidad de Cantabria

Santander 2022

UNIVERSIDAD DE CANTABRIA

TESIS DOCTORAL

Aportaciones a la Economía de la Innovación y del Cambio Tecnológico: el proceso inventivo desde la Ciencia de Redes

Autor:

José Alejandro Mendoza Mejía

Director:

Faustino Prieto Mendoza

José María Sarabia Alegría

Fecha: 03 de junio de 2022

Agradecimientos

A mis directores de tesis, Faustino Prieto Mendoza y José María Sarabia Alegría, mi gratitud por su apoyo, confianza y dedicación. A Faustino quiero expresarle mi más sincera admiración, conocer personas que aman su labor y encuentran en ella como impactar en la vida de los demás, es uno de los aprendizajes y ejemplos más valiosos que llevaré conmigo de esta etapa de mi vida.

Desde la distancia recordándoles con cariño, y esperando verles pronto, dedico esta tesis a mi familia. A mis abuelos, mi madre, mis hermanas, mis tías y a mi pequeña sobrina, Victoria. A José Ángel, por su inconmensurable paciencia y comprensión. Siempre presentes en mi corazón.

A Sandra, Lizeth, Diana y José Ignacio, esas personas valiosas que siempre estarán felices por mis triunfos.

A quienes ya no están y viven en mis recuerdos, mi amigo, compañero y colega, Fernando Velázquez; y a Ute Krogmann, una persona maravillosa que un día de septiembre de 2017 me vió llegar a España, en el inicio de una aventura que ahora llamo mi vida.

Muchas gracias a todos.

Índice general

1. Introducción **1**

1.1. La Economía de la Innovación 1
1.2. Proceso Combinatorio y Novedad Inventiva 3
1.3. Preguntas y Objetivos de Investigación 5
1.4. Software Utilizado 6
1.5. Aportaciones Originales 6
1.6. Esquema de la Tesis Doctoral 8

2. Datos **11**

2.1. Base de datos 11
2.1.1. Base de datos de patentes 11
2.1.2. Conceptualización de las variables. 20
2.1.3. Software estadístico 26
2.1.4. Procesamiento y manipulación de datos 27
2.2. Apéndice 48

3. Desigualdad en la frecuencia de uso de códigos tecnológicos. **53**

3.1. Introducción 53
3.2. Datos y metodología 56
3.2.1. Datos 56
3.2.2. Frecuencia de uso de códigos IPC y CPC 57
3.2.3. Estadística descriptiva 58
3.2.4. Medidas de desigualdad 62
3.2.5. Fuentes de la novedad tecnológica 64
3.2.6. Descomposición de la desigualdad por actividades económicas productivas 65
3.3. Resultados 69
3.4. Conclusiones 81

4. Modelización estadística de la frecuencia de uso de códigos tecnológicos 83
4.1. Introducción 83
4.2. Datos y metodología 85
4.2.1. Datos 85
4.2.2. Metodología 87
4.3. Resultados 106
4.3.1. Distribución de probabilidad de la frecuencia de uso de códigos tecnológicos IPC 106
4.3.2. Distribución de probabilidad de la frecuencia de uso de códigos tecnológicos CPC 123
4.3.3. Test de bondad de ajuste Chi-Cuadrado para los modelos con mejor ajuste de 1 y 2 parámetros 141
4.3.4. Análisis de modelos de distribución de probabilidad de 3 parámetros 146
4.4. Conclusiones 170

5. Redes de códigos tecnológicos en patentes. 171
5.1. Introducción 171
5.2. Datos y metodología 172
5.3. Resultados 175
5.4. Conclusiones 185

Índice de figuras

3.1. Número de patentes PCT(EPO), y el número de diferentes códigos IPC y CPC por año (izquierda). Número de veces en un año que los códigos IPC y CPC son usados (derecha), para el periodo 1977-2018. 56
3.2. Estadística descriptiva. Frecuencia de uso de códigos Tecnológicos IPC y CPC. 1977 - 2018 61
3.3. Curva de Lorenz. Distribución de ingresos en una población. 62
3.4. Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada con el Coeficiente de Gini (izquierda) y el índice de Theil (derecha). Periodo: 1977 - 2018. 69
3.5. Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada con el Coeficiente de Gini (izquierda) y el índice de Theil (derecha) entre los años 2001 y 2018. 70
3.6. Crecimiento (%) de la desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada a partir del coeficiente de Gini (izquierda y el índice de Theil (derecha). Periodo:1979-2018. 71
3.7. Curva de Lorenz, correspondiente a los códigos tecnológicos IPC para los años seleccionados: 1977, 1980, 1990, 2000, 2010 y 2018. 72
3.8. Curva de Lorenz, correspondiente a los códigos tecnológicos CPC para los años seleccionados: 1977, 1980, 1990, 2000, 2010 y 2018. 73
3.9. Porcentaje (%) de frecuencia de uso para el primer decil, P90-100 (arriba) y los fractiles P99 - 100, P95 - 99 y P90 - 95 (abajo); para códigos IPC (izquierda) y CPC (derecha). Periodo: 1977 - 2018. . . 74
3.10. Porcentaje de aplicaciones de patentes por año (en base anual), clasificadas por las fuentes de novedad tecnológica (origen, combinación novedosa, combinación, y refinamiento), para códigos IPC (izquierda.) y códigos CPC (derecha.). Periodo: 1977-2018. 76
3.11. Porcentaje de aplicaciones de patentes por año (en base anual), agrupadas por fuentes de novedad tecnológica: (1) origen + combinaciones novedosas; (2) combinaciones + refinamientos;, para códigos IPC (izquierda.) y códigos CPC (derecha.). Periodo: 1977-2018. 77

3.12. Porcentaje de códigos tecnológicos más usados en un año que mantienen este estatus de un año a otro: considerando códigos con frecuencia de uso en los fractiles $P95-100$ (línea A en color rojo) y en el fractil $P90-100$ (línea B en color azul); para códigos IPC (izquierda) y CPC (derecha). Periodo: 1977-2018. 77
3.13. Ranking de divisiones NACE en los años 2000 y 2018 en términos de la desigualdad en la frecuencia de uso de códigos IPC, estimados por el coeficiente de Gini (izquierda) y el índice de Theil (derecha). . . . 79
3.14. Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC, estimada con el índice de Theil (Izquierda), y descomposición de la desigualdad en la frecuencia de uso de códigos IPC, estimada con el índice de Theil (dcha), periodo 1977 - 2018. 80

4.1. Ajuste de los modelos de distribución de probabilidad Geométrica (arriba) y Binomial Negativa Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 113
4.2. Ajuste de los modelos de distribución de probabilidad Zeta (arriba) y Zipf-Mandelbrot (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 114
4.3. Ajuste de los modelos de distribución de probabilidad Yule-Simon (arriba) y Beta-Geométrica (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 115
4.4. Ajuste de los modelos de distribución de probabilidad Logarítmica (arriba) y Poisson Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 116
4.5. Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto II (arriba) y Discreta Generalizada de Pareto III (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 117
4.6. Ajuste de los modelos de distribución de probabilidad Geométrica (arriba) y Binomial Negativa Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018 131
4.7. Ajuste de los modelos de distribución de probabilidad Zeta (arriba) y Zipf-Mandelbrot (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018. 132
4.8. Ajuste de los modelos de distribución de probabilidad Yule-Simon (arriba) y Beta-Geométrica (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018. 133
4.9. Ajuste de los modelos de distribución de probabilidad Logarítmica (arriba) y Poisson Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018. 134

4.10. Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto II (arriba) y Discreta Generalizada de Pareto III (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018. 135
4.11. Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto IV (arriba) y Discreta Generalizada de Waring (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018. 153
4.12. Ajuste de los modelos de distribución de probabilidad Discreta Gneralizada de Pareto IV (arriba) y Discreta Generalizada de Waring (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018. 154

5.1. Ranking de divisiones NACE en términos de la densidad de la red de códigos tecnológicos IPC. Periodo 1977 -2018. 179
5.2. Subred correspondiente a la división NACE 16 Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting (nivel 2 - dígitos). Enlaces de los 5 principales nodos en términos del Grado de Centralidad. Periodo 1977 - 2018. 182
5.3. Subred correspondiente a la división NACE 42 Civil Engineering (nivel 2 - dígitos). Enlaces de los 5 principales nodos, en términos del Grado de Centralidad. Periodo 1977 - 2018. 183
5.4. Subred correspondiente a la división NACE 18 Printing and Reproduction of Recorded Media (nivel 2 - dígitos). Enlaces a los 5 principales nodos, en términos del Grado de Centralidad. Periodo 1977 - 2018. 184
5.5. Red de códigos tecnológicos IPC conformada a partir de los enlaces entre los nodos principales (mayor grado de centralidad) de las divisiones NACE analizadas. Periodo 1977 - 2018. 185

Índice de cuadros

2.1. Variables de identificación de las aplicaciones de patentes PCT, en común en los ficheros de datos de aplicantes e inventores, en la base de datos REGPAT. 12
2.2. Listado de variables del fichero de datos de aplicantes de patentes PCT en REGPAT. 13
2.3. Listado de variables del fichero de datos de inventores de aplicaciones de patentes PCT en REGPAT. 14
2.4. Listado de variables del fichero de datos de códigos tecnológicos IPC de aplicaciones de patentes PCT en REGPAT. 15
2.5. Secciones de la Clasificación Internacional de Patentes IPC 16
2.6. Ejemplo código tecnológico "$A23K50/15$" en la Clasificación Internacional de Patentes IPC. 18
2.7. Secciones de la Clasificación Cooperativa de Patentes CPC 19
2.8. Esquema de clasificación de combinaciones de códigos tecnológicos. 47

3.1. Tabla de frecuencia de uso de códigos tecnológicos IPC y CPC correspondientes al año 1978. 58
3.2. Resumen estadística descriptiva de datos. Frecuencia de uso de códigos tecnológicos IPC. 1977 - 2018. 59
3.3. Resumen estadística descriptiva de datos. Frecuencia de uso de códigos tecnológicos CPC. 1977 - 2018. 60
3.4. Frecuencia de uso de códigos tecnológicos IPC en aplicaciones de patentes PCT. 1977 - 2018. División NACE (Nivel - 2 dígitos.) . . . 67

4.1. Funciones de densidad y notación de los modelos de distribución de probabilidad considerados. 88
4.2. Parámetros estimados y errores estándar para los modelos de distribución de probabilidad 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018. 106

4.3. Parámetros estimados y errores estándar para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018. 109
4.4. Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977-2018. . 119
4.5. $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977-2018. . 121
4.6. Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018. 123
4.7. Parámetros estimados y errores estándar para los modelos de distribución de probabilidad considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018. . 124
4.8. Parámetros estimados y errores estándar para los modelos de distribución de probabilidad considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977 - 2018. 127
4.9. Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018. . 137
4.10. $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018. . 139
4.11. Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977 - 2018. 141
4.12. Resultados del test de ajuste Chi- Cuadrado para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados.Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 -2018. 142
4.13. Resultados del test de bondad de ajuste Chi- Cuadrado para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados. Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 -2018 144
4.14. Años en los que no se puede rechazar la hipótesis nula H_0*: los datos siguen el modelo de distribución seleccionado* al 5 por ciento de significancia, para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados. 145

4.15. Parámetros estimados y errores estándar para los modelos de distribución de probabilidad Pareto discreta generalizada IV y Waring generalizada desplazada. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 - 2018. 147
4.16. Parámetros estimados y errores estándar para los modelos de distribución de probabilidad Pareto discreta generalizada IV y Waring generalizada desplazada. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 - 2018. 150
4.17. Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978-2018. 155
4.18. $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978-2018. . 157
4.19. Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978-2018.159
4.20. $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978-2018. . 161
4.21. Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 - 2018. 164
4.22. Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 - 2018. 165
4.23. Resultado de test de bondad de ajuste Chi- Cuadrado para modelos de distribución de probabilidad de 3 parámetros.Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 -2018. 165
4.24. Resultado de test de bondad de ajuste Chi- Cuadrado para modelos de distribución de probabilidad de 3 parámetros.Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 -2018. 166
4.25. Años en los que no se puede rechazar la hipótesis nula H_0*: los datos siguen el modelo de distribución seleccionado* al 5 por ciento de significancia, para los modelos de distribución de probabilidad de 2 y 3 parámetros con mejor ajuste a los datos. 169

5.1. Número de nodos y enlaces de cada subred NACE (nivel 2-dígitos) de códigos IPC. Periodo 1977-2018. 177
5.2. Densidad de cada subred NACE (nivel 2-dígitos) de códigos IPC. Periodo 1977-2018. 178
5.3. Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018. 179
5.3. Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018. 180
5.3. Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018. 181

Capítulo 1

Introducción

1.1. La Economía de la Innovación

El surgimiento de la Economía de la Innovación como disciplina y campo de investigación es relativamente nuevo. Los estudios que buscan crear una base de conocimiento que ayude al diseño de políticas de innovación pueden considerarse escasos, si son comparados con otras disciplinas de la Economía (Fagerberg y Verspagen, 2009). Pese a ello, la creciente popularidad de la innovación ha dado lugar a la aparición de diferentes líneas y temáticas de investigación cuyas aportaciones no se ven limitadas a la Economía, generando un importante impacto que trasciende a otras áreas del conocimiento como la Administración, la Organización Industrial o la Sociología (Martin, 2012).

El estudio de la relación entre la Economía y la Innovación en la literatura nos traslada a la mitad del siglo XX donde emerge como figura central Joseph Schumpeter (1934 y 1942) quien brindó valiosas aportaciones acerca del papel de la innovación en el desarrollo económico, centradondose en la interacción del innovador o *"emprendor"* con su entorno, e introduciendo el concepto de "destrucción creativa". Más adelante, en las primeras decadas de la segunda mitad del siglo XX, fueron importantes los esfuerzos dedicados a estudiar el papel de la tecnología y el cambio técnico en el crecimiento económico.

Los modelos de crecimiento que incorporan la tecnología, van desde las aportaciones de Solow (1956 y1957) que sitúan el progreso tecnológico como un factor exogeno del crecimiento, a los modelos de crecimiento endógeno (Romer, 1986; Lucas,1988) que entienden el crecimiento como un resultado mismo del sistema económico, y no de fuerzas externas, en donde el progreso técnico es el resultado del conocimiento adquirido por los trabajadores en la empresas, en alusión al concepto de learning - by- doing (Arrow, 1962).

Otros modelos de crecimiento, denóminados modelos de I + D (Romer,1990; Grossman y Helpman,1991; Jones,1995) incorporan la importancia de las actividades de I + D, en cuanto estas actividades permiten una diferenciación de los productos y un poder de mercado. Algunos modelos incluyen en la función de crecimiento el capital humano (Lucas,1993), mientras que otros plantean nuevas perspectivas para entender el crecimiento como lo es el enfoque evolucionista (Nelson y Winter, 1982).

Es preciso resaltar la importancia dada en el estudio de la relación Economía e Innovación, a la comprensión del proceso de difusión de la innovación y la tecnología (Rogers, 1962; Dosi, 1991; Baptista 2001), la internacionalización tecnológica (Archibugi y Michie, 1995; Cantwell, 1995; Alvarez y Molero, 2005), y los efectos de la innovación en la competitividad y la productividad (Fagerberg et al, 2007; Crespi y Pianta, 2008; Audretsh y Keilbach, 2008; Lang, 2009; Bogliciano y Pianta, 2010; Ejermo et al, 2011).

Una área de relevancia en el estudio de la la Economía de la innovación es aquella orientada a la comprensión del proceso de innovación y sus factores determinantes. En esta línea podemos situar aportaciones de diversa índole que abordan, entre otros temas, la gestión de la innovación y los modelos interactivos del proceso de innovación (Kline y Rosenberg, 1986; Teece, 1986), los modelos y patrones de innovación (Utterback y Abernathy, 1975; Abernathy, 1976; Abernathy y Utterback, 1978), las tipologías de innovación (Pavitt,1984), los flujos de comunicación en organizaciones de I+ D (Allen, 1977), la innovación basada en los recursos de las empresas (Cohen y levinthal, 1989; Grant, 1991), la gestión del conocimiento y el aprendizaje organizacional (Nonaka, Takeuchi y Umemoto, 1996).

Franjas conceptuales abordadas en los estudios de innovación, que pueden considerarse más actuales, tratan acerca de las redes de colaboración y la innovación abierta (Chesbrough, 2003; von Hippel, 2006; Laursen y Salter, 2006), y los sistemas de innovación (Lundvall, 1992; Tidd, Besaant y Pavitt, 1997; Malerba, 2002), que comprende el estudio de los sistemas nacionales y sectoriales de innovación, y la relación entre los actores que confluyen en estos sistemas, por ejemplo, la Triple Hélice (Etzkowitz y Leydesdorff, 2000), entre otras.

1.2. Proceso Combinatorio y Novedad Inventiva

El estudio de las fuentes de novedad en la invención y el cambio tecnológico ha sido abordado principalmente desde 3 perspectivas, de acuerdo con Strumsky y Lobo (2015): la perspectiva combinatoria que supone la construcción de nuevas tecnologías a partir de otras tecnologías existentes; la perspectiva del desarrollo tecnológico humano que se refiere al gradualismo o lenta acumulación de variaciones o mejoras en las tecnologías que son el resultado de modificaciones deliberadas y accidentales; y una tercera perspectiva que sitúa el origen de las nuevas tecnologías en accidentes que generan nuevos caminos, en vez de la ampliación de los caminos ya conocidos.

En esta investigación entendemos el cambio tecnológico como un proceso de combinación y recombinación de tecnologías, nuevas y existentes, que integradas dan como resultado una novedad tecnológica (Parraguez et al., 2020). La dinámica de este proceso de recombinación de tecnologías ha sido descrita, normalmente, desde los criterios de familiaridad de los componentes usados y la novedad de las combinaciones utilizadas (Arts y Veugelers, 2015). Las tecnologías que son definidas como "*artefactos, dispositivos, métodos y materiales disponibles para los seres humanos para realizar tareas específicas*"(Youn et al., 2015, p.1) se convierten en bloques para la construcción de otras tecnologías (Strumsky y Lobo, 2015), y por tanto, la forma como las tecnologías son configuradas para lograr una función deseada en una invención constituye un elemento central del cambio tecnológico.

De este modo, una invención integra una variedad de funcionalidades tecnológicas diferentes (fruto de la combinación), y cuando una nueva funcionalidad tecnológica es incorporada dentro del reservorio de tecnologías disponibles ocurre la novedad tecnológica (Strumsky y Lobo, 2015). Los autores citados precisan en hacer una distinción entre la novedad inventiva y la novedad tecnológica, señalando que una novedad inventiva (ej. un nuevo artefacto, dispositivo...) no siempre supone la introducción de una novedad tecnológica.

Las definiciones de tecnología e invención propuestas por Youn et al., (2015) y Strumsky y Lobo (2015) buscan facilitar la cuantificación de la invención como un proceso combinatorio mediante el uso de los códigos tecnológicos en patentes, que según los autores cumplen su función:

> *"como identificadores de distintas funcionalidades tecnológicas hacen posible describir, y categorizar, la novedad inventiva como el resultado del desarrollo de nuevas capacidades tecnológicas, la combinación de capacidades tecnológicas en formas que no han sido previamente presenciada*

> *en el registro de patentes, o de la reutilización de tecnologías* (Strumsky y Lobo, 2015, p.4)".

La invención supone un proceso de búsqueda por parte del inventor, quien debe tomar la decisión de explorar o explotar. El concepto de exploración puede asociarse al uso de combinaciones tecnológicas novedosas (nuevos conocimientos), mientras que si el inventor hace uso de combinaciones existentes (conocimiento previo) se puede afirmar que las invenciones proceden de la explotación de tecnologías conocidas. Este enfoque es abordado por Wang et al.(2017), quienes encuentran una relación entre el balance de las actividades de exploración y explotación con la calidad de la invención. Las combinaciones pueden involucrar diferentes fuentes y conducir a diferentes resultados en función de su uso previo o distancia tecnológica (ej. Uzzi et al., 2013; Kim et al., 2016; Art y Veugelers, 2017).

El reto de entender los mecanismos del cambio tecnológico no parece ser una tarea sencilla. Existe un vacío en la caracterización formal y cuantitativa del proceso combinatorio, tanto en el análisis de las características del proceso combinatorio (Youn et al.,2015) como en la exploración de nuevas metodologías que proporcionen una perspectiva temporal, integren fuentes de información y posibiliten nuevos niveles de análisis (Parraguez et al.,2020).

La identificación, medición y monitoreo de estos procesos complejos requiere de objetos de estudio bien definidos, de fácil codificación y sistematización; y en esta línea, recientes estudios basados en la ciencia de las redes (ej. Mun et al., 2019) plantean el uso de información tecnológica de patentes como un proxy de las invenciones. La descripción cuantitativa del proceso combinatorio requiere discretizar las tecnologías para mapear, analizar y representar dominios tecnológicos, siendo una de las manera más habituales, el uso de códigos tecnológicos (Parraguez et al.,2020). La información de patentes han sido empleada, entre otros fines, para predecir tecnologías prometedoras (Kim y Bae, 2017), identificar tendencias (Jeon et al., 2019) o mapear el paisaje tecnológico (Kay et al, 2014).

Los sistemas de clasificación de patentes pueden ser considerados como "*la organización intelectual de la base de datos de nuevos productos y procesos de valor económico*" (Leydesdorff et al.,2008, p.1574). El uso de los sistemas de clasificación de patentes es importante para establecer una relación entre las tecnologías y las actividades industriales (Lafond y Kim, 2019). Una tarea que según los autores es complicada dado que las clasificaciones de patentes son construidas con base a criterios de funcionalidad, y no de aplicación. Lafond y Kim (2019) sugieren además que el estudio de la evolución histórica de los sistemas de clasificación ofrece información relevante del cambio tecnológico, y destacan que en la literatura son pocos

los estudios que advierten de sesgos en la asignación de los códigos tecnológicos. Aunque por la naturaleza estable de los códigos tecnológicos existen posibles limitaciones para que estos puedan capturar los cambios continuos en la naturaleza de las invenciones (Venugopalan y Rai, 2015).

1.3. Preguntas y Objetivos de Investigación

En esta tesis doctoral, hemos intentado responder a las siguientes preguntas de investigación. En cuanto al tercer capítulo, ¿Es la invención un proceso combinatorio basado en el uso generalizado de pocas tecnologías?. De ser cierto, ¿Cómo han evolucionado en el tiempo las disparidades en la frecuencia de uso de estas tecnologías?. En el cuarto capítulo, la pregunta de investigación fue: ¿Qué modelo de distribución paramétrica, entre los más conocidos, es el más adecuado para modelizar la frecuencia de uso de tecnologías (códigos tecnológicos) en patentes?; y por último, en el quinto capítulo, fueron planteados los siguientes interrogantes: ¿Cuáles son las características estructurales de la red de códigos tecnológicos en patentes? ¿Es la red de códigos tecnológicos en patentes una red cohesionada y con alta complejidad? ¿Converge esta red en nodos específicos?. Partiendo de estas preguntas de investigación, el objetivo de la tesis doctoral es ofrecer relevantes aportaciones acerca de la invención como un proceso combinatorio de tecnologías, que permita expandir nuestra comprensión del proceso inventivo y el cambio tecnológico.

Igualmente, podemos enumerar otros objetivos específicos de la tesis doctoral:

- Estimar la desigualdad total en la frecuencia de uso de códigos tecnológicos en patentes, a partir de las medidas de desigualdad, Coeficiente de Gini e Índice de Theil, ampliamente conocidas en las ciencias económicas.
- Establecer un ranking de actividades económicas productivas en términos de la desigualdad en la frecuencia de uso de tecnologías (códigos tecnológicos)
- Relacionar la desigualdad en la frecuencia de uso de códigos tecnológicos y la evolución de las fuentes de la novedad tecnológica.
- Determinar un modelo de distribución parámetrica adecuado para modelar la frecuencia de uso de tecnologías en patentes.
- Extender el esfuerzo realizado en la literatura en el análisis de la frecuencia de aparición de las palabras en el lenguaje humano, de forma interdisciplinar, al ámbito de la frecuencia de aparición de los códigos en el "lenguaje"de las combinaciones de tecnologías en la invención.

- Caracterizar la estructura de las redes complejas de códigos tecnológicos en patentes, en cuanto a su cohesión (intensidad de la red) y convergencia (centralidad de los nodos).

1.4. Software Utilizado

Como es común en los estudios que requieren un manejo masivo de datos, fue necesario el uso de un software estadístico para el manejo y manipulación de las bases de datos de patentes. En esta investigación se utilizó el software R de libre acceso, versión R-4.1.3.tar.gz, con fecha de publicación del 10 de marzo de 2022.

Dentro de las posibilidades que ofrece este software se usó principalmente, la reciente versión 1.0.7 de junio de 2021 del paquete *dplyr* que es una herramienta rápida y estable, que consiste en una gramática (conjunto de verbos) que facilita la manipulación de bases de datos, entre los cuales destacar: *filter* (permite obtener un conjunto de observaciones usando condiciones lógicas), *arrange*(reordenar filas), *summarise* (generación de resúmenes de variables), *mutate* (añadir nuevas variables), *pipe* (concatenar acciones), entre otros verbos, que operan naturalmente en conjunto con *group_by* para realizar operaciones por grupos de variables.

Finalmente indicar que, el paquete *ggplot2* fue empleado para la representación gráfica de las estimaciones, y el paquete *igraph* para generar las matrices de adyacencia y la redes de códigos tecnológicos, así como su interfaz visual.

1.5. Aportaciones Originales

- Se estima por primera vez, desde lo que conocemos, la desigualdad que emerge de la frecuencia de uso de códigos tecnológicos en patentes, su evolución y como se relaciona esta desigualdad con las fuentes de la novedad tecnológica.

  Como resultado de dicho estudio, se encuentra que la desigualdad total en la frecuencia de uso de códigos tecnológicos tiene dos fases claramente diferenciadas. Durante la primera fase (1977-2000), la desigualdad total muestra un patrón de crecimiento. En la segunda fase (2001-2018), la desigualdad total muestra una relativa estabilización. Además, se evidencia que la desigualdad total anual en la frecuencia de uso de códigos tecnológicos tiende a ser persistentemente alta, sin importar si son considerados los códigos IPC o CPC, o si la estimación es realizada usando el coeficiente de Gini o el índice de Theil.

Así mismo, se encuentra que la evolución anual de la desigualdad total en la frecuencia de uso de códigos tecnológicos parece estar relacionada con los cambios que se presentan en la evolución de las fuentes de novedad inventiva en donde las patentes, en los años iniciales, son en su mayoría constituidas por códigos tecnológicos nuevos, y luego, se da paso a una segunda etapa en la cual predominan las patentes descritas por la combinación de códigos tecnológicos previamente usados.

Adicionalmente, se encuentra que dicha evolución anual de la desigualdad total estaría relacionada con un mecanismo de "el rico se hace más rico" ("rich-get-richer"), haciendo analogía de códigos tecnológicos ricos a aquellos más usados.

Finalmente, tras descomponer dicha desigualdad por actividades económicas, tomando como referencia la clasificación NACE, se encuentra que la desigualdad total está conducida principalmente por las diferencias en la frecuencia de uso entre los códigos que conforman cada grupo.

- Se analiza, por primera vez en la literatura hasta lo que se conoce, la distribución de frecuencia de uso de códigos tecnológicos de patentes, considerando todo el rango de su distribución, y teniendo en cuenta la naturaleza discreta de dicha variable aleatoria.

  Dicho análisis se realiza considerando 13 modelos de distribución de probabilidad paramétricos, de 1,2 y 3 parámetros, conocidos en la literatura. Y en particular, como una extensión interdisciplinar del ámbito del estudio de la frecuencia de aparición de las palabras en el lenguaje al ámbito de la Economía de la Innovación, se incorporan modelos de distribución de la jerarquía de distribuciones discretas de Pareto: tipo III y IV, y modelos de distribución de 3 parámetros como el modelo Pareto tipo IV anterior y la distribución de Waring generalizada desplazada, novedosos hasta lo que se conoce en este ámbito de investigación.

  Como resultado de dicho análisis, se encuentra que la distribución Pareto discreta generalizada de tipo IV ha sido el modelo de distribución que mejor ajuste ha proporcionado para dicha variable de interés, tanto para códigos IPC como para códigos CPC, durante el periodo 1978-2018, en comparación con otros 12 modelos de distribución considerados.

- Dentro del marco general de las Ciencias de la Complejidad, y en particular dentro de la Teoría de la Redes Socioeconómicas Complejas, se exploran las características estructurales de las redes de códigos tecnológicos en patentes clasificadas por actividades económicas productivas.

Se analizan para cada división NACE, la red cuyos vértices son los códigos tecnológicos, y los enlaces el haber formado parte dichos códigos de una misma patente, y se estiman (por primera vez en este ámbito, según se conoce) dos medidas: por un lado, una medida global, la Densidad de la Red (Network Density); y por otro lado, una medida local, el Grado de Centralidad (Degree Centrality).

1.6. Esquema de la Tesis Doctoral

En esta tesis hemos desarrollado 4 capítulos que intentan ayudarnos a comprender mejor el proceso de invención y del cambio tecnológico. El esquema de la misma es el siguiente:

- El capítulo 2 es un capítulo instrumental que muestra, de forma detallada, el manejo y procesamiento de datos masivos de patentes correspondientes a la base de datos regionalizada de patentes REGPAT, publicada por la Organización para la Cooperación y el Desarrollo Económico OCDE (versión Julio 2021), para el periodo 1977 - 2018. A partir de los ficheros de datos de las aplicaciones de patentes solicitadas bajo el Tratado de Cooperación en materia de Patentes PCT, en fase internacional, con la Oficina Europea de Patentes EPO como oficina designada, fueron construidas las variables consideradas en las estimaciones que soportan los resultados de los capítulos 3, 4 y 5.

- En el capítulo 3 se aborda por primera vez, desde lo que conocemos, la desigualdad en la frecuencia de uso de códigos tecnológicos. Considerando los códigos tecnológicos de la Clasificación Internacional de Patentes (IPC) y la Clasificación Cooperativa de Patentes (CPC) en aplicaciones de patentes PCT (periodo 1977 -2018) se estima la desigualdad en la frecuencia de uso de códigos tecnológicos, a nivel global y desagregada por actividades económicas productivas segun la Clasificación Europea de Actividades Económicas Productivas NACE (nivel de 2 - dígitos), a partir de dos medidas de desigualdad ampliamente conocidas en la Economía, el Coeficiente de Gini y el Índice de Theil. Además, se evalua la relación de la desigualdad estimada con la evolución de las fuentes de la novedad tecnológica, y con la probabilidad de los códigos tecnológicos más usados en mantener este estatus de un año a otro.

- En el capítulo 4, se ajustan los modelos de distribución paramétrica seleccionados (de 1, 2 y 3 parámetros), en un análisis anual, a las variables aleatorias discretas: frecuencia de uso de códigos tecnológicos IPC y CPC, mediante el Método de Máxima Verosmilitud (MLE, por sus siglas en inglés). Los resultados de las estimaciones de los parámetros para los modelos de distribución son comparados

entre sí usando el Criterio de Información de Akaike (AIC), y luego tras seleccionar las distribuciones con un mejor desempeño, se efectúa el test de bondad de ajuste Chi-Cuadrado para testar dichos modelos, dentro del periodo considerado.

- En el capítulo 5 se caracteriza la estructura de las redes complejas de códigos tecnológicos en patentes. Para ello, se hace uso de una medida global de la red, *Intensidad de la Red*, que denota su complejidad y cohesión; y una medida de análisis local, *Grado de Centralidad*, que indaga acerca de la convergencia de las redes en nodos (códigos tecnológicos) específicos. Estas medidas de la red fueron estimadas para las redes códigos tecnológico IPC clasificadas según las divisiones NACE para el periodo 1977-2018.

Capítulo 2

Datos

2.1. Base de datos

2.1.1. Base de datos de patentes

En este estudio se utilizó información de patentes extraída de la base de datos REGPAT de la Organización para la Cooperación y el Desarrollo Económico OCDE, versión de julio de 2021. La consulta de la base de datos fue realizada en septiembre de 2021. La base de datos REGPAT fue creada por la OCDE con el interés de estimular la investigación y contribuir a una mejor comprensión de la dimensión regional de la innovación (Maraut, S. et al., 2008). La principal característica de la base de datos REGPAT es que relaciona la información de patentes con el lugar de residencia de los aplicantes e inventores. Este proceso de regionalización [*] permite que los datos de patentes puedan ser usados en conexión con otras variables desde una perspectiva regional.

La base de datos REGPAT se actualiza bianualmente y puede ser solicitada a la Dirección de Ciencia, Tecnología e Innovación de la OCDE mediante el diligenciamiento de un formulario en línea, disponible en `https://survey2018.oecd.org/Survey.aspx?s=03d8572cdd7f492aa1ed5350b7b9f044`. La información de REGPAT proviene fundamentalmente de la base de datos de estadísticas de patentes de la Oficina Europea de Patentes (EPO), PATSTAT, y comprende dos conjuntos de archivos: las aplicaciones de patentes presentadas ante la EPO y las aplicaciones de patentes bajo el Tratado de Cooperación en materia de Patentes PCT, en fase internacional.

[*] La metodología para la identificación de las regiones de residencia de aplicantes e inventores es descrita por Maraut, S. et al., (2008) en la publicación The OCDE REGPAT database: A presentation. Disponible para consulta en: `https://doi.org/10.1787/241437144144`

Para el desarrollo de la investigación se accedió a los archivos correspondientes a las aplicaciones de patentes PCT en fase internacional. Una aplicación PCT o aplicación de patente internacional, es aquella en donde el aplicante ha usado el Tratado de Cooperación en Patentes como vía para la protección de sus derechos de patentes en un número considerable de países, a través del trámite de una única aplicación internacional.

La base de datos REGPAT proporciona tres ficheros de datos. Estos ficheros de datos incluyen información acerca de los aplicantes, inventores y códigos tecnológicos asignados a cada una de las aplicaciones de patente PCT. Estas bases se relacionan entre sí por medio de una variable en común, el número de publicación PCT (*pct_nbr*). Además, los fichero de datos de aplicantes e inventores de aplicaciones de patente PCT comparten las variables: Número de aplicación PCT (*internat_appln_nr*) y Número de identificación de la aplicación PCT (*appln_id*), que se listan en el cuadro 2.1.

Cuadro 2.1: Variables de identificación de las aplicaciones de patentes PCT, en común en los ficheros de datos de aplicantes e inventores, en la base de datos REGPAT.

Variable	**Nomenclatura**
Número de publicación PCT	*pct_nbr*
Número de aplicación PCT	*internat_appln_nr*
Número de identificación de la aplicación PCT	*appln_id*

Los ficheros de datos REGPAT están estructurados en formato largo (long format). Por este motivo, la información correspondiente a cada una de las aplicaciones de patente PCT puede encontrarse en una o varias filas u observaciones. El número de filas con información correspondiente a una aplicación de patente PCT viene determinado por el número de valores o elementos que puedan tomar las variables en el fichero de datos. A continuación, se realiza una descripción de las variables de los ficheros de datos de REGPAT para aplicaciones de patente PCT:

- **Fichero de datos de aplicantes de patentes PCT en la base de datos REGPAT.**

En el cuadro 2.2 se listan las variables con información acerca de los aplicantes de patentes PCT en la base de datos REGPAT. En el cuadro mencionado se ofrece un ejemplo del fichero de datos de aplicantes tomando como referencia la aplicación de patente PCT con número de publicación PCT (*pct_nbr*) “WO1978000017”, número de aplicación PCT (*internat_appln_nr*) “WO1978CH00001”, y número de aplicación

o ID de la aplicación (*appln_id*) 6508281.

Tal como se puede observar en el cuadro 2.2, el fichero de datos de aplicantes de patentes PCT contiene datos acerca del nombre del aplicante (*app_name*), dirección (*address*), región (*reg_code*) y país de localización (*ctry_code*). Adicionalmente, el fichero incluye las variables de asignación múltiple regional (*reg_share*) y participación porcentual del aplicante en la invención (*app_share*). Esta última variable es útil al momento de realizar conteos fraccionales en aquellos casos en que existe más de un aplicante por patente, siendo en todo caso, $app_share \leq 1$.

Cuadro 2.2: Listado de variables del fichero de datos de aplicantes de patentes PCT en REGPAT.

Variables	**Nomenclatura**	**Aplicante 1**	**Aplicante 2**
Número de publicación PCT	*pct_nbr*	WO1978000017	WO1978000017
Número de aplicación PCT	*internat_appln_nr*	WO1978CH0001	WO1978CH0001
Identificador de la aplicación	*appln_id*	6508281	6508281
Nombre del aplicante	*app_name*	Batelle Memorial Institute	Solagro S.A
Dirección del aplicante	*address*	CH-1227 Cauroge/ Geneve	B-1050 Bruxelles
Código de región	*reg_code*	CH	BE
Código país	*ctry_code*	CH013	BE100
Múltiple localización regional	*reg_share*	1.00	1.00
Conteo fraccional aplicantes	*app_share*	0.50	0.50

En el ejemplo citado, la aplicación de patente PCT "WO1978000017"fue presentada por dos aplicantes: el aplicante 1, Batelle Memoria Institute; y el aplicante 2, Solagro S.A. El aplicante 1 está localizado en la dirección CH-1227 Carouge/Geneve, que corresponde en la Nomenclatura de Unidades Territoriales para Estadísticas (NUTS, por sus siglas en inglés) al código CH013 en Suiza (CH). Por su parte, el aplicante 2 se situa en la dirección B-1050 Bruxelles, en la NUTS BE100 en Bélgica (BE). La aplicación PCT tiene un *app_share*=0.5, resultado de dividir 1 entre el número total de aplicantes.

■ **Fichero de datos de inventores de aplicaciones de patentes PCT en REGPAT.**

En el cuadro 2.3 se muestran las variables del fichero de datos de inventores de aplicaciones de patentes PCT en la base de datos REGPAT. El fichero de datos de inventores comprende, al igual que el fichero de aplicantes, las variables: Número de publicación PCT (*pct_nbr*), Número de aplicación PCT (*internat_appln_nr*) y Número de aplicación o ID de aplicación (*appln_id*). Además incluye información acerca del nombre del inventor (*inv_name*), la dirección de residencia (*address*), así como la región (*reg_code*) y su país de residencia (*ctry_code*). Por último, se encuentran las variables de múltiple localización regional (*reg_share*) y la participación porcentual de inventor (*inv_share*), donde $inv_share \leq 1$.

Cuadro 2.3: Listado de variables del fichero de datos de inventores de aplicaciones de patentes PCT en REGPAT.

Variable	**Nomenclatura**	**Inventor 1**	**Inventor 2**
Número de publicación PCT	*pct_nbr*	WO1978000017	WO1978000017
Número de aplicación PCT	*internat_appln_nr*	WO1978CH0001	WO1978CH0001
Identificador de la aplicación	*appln_id*	6508281	6508281
Nombre del inventor	*inv_name*	SACHETTO, Jean - Pierre	REGNAULT, Alain
Dirección del inventor	*address*	F-74160 Saint- Julien -en-Genevois	F-01210 Par Ornex
Código de región	*reg_code*	FR718	FR711
Código país	*ctry_code*	FR	FR
Múltiple localización regional	*reg_share*	1.00	1.00
Conteo fraccional inventores	*inv_share*	0.50	0.50

Siguiendo con el ejemplo anterior, se seleccionó en el fichero de datos de inventores la aplicación de patente PCT con número de publicación PCT "$WO1978000017$". Tal como se ilustra en el cuadro 2.3, la aplicación de patente PCT tiene dos inventores. El inventor 1: SACHETTO, Jean- Pierre, y el inventor 2: REGNAULT, Alain. Ambos inventores residen en Francia (FR), en las regiones con NUTS FR718 y FR711, respectivamente. La aplicación de patente PCT tiene una participación porcentual de inventores $inv_share = 0.5$, resultado de dividir 1 entre el número total de inventores, en este caso, 2.

■ **Fichero de datos de códigos tecnológicos IPC en aplicaciones de patentes PCT de la base de datos REGPAT.**

El tercer fichero de datos para aplicaciones de patente PCT en REGPAT contiene las variables: Número de publicación PCT (*pct_ nbr*), Año de prioridad (*prio_year*), Año de aplicación (*app_year*) y Códigos tecnológicos según la Clasificación Internacional de Patentes IPC (*IPC*). En el cuadro 2.4 se proporciona un ejemplo de la estructura del fichero de datos de códigos IPC para una aplicación de patente PCT.

Cuadro 2.4: Listado de variables del fichero de datos de códigos tecnológicos IPC de aplicaciones de patentes PCT en REGPAT.

Variable	**Nomenclatura**	**Ejemplo aplicación PCT**
Número de publicación PCT	*pct_nbr*	WO1978000017
Año de prioridad	*prio_year*	1977
Año de aplicación	*app_year*	1978
Lista de códigos IPC	*IPC*	A23K50/15, C07H13/12

La variable año de prioridad (*prio_year*) corresponde al año en que se efectúa la primera solicitud de patente para cada una de las aplicaciones de patente PCT. El año de prioridad es comunmente entendido como un periodo de 12 meses contado a partir de la fecha de prioridad (primera solicitud). Durante este periodo, el aplicante puede solicitar de forma directa la protección de su invención en una oficina de patente nacional, por vía regional (por ejemplo, ante la EPO) o internacional mediante el procedimiento PCT.

La fecha de prioridad es la primera fecha de solicitud o aplicación de una patente ante cualquier oficina de patente a nivel mundial, con el ánimo de proteger una invención. La fecha de prioridad, en nuestro caso el año de prioridad, puede considerarse como la fecha más cercana al momento en que ocurre una invención. Por ello, cuando se compilan estadísticas de patentes con el objetivo de reflejar la actividad inventiva, se recomienda su uso como fecha de referencia (OCDE, 2009, p.53).

El año de aplicación (*app_year*) corresponde en el fichero de datos al año de aplicación o solicitud de patente ante la EPO. La base de datos de patentes PCT en REGPAT comprende las aplicaciones de patente PCT en las cuales la EPO es una oficina designada. El año de aplicación puede ser igual o posterior al año de prioridad, pero en ningún caso anterior.

Por último, el fichero de datos contiene los códigos tecnológicos de acuerdo a la Clasificación Internacional de Patentes IPC asignados a cada una de las aplicaciones de patente PCT. Para la comprensión de los códigos tecnológicos IPC, se detalla a continuación la estructura jerárquica y el diseño de los símbolos en la Clasificación Internacional de Patente IPC. Con esta finalidad, se consultó la más reciente versión de la Guia IPC (año 2022), documento publicado y actualizado anualmente por la Organización Mundial de la Propiedad Intelectual WIPO.

En el primer nivel de la clasificación IPC encontramos las secciones que "*representan el cuerpo de conocimientos que pueden considerarse propios del campo de la patente*"(WIPO, 2019, p.3). Tal como se muestra en el cuadro 2.5, en la Clasificación Internacional de Patentes existen ocho (8) secciones, cada una de ellas representada por una letra mayúscula.

Cuadro 2.5: Secciones de la Clasificación Internacional de Patentes IPC

Sección:	**Nombre de la sección**
A	Necesidades humanas
B	Operaciones; Transporte
C	Química; Metalurgia
D	Textiles; Papel
E	Construcciones fijas
F	Ingeniería mecanica; Iluminación; Calefacción; Armas; Voladura.
G	Física
H	Electricidad

Para ofrecer un ejemplo de la estructura de los códigos tecnológicos IPC hemos tomado como referencia el código IPC, A23K50/15, que describe a la aplicación de patente PCT "WO1978000017". Este código tecnológico IPC corresponde a la sección "A"denominada "Necesidades humanas".

$$\underset{\text{Sección}}{\underbrace{A}}\ 23K50/15$$

En el segundo y tercer nivel jerárquico tenemos las clases y las subclases tecnológicas. La clase tecnológica es la gran categoría de la invención y se identifica con una letra en mayúscula seguida de un número de dos dígitos:

$$\underbrace{A23}_{\text{Clase tecnológica}} \qquad K50/15$$

Las subclases tecnológicas "*delinean procesos, características estructurales y especificaciones funcionales de la clase tecnológica*"(Youn et al, 2014, p.3), y su símbolo está compuesto por el símbolo de la clase tecnológica seguido de una letra mayúscula:

$$\underbrace{A23K}_{\text{Subclase tecnológica}} \qquad 50/15$$

Una subclase tecnológica se divide en grupos. Estos grupos pueden ser grupos principales o subgrupos. El símbolo del grupo consiste en el símbolo de la subclase tecnológica seguido por dos números separados por una línea oblicua. El símbolo del grupo principal tiene la misma estructura de grupo, pero en este caso la línea oblicua es seguida por dos ceros "00":

$$\overbrace{\underbrace{A23K}_{\text{Subclase tecnológica}} \qquad 50/00}^{\text{Grupo principal}}$$

El subgrupo es el nivel inferior de la clasificación IPC y depende del nivel del grupo principal. El símbolo del subgrupo consiste del símbolo de la subclase tecnológica seguido por, entre 1 y 3 números del grupo principal, una línea oblicua y al menos dos números diferentes de "00":

$$\overbrace{\underbrace{A23K}_{\text{Subclase tecnológica}} \qquad 50/15}^{\text{Subgrupo}}$$

En el cuadro 2.6 se muestran los niveles jerárquicos, símbolos y descripción del código tecnológico "$A23K50/15$" en la clasificación IPC. La descripción del código tecnológico IPC en el cuadro 2.6 conserva la grafía de la clasificación IPC disponible para la consulta de usuarios en `https://www.wipo.int/classifications/ipc`.

Cuadro 2.6: Ejemplo código tecnológico "$A23K50/15$" en la Clasificación Internacional de Patentes IPC.

Nivel jerárquico	**Símbolo**	**Descripción**
Sección:	A	Necesidades humanas
Clase:	A23	Alimentos o productos alimenticios; su tratamiento, no cubierto por otras clases.
Subclase:	A23K	Alimentos especialmente adaptados para animales; métodos especialmente adaptados para su producción.
Grupo principal:	A23K 50/00	Piensos especialmente adaptados para animales particulares.
Subgrupo:	A23K 50/15	...que contienen sustancias que se convierten metabólicamente en proteínas.

El fichero de datos de códigos tecnológicos IPC en REGPAT considerado en este investigación contiene 14050714 observaciones con información completa, para 3789945 aplicaciones de patentes PCT, con fecha de prioridad entre los años 1977 y 2018. Estas aplicaciones de patente PCT comprenden 72619 códigos tecnológicos diferentes.

■ Fichero de datos de Códigos tecnológicos CPC en la base de datos REGPAT

La base de datos REGPAT proporciona un fichero de datos independiente que se corresponde con la Clasificación Cooperativa de Patentes, CPC. Esta clasificación es el resultado del esfuerzo conjunto de las Oficinas de Marcas y Patentes de patentes de Estados Unidos (USPTO) y la Oficina Europea de Patentes (EPO) para armonizar los sistemas de clasificación de patentes, y lograr una estructura similar a la Clasificación Internacional de Patentes IPC administrada por la Organización Mundial de la Propiedad Intelectual (OMPI).

De acuerdo con la Guia de la Clasificación Cooperativa de Patentes, versión 2017, el esquema y las definiciones de la CPC pueden considerarse, en términos generales, similares a la clasificación IPC. En la publicación se distinguen 3 partes o compartimentos en el esquema de la clasificación CPC. En primer lugar, encontramos el compartimento principal denominado "Main Trunk", en donde una proporción

Cuadro 2.7: Secciones de la Clasificación Cooperativa de Patentes CPC

Sección:	**Nombre de la sección**
A	Necesidades humanas
B	Operaciones; Transporte
C	Química; Metalurgia
D	Textiles; Papel
E	Construcciones fijas
F	Ingeniería mecanica; Iluminación; Calefacción; Armas; Voladura.
G	Física
H	Electricidad
Y	Etiquetado general de nuevos desarrollos tecnológicos; etiquetado general de tecnologías transversales que comprenden varias secciones IPC; temas técnicos cubiertos por antiguas colecciones de arte de referencia cruzada de la USPC [XRACS] y resúmenes.

considerable de los códigos CPC son un refinamiento de los actuales códigos tecnológicos IPC, incluyendo más subdivisiones.

En esta parte del esquema de la clasificación CPC, los símbolos de los códigos tecnológicos se corresponden con los símbolos estandar (números y letras) usados para idénticos niveles en la clasificación IPC. Los códigos asignados a una patente bajo la clasificación CPC en el Main trunk contienen los aspectos relacionados con la novedad de la invención.

Los restantes compartimentos de la clasificación CPC denominados códigos indexados y la sección Y, sección adicional en comparación a la clasificación IPC (ver cuadros 2.5 y 2.7), son reservados para la codificación de las características de la invención que con anterioridad hacen parte del estado de la técnica. Más detalle de las definiciones esenciales y diferencias de la clasificación CPC con respecto a la clasificación IPC pueden ser consultados en la Guia CPC (2017), última versión disponible en `https://www.cooperativepatentclassification.org/`

El fichero de datos de códigos CPC de REGPAT incluye dos variables: Número de aplicación PCT (*appln_id*) que sirve para relacionar este fichero con otros ficheros de datos, y los códigos tecnológicos CPC (*CPC_class*). Este fichero considerado en la investigación contiene 51169333 observaciones correspondientes a 241455 códigos tecnológicos diferentes.

2.1.2. Conceptualización de las variables.

En este apartado, se presenta un ejemplo estilizado que facilita la comprensión de la invención como resultado del proceso combinatorio de tecnologías (Clancy, 2018; Fleming y Sorenson, 2001; Parraguez et al., 2020; Solé, Amor y Valverde, 2016; Strumsky y Lobo, 2015; Taalbi, 2017; Youn et al., 2015). En este ejemplo se tomó como referencia la propuesta metodológica de Youn et al., (2015) que detalla la forma como se acumulan las patentes (P), Códigos tecnológicos (T) y Combinaciones (C) en el sistema o registro de patentes.

Las aplicaciones de patentes PCT seleccionadas para el ejemplo fueron extraídas de la base de datos REGPAT y cumplen la función de orientar e ilustrar al lector. Se entenderá que las aplicaciones de patente PCT consideradas en este ejemplo, corresponderán a la totalidad de aplicaciones PCT de un año i .

Ahora bien, supongamos que en dicho año i, hay 2 aplicaciones de patentes PCT. Cada una de estas aplicaciones se identifica por su número de publicación de patente PCT (*pct_nbr*). A su vez, estas aplicaciones de patente PCT son descritas por un conjunto de códigos tecnológicos IPC:

$$\begin{aligned} \textit{Aplicación de patente PCT 1} &= \text{WO1978000001} \\ \textit{Aplicación de patente PCT 2} &= \text{WO1978000003} \end{aligned}$$

$$\text{Códigos IPC patente PCT 1} = \left\{ \begin{array}{c} \text{Código 1: } B01J037/02 \\ \text{Código 2: } B01D017/02 \\ \text{Código 3: } C12H001/044 \\ \text{Código 4: } B01J020/28 \\ \text{Código 5: } G21F009/12 \\ \text{Código 6: } B01J035/06 \\ \text{Código 7: } C03C023/00 \\ \text{Código 8: } C03B019/12 \\ \text{Código 9: } C03C013/04 \\ \text{Código 10: } C02F001/68 \\ \text{Código 11: } C08J005/22 \\ \text{Código 12: } C03B037/016 \\ \text{Código 13: } C03C011/00 \\ \text{Código 14: } C03C003/00 \\ \text{Código 15: } B01D039/20 \end{array} \right\}$$

$$\text{Códigos IPC aplicación 2} = \left\{ \begin{array}{c} \text{Código 1: } F16B007/08 \\ \text{Código 2: } F16L003/12 \\ \text{Código 3: } F16L003/13 \\ \text{Código 4: } F16B002/20 \\ \text{Código 5: } E05C019/06 \end{array} \right\}$$

Supongamos también que en el año $i+1$, hay 3 aplicaciones de patentes PCT adicionales, que tambien son descritas por códigos tecnológicos IPC :

$$\begin{array}{c} \textit{Aplicación de patente PCT 3} = WO1979000088 \\ \textit{Aplicación de patente PCT 4} = WO1979000418 \\ \textit{Aplicación de patente PCT 5} = WO1980000419 \end{array}$$

$$\text{Códigos IPC patente PCT 3} = \left\{ \begin{array}{c} \text{Código 1: } F16H041/26 \\ \text{Código 2: } F16H061/56 \end{array} \right\}$$

$$\text{Códigos IPC patente PCT 4} = \left\{ \begin{array}{c} \text{Código 1: } G09C001/00 \\ \text{Código 2: } H04L009/06 \\ \text{Código 3: } G09C001/10 \\ \text{Código 4: } H04L009/00 \end{array} \right\}$$

$$\text{Códigos IPC patente PCT 5} = \left\{ \begin{array}{c} \text{Código 1: } B01D029/11 \\ \text{Código 2: } B01D039/16 \\ \text{Código 3: } B01D039/20 \\ \text{Código 4: } D02G003/36 \\ \text{Código 5: } D02G003/44 \end{array} \right\}$$

El total de aplicaciones de patentes PCT en los dos años anteriores, estaría representado por el conjunto *Número de patentes PCT*. Este conjunto contiene un total de 5 aplicaciones de patentes PCT:

$$\textit{Número de patentes PCT} = \left\{ \begin{array}{c} \text{Aplicación de patente PCT 1: } WO1978000001 \\ \text{Aplicación de patente PCT 2: } WO1978000003 \\ \text{Aplicación de patente PCT 3: } WO1979000088 \\ \text{Aplicación de patente PCT 4: } WO1979000418 \\ \text{Aplicación de patente PCT 5: } WO1980000419 \end{array} \right\}$$

El total de códigos tecnológicos, de dichas aplicaciones de patentes PCT, en los dos años anteriores, estaría representado por el conjunto *Número de códigos IPC*. En nuestro ejemplo, este conjunto contiene un total de 30 códigos IPC:

$$\textit{Número de códigos IPC} = \left\{ \begin{array}{ccc} B01J037/02 & B01D017/02 & C12H001/044 \\ B01J020/28 & G21F009/12 & B01J035/06 \\ C03C023/00 & C03B019/12 & C03C013/04 \\ C02F001/68 & C08J005/22 & C03B037/016 \\ C03C011/00 & C03C003/00 & B01D039/20 \\ F16B007/08 & F16L003/12 & F16L003/13 \\ F16B002/20 & E05C019/06 & F16H041/26 \\ F16H061/56 & G09C001/00 & H04L009/06 \\ G09C001/10 & H04L009/00 & B01D029/11 \\ B01D039/16 & D02G003/36 & D02G003/44 \end{array} \right\}$$

A continuación, si observamos en particular los conjuntos de códigos tecnológicos IPC asignados a la aplicación de patente PCT 1 y la aplicación de patente PCT 5, el conjunto de códigos tecnológicos de las aplicaciones de patente PCT mencionadas tienen un elemento en común, el código IPC $B01D039/20$. La primera aparición o uso del código tecnológico IPC, $B01D039/20$, ocurre en la aplicación de patente PCT 1 en el año i. Por su parte, el segundo uso del código tecnológico IPC $B01D039/20$ en la descripción de una aplicación de patente PCT se efectuó en el año $i+1$, especificamente, en la aplicación de patente PCT 5. En este caso, el código tecnológico aparecerá en el conjunto *Número de códigos IPC* una sola vez, y se asociará al año de su primera aparición.

Por tanto, a partir de los conjuntos de códigos tecnológicos IPC que describen las aplicaciones de patente PCT, se puede además determinar si los códigos IPC usados, son nuevos o reutilizados. Si en una aplicación de patente PCT es empleado un código tecnológico IPC no usado previamente, este código IPC es considerado como un *Código IPC nuevo*. Si en la descripción de una aplicación de patente PCT se utiliza un código tecnológico IPC que ha sido usado previamente en una aplicación de patente PCT, dicho código IPC será considerado un *Código IPC reutilizado*.

$$\text{Códigos IPC patente PCT 1} = \left\{ \begin{array}{c} \text{Código 1: } B01J037/02 \\ \text{Código 2: } B01D017/02 \\ \text{Código 3: } C12H001/044 \\ \text{Código 4: } B01J020/28 \\ \text{Código 5: } G21F009/12 \\ \text{Código 6: } B01J035/06 \\ \text{Código 7: } C03C023/00 \\ \text{Código 8: } C03B019/12 \\ \text{Código 9: } C03C013/04 \\ \text{Código 10: } C02F001/68 \\ \text{Código 11: } C08J005/22 \\ \text{Código 12: } C03B037/016 \\ \text{Código 13: } C03C011/00 \\ \text{Código 14: } C03C003/00 \\ \textbf{Código 15: B01D039/20} \end{array} \right\}$$

$$\text{Códigos IPC patente PCT 5} = \left\{ \begin{array}{c} \text{Código 1: } B01D029/11 \\ \text{Código 2: } B01D039/16 \\ \textbf{Código 3: B01D039/20} \\ \text{Código 4: } D02G003/36 \\ \text{Código 5: } D02G003/44 \end{array} \right\}$$

Siguiendo el ejemplo planteado, los códigos tecnológicos IPC que describen la aplicación de patente PCT 1, son nuevos en su totalidad (considerando que no aparecieron en años previos). Esta aplicación de patente PCT contiene un total de 15 códigos IPC nuevos. Por su parte, la aplicación de patente PCT 5 del año $i + 1$, tiene un total de 5 códigos IPC, 4 códigos IPC nuevos y 1 código IPC reutilizado. El código reutilizado en la aplicación de patente PCT 5 es el código tecnológico IPC $B01D039/20$, resaltado en negrita.

Por último, se entenderá como una Combinación aquel código tecnológico IPC (en solitario) o el conjunto de códigos tecnológicos IPC asignados a una aplicación de patente PCT. Toda aplicación de patente PCT es descrita por una combinación de códigos tecnológicos IPC. Se presenta a modo de ejemplo, la combinación de códigos tecnológicos IPC para la aplicación de Patentes PCT 1 del año 1, correspondiente al número de publicación PCT "WO1978000001".

$$\text{Combinación patente PCT 1: } = \left\{ \begin{array}{c} B01D017/02, B01D039/20, B01J020/28, \\ B01J035/06, B01J037/02, C02F001/68, \\ C03B019/12, C03B037/016, C03C003/00, \\ C03C011/00, C03C013/04, C03C023/00, \\ C08J005/22, C12H001/044, G21F009/12 \end{array} \right\}$$

El total de combinaciones, de dichas aplicaciones de patentes PCT, en los dos años anteriores, estaría representado por el conjunto *Número de combinaciones*. En nuestro ejemplo, este conjunto contiene un total de 5 combinaciones, dado que son diferentes entre ellas, y por tanto, en este caso, el conjunto *Número de combinaciones* sería:

$$\textit{Número de combinaciones}= \left\{ \begin{array}{c} \text{Combinación aplicación 1} \\ \text{Combinación aplicación 2} \\ \text{Combinación aplicación 3} \\ \text{Combinación aplicación 4} \\ \text{Combinación aplicación 5} \end{array} \right\}$$

Ahora bien, supongamos que, existiesen aplicaciones de patente PCT con igual o diferente año de prioridad (o de invención) pero con una misma combinación de códigos tecnológicos. Una combinación siempre será contada en una única ocasión, por ello, bajo el supuesto anterior, la combinación será asignada a la primera aplicación de patente PCT que contenga dicha combinación en el orden establecido en la base de datos. El orden de las aplicaciones de patentes PCT en la base de datos se determina por el año de prioridad (orden cronológico), y número de publicación PCT (*pct_nbr*), de menor a mayor.

Finalmente, indicar que las combinaciones que describen las aplicaciones de patente PCT pueden ser clasificadas en función de los códigos IPC que las constituyen. En este sentido, (Strumsky y Lobo, 2015; Lobo y Strumsky, 2019) proponen una taxonomía de la novedad inventiva que proporciona una serie de definiciones útiles para la identificación de los recursos de novedad tecnológica y la categorización de la novedad de las patentes de invención a partir de los códigos tecnológicos IPC que las describen. Los autores mencionan 4 definiciones: origen, combinación novedosa, refinamiento y combinación. Para los efectos de esta investigación, se entenderán cada una de estas definiciones de la siguiente manera:

- **Origen**: son aquellas combinaciones que contienen exclusivamente *Códigos IPC nuevos*. En estas combinaciones, todos los códigos tecnológicos IPC de la combinación no han sido utilizados previamente en el sistema de patentes considerado.
- **Combinación novedosa**: son aquellas combinaciones de códigos tecnológicos IPC que comprenden tanto códigos tecnológicos nuevos como reutilizados.
- **Combinación**: la combinación está constituida por códigos IPC reutilizados, y además, esta combinación de códigos no ha sido usada previamente.
- **Refinamiento**: la combinación de códigos IPC reutilizados que describe una aplicación de patente PCT ha sido usada previamente en el sistema de patentes.

Por tanto, las combinaciones de códigos tecnológicos IPC que describen una aplicación de patente PCT pueden contener sólo códigos IPC nuevos (combinación origen), códigos nuevos y códigos reutilizados (combinación novedosa), sólo códigos reutilizados cuya combinación aparece por primera vez (combinación), o finalmente, sólo códigos reutilizados cuya combinación ya apareció previamente (refinamiento).

Siguiendo con nuestro ejemplo, y teniendo en cuenta las definiciones previamente expuestas, clasificamos las 5 combinaciones de códigos IPC en aplicaciones de patente PCT que corresponden a los años i e $i+1$. Las aplicaciones de patente PCT 1, 2, 3 y 4 son descritas por una combinación de códigos IPC que aparecen por primera vez en el registro de patentes, y por ello, estas combinaciones son consideradas Origen.

Por otro lado, la combinación de códigos tecnológicos asociada a la aplicación de patente PCT 5, incluye el código IPC $B01D039/20$ previamente usado en la aplicación de patente PCT 1. En este caso particular, el código IPC $B01D039/20$ en su segundo uso (en la aplicación de patente PCT 5) es acompañado de códigos IPC nuevos, y por ende, diferentes a aquellos códigos IPC de la aplicación de patente PCT en la que se efectuó su primer uso. Esta combinación es clasificada como una combinación novedosa.

2.1.3. Software estadístico

En el procesamiento de las bases de datos se utilizó el Software R. Este lenguaje de programación y software libre para la computación estadística y representación gráfica de datos permite la reproducibilidad y replicación de análisis de datos al ser un programa conducido por comandos en línea. El software R posibilita la experimentación y el aprendizaje desde diferentes enfoques analíticos basados en un mismo conjunto de datos.

El software R puede considerarse extensible, al encontrarse en constante actualización. Usuarios procedentes de diversas áreas de conocimiento desarrollan *paquetes*, que son un conjunto de funciones, usadas para conducir todo tipo de análisis. El sistema de paquetes del Software R organiza y coordina su desarrollo a través de 3 relaciones: relación de dependencia, relación de importación y relación de recomendación (Chen et al., 2019).

De acuerdo con los autores, la relación de dependencia se establece cuando un nuevo paquete de programación requiere para su uso la ejecución de paquetes de software existentes. Esta relación constituye en sí misma la fuerza que motiva la expansión de R. La relación de importación evita posibles conflictos entre diferentes paquetes que cumplen una misma función, determinando la autoría del paquete de software. Entre tanto, la relación de recomendación sugiere al desarrollador el uso de varios paquetes de software para la solución de problemas complejos, de forma tal que los usuarios puedan tomar las mejores decisiones.

En el repositorio de documentos del Software R, Comprehensive R Archive Network (CRAN) se encuentran disponibles para descarga las actualizaciones del software y paquetes. Como referencia, en la reciente actualización del software R correspondiente a la versión R-4.1.3.tar.gz, con fecha de publicación del 10 de marzo de 2022, se encuentran disponibles para consulta y descarga más de 18000 paquetes del Software R.

En el procesamiento de los datos se utilizó el paquete *dplyr*. La versión 0.1.1 del paquete *dplyr*, autoria de Hadley Wickham y Romain Francois, se publicó en enero de 2014. En la reciente versión 1.0.7 de junio de 2021, se añaden dos nuevos autores, Lionel Henry y Kirill Muller. El paquete *dplyr* es una herramienta rápida y estable, que consiste en una gramática (conjunto de verbos) que facilita la manipulación de bases de datos.Entre los verbos destacados en este paquete del Software R se encuentran: *filter* (permite obtener un conjunto de observaciones usando condiciones lógicas), *arrange*(reordenar filas), *summarise* (generación de resúmenes de

variables), *mutate* (anãdir nuevas variables), *pipe* (concatenar acciones), entre otros verbos, que operan naturalmente en conjunto con *group_by* para realizar operaciones por grupos de variables.

2.1.4. Procesamiento y manipulación de datos

Lectura y depuración de la base de datos IPC

En primer lugar, se hizo la lectura del fichero de datos de códigos tecnológicos IPC asignados a las aplicaciones de Patente PCT (*202107_PCT_IPC.txt*). Tal como se mencionó en el apartado 2.1.1, el fichero de datos IPC de REGPAT proporciona para cada aplicación de patente PCT, información acerca de 4 variables: Número de publicación PCT *(pct_nbr)*, Año de prioridad *(prio_year)*, Año de aplicación *(app_year)* y Códigos tecnológicos IPC asignados *(IPC)*.

```
rm(list=ls())
library(dplyr)
first.year<-1977
last.year<-2018

numyears<-last.year-first.year+1
> numyears
[1] 42
```

Siguiendo la recomendación de la OCDE (2009) se usó el año de prioridad *(prio_year)* en vez del año de aplicación *(app_year)* como fecha de referencia. El año de prioridad es la fecha más adecuada para reflejar el periodo en que se origina una invención. La base de datos de REGPAT, versión de julio de 2021, comprende aplicaciones de patente PCT con año de prioridad entre 1977 y 2020.

Para definir el periodo de análisis de 42 años, se tuvo en cuenta el tiempo de 18 meses que puede transcurrir, desde la solicitud de una patente PCT hasta el momento de su publicación. Por ello, buscando garantizar la fiabilidad de los resultados, se estimó prudente considerar en el procesamiento de datos [**] aquellas aplicaciones de patente PCT con año de prioridad *(prio_year)* entre 1977 (año inicial o

[**]En los apartados que constituyen el capítulo de Datos y Metodología se intenta mostrar al lector, paso a paso, la ejecución del código de programación en el entorno R Studio. El lector puede visualizar fragmentos del código de programación empleado. Estos fragmentos de código conservan el tipo de letra propia del entorno de programación. De igual manera, dentro del fragmento del código de programación, se resaltan en color azul aquellas palabras que suponen la ejecución de comandos o funciones especiales, y en color verde, los comentarios, a modo de aclaración, agregados por el autor. En el documento los fragmentos de códigos de programación son acompañados por ejemplos. Estos ejemplos consisten en ordenes dadas por el autor para obtener un resultado y se identifican al estar precedidas por el símbolo "Mayor que"(>) situado al inicio de la línea de texto.

`first.year`) y 2018 (año final o `last.year`).

En el siguiente fragmento del código de programación se muestra la lectura del fichero de datos:

```
DatosIPC<-read.table(202107_PCT_IPC.txt, sep =|,
header = TRUE, dec =.)
```

En un primer ejercicio de exploración mediante la función `dim` se logró determinar la dimensión de la base de datos. La base de datos IPC inicial tiene un total de 14050714*** observaciones, y 4 columnas con información correspondiente a las variables previamente mencionadas.

```
> dim(DatosIPC)
[1] 14050714        4

> colnames(DatosIPC)
[1] pct_nbr  prio_year app_year  IPC
```

Para conocer la estructura inicial de la base de datos de códigos tecnológicos IPC, se realizó una visualización de las primeras observaciones utilizando la función `head`. Las variables se situan en la base de datos de la siguiente manera: en la columna 1, el número de publicación PCT *(pct_nbr)*; en la columna 2, el año de prioridad *(prio_year)*; en la columna 3, el año de aplicación *(app_year)*; y por último, en la columna 4, los códigos tecnológicos IPC *(IPC)*.

```
> head(DatosIPC)
       pct_nbr prio_year app_year         IPC
1 WO2001028908      1999     2000  B65H054/84
2 WO2001028908      1999     2000  B65H054/76
3 WO2001030055      1999     2000 H04M001/725
4 WO2001030055      1999     2000  H04M001/57
5 WO2001030055      1999     2000 H04M001/677
6 WO2001030055      1999     2000 H04M001/665
```

Paso seguido, la base de datos fue depurada eliminando aquellas observaciones con información incompleta y/o con fecha de prioridad fuera del periodo de estudio. Para obtener las observaciones correspondientes a las aplicaciones de patente PCT dentro del periodo de estudio, se determinaron los valores que podría tomar la variable año de prioridad *(prio_year)* en la columna 2 de la base de datos.

***Teniendo en cuenta que el procesamiento y análisis de la base de datos se efectua en el lenguaje de programación R, se decidió prescindir del uso del signo punto (.) para indicar cantidades en miles. En este caso, el signo punto (.) fue usado para indicar las cifras decimales.

La variable Año de prioridad *(prio_year)* tomó valores que podrían ser, mayores que o igual (>=) al año inicial o `first.year`, 1977, y menores que o igual (<=) al año final o `last.year`, 2018. El número de observaciones en la base de datos con *prio_year* entre 1977 y 2018 fue 13625614:

```
DatosIPC<-DatosIPC[DatosIPC[,2]>=first.year,]
DatosIPC<-DatosIPC[DatosIPC[,2]<=last.year,]
```

A continuación, se eliminaron aquellas observaciones con valores NAs y/o sin datos (vacías). En la depuración se usó la función `is.na` para identificar valores perdidos, `NAs`, en las columnas 1, 2 y 4. Con esta función se obtiene un vector lógico con valores `TRUE` o `FALSE`, en donde `TRUE` representa los valores perdidos `NAs` en los vectores de datos. Los valores `NAs` en la base de datos son eliminados con la expresión `!is.na`. En este caso, se encontró que la base de datos no contenía valores perdidos `NAs`.

```
DatosIPC<-DatosIPC[!is.na(DatosIPC[,1]),] # pct_nbr
DatosIPC<-DatosIPC[!is.na(DatosIPC[,2]),] # prio_year
DatosIPC<-DatosIPC[!is.na(DatosIPC[,4]),] # IPC
```

En el proceso de depuración de la base de datos fueron eliminadas las observaciones en donde al menos una de la columnas de interés no contuviese datos (vacías). En la columna 1, Número de publicación PCT *(pct_nbr)* y en la columna 2, Año de prioridad *(prio_year)* no se encontraron observaciones sin datos. Entre tanto, en la columna 4 correspondiente a la variable Códigos tecnológicos IPC *(IPC)*, el número de observaciones sin datos fue de 2122. Estas observaciones fueron eliminadas y la dimensión de base de datos se redujo a 13623492 observaciones:

```
DatosIPC<-DatosIPC[DatosIPC[,1]!=,]
DatosIPC<-DatosIPC[DatosIPC[,2]!=,]
DatosIPC<-DatosIPC[DatosIPC[,4]!=,]

> dim(DatosIPC)
[1] 13623492        4
```

Para identificar posibles inconsistencias en la base de datos se efectuó una comprobación que consistió en determinar si cada una de las aplicaciónes de patente PCT estaba asociada a un único Año de prioridad *(prio_year)*. Con el operador *pipe* (%>%) que permite la concatenación de acciones, se creó un objeto en donde fueron agrupadas las observaciones de la base de datos según la columna 1, Número de publicación PCT *(pct_nbr)*. Además, se generó una nueva variable o columna denominada *Comprobación* que es resultado de calcular la longitud de valores únicos de Año de prioridad *(prio_year)* para cada aplicación PCT.

```
Comprobacion<-DatosIPC %>% %group_by(pct_nbr) %>% %
                  summarise(comprobar=length(unique(prio_year)))
```

```
> head(Comprobacion)

  pct_nbr       comprobar
  <chr>             <int>
1 WO1978000001          1
2 WO1978000003          1
3 WO1978000004          1
4 WO1978000005          1
5 WO1978000006          1
6 WO1978000007          1
```

El vector aux fue creado para almacenar aquellas aplicaciones de patente PCT asociadas a más de un valor en la variable Año de prioridad *(prio_year)*. En otras palabras, en el vector aux se almacenaron aquellas observaciones o aplicaciones de patentes PCT del objeto Comprobación para las cuales la columna 2 toma un valor mayor que uno (> 1). De este modo, se identificaron 20 aplicaciones de patentes PCT con inconsistencias.

```
aux<-Comprob$pct_nbr[Comprob$comprobar>1]
```

Las aplicaciones de patente PCT del vector aux fueron eliminadas de la base de datos DatosIPC que ve reducida su dimensión en 126 observaciones, pasando de 13623492 a 13623366.

```
DatosIPC[DatosIPC$pct_nbr==aux[1],]
DatosIPC<-DatosIPC[which(!DatosIPC$pct_nbr %in %aux),]
> dim(DatosIPC)
[1] 13623366        4
```

A continuación, se generó la base de datos denominada como BaseLarga_IPC tomando las columnas 1, 2 y 4 de la base de datos DatosIPC. En esta base de datos de formato largo (cada aplicación de patente PCT tiene tantas filas como códigos IPC la componen), se tiene información de los códigos tecnológicos IPC (*IPC*) y año de prioridad *(prio_year)* para cada número de publicación PCT (*pct_nbr*).

```
BaseLarga_IPC<-DatosIPC[,c(1,2,4)]
> dim(BaseLarga_IPC)
[1] 13623366        3

head(BaseLarga_IPC)
> head(BaseLarga_IPC)
       pct_nbr prio_year          IPC
1 WO2001028908      1999  B65H054/84
2 WO2001028908      1999  B65H054/76
```

```
3 WO2001030055      1999 H04M001/725
4 WO2001030055      1999  H04M001/57
5 WO2001030055      1999 H04M001/677
6 WO2001030055      1999 H04M001/665
```

Con el objetivo de facilitar más adelante la intersección de las bases de datos, se efectuó una agrupación de las observaciones en la base de datos `DatosIPC` según la variable *pct_nbr*, y se eliminaron posibles inconsistencias en la variable Año de prioridad (*prio_year*). Si un *pct_nbr* estuviese asociado a más de un valor *prio_year*, se tomaría el *prio_year* menor (más antiguo). Se determinó que `DatosIPC` contaba con información de 3789945 aplicaciones de Patente PCT.

```
DatosIPC<-DatosIPC[,c(1,2)] #Se quitan duplicados
DatosIPC<-DatosIPC %>% %group_by(pct_nbr) %>% %
summarise(prio_year=min(prio_year))
```

```
> dim(DatosIPC)
[1] 3789945       2
> head(DatosIPC)
# A tibble: 6 x 2
  pct_nbr      prio_year
  <chr>            <int>
1 WO1978000001      1977
2 WO1978000003      1977
3 WO1978000004      1977
4 WO1978000005      1977
5 WO1978000006      1977
6 WO1978000007      1977
```

Lectura y depuración de la base de datos CPC

Para obtener la base de datos de códigos tecnológicos CPC se procedió inicialmente a la lectura del fichero en la base de datos REGPAT (*202107_CPC_Classes.txt*). Tal como se puede observar en el siguiente fragmento del código de programación, la base de datos de códigos tecnológicos CPC consiste de 51169333 observaciones y 2 columnas. En la columna 1 de la base de datos se encuentra la variable Número de identificación de la aplicación PCT (*appln_id*), y en la columna 2, los Códigos tecnológicos CPC (*CPC_Class*):

```
DatosCPC<-read.table(202107_CPC_Classes.txt,
                                sep =|,header = TRUE,dec =.)
```

```
> dim(DatosCPC)
[1] 51169333        2

> colnames(DatosCPC)
[1] appln_id   CPC_Class
```

Al igual que en la base de datos de códigos IPC, se realizó un proceso de depuración. Fueron eliminadas inicialmente aquellas observaciones con valores perdidos NAs, y luego, las observaciones sin datos en las columnas 1 y 2. En el proceso de depuración no se encontraron observaciones con NAs ni datos faltantes, y por tanto, la dimensión de la base de datos se mantuvo inalterada.

```
DatosCPC<-DatosCPC[!is.na(DatosCPC[,1]),] # appln_id
DatosCPC<-DatosCPC[!is.na(DatosCPC[,2]),] # CPC_Class
DatosCPC<-DatosCPC[DatosCPC[,1]!=,]
DatosCPC<-DatosCPC[DatosCPC[,2]!=,]
```

En la columna 2 se observó que algunos códigos tecnológicos CPC tenian espacios intermedios que podrían suponer, más adelante, algun grado de dificultad en el procesamiento y análisis de la base de datos. Por esta razón, se instaló la libreria `stringr` que permite la manipulación de vectores de caracteres. Así, con la función `str_replace_all` se eliminaron los espacios intermedios en los códigos tecnológicos CPC:

```
library(stringr)
x<-DatosCPC[,2]
DatosCPC[,2]<-str_replace_all(x, fixed( ), )
```

Lectura y depuración de la base de datos de aplicantes PCT

La base de datos de aplicantes PCT de REGPAT es empleada en este estudio para relacionar las bases de datos de códigos tecnológicos a partir de la variable Número de identificación de la aplicación (*appln_id*). En primer lugar, se realizó la lectura de la base de datos de aplicantes PCT de REGPAT. Con la función `dim` se estableció la dimensión de la base de datos que consistia de un total de 4365343 observaciones y 11 columnas. Cada una de las columnas se incluía información acerca de las variables de la base de datos de aplicantes PCT detalladas en el apartado 2.1.1.

```
DatosPCT<-read.table(202107_PCT_App_reg.txt,sep =|,
header = TRUE,dec =.)
>  dim(DatosPCT)
[1] 4365343      11
```

Seguidamente, en la etapa de depuración de la base de datos fueron eliminadas las observaciones con valores perdidos NAs y/o con datos faltantes en las columnas de interés para el estudio. Especificamente, la verificación de inconsistencias en la base de datos de aplicantes PCT fue realizada en las columnas 1 y 3, que contienen información de las variables Número de publicación PCT (*pct_nbr*) PCT y Número de identificación de la aplicación (*appln_id*), respectivamente.

```
DatosPCT<-DatosPCT[!is.na(DatosPCT[,1]),] # pct_nbr
DatosPCT<-DatosPCT[!is.na(DatosPCT[,3]),] # appln_id
DatosPCT<-DatosPCT[DatosPCT[,1]!=,]
DatosPCT<-DatosPCT[DatosPCT[,3]!=,]

>  dim(DatosPCT)
[1] 4360392      11

>  colnames(DatosPCT[,c(1,3)])
[1] pct_nbr appln_id
```

La dimensión de la base de datos de aplicantes PCT se ve reducida en 4951 observaciones. Dicha reducción obedece a la existencia de valores perdidos `NAs` en la columna 3 de la variable Número de identificación de la aplicación (*appln_id*). Tras este paso, la dimensión de la base de datos es de 4360392 observaciones y 11 columnas.

Una comprobación final fue realizada creando dos objetos en los cuales se concatenaron acciones con el operador `pipe`. La finalidad que perseguían estas acciones era constatar que ningún número de publicación PCT (*pct_nbr*) estuviese asociado a más de un número de aplicación (*appln_id*), y viceversa.

En un primer objeto, las observaciones fueron inicialmente agrupadas según la variable Número de publicación PCT (*pct_nbr*). Después, se generaba una nueva variable que indicaba la longitud del vector de valores únicos de la variable Número de identificación de la aplicación PCT (*appln_id*) asociadas a cada *pct_nbr*. En aquellos casos en que la nueva variable tomase un valor mayor que uno (> 1), el *pct_nbr* relacionado era incorporado en el vector `aux`, y posteriormente, sus correspondientes observaciones eliminadas de la base de datos.

En el siguiente extracto de código de programación se puede observar la concatenación de acciones para la verificación de la base de datos `DatosPCT`. Asimismo, se presenta la estructura del objeto `Comprob` con la función `head`. En la columna 1 del objeto se encuentran las diferentes aplicaciones de patente PCT *pct_nbr* de la base de datos, y en la columna 2, la cantidad de *appln_id* que le son asociados a cada una de ellas:

```
Comprob<-DatosPCT %>%group_by(pct_nbr)  %>%
                  summarise(comprobar=length(unique(appln_id)))
> head(Comprob)
# A tibble: 6 x 2
  pct_nbr       comprobar
  <chr>             <int>
1 WO1978000001          1
2 WO1978000003          1
```

```
3 WO1978000004          1
4 WO1978000005          1
5 WO1978000006          1
6 WO1978000007          1
```

En el vector `aux` se almacenaron aquellas aplicaciones de patente PCT asociadas a más de un valor en la variable Número de identificación de la aplicación *(appln_id)*. En el vector `aux` se guardaron aquellas observaciones o aplicaciones de patentes PCT del objeto, `Comprob`, para las cuales la columna 2 `comprobar` tomó un valor mayor que uno (> 1). La verificación arrojó como resultado que 3 aplicaciones de patente PCT (identificadas por su *pct_nbr*) tenían asociado más de un valor de la variable Número de identificación de la aplicación (*appln_id*).

```
aux<-Comprob$pct_nbr[Comprob$comprobar>1]
> aux
[1] WO2000000662  WO2000008114  WO2019083002
```

Las aplicaciones de patente PCT con *pct_nbr* "WO2000000662", "WO2000008114"y "WO2019083002" contenidas en el vector `aux` fueron eliminadas de la base de datos. Con la eliminación de estas aplicaciones de patente PCT, la dimensión de la base de datos se redujo en 6 observaciones, situandose en 4360386 observaciones.

```
DatosPCT[DatosPCT$pct_nbr==aux[1],]
DatosPCT<-DatosPCT[which(!DatosPCT$pct_nbr %in %aux),]

> dim(DatosPCT)
[1] 4360386      11
```

En el segundo objeto, las observaciones son agrupadas según la variable Número de identificación de la aplicación PCT (*appln_id*). En esta ocasión, se crea una nueva variable denominada `comprobar` que indica cuantos valores único de la variable *pct_nbr* asociado a cada *appln_id*.

```
Comprob<-DatosPCT %>% %group_by(appln_id)    %>% %
           summarise(comprobar=length(unique(pct_nbr)))

> head(Comprob)

  appln_id comprobar
     <int>     <int>
1     6697         1
2     6699         1
3     6701         1
4     6703         1
5     6705         1
6     6707         1
```

Siguiendo la misma lógica aplicada en el primer objeto, en aquellos casos en que la variable `comprobar` tomase un valor mayor que uno (> 1), dichas observaciones serían incorporadas en el vector `aux`, y posteriormente, eliminadas de la base de datos. Los resultados de la comprobación muestran que el vector `aux` no contiene elementos, y por lo tanto, no se requiere eliminar observaciones.

```
aux<-Comprob$appln_id[Comprob$comprobar>1]
> aux
integer(0)
```

Finalizada la depuración de la base de datos de aplicantes PCT, se reduce el número de columnas del objeto que contiene la base de datos `DatosPCT` seleccionando tan solo las columnas 1, *pct_nbr*, y la columna 3, *appln_id*. Más adelante, se ilustra el por qué estas variables son necesarias para la unión de las bases de datos de códigos tecnológicos IPC y CPC.

```
DatosPCT<-DatosPCT[,c(1,3)]

> dim(DatosPCT)
[1] 4360392       2

       pct_nbr appln_id
1 WO1978000001 46895317
2 WO1978000003 15648845
3 WO1978000004 22653861
4 WO1978000005 43420652
5 WO1978000006  6508282
6 WO1978000007 46895325
```

Por último, las observaciones son agrupadas según el Número de publicación PCT (*pct_nbr*), y después, a cada *pct_nbr* le es asignado su *appln_id*. Este procesamiento implica que el número de observaciones en la base de datos `DatosPCT` es igual al número de aplicaciones de patente PCT, 3922999, para las cuales se tiene información completa en las variables *pct_nbr* y *appln_id*:

```
DatosPCT<-DatosPCT %>% group_by(pct_nbr)  %>%
                  summarise(appln_id=min(appln_id))

> head(DatosPCT)

  pct_nbr       appln_id
  <chr>            <int>
1 WO1978000001 46895317
2 WO1978000003 15648845
3 WO1978000004 22653861
4 WO1978000005 43420652
5 WO1978000006  6508282
6 WO1978000007 46895325

> dim(DatosPCT)
[1] 3922999       2
```

Intersección de las bases de datos

El proceso de manipulación y depuración de las bases de datos de REGPAT seleccionadas, nos condujo a 3 ficheros o bases de datos a partir de los cuales se soportó la etapa de combinación y unificación de variables:

Base de datos de códigos tecnológicos IPC (`DatosIPC`): esta base de datos consiste de 13623366 observaciones y 3 columnas que incluyen información del Número de publicación PCT (*pct_nbr*), el Año de prioridad (*prio_year*), y los Códigos tecnológicos IPC (*IPC*) asignados a las aplicaciones de patente PCT dentro del periodo en estudio.

Base de datos de códigos tecnológicos CPC (`DatosCPC`): comprende 51169333 observaciones y 2 columnas con información de Códigos tecnológicos CPC (*CPC_Class*). En esta base de datos, los códigos CPC son relacionados con la variable Número de identificación de la aplicación PCT (*appln_id*).

Base de datos de aplicantes PCT (`DatosPCT`): Esta base de datos incluye 3922999 observaciones que corresponden a igual número de aplicaciones PCT. Para estos registros, la base de datos brinda información de las variables Número de publicación PCT (*pct_nbr*) y Número de identificación de la aplicación PCT (*appln_id*).

Para la combinación o intersección de las bases de datos se utilizó la función `join` del paquete dplyr. Dentro de la familia de la función `join` se usó el tipo `inner_join` que permite desde un punto inicial, en donde se tienen dos bases de datos que comparten un vector o columna (con igual denominación/nombre), generar una nueva base de datos que mantiene los registros comunes.

Inicialmente, se realizó la intersección de las bases de datos `DatosIPC` y `DatosPCT`. Dicha intersección se llevó a cabo a partir del vector o columna de la variable Número de publicación PCT (*pct_nbr*), común para ambas bases de datos. Tras esta operación, se obtuvo la base de datos `Base1` en donde fue posible relacionar los códigos tecnológicos IPC con la variable *appln_id*.

Es importante precisar que la variable *appln_id* era necesaria para establecer una relación entre las bases de datos `DatosIPC` y `DatosCPC`. Además, esta intersección proporciona el Año de prioridad (*prio_year*), que proviene de la base de datos `DatosIPC`, correspondiente a cada Número de publicación PCT (*pct_nbr*), posibilitando el desarrollo de análisis temporales. La base de datos `DatosPCT` tiene información para 3922999 aplicaciones de patente PCT, mientras que `DatosIPC` para un total de 3789945 de aplicaciones PCT. La intersección de estas dos bases

de datos resultó en una base de datos con información para 3788949 aplicaciones de patentes PCT.

```
Base1<-inner_join(DatosIPC,DatosPCT,by=pct_nbr)
```

```
> head(Base1)
# A tibble: 6 x 3
  pct_nbr      prio_year appln_id
  <chr>            <int>    <int>
1 WO1978000001      1977 46895317
2 WO1978000003      1977 15648845
3 WO1978000004      1977 22653861
4 WO1978000005      1977 43420652
5 WO1978000006      1977  6508282
6 WO1978000007      1977 46895325
> dim(Base1)
[1] 3788949       3
```

Por otro lado, la base de datos `DatosCPC` tan sólo contenía 2 variables, los Códigos tecnológicos CPC (`CPC_Classes`) y el Número de identificación de la aplicación (*appln_id*), siendo necesario agregar a estas columnas las variables *pct_nbr* y *prio_year*. Por este motivo, la base de datos `DatosCPC` fue intersectada con `Base1` a partir de la columna en cómun *appln_id*, obteniendo así la base de datos `BaseLarga_CPC` con 3786795 observaciones o aplicaciones de Patente PCT.

```
BaseLarga_CPC<-inner_join(Base1,DatosCPC,by=appln_id)
> dim(unique(BaseLarga_CPC[,3]))
 [1] 3786795
```

Persiguiendo la finalidad de que las bases de datos de códigos tecnológicos sean comparables, se decidió que ambas contuviesen información del mismo conjunto de aplicaciones de Patente PCT (por su pct_nbr). De esta manera, se procedió a la creación de los objetos `PatenteIPC` y `PatenteCPC`. En estos objetos fueron almacenados los valores únicos de la variable Número de publicación PCT (*pct_nbr*).

Posteriormente, se aplicó un `inner_join` para intersectar los objetos `PatenteIPC` y `PatenteCPC`. Se obtuvo el objeto `Patentes` que consistia en el conjunto de elementos únicos de la variable Numero de publicación PCT (*pct_nbr*) en común para ambas bases de datos.

```
PatentesIPC<-data.frame(pct_nbr=unique(BaseLarga_IPC[,1]))
>  dim(PatentesIPC)[1]
[1] 3789945

PatentesCPC<-unique(BaseLarga_CPC[,1])
>  dim(PatentesCPC)[1]
[1] 3786795
```

```
Patentes<-inner_join(PatentesIPC,PatentesCPC,by=pct_nbr)
> dim(Patentes)
[1] 3786795
```

Se creó la `BaseLarga_IPC_Final`, con un inner_join, de forma tal que se mantuviesen aquellos elementos con pct_nbr dentro del objeto `Patentes`. La verificación constató que las bases de datos de formato largo para los códigos tecnológico IPC y CPC contenian igual número de niveles, 3786795 aplicaciones de patente PCT.

La dimensión de las base larga final de códigos tecnológicos IPC es de 13615554 observaciones y 3 columnas:

```
BaseLarga_IPC_Final<-inner_join(BaseLarga_IPC,Patentes,by=pct_nbr)
> dim(unique(BaseLarga_IPC_Final[,1]))
[1]   3786795

> dim(BaseLarga_IPC_Final)
[1] 13615554    3
```

La dimensión de la base larga final de códigos tecnológicos CPC es de 25227358 observaciones y 3 columnas:

```
BaseLarga_CPC_Final<-BaseLarga_CPC[,c(1,2,4)]
> dim(unique(BaseLarga_CPC_Final[,1]))
[1]   3786795

> dim(BaseLarga_CPC_Final)
[1]  25227358   3
```

En la ordenación de las observaciones de la bases de datos, se utilizó la función `order` aplicando dos criterios. En primer lugar, las observaciones de las bases de datos fueron ordenadas según la columna 2, Año de prioridad (*prio_year*), y posteriormente, según la columna 1, Número de publicación PCT *(pct_nbr)*.

```
BaseLarga_IPC_Final<-BaseLarga_IPC_Final[
order(BaseLarga_IPC_Final$prio_year,BaseLarga_IPC_Final$pct_nbr),]

BaseLarga_CPC_Final<-BaseLarga_CPC_Final[
order(BaseLarga_CPC_Final$prio_year,BaseLarga_CPC_Final$pct_nbr),]
```

Se guardan los ficheros de las bases de datos finales para su posterior lectura y procesamiento:

```
write.table(BaseLarga_IPC_Final,file=BaseLarga_IPC_Final.txt,
  row.names=FALSE, quote=FALSE)

write.table(BaseLarga_CPC_Final,file=BaseLarga_CPC_Final.txt,
  row.names=FALSE, quote=FALSE)
```

Conteo de códigos tecnológicos.

Para efectuar el conteo de códigos tecnológicos se diseñó una estrategia de programación que permitiese obtener resultados de una forma más clara y eficiente. En primer lugar, se procedió a la lectura de las bases de datos de códigos tecnológicos IPC y CPC:

```
# Lectura de la bases de datos de c\'odigos tecnol\'ogicos IPC
BaseLarga_IPC<-read.table(BaseLarga_IPC_Final.txt,
sep =,header=TRUE)

# Lectura de la bases de datos de c\'odigos tecnol\'ogicos CPC
BaseLarga_CPC<-read.table(BaseLarga_CPC_Final.txt,
sep =,header=TRUE)
```

Para facilitar el procesamiento de las bases de datos y simplificar el código de programación, se definió el uso de un selector que dependiendo el valor que le fuese asignado, estipulaba la lectura de una base de datos u otra. Si a este selector le era asignado un valor igual a 1, se efectuaba automáticamente la lectura de la base de datos de códigos IPC. En el caso en que se indicase que el selector tomaba un valor igual a 2, la lectura y procesamiento de datos correspondía a la base de datos de códigos tecnológicos CPC. Con la finalidad de unificar el código de programación, en ambas bases de datos fue modificado el nombre de la columna que contenía los códigos tecnológicos, bajo la nueva denominación `Codes`.

```
Selector<-1
if (Selector==1){
  data<-BaseLarga_IPC
} else if (Selector==2){
  data<-BaseLarga_CPC
}

colnames(data)[3]<-Codes
```

En un siguiente paso, se creo el objeto `registros` - tipo integer- que tomó un valor númerico entero igual al número de observaciones de la base de datos `data`. La dimensión de la base de datos `data` es igual a la dimensión de base de datos larga final de códigos tecnológicos IPC o CPC, según corresponda.

```
registros<-dim(data1)[1]

> class(registros)
[1] integer

> dim(data) # selector ==1
[1] 13615554        3
```

```
> dim(data) # selector ==2
[1] 25227358        3
```

Posteriormente, fueron creados cuatro vectores de ceros (0), de igual longitud que el objeto `registros`. Estos vectores de ceros son creados con la finalidad de almacenar los valores que toman las variables de Códigos Nuevos (*Codigos_Nuevos*), Códigos Reutilizados (*Codigos_Reutilizados*), Número de códigos reutilizados acumulados *(Acum_Codigos_Reutilizados)* y Número de códigos nuevos acumulados *(Acum_Codigos_Nuevo)*.

```
Codigos_Nuevos=rep(0,registros)
Codigos_Reutilizados=rep(0,registros)
Acum_Codigos_Nuevos=rep(0,registros)
Acum_Codigos_Reutilizados=rep(0,registros)
```

Un código tecnológico tan solo puede ser usado en una única ocasión para describir una misma aplicación de patente PCT. A partir de esto, se pudo inferir que el código tecnológico situado en la posición $i = 1$ de la base de datos, es un código IPC nuevo. Este código tecnológico es almacenado en un objeto denominado `CodigosDiferentes`.

En el caso de la base de datos de códigos tecnológicos IPC, la posición $i = 1$ correspondía al código tecnológico 'C03C013/04"de la aplicación de patente PCT con *pct_nbr* "WO197800001". Para la base de datos de códigos tecnológicos CPC, el código almacenado en el objeto `CodigosDiferentes` fue 'B01D17/0202", igualmente asociado a la aplicación de patente "WO197800001". Este objeto almacenará como su nombre lo indica, el conjunto de diferentes códigos tecnológicos en el registro de patentes PCT.

```
CodigosDiferentes<-data$Codes[1]
```

```
> CodigosDiferentes  # selector == 1   C\'odigos IPC
[1] C03C013/04

> CodigosDiferentes  # selector == 2   C\'odigos CPC
[1] B01D17/0202
```

Partiendo de que el código tecnológico en la posición $i = 1$ de la base de datos `data` es un codigo nuevo, fue posible inferir que el número de códigos nuevos acumulados *(Acum_Codigos_Nuevo)* en esta posición era igual a 1, y los códigos reutilizados acumulados *(Acum_Codigos_Reutilizados)*, igual 0. Se muestra a continuación como se fijaron los valores que toman las variables para la posición $i = 1$ de la base de datos `data`:

```
Codigos_Nuevos[1]<-1
Codigos_Reutilizados[1]<-0
Acum_Codigos_Nuevos[1]<-1
Acum_Codigos_Reutilizados[1]<-0
```

Una vez se definieron los valores que toman las variables en las columnas o variables para la primera posición de la base de datos, se procedió al cálculo del número de códigos tecnológicos para las posiciones que van, desde $i = 2$ hasta $i = registros$. Para esto, se utilizó un bucle con el comando `for`, con las siguientes instrucciones:

1. Si el código tecnológico en la posición `i` se encuentra en el vector de elementos del objeto `CódigosDiferentes`, se entiende que el código tecnológico no es nuevo, es un código reutilizado. Por lo tanto, tenemos que el número de códigos nuevos *Codigos_Nuevos* para la posición `i` es igual a cero y el número de códigos reutilizados *Codigos_Reutilizados* es igual a 1. Entre tanto, el número de códigos nuevos acumulados, *Acum_Codigos_Nuevos*, es igual al valor acumulado en la posición anterior `i-1`, y el número de códigos tecnológicos reutilizados acumulados, *Acum_Codigos_Reutilizados*, es igual al valor acumulado en la posición `i-1` más 1.

```
if (data$Codes[i] %in %CodigosDiferentes){
Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]+1
Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]
Codigos_Reutilizados[i]<-1
Codigos_Nuevos[i]<-0
```

2. Si el código tecnológico en la posición `i` no se encuentra en el conjunto de elementos del objeto `CódigosDiferentes`, el código tecnológico es almacenado en el objeto. El código tecnológico en la posición `i` es nuevo. Por ende, el número de códigos IPC nuevos acumulados, *Acum_Codigos_Nuevos*, en la posición `i` es igual al número de códigos nuevos acumulados en la posición `i-1` más 1. Por su parte, el número de códigos IPC reutilizados acumulados, *Acum_Codigos_Reutilizados*, en la posición `i` no varía, siendo igual al valor registrado en la posición `i-1`.

```
} else {
 CodigosDiferentes<-c(CodigosDiferentes,data$Codes[i])
 Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]
 Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]+1
 Codigos_Reutilizados[i]<-0
 Codigos_Nuevos[i]<-1
  }
}
```

Las instrucciones del bucle `for` programado se repitieron en sucesivas iteraciones hasta alcanzar el número total de observaciones de la base datos `data`. A continuación, se muestra el código de programación completo del bucle `for`:

```
for (i in 2:registros){
  if (data$Codes[i] %in %CodigosDiferentes){
  Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]+1
  Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]
  Codigos_Reutilizados[i]<-1
  Codigos_Nuevos[i]<-0
  } else {
  CodigosDiferentes<-c(CodigosDiferentes,data$Codes[i])
  Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]
  Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]+1
  Codigos_Reutilizados[i]<-0
  Codigos_Nuevos[i]<-1
  }
}
```

Una vez se calculó el número de códigos tecnológicos para cada posición `i` de la base de datos, los resultados incorporados en los vectores de ceros fueron añadidos con la función `cbind` como nuevas columnas en el objeto `data2`:

```
data2<-cbind(data,
             Acum_Codigos_Reutilizados,Acum_Codigos_Nuevos,
             Codigos_Reutilizados,Codigos_Nuevos)
```

Generación de combinaciones de códigos tecnológicos

Para obtener las combinaciones de códigos tecnológicos que describen a cada una de las aplicaciones de patente PCT, se realizó una transformación de la estructura de la base de datos `data2`. Desde un formato largo donde cada código tecnológico ocupaba una fila de la columna 3 de la base de datos, pasamos a una base de datos de formato ancho. La columna 3 de códigos tecnológicos es reemplazada por las combinaciones de códigos tecnológicos (agrupados) que describen cada una de las aplicaciónes de patente PCT. La longitud de la nueva base de datos es igual al número de niveles de la columna 1, *pct_nbr*, es decir, el número de aplicaciones de patente PCT.
En el código de programación utilizado para generar las combinaciones de códigos tecnológicos se concatenaron operaciones del paquete *dplyr* mediante el operador *pipe* ($\%>\%$). La primera operación consistió en agrupar el conjunto de filas según la columna 1 (*pct_nbr*) de la base de datos, usando la función *group_by*. Una vez los datos fueron agrupados se usó la función *summarise* para definir nuevas variables, y de este modo, generar una nueva base de datos.

La columna Combinaciones de códigos tecnológicos (*Combin*) se obtuvo juntando los elementos del vector de códigos tecnológicos `Codes` de la base de datos larga en una misma fila, mediante la función *paste*. En estas combinaciones, los códigos tecnológicos fueron ordenados alfabeticamente y separados por una coma (,), usando las funciones *sort* y *collapse*. El código de programación usado para obtener las combinaciones de códigos tecnológicos fue el siguiente:

```
CombinacionesFichero<-data2 %> %group_by(pct_nbr) %> %summarise(
       prio_year=min(prio_year),
       Combin=paste(sort(Codes),collapse=,),
       Acum_Codigos_Reutilizados=max(Acum_Codigos_Reutilizados),
       Acum_Codigos_Nuevos=max(Acum_Codigos_Nuevos),
       Codigos_Reutilizados=sum(Codigos_Reutilizados),
       Codigos_Nuevos=sum(Codigos_Nuevos) )
```

Las combinaciones de códigos tecnológicos de la base de datos `data3` fueron ordenadas según el año de prioridad *prio_year* y el número de publicación PCT *pct_nbr*:

```
data3<-CombinacionesFichero[order(CombinacionesFichero$prio_year,
               CombinacionesFichero$pct_nbr),]
```

Una vez ordenadas las combinaciones de códigos tecnológicos para cada aplicación de patente PCT, se generó el objeto `CódigosTotal` cuyos valores resultan de la suma de los códigos tecnológicos nuevos y reutilizados para cada aplicación de patente PCT. Este vector fue añadido a la base de datos con la función `cbind` en el objeto `data4`:

```
CodigosTotal<-data3$Codigos_Reutilizados+data3$Codigos_Nuevos

data4<-cbind(data3,Codigos_Total=CodigosTotal)
```

La base de datos `data4` tiene una dimensión de 3789945 observaciones (número de diferentes niveles de aplicaciones de patente PCT) y 8 columnas. Tal cómo se puede observar a continuación:

```
> dim(data4)
[1] 3786795       8

> colnames(Combi)
[1] pct_nbr    prio_year    Combin   Acum_Codigos_Reutilizados
[5] Acum_Codigos_Nuevos    Codigos_Reutilizados   Codigos_Nuevos
Codigos_Total
```

Conteo de combinaciones de códigos tecnológicos

Para el conteo de combinaciones de códigos tecnológicos IPC se creó un objeto tipo `integer` denominado `registros2` cuya longitud es igual al número de observaciones de la base de datos `data4`. Además, se crearon 2 vectores de ceros utilizando la función `rep`. Estos vectores tenían una longitud igual al valor numérico del objeto `registros2`, y fueron empleados para almacenar información acerca de las nuevas variables: Nueva combinación (*NuevaComb_Si_1*) y Combinaciones acumuladas (*Acum_Combinaciones*).

```
registros2<-dim(data4)[1]
registros2
NuevaComb_Si_1=rep(0,registros2)
Acum_Combinaciones=rep(0,registros2)
```

La variable Nueva combinación (*NuevaComb_Si_1*), es una variable dicotómica codificada con valores 0 y 1. Cuando la variable toma un valor igual a 1, indica que la combinación de códigos tecnológicos es nueva. Cuando esta variable toma un valor igual a 0 significa que la correspondiente combinación de códigos tecnolgícos fue utilizada previamente. La segunda variable, Combinaciones acumuladas (*Acum_Combinaciones*) muestra el número de combinaciones de códigos tecnológicos acumuladas para cada posición `i` en la base de datos.

Luego, en el objeto `Combinaciones` se guardaron los datos de la variable Combinaciones de códigos tecnológicos de la base de datos `data4`. Además, para almacenar las combinaciones de códigos tecnológicos diferentes en el registro de patentes, se creó un segundo objeto llamado `CombinacionesDiferentes`. El primer elemento del objeto `CombinacionesDiferentes` fue la combinación de códigos tecnológicos situada en la primera posición del objeto `Combinaciones`, correspondiente a su vez, a la primera aplicación de patente PCT en la base de datos `data4`.

```
Combinaciones<-data4$Combin
CombinacionesDiferentes<-Combinaciones[1]
NuevaComb_Si_1[1]<-1
Acum_Combinaciones[1]<-1
```

Se fijaron los valores que toman las variables Nueva combinación (*NuevaComb_Si_1*) y Combinaciones acumuladas (*Acum_Combinaciones*) para la primera posición del objeto `Combinaciones`. La lógica nos indica que la combinación en la posición `i=1` es nueva, puesto que su aparición supone la primera vez que es usada en la descripción de una aplicación de patente PCT. Se determinó que ambas variables tomaban un valor igual a 1.

Se utilizó un bucle con el comando `for` para determinar si las combinaciónes de códigos tecnológicos correspondientes a las aplicaciones de patente PCT con posición, desde $i = 2$ hasta $i = registros2$, de la base de datos `Combinaciones`, eran nuevas o no. En primer lugar, se verificaba si la combinación en la posición `i` estaba incluida en el objeto `CombinacionesDiferentes`.

De esta manera, si una combinación se encontraba en el objeto `Combinaciones Diferentes`, la variable Nueva combinación *NuevaComb_Si_1*, tenia un valor igual a cero, y el número de combinaciones acumuladas (*Acum_Combinaciones*) era igual al valor obtenido para la posición `i-1`. Si una combinación no estaba incluida en el objeto `Combinaciones Diferentes` era considerada nueva e incluida en el objeto. Para esta combinación, la variable Nueva combinación *NuevaComb_Si_1* tomaba un valor igual a 1, y veia incrementado el número de combinaciones acumuladas, siendo el valor de la variable Combinaciones acumuladas (*Acum_Combinaciones*) igual al valor de la posición `i-1` más 1.

```
for (i in 2:registros2){
    if (Combinaciones[i]%in%CombinacionesDiferentes){
    NuevaComb_Si_1[i]<-0
    Acum_Combinaciones[i]<-Acum_Combinaciones[i-1]
  } else {
    CombinacionesDiferentes<-c(CombinacionesDiferentes,
    Combinaciones[i])
    NuevaComb_Si_1[i]<-1
    Acum_Combinaciones[i]<-Acum_Combinaciones[i-1]+1
  }
}

data5<-cbind(data4,NuevaComb_Si_1,Acum_Combinaciones)
```

Tipo de combinaciones de códigos tecnológicos

Se establece una clasificación de las combinaciones de códigos tecnológicos que describen las aplicaciones de patentes PCT. Esta clasificación se elaboró siguiendo las definiciones propuestas por Strumsky y Lobo (2015) para la identificación de los recursos de novedad tecnológica y la categorización de la novedad en un patente de invención.

Inicialmente, se crea un objeto denominado `TipoCombinacion` que es un vector de ceros (0) de igual longitud al número de observaciones en la base de datos `data5`. Luego, para determinar el tipo de combinación de códigos tecnológicos para cada aplicación de patente PCT, se realizó un bucle con el comando `for`. En este bucle se implementó una estructura condicional múltiple `if - else if - else`, con

las instrucciones que se detallan a continuación:

```
TipoCombinacion=rep(0,registros2)
```

Condición 1: Si la columna 6 correspondiente a la variable códigos reutilizados en una aplicación de patente PCT de la base de datos tiene valor igual a cero, entonces, la totalidad de los códigos tecnológicos en la combinación son nuevos. La combinación es clasificada como origen.

```
for (i in 1:registros2){
  if (data5[i,6]==0){
    TipoCombinacion[i]<-1
```

Condición 2: Si la columna 6 de códigos reutilizados es mayor que cero (no se cumple la condición 1), y además, la columna 7 de códigos nuevos toma un valor mayor que cero, la combinación de códigos tecnológicos que describe la aplicación de patente PCT es una combinación novedosa. En este tipo de combinaciones se hace uso simultáneo de códigos nuevos y reutilizados.

```
 } else if (data5[i,7]>0){
   TipoCombinacion[i]<-2
```

Condición 3: En un aplicación de patente PCT, si el número de códigos reutilizados en columna 6 es mayor que cero, y el número de códigos nuevos en la columna 7 es igual a cero, no se cumplen las condiciones 1 y 2. Se entiende que, en este supuesto, la combinación que describe la aplicación de patente PCT contiene exclusivamente códigos tecnológicos previamente usados. Si la combinación de códigos tecnológicos utilizados es nueva, es decir, la columna 9 de la variable Nueva Combinación (*NuevaComb_Si_1*) tiene un valor igual a la unidad, la combinación es clasificada como combinación.

```
  } else if (data5[i,9]==1){
    TipoCombinacion[i]<-3
```

Condición 4: Si no se cumplen las condiciones previas, la combinación de códigos tecnológicos que describe la aplicación de patente PCT es clasificada como refinamiento. Este tipo de combinación contiene exclusivamente códigos tecnológicos reutilizados, y no resulta ser novedosa. Para esta aplicación de patente PCT, la variable Nueva Combinación (*NuevaComb_Si_1*) es igual a cero.

```
} else {
  TipoCombinacion[i]<-4
 }
```

Por último, la base de datos final denominada `FicheroFinal` fue construida añadiendo, con la función `cbind`, la columna `TipoCombinacion` a la base de datos `data5`:

```
FicheroFinal<-cbind(data5,TipoCombinacion)
```

Los ficheros de datos fueron generados para cada clasificación de códigos tecnológicos usando un selector. Cuando el selector tomase un valor igual a 1, se generaba la base de formato ancho para los códigos tecnológicos IPC, y cuando el valor asignado fuese 2, para los códigos tecnológicos CPC:

```
if (Selector==1){
  Filename<-BaseAnchaFinal_IPC.txt
} else if (Selector==2){
  Filename<-BaseAnchaFinal_CPC.txt
}
write.table(FicheroFinal,file=Filename,row.names=FALSE,
col.names=FALSE,quote=FALSE)
```

En el cuadro 2.8 se ofrece un esquema, a modo de resumen, que facilita la comprensión de la clasificación de las combinaciones de códigos tecnológicos. La lógica de clasificación propuesta esta basada en los valores que toman las variables de Códigos nuevos, Códigos reutilizados y Nueva combinación.

Cuadro 2.8: Esquema de clasificación de combinaciones de códigos tecnológicos.

Definición	**Códigos nuevos**	**Códigos reutilizados**	**Nueva combinación**
Origen	SI	NO	SI
Combinación novedosa	SI	SI	SI
Combinación	NO	SI	SI
Refinamiento	NO	SI	NO

En el apartado 2.2 Apéndice, a continuación, se presenta el código de programación utilizado para la lectura, depuración y procesamiento de los ficheros de datos de la base de datos REGPAT.

2.2. Apéndice

```
# Lectura del paquete dplyr
rm(list=ls())
library(dplyr)

# Definici\'on del periodo de an\'alisis
first.year<-1977
last.year<-2018
numyears<-last.year-first.year+1

# Lectura de las bases de datos:

DatosPCT<-read.table(202107_PCT_App_reg.txt,
sep =|,header = TRUE,dec =.)

DatosIPC<-read.table(202107_PCT_IPC.txt,
            sep =|,header = TRUE,dec =.)

DatosCPC<-read.table(202107_CPC_Classes.txt,
            sep =|,header = TRUE,dec =.)


## Depuracion base de datos de aplicantes:
DatosPCT<-DatosPCT[!is.na(DatosPCT[,1]),] # pct_nbr
DatosPCT<-DatosPCT[!is.na(DatosPCT[,3]),] # appln_id
DatosPCT<-DatosPCT[DatosPCT[,1]!=,]
DatosPCT<-DatosPCT[DatosPCT[,3]!=,]

Comprob<-DatosPCT %>% %group_by(pct_nbr) %>% %
summarise(comprobar=length(unique(appln_id)))

aux<-Comprob$pct_nbr[Comprob$comprobar>1]
aux
DatosPCT[DatosPCT$pct_nbr==aux[1],]
DatosPCT<-DatosPCT[which(!DatosPCT$pct_nbr %in %aux),]

DatosPCT<-DatosPCT[,c(1,3)]

DatosPCT<-DatosPCT %>% %group_by(pct_nbr) %>% %
summarise(appln_id=min(appln_id))


##Depuracion base de datos IPC:
DatosIPC<-DatosIPC[DatosIPC[,2]>=first.year,]
DatosIPC<-DatosIPC[DatosIPC[,2]<=last.year,]

DatosIPC<-DatosIPC[!is.na(DatosIPC[,1]),] # pct_nbr
DatosIPC<-DatosIPC[!is.na(DatosIPC[,2]),] # prio_year
DatosIPC<-DatosIPC[!is.na(DatosIPC[,4]),] # IPC
DatosIPC<-DatosIPC[DatosIPC[,1]!=,]
DatosIPC<-DatosIPC[DatosIPC[,2]!=,]
```

```
DatosIPC<-DatosIPC[DatosIPC[,4]!=,]

Comprob<-DatosIPC %>% %group_by(pct_nbr) %>% %
summarise(comprobar=length(unique(prio_year)))
head(Comprob)
aux<-Comprob$pct_nbr[Comprob$comprobar>1]
aux

BaseLarga_IPC<-DatosIPC[,c(1,2,4)] # Se crea la BaseLarga_IPC

DatosIPC[DatosIPC$pct_nbr==aux[1],]
DatosIPC<-DatosIPC[which(!DatosIPC$pct_nbr %in %aux),]

#Depuracion base de datos CPC:
DatosCPC<-DatosCPC[!is.na(DatosCPC[,1]),] # appln_id
DatosCPC<-DatosCPC[!is.na(DatosCPC[,2]),] # CPC_Class
DatosCPC<-DatosCPC[DatosCPC[,1]!=,]
DatosCPC<-DatosCPC[DatosCPC[,2]!=,]
library(stringr) #Se eliminan espacios en codigos CPC
x<-DatosCPC[,2]
DatosCPC[,2]<-str_replace_all(x, fixed( ), )


# Interseccion de las bases de datos:

Base1<-inner_join(DatosIPC,DatosPCT,by=pct_nbr)

BaseLarga_CPC<-inner_join(Base1,DatosCPC,by=appln_id)

PatentesIPC<-data.frame(pct_nbr=unique(BaseLarga_IPC[,1]))
head(PatentesIPC)
dim(PatentesIPC)[1]    # 3789945
PatentesCPC<-unique(BaseLarga_CPC[,1])
head(PatentesCPC)
dim(PatentesCPC)[1]    # 3786795
Patentes<-inner_join(PatentesIPC,PatentesCPC,by=pct_nbr)
dim(Patentes)          # 3786795

BaseLarga_IPC_Final<-inner_join(BaseLarga_IPC,Patentes,
                    by=pct_nbr)
BaseLarga_CPC_Final<-BaseLarga_CPC[,c(1,2,4)]
length(unique(BaseLarga_IPC_Final[,1])) # 3786795
dim(unique(BaseLarga_CPC_Final[,1]))    # 3786795

## Ordenaci\'on de las bases de datos:

BaseLarga_IPC_Final<-BaseLarga_IPC_Final[order
(BaseLarga_IPC_Final$prio_year,BaseLarga_IPC_Final$pct_nbr),]

BaseLarga_CPC_Final<-BaseLarga_CPC_Final[order
(BaseLarga_CPC_Final$prio_year,BaseLarga_CPC_Final$pct_nbr),]
#
```

```
#Creaci\'on de vectores para el conteo de c\'odigos tecnol\'ogicos:

Selector<-2   #   <<<<----------------
if (Selector==1){
  data<-BaseLarga_IPC
} else if (Selector==2){
  data<-BaseLarga_CPC
}
colnames(data)[3]<-Codes

registros<-dim(data)[1]
registros
Acum_Codigos_Reutilizados=rep(0,registros)
Acum_Codigos_Nuevos=rep(0,registros)
Codigos_Reutilizados=rep(0,registros)
Codigos_Nuevos=rep(0,registros)
#
CodigosDiferentes<-data$Codes[1]
View(CodigosDiferentes)

Acum_Codigos_Reutilizados[1]<-0
Acum_Codigos_Nuevos[1]<-1
Codigos_Reutilizados[1]<-0
Codigos_Nuevos[1]<-1
for (i in 2:registros){
  if (round((i/1000000))==i/1000000){message(i)}
  if (data$Codes[i] %in %CodigosDiferentes){
    Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]+1
    Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]
    Codigos_Reutilizados[i]<-1
    Codigos_Nuevos[i]<-0
  } else {
    CodigosDiferentes<-c(CodigosDiferentes,data$Codes[i])
    Acum_Codigos_Reutilizados[i]<-Acum_Codigos_Reutilizados[i-1]
    Acum_Codigos_Nuevos[i]<-Acum_Codigos_Nuevos[i-1]+1
    Codigos_Reutilizados[i]<-0
    Codigos_Nuevos[i]<-1
  }
}
data2<-cbind(data,
             Acum_Codigos_Reutilizados,Acum_Codigos_Nuevos,
             Codigos_Reutilizados,Codigos_Nuevos)

#Generaci\'on de combinaciones de c\'odigos y transformaci\'on
#de base de datos de formato largo a ancho:

CombinacionesFichero<-data2 %> %group_by(pct_nbr) %> %summarise(
          prio_year=min(prio_year),
          Combin=paste(sort(Codes),collapse=,),
          Acum_Codigos_Reutilizados=max(Acum_Codigos_Reutilizados),
          Acum_Codigos_Nuevos=max(Acum_Codigos_Nuevos),
          Codigos_Reutilizados=sum(Codigos_Reutilizados),
          Codigos_Nuevos=sum(Codigos_Nuevos) )
```

```
#se ordena por a\~no de prioridad y pct_nbr
data3<-CombinacionesFichero[order(CombinacionesFichero$prio_year,
           CombinacionesFichero$pct_nbr),]

dim(data3)
#
CodigosTotal<-data3$Codigos_Reutilizados+data3$Codigos_Nuevos
data4<-cbind(data3,Codigos_Total=CodigosTotal)
sum(data4$Codigos_Total)


# Conteo de combinaciones

registros2<-dim(data4)[1]
registros2
NuevaComb_Si_1=rep(0,registros2)
Acum_Combinaciones=rep(0,registros2)
#
Combinaciones<-data4$Combin
CombinacionesDiferentes<-Combinaciones[1]
NuevaComb_Si_1[1]<-1
Acum_Combinaciones[1]<-1
for (i in 2:registros2){
    if (Combinaciones[i]%in%CombinacionesDiferentes){
    NuevaComb_Si_1[i]<-0
    Acum_Combinaciones[i]<-Acum_Combinaciones[i-1]
  } else {
    CombinacionesDiferentes<-c(CombinacionesDiferentes,
    Combinaciones[i])
    NuevaComb_Si_1[i]<-1
    Acum_Combinaciones[i]<-Acum_Combinaciones[i-1]+1
  }
}
data5<-cbind(data4,NuevaComb_Si_1,Acum_Combinaciones)


# Creaci\on vector para incluir datos de tipo de combinaci\'on
# para cada aplicaci\'on de patente PCT:
 TipoCombinacion=rep(0,registros2)

# Bucle con comando for para determinar tipo de
#combinaci\'on:

# 1: Origen
# 2: Combinaci\'on novedosa
# 3: Combinaci\'on
# 4: Refinamiento
for (i in 1:registros2){
  if (data5[i,6]==0){
    TipoCombinacion[i]<-1
  } else if (data5[i,7]>0){
    TipoCombinacion[i]<-2
```

```
  } else if (data5[i,9]==1){
    TipoCombinacion[i]<-3
  } else {
    TipoCombinacion[i]<-4
  }
}


if (Selector==1){
  Filename<-BaseAnchaFinal_IPC.txt
} else if (Selector==2){
  Filename<-BaseAnchaFinal_CPC.txt
}

#Incorporaci\'on de la columna Tipo de combinaci\'on
a la base de datos.
FicheroFinal<-cbind(data5,TipoCombinacion)

# Conversi\'on de la base de datos en un archivo de texto:
write.table(FicheroFinal,file=Filename,row.names=FALSE,
col.names=FALSE,quote=FALSE)
```

Capítulo 3

Desigualdad en la frecuencia de uso de códigos tecnológicos.

3.1. Introducción

Los códigos tecnológicos son usados para clasificar una nueva patente dentro de un categoría técnica específica e identificar los componentes de novedad de una invención (Youn et al., 2015). Los códigos tecnológicos pueden además ser usados, por ejemplo, para analizar la productividad de la innovación (Strumsky, Lobo y Tainter, 2010), estudiar el cambio tecnológico (Strumsky, Lobo y Van der Leeuw, 2012), identificar los recursos de la novedad tecnológica (Strumsky y Lobo, 2015), determinar las tendencias de desarrollo tecnológico (Chang, 2017), extraer las tendencias de invenciones tecnológicas agrupadas por aplicantes (Chae y Gim, 2019) o hacer un seguimiento a la composición tecnológica de las industrias (Goldschlag, Lybbert y Zolas, 2020).

Sin importar donde se origina una invención, y donde su inventor decide patentarla, uno o más códigos tecnológicos son asignados a la patente, reflejando un proceso de innovación basado principalmente en el conocimiento de tecnologías disponibles, la combinación de estas tecnologías, y ocasionalmente la creación de nuevas tecnologías (Weitzman, 1998; Arthur, 2009). Sin embargo, no todos los códigos tecnológicos son usados con la misma frecuencia (Huang, 2016; Smojver, Storga y Potocki, 2016). Por un lado, podemos encontrar las tecnologías convencionales que siendo parte de muchas patentes, pueden ser asociadas con áreas tecnológicas con alto potencial de mercado, y por otro lado, hay tecnologías poco comunes siendo parte de combinaciones tecnológicas inusuales (Harrigan y Fang, 2019).

El objetivo de este capítulo es medir la desigualdad en la frecuencia de uso de códigos tecnológicos de patentes. Tomando como referencia todas las posibilidades de combinación, de nuevas y conocidas tecnologías disponibles en un momento específico, y a partir de la estimación de cómo la desigualdad en la frecuencia de uso de códigos tecnológicos cambia en el tiempo, se puede determinar: si la invención como un proceso combinatorio, esta basado (o no) en el uso generalizado de pocas tecnologías; si las disparidades en esa frecuencia de uso son más altas ahora que en el pasado; y ésto puede ayudarnos a entender mejor el proceso de innovación y cambio tecnológico (una amplia encuesta acerca de cómo las estadísticas de patentes pueden ser consideradas como indicadores económicos puede ser encontrada en (Griliches,1998)).

La desigualdad es un concepto clave en Economía (Sen, 1997), con una alto porcentaje de la literatura dedicado a la desigualdad de ingresos (Piketty y Saez 2003; Sarabia, Jordá,y Prieto, 2019), pero tambien enfocado en la desigualdad del bienestar (Cagetti y De Nardi, 2008; Zucman, 2019), desigualdad en la educación (Tilak, 1987; Jordá y Alonso, 2017; Coco, Lagravinese y Resce, 2020), o desigualdad en productividad (Ezcurra, Pascual y Rapún 2007), entre otros, y que puede ser extendida a otras variables. En particular, en la Economía de la Innovación, los esfuerzos han estado concentrados, entre otros, en el impacto de la innovación en el crecimiento económico (Adrián Risso y Sánchez Carrera, 2019), los retornos de la innovación (Marsili y Salter 2005), la especialización tecnológica en áreas metropolitanas (O Huallacháin y Lee 2011), o en países (Mancusi, 2003) o el rol de las instituciones (Biurrun, 2020). Sin embargo, desde lo que conocemos, no hay esfuerzos previos para medir la desigualdad que emerge del surgimiento de códigos tecnológicos entre diferentes patentes.

Para este análisis, se consideró la base de datos de patentes REGPAT de la OCDE, en la cual están disponibles los códigos tecnológicos de la Clasificación Internacional de Patentes (IPC) y la Clasificación Cooperativa de Patentes (CPC). Fue estimada la desigualdad total en la frecuencia de uso (anual) de estos códigos tecnológicos, durante el periodo 1977 - 2018. Para esto, se usaron dos medidas de desigualdad ampliamente conocidas: el Coeficiente de Gini y el Índice de Theil. Adicionalmente, se exploró la relación entre la evolución de esta desigualdad total y las diferentes fuentes de la novedad inventiva: origen, combinaciones novedosas, combinaciones y refinamientos (Strumsky y Lobo, 2015); y con la probabilidad de los códigos más usados de mantener este estatus de un año a otro.

Finalmente, basados en la descomposición del Índice de Theil, y usando la correspondencia entre lós códigos tecnológicos IPC y la clasificación europea de actividades económicas NACE, se estudió la descomposición de la desigualdad total por

actividades económicas productivas, y se calculó la contribución de los componentes intra-grupo e inter-grupo en la desigualdad total previamente estimada.

El capítulo esta organizado de la siguiente manera: en el apartado 3.2 se describe la base de datos y la metodología utilizada. Los resultados de las estimaciones y análisis de la desigualdad en la frecuencia de uso de códigos tecnológicos son presentados en el apartado 3.3; y por último, en el apartado 3.4, se presentan las conclusiones.

3.2. Datos y metodología

3.2.1. Datos

En este capítulo se usó la información de patentes obtenida de la base de datos de patentes REGPAT (Maraut, S et al.,2008), versión de julio de 2021, publicada por la Organización para el Desarrollo y la Cooperación Económica OCDE, y de libre acceso disponible bajo el diligenciamiento de un formulario en línea. Fueron consideradas las aplicaciones de patentes solicitadas bajo el Tratado de Cooperación de Patentes (PCT, por sus sigles en inglés) en fase internacional, con la Oficina Europea de Patentes (EPO) como oficina designada, y con año de prioridad comprendido entre 1977 y 2018. La fecha de prioridad se utilizó como año de referencia para cada aplicación de patente PCT, siguiendo la recomendación de la OCDE (ver OCDE, 2009, p.62) como la fecha más cercana al momento en que la invención fue concebida. Teniendo en cuenta que una aplicación de patente puede ser publicada hasta 18 meses después, contados a partir de la fecha de solicitud, se seleccionó el periodo de 42 años, desde 1977 a 2018.

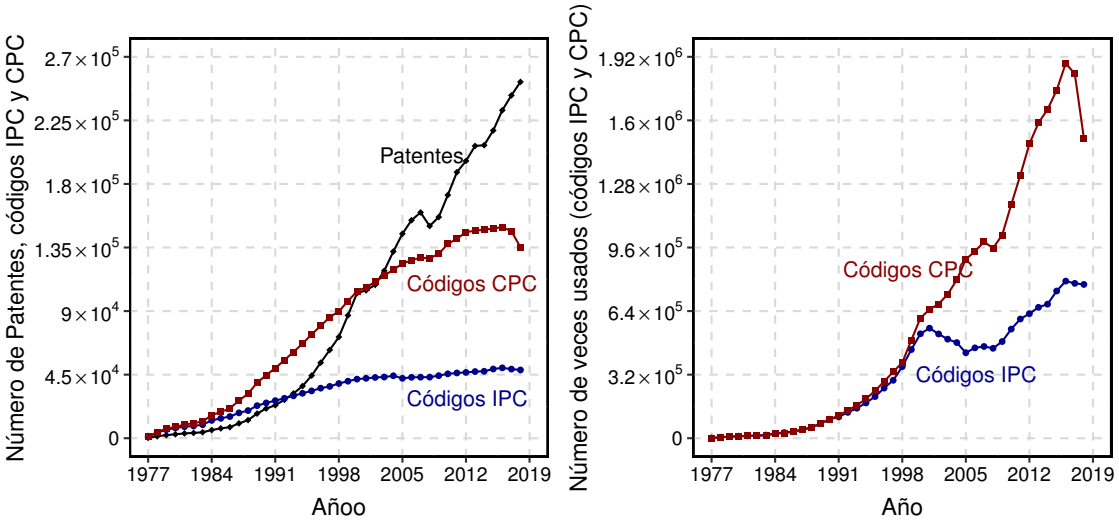

Figura 3.1: Número de patentes PCT(EPO), y el número de diferentes códigos IPC y CPC por año (izquierda). Número de veces en un año que los códigos IPC y CPC son usados (derecha), para el periodo 1977-2018.

La frecuencia de uso de códigos tecnológicos en patentes, se analizó anualmente, para los códigos IPC y CPC (correspondientes a los sistemas de Clasificación Internacional de Patentes y la Clasificación Cooperativa de Patentes, respectivamente). Para ello, se obtuvo la información acerca de los códigos tecnológicos IPC del fichero

"202107_PCT_IPC.txt" y la información de los códigos tecnológicos CPC usando los ficheros "202107_PCT_App_reg.txt" y "202107_CPC_Classes.txt". El procesamiento y manipulación de los ficheros de datos se detalló en el apartado 2.1.4 del capítulo de datos.

En resumen, el conjunto de datos utilizado estaba compuesto por 3786795 aplicaciones de patentes PCT, con 72607 códigos tecnológicos IPC que fueron usados 13615554 veces en el periodo 1977- 2018. Además, estas aplicaciones de patente PCT incluían 234782 códigos tecnológicos CPC diferentes que fueron usados 25227358 durante el periodo de estudio. La figura 3.1 muestra, en la izquierda, el número de aplicaciones de patente PCT y el número de códigos tecnológicos IPC y CPC diferentes por año; y en la derecha, el número de veces que los códigos tecnológicos IPC y CPC fueron usados.

3.2.2. Frecuencia de uso de códigos IPC y CPC

La variable aleatoria de interés es la frecuencia de uso anual de los códigos tecnológicos IPC (X_1) y códigos tecnológicos CPC (X_2). La variable aleatoria es discreta y tan solo puede tomar valores númericos enteros positivos. Esta característica de la variable aleatoria obedece a que el número de códigos tecnológicos que pueden ser asignados a una patente o invención es un número entero, que en todo caso, es mayor o igual que 1. Un código tecnológico solamente puede ser usado en una única ocasión al momento de describir una invención. De ahí que el número de veces que un código tecnológico es utilizado en el sistema de patentes es igual al número de patentes a las cuales dicho código tecnológico le es asignado.

Considerando que un análisis completo del periodo comprendido entre 1977 y 2018 incluiría códigos tecnológicos con diferente antigüedad, se determinó conducir un análisis anual, agrupando aquellos códigos usados en un mismo año. Como referencia, el cuadro 3.1 muestra para los códigos IPC y CPC, las tablas de frecuencia para el conjunto de datos correspondientes al año 1978. En este año, 3260 códigos IPC y 3791 códigos CPC fueron usados en una ocasión (en una sola patente). Entre tanto, los valores máximos en la tabla de frecuencia nos indican que 1 código IPC fue usado 18 veces, mientras que 1 código CPC fue usado hasta en 15 patentes diferentes.

Cuadro 3.1: Tabla de frecuencia de uso de códigos tecnológicos IPC y CPC correspondientes al año 1978.

Códigos IPC. Año: 1978		Códigos CPC. Año: 1978	
x_{1i} (Frecuencia de uso)	n_{1i} (Número de códigos)	x_{2i} (Frecuencia de uso)	n_{2i} (Número de códigos)
1	3260	1	3791
2	521	2	338
3	116	3	55
4	48	4	54
5	19	5	14
6	4	6	6
7	2	7	4
8	3	8	2
9	3	9	2
12	2	11	1
18	1	13	1
		15	1

3.2.3. Estadística descriptiva

En los cuadros 3.2 y 3.3 se presentan las principales características empíricas para la variable aleatoria, frecuencia de uso de códigos tecnológicos IPC y CPC, respectivamente. Fueron calculadas para cada año en el periodo de estudio: la media, el valor mínimo, el valor máximo, la varianza, el coeficiente de variación, dispersión, la simetría y la kurtosis.

Además, en la figura 3.2 se representa graficamente la evolución de algunas de las carácteristicas empíricas de las variables aleatorias de interés: frecuencia de uso de códigos tecnológicos IPC (en color azul) y códigos CPC (en color rojo). Se puede destacar que los códigos IPC fueron usados en promedio más que los códigos CPC para todos los años considerados. La media de la frecuencia de uso de los códigos IPC y códigos CPC, fue de aproximadamente 16 y 11 veces, respectivamente. Ambas variables exhiben una simetría positiva y una alta kurotsis. Entre tanto, los valores del coeficiente de variación (dispersión relativa de los datos) de la frecuencia de uso de códigos IPC y CPC son similares.

Cuadro 3.2: Resumen estadística descriptiva de datos. Frecuencia de uso de códigos tecnológicos IPC. 1977 - 2018.

Año	Cantidad	Media	Min	Max	Varianza	CV	Disp	Simetría	kurtosis
1977	1019	1.08	1	3	0.09	0.28	0.08	4.03	16.95
1978	3979	1.27	1	18	0.62	0.62	0.49	6.66	84.25
1979	6011	1.38	1	25	1.02	0.73	0.74	7.15	102.38
1980	7273	1.42	1	22	1.02	0.71	0.72	5.59	61.32
1981	8173	1.50	1	36	1.63	0.85	1.09	8.33	147.60
1982	8684	1.53	1	37	1.79	0.87	1.17	7.86	122.47
1983	9572	1.63	1	43	2.74	1.01	1.68	8.30	122.86
1984	12640	1.94	1	117	10.37	1.66	5.35	16.54	414.16
1985	14063	2.02	1	108	9.16	1.50	4.54	13.93	319.78
1986	15338	2.21	1	150	12.13	1.58	5.50	14.56	393.95
1987	17974	2.47	1	257	22.73	1.93	9.19	19.29	698.21
1988	19577	2.76	1	327	34.38	2.13	12.47	20.54	767.50
1989	23051	3.24	1	549	69.94	2.58	21.61	26.25	1193.67
1990	25006	3.83	1	758	116.76	2.82	30.51	28.26	1430.53
1991	26609	4.04	1	811	126.23	2.78	31.23	26.22	1329.40
1992	28285	4.60	1	1017	189.61	2.99	41.20	28.52	1514.55
1993	30002	5.04	1	1187	258.07	3.19	51.21	28.85	1527.22
1994	31805	5.59	1	1419	349.58	3.35	62.58	31.62	1766.30
1995	33499	6.26	1	1648	474.35	3.48	75.82	32.08	1761.45
1996	35430	7.12	1	1965	704.14	3.73	98.88	31.29	1658.41
1997	36672	7.95	1	2221	954.24	3.88	120.01	32.03	1686.48
1998	38747	9.30	1	2516	1527.97	4.20	164.33	29.45	1373.14
1999	40366	11.06	1	2919	2433.87	4.46	220.11	26.12	1033.94
2000	41844	12.56	1	3611	3604.25	4.78	286.87	26.97	1061.96
2001	42417	13.06	1	3155	3259.15	4.37	249.52	24.38	909.92
2002	43034	12.20	1	2325	2150.36	3.80	176.31	21.23	728.50
2003	43294	11.52	1	2139	1695.12	3.57	147.18	22.00	804.43
2004	44222	10.89	1	2008	1266.96	3.27	116.39	19.22	667.42
2005	42395	10.15	1	1767	1150.69	3.34	113.35	20.04	720.75
2006	43117	10.55	1	1905	1271.90	3.38	120.55	20.85	763.59
2007	43201	10.69	1	2043	1296.25	3.37	121.25	21.56	849.15
2008	43236	10.47	1	1863	1139.82	3.22	108.85	19.97	758.70
2009	44279	10.99	1	1821	1215.69	3.17	110.59	17.90	605.66
2010	45593	12.05	1	1936	1549.57	3.27	128.60	17.58	563.99
2011	46206	12.99	1	2227	1827.15	3.29	140.68	17.67	579.35
2012	46586	13.46	1	2595	1998.00	3.32	148.45	18.73	665.47
2013	47194	13.96	1	2872	2159.06	3.33	154.67	20.33	797.69
2014	47377	14.25	1	2845	2337.96	3.39	164.10	20.74	802.53
2015	48988	15.13	1	3121	2833.90	3.52	187.34	21.66	849.03
2016	49819	15.88	1	2874	3075.53	3.49	193.71	20.32	722.74
2017	48930	15.93	1	2899	3175.75	3.54	199.33	20.35	734.72
2018	48327	16.02	1	3456	3250.96	3.56	202.92	20.49	766.91

Cuadro 3.3: Resumen estadística descriptiva de datos. Frecuencia de uso de códigos tecnológicos CPC. 1977 - 2018.

Año	Cantidad	Media	Min	Max	Varianza	CV	Disp.	Simetría	Kurtosis
1977	1043	1.05	1	3	0.06	0.23	0.06	5.40	31.30
1978	4269	1.18	1	15	0.49	0.59	0.41	7.39	87.40
1979	6780	1.18	1	18	0.41	0.54	0.35	8.97	156.44
1980	8011	1.19	1	19	0.45	0.56	0.38	8.45	136.18
1981	9636	1.22	1	23	0.54	0.60	0.44	9.48	175.35
1982	10699	1.23	1	28	0.65	0.65	0.53	11.55	264.26
1983	11845	1.26	1	34	0.81	0.71	0.64	13.16	336.28
1984	16332	1.35	1	79	1.40	0.88	1.04	23.04	1214.57
1985	18578	1.40	1	75	1.59	0.90	1.14	16.56	694.68
1986	21177	1.49	1	101	2.69	1.10	1.80	18.12	772.56
1987	26616	1.58	1	184	4.64	1.37	2.94	32.27	2158.65
1988	31483	1.72	1	264	7.56	1.60	4.40	38.72	2986.83
1989	39338	1.89	1	411	12.90	1.90	6.84	50.07	4715.44
1990	44682	2.12	1	620	25.09	2.36	11.81	58.48	5979.76
1991	49572	2.30	1	649	28.74	2.33	12.51	50.21	4879.81
1992	55291	2.55	1	814	44.12	2.61	17.33	52.70	5055.93
1993	61024	2.72	1	918	57.47	2.79	21.16	53.07	4999.23
1994	66925	2.96	1	1110	80.72	3.04	27.28	58.07	5636.30
1995	73531	3.25	1	1306	101.91	3.11	31.40	58.55	5858.51
1996	79495	3.59	1	1531	143.03	3.33	39.86	56.62	5537.49
1997	85805	3.89	1	1726	189.34	3.53	48.63	62.47	6527.87
1998	89760	4.24	1	2146	291.70	4.02	68.74	59.00	5605.23
1999	97119	5.09	1	2731	553.00	4.62	108.67	54.94	4708.52
2000	103945	5.82	1	3292	821.84	4.93	141.22	55.24	4820.03
2001	106820	6.07	1	3606	792.08	4.64	130.49	54.83	5002.64
2002	110688	6.10	1	3933	851.42	4.78	139.53	57.64	5470.10
2003	115078	6.30	1	4350	948.72	4.89	150.66	60.93	6124.12
2004	119338	6.68	1	4694	1014.10	4.77	151.91	62.52	6538.94
2005	123497	7.29	1	4759	1129.27	4.61	154.87	59.08	5950.82
2006	126037	7.47	1	4501	1111.83	4.47	148.92	56.97	5502.29
2007	127776	7.74	1	4226	1207.01	4.49	156.01	50.64	4290.96
2008	127426	7.51	1	3958	957.84	4.12	127.53	50.09	4425.91
2009	130478	7.84	1	3684	931.65	3.89	118.86	43.84	3495.14
2010	137569	8.55	1	3284	905.22	3.52	105.87	37.52	2704.60
2011	141182	9.38	1	3593	1023.54	3.41	109.14	36.92	2764.71
2012	145924	10.15	1	3581	1086.76	3.25	107.02	31.77	2199.31
2013	146838	10.81	1	3474	1208.27	3.22	111.78	29.76	1963.55
2014	147671	11.21	1	3300	1202.47	3.09	107.27	26.19	1620.00
2015	148415	11.80	1	3734	1320.32	3.08	111.86	24.99	1601.89
2016	149521	12.62	1	4121	1585.82	3.16	125.64	26.35	1813.70
2017	146524	12.52	1	4763	1689.98	3.28	135.00	28.36	2074.60
2018	135246	11.16	1	4821	1571.74	3.55	140.79	33.94	2817.67

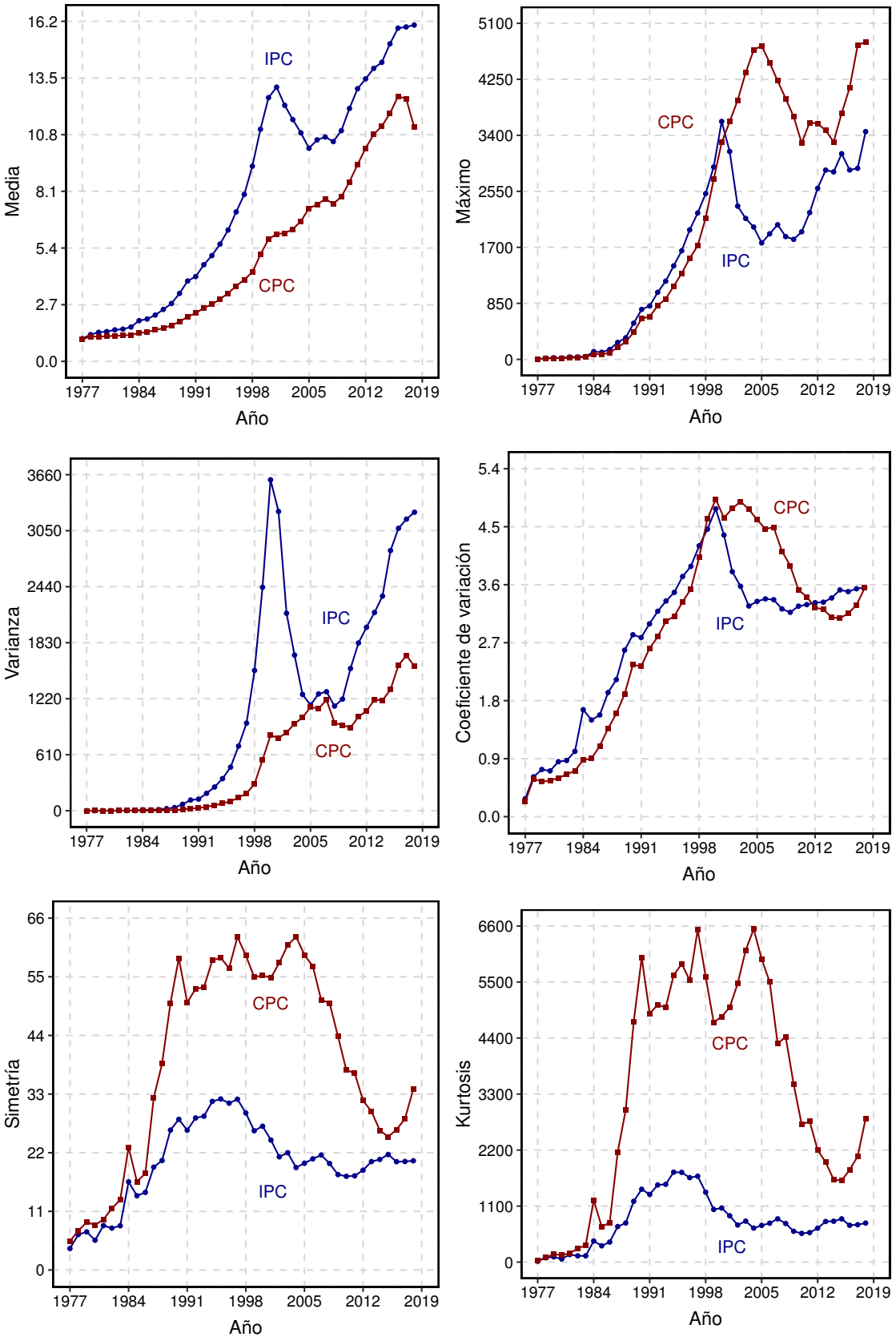

Figura 3.2: Estadística descriptiva. Frecuencia de uso de códigos Tecnológicos IPC y CPC. 1977 - 2018

3.2.4. Medidas de desigualdad

Para el análisis, fueron usados los índices de Gini y Theil (Arnold y Sarabia, 2018), dos de las medidas de desigualdad más conocidas.

Coeficiente de Gini

El coeficiente de Gini es uno de los indicadores de desigualdad más utilizados para medir la concentración de ingresos en una población. Este indicador toma valores comprendidos entre 0 y 1, donde 0 corresponde a un situación de perfecta equidad en la distribución de ingresos. Cuando este indicador toma un valor igual a 1, existe el máximo nivel de desigualdad posible (Sitthiyot y Holasut, 2021), donde una persona o grupo concentra la totalidad del ingreso.

El primer paso para calcular el Coeficiente de Gini es construir la curva de Lorenz. Esta curva, de fácil construcción e interpretación intuitiva, supone una aproximación alternativa a la representación de la dispersión de una distribución de renta o ingresos (Gisbert y Notario, 2009). En la figura 3.3, se ofrece una representación gráfica de la curva de lorenz (en color azul), tomando como ejemplo la distribución de ingresos. En esta figura se encuentran en el eje de las abscisas (horizontal), ordenados de menor a mayor, los porcentajes acumulados de población, y en el eje de las ordenadas (vertical), los porcentajes acumulados de ingresos. Cada punto a lo largo de la curva de Lorenz representa el porcentaje acumulado de ingresos correspondiente a cada porcentaje acumulado de la población.

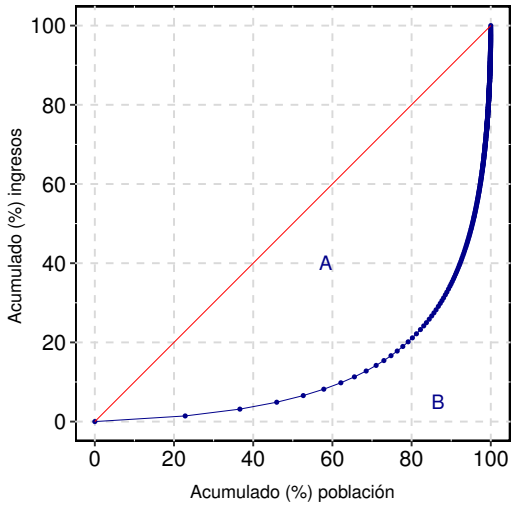

Figura 3.3: Curva de Lorenz. Distribución de ingresos en una población.

La estimación e interpretación geométrica del coeficiente de Gini a partir de la cons-

trucción de la curva de Lorenz es sencilla. El coeficiente de Gini es normalmente calculado como un ratio de las áreas de la curva de Lorenz. En la figura 3.3 se observan dos áreas claramente definidas: el área B por debajo de la curva de Lorenz, y el área A entre la curva de Lorenz y la línea diagonal de 45° en color rojo, y trazada desde el origen que representa la distribución uniforme del ingreso. En su representación geométrica, el coeficiente de Gini es igual al cociente entre el área A y el área de forma triangular conformada por la suma de las áreas A y B:

$$G = \frac{A}{A + B} \tag{3.1}$$

Ahora tenemos que, si $A + B = 1/2$, entonces:

$$G = 2A = 1 - 2B \tag{3.2}$$

La lógica de la interpretación gráfica de la curva de Lorenz y su relación con el coeficiente de Gini sugiere que, entre más cerca se encuentren los valores estimados representados en la curva de Lorenz a la línea de igualdad, menor será el valor del coeficiente de Gini. Por otro lado, si la curvatura es pronunciada, evidenciando una mayor convexidad, existe un alto grado de desigualdad en la distribución, y por ende, el valor del coeficiente de Gini es mayor.

En este trabajo, el coeficiente de Gini fue utilizado para medir la concentración de la frecuencia de uso de códigos tecnológicos IPC y CPC, y fue calculado usando la conocida formula de Brown (1994) en su notación original, que viene dada por la siguiente expresión:

$$G = 1 - \sum_{i=1}^{k} (X_i - X_{i-1})(Y_i + Y_{i-1}) \tag{3.3}$$

En la expresión 3.3, Y es el porcentaje acumulado de la frecuencia de uso de códigos tecnológicos en un año, y X es el porcentaje acumulado de códigos tecnológicos, sobre k valores de la frecuencia de uso ordenados desde el código tecnológico usado una sola vez (usado sólo en 1 aplicación de patente) al código más usado en el año considerado.

Para el cálculo del coeficiente de Gini, en el software R, se utilizó el siguiente código de programación:

```
DifX<-X[2:(N+1)]-X[1:N]
SumY<-Y[2:(N+1)]+Y[1:N]
GiniIndex_a[i]<-abs(1-sum(DifX*SumY))
```

Adicionalmente, a modo de comprobación, tambien fueron usadas las librerias `ineq` y `reldist` que permiten realizar la estimación directa del coeficiente de Gini:

```
library(reldist)
GiniIndex_b[i]<-gini(data$xi,data$Oi)

library(ineq)
GiniIndex_c[i]<-ineq(databis,type=Gini)
```

Índice de Theil.

El Índice de Theil (Allison, 1978; Shorrocks, 1980) fue calculado según la expresión a continuación:

$$T(\boldsymbol{y}, n) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i}{\mu} \right) \log \left(\frac{y_i}{\mu} \right) \tag{3.4}$$

Donde $\boldsymbol{y} = (y_1, \ldots, y_i, \ldots, y_n)$ es la frecuencia de uso de un vector anual para una población de n códigos tecnológicos (ver figura 3.1, en la izquierda), μ es el valor promedio o media de la frecuencia de uso por año (ver figura 3.2), y en donde logaritmos naturales son usados.

Adicionalmente, se obtuvo la serie histórica del porcentaje de la frecuencia de uso de códigos tecnológicos anual, correspondiente al primer decil (Top 10 %) desde 1977 hasta 2018. En particular, dentro del primer decil (P90-100) también fue analizado el porcentaje del total de la frecuencia de uso para los fractiles $P99-P100$ (0,99 − 1,00 fractiles), $P95 - P99$ y $P90 - P95$.

3.2.5. Fuentes de la novedad tecnológica

En este capítulo, se analizó la relación entre la evolución anual de la desigualdad en la frecuencia de uso de códigos tecnológicos y las fuentes de la novedad tecnológica. Por un lado, siguiendo la clasificación de las diferentes fuentes de la novedad tecnológica en invenciones, definidas por Strumsky y Lobo (2015) en base anualizada (tomando una ventana de ańalisis de 1 año), se clasificaron las patentes de la base de datos en 4 categorías que se corresponden con el esquema de clasificación de las combinaciones de códigos tecnológicos que se detalla en el apartado 2.1.4 (ver cuadro 2.8), y que se describe a continuación:

- *Origen*: todos los códigos de la patente son nuevos (no usado en años previos);
- *Combinación novedosa*: la patente tiene al menos un nuevo código tecnológicos, en combinación con otros códigos usados en años previos;
- *Combinación*: los códigos de la patente no son nuevos, sin embargo, la combinación de estos códigos es nueva.
- *Refinamiento*: los códigos de la patente y la combinación de ellos no son nuevas.

Las patentes clasificadas como origen y combinaciones novedosas proporcionan nuevos códigos con un alto potencial de uso, que pueden contribuir al total de la desigualdad, reduciéndola (en la medida en que algunos de estos códigos pasan a ser ampliamente usados, incrementando la proporción de uso asociada a códigos nuevos), o incrementándola (cuando estos nuevos códigos concentran la frecuencia de uso, convirtiéndose uno de estos códigos tecnológicos en uno de los más usados, en detrimento de otros códigos).

Así mismo, se evaluó la proporción de los códigos usados con mayor frecuencia que mantienen su estatus de un año a otro. Un patrón de incremento en esta proporción, en un contexto de incremento en el número de patentes y en la frecuencia total de uso, puede ser interpretado en términos de un mecanismo de "el rico se hace más rico"("rich-get-richer"). Haciendo la analogía de rico con aquellos códigos tecnológicos más usados, un escenario en donde la probabilidad de que los códigos ricos se mantengan ricos crezca de un año a otro, estará directamente relacionada con el incremento de la desigualdad total en la frecuencia de uso de códigos tecnológicos. Para este análisis, se consideraron los códigos tecnológicos con frecuencia de uso en los fractiles $P95 - 100$ y $P90 - 100$.

3.2.6. Descomposición de la desigualdad por actividades económicas productivas

NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) es el esquema de clasificación europeo de actividades económicas productivas (Eurostat, 2008). Esta clasificación tiene una estructura jerárquica con 4 niveles: secciones, divisiones, grupos y clases, identificadas por un código alfabético o un código numérico de dos, tres o cuatro dígitos, respectivamente.

En esta investigación, se consideró la concordancia entre las divisiones NACE (nivel - 2 dígitos) y los códigos IPC (ver Van Looy, Vereyen y Schmoch 2015, pp. 8-11, table 2) para analizar la descomposición de la desigualdad en la frecuencia de uso de

códigos tecnológicos en patentes por actividades económicas productivas. Algunos aspectos claves que deben ser abordados:

1. Los códigos IPC con símbolo de subclase tecnológica C07B, C07C, C07F, C07G, C12S or C40B, corresponden a la división NACE 20. Los códigos IPC con símbolo de subclase tecnológica C12M corresponden a la división NACE 32, cuando son usados en una aplicación de patente junto con un código tecnológico IPC del grupo A61K008/. En otros casos, todos estos códigos corresponden a la división NACE 21. Estos códigos IPC pueden corresponder a dos diferentes divisiones NACE, lo que significa que pueden ser considerados como códigos diferentes en términos de la correspondencia NACE.

2. Existe un grupo Co-IPC en la lista de las divisiones NACE para todos aquellos códigos IPC que no han sido clasificados.

3. Nuestro conjunto de datos de códigos tecnológicos IPC, basado en la correspondencia con las divisiones NACE, puede ser agrupado en 27 subconjuntos que unidos son iguales al conjunto de datos. Entonces, estos subconjuntos proporcionan una partición del conjunto de datos.

El cuadro 3.4 muestra el total de la frecuencia de uso de códigos tecnológicos IPC para cada división NACE en el periodo 1977-2018, expresada en el número de veces y porcentaje (como se indicó en el apartado 3.2.1, en nuestro conjunto de datos tenemos 72607 diferentes códigos IPC que fueron usados 13615554 veces en total, en 3786795 aplicaciones de patentes entre 1977 y 2018). Se puede resaltar que la división NACE más popular (considerando todo el periodo 1977-2018) fue la división NACE 26 (Manufacture of Computer, Electronic and Optical Products), con un 26.28 por ciento de toda la actividad total, y la menos popular la división NACE 16 (Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials), con un 0.03 por ciento.

Cuadro 3.4: Frecuencia de uso de códigos tecnológicos IPC en aplicaciones de patentes PCT. 1977 - 2018. División NACE (Nivel - 2 dígitos.)

NACE	Descripción	Códigos (veces usados)	Porcentaje (%)
10	Manufacture of Food Products	111635	0.82
11	Manufacture of Beverages	14525	0.11
12	Manufacture of Tobacco Products	13184	0.10
13	Manufacture of Textiles	55587	0.41
14	Manufacture of Wearing Appareal	18107	0.13
15	Manufacture of Leather and Related Products	20072	0.15
16	Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials	4619	0.03
17	Manufacture of Paper and Paper Products	33574	0.25
18	Printing and Reproduction of Recorded Media	31817	0.23
19	Manufacture of Coke and Refined Petroleum Products	46484	0.34
20	Manufacture of Chemicals and Chemical Products	1648487	12.11
21	Manufacture of Basic Pharmaceutical Products and Pharmaceutical preparations	2383402	17.50
22	Manufacture of Rubber and Plastic Products	192728	1.42
23	Manufacture of Other Non-Metallic Mineral Products	234959	1.73
24	Manufacture of Basic Metals	143390	1.05
25	Manufacture of Fabricated Metal Products, except Machinery and Equipment	176859	1.30
26	Manufacture of Computer, Electronics and Optical products	3578270	26.28
27	Manufacture of Electrical Equipment	880316	6.47
28	Manufacture of Machinery and Equipment N.E.C	1947629	14.30
29	Manufacture of Motor Vehicles, Trailers and Semi-Trailers	460651	3.38
30	Manufacture of Other Transport Equipment	102954	0.76
31	Manufacture of Furniture	40781	0.30
32	Other Manufacturing	1138554	9.36
42	Civil Engineering	12501	0.09
43	Specialised Construction Activities	79148	0.58
62	Computer Programming, Consultancy and Related Activities	139336	1.02
Co_IPC	IPC codes not elsewhere classified	105985	0.78
	Total	13615554	100.00

Para el análisis de la descomposición de la desigualdad total en la frecuencia de uso de códigos tecnológicos por actividades económicas productivas, y basados en la partición del conjunto de datos descrito previamente, se consideró la descomposición del índice de Theil en dos términos (Shorrocks, 1980; Shorrocks y Wan, 2005), como la suma del término de desigualdad intra-grupo, within-group $T(\boldsymbol{y}, n)_W$ y el término de desigualdad inter-grupo, between-group $T(\boldsymbol{y}, n)_B$, tal como se presenta a continuación:

$$T(\boldsymbol{y}, n) = T(\boldsymbol{y}, n)_W + T(\boldsymbol{y}, n)_B = \sum_{g=1}^{G} \frac{n_g \mu_g}{n\mu} T(\boldsymbol{y}^g, n_g) + \sum_{g=1}^{G} \frac{n_g \mu_g}{n\mu} \log\left(\frac{\mu_g}{\mu}\right) \quad (3.5)$$

En la expresión 3.5, G en el número de subconjuntos de datos (en nuestro caso, $G = 27$ divisiones NACE), y para cada subconjunto g, tenemos que: $\boldsymbol{y}^g = (y_1, \ldots, y_{n_g})$ es la frecuencia de uso de un vector anual para una población de n_g códigos, con un valor de media, μ_g; $\frac{n_g \mu_g}{n\mu}$ es la proporción de la frecuencia de uso anual; y $T(\boldsymbol{y}^g, n_g) = \frac{1}{n_g} \sum_{i=1}^{n_g} \left(\frac{y_i}{\mu_g}\right) \log\left(\frac{y_i}{\mu_g}\right)$ es el índice de Theil para ese subconjunto.

Por un lado, el componente de la desigualdad intra-grupo, within-group inequality, $T(\boldsymbol{y}, n)_W$, nos da una suma ponderada de la desigualdad en la frecuencia de uso anual de códigos tecnológicos IPC para cada subconjunto NACE (una suma ponderada del índice de Theil $T(\boldsymbol{y}^g, n_g)$ para cada división NACE), y por otro lado, el componente de desigualdad inter-grupo, the between-group inequality $T(\boldsymbol{y}, n)_B$, nos proporciona la desigualdad en la frecuencia de uso de códigos tecnológicos en un año entre los diferentes subconjuntos NACE reemplazando la frecuencia de uso de cada código IPC por la media de la frecuencia de uso μ_g de su respectiva división NACE en ese año.

3.3. Resultados

Desigualdad total en la frecuencia de uso de códigos tecnológicos

La figura 3.4 muestra la desigualdad total en la frecuencia de uso anual de códigos tecnológicos IPC y CPC estimada a partir del coeficiente de Gini (ver expresión 3.3) y el índice de Theil (ver expresión 3.4) desde 1977 hasta 2018. Se puede observar que mayores valores son alcanzados para los códigos IPC en ambas medidas en todos los años considerados, y que estas diferencias parecen estar disminuyendo con los años.

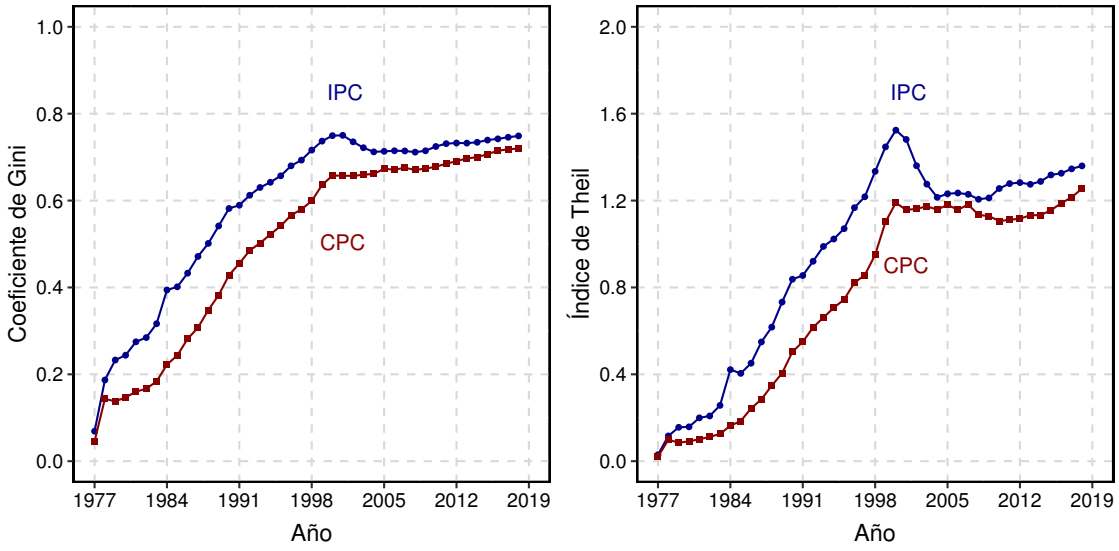

Figura 3.4: Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada con el Coeficiente de Gini (izquierda) y el índice de Theil (derecha). Periodo: 1977 - 2018.

La evidencia empírica sugiere que la evolución de la desigualdad en la frecuencia de uso de códigos tecnológicos tiene dos fases claramente diferenciadas, sin importar si son considerados los códigos IPC o CPC, el coeficiente de Gini o el índice de Theil:

- En una primera fase, desde 1977 hasta el año 2000, la desigualdad total muestra un patrón de crecimiento. El coeficiente de Gini crece de forma continua hasta alcanzar valores de 0.75 (códigos IPC) y 0.66 (códigos CPC) en el año 2000. El índice de Theil tambien crece a valores de 1.52 (códigos IPC) y 1.19 (códigos CPC) en ese mismo año.

- En una segunda fase, desde 2001 hasta 2018, la desigualdad total muestra una relativa estabilización en comparación con la primera fase. La desigualdad exhibe un patrón en forma de U en el caso de los códigos IPC (tanto en el coeficiente de Gini como en el índice de Theil), un suave incremento en el caso del coeficiente de Gini en los códigos CPC, y un patron en forma de U (despues de algunos años) en el caso del índice de Theil para los códigos CPC.

La figura 3.5 muestra una imagen ampliada de la segunda fase descrita en donde puede observarse con mayor claridad los patrones en forma de U de la desigualdad total en la frecuencia de uso de códigos estimada con el coeficiente de Gini (izquierda) y el índice de Theil (derecha). Además, la evidencia sugiere que la desigualdad total en la frecuencia de uso de códigos tecnológicos despues de estas dos fases, tiende a ser persistentemente alta.

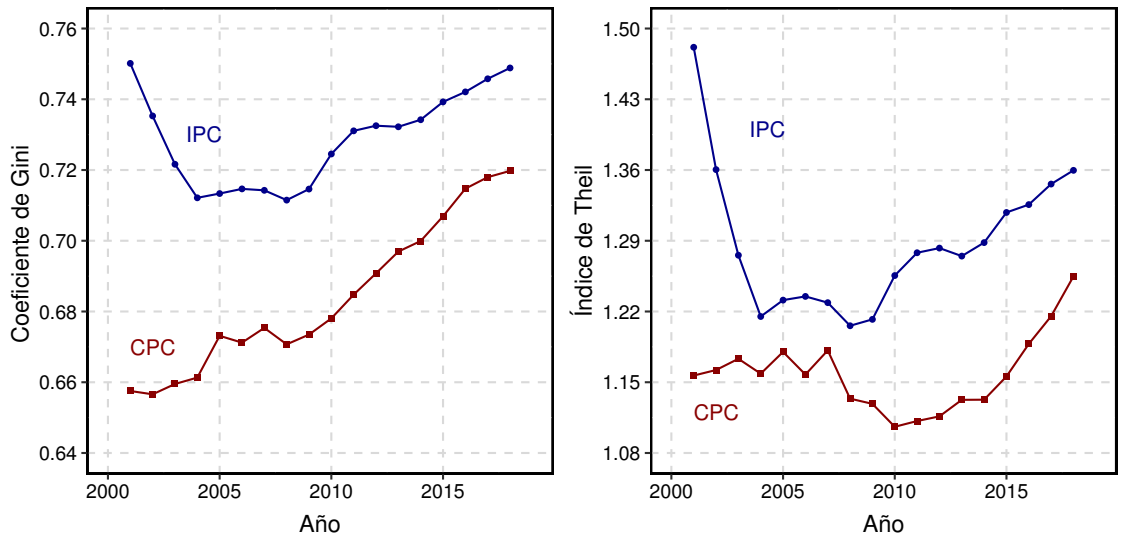

Figura 3.5: Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada con el Coeficiente de Gini (izquierda) y el índice de Theil (derecha) entre los años 2001 y 2018.

En la figura 3.6 se presenta el crecimiento de la desigualdad en la frecuencia de uso, expresada en términos del porcentaje de variación en el total de la desigualdad de un año a otro para los códigos tecnológicos IPC y CPC, basados en los índice de desigualdad considerados, para el periodo comprendido entre 1979 y 2018 (el año 1978 no fue incluido en la gráfica para que esta fuese más clara). En la figura 3.6 puede constatarse con facilidad la existencia de dos fases en la evolución de la desigualdad en la frecuencia de uso de los códigos: una primera fase caracterizada por valores de tasas de crecimiento positivos en la mayoria de los años, y una

segunda fase de relativa estabilización con valores de la tasa de crecimiento de la desigualdad cercanos a cero.

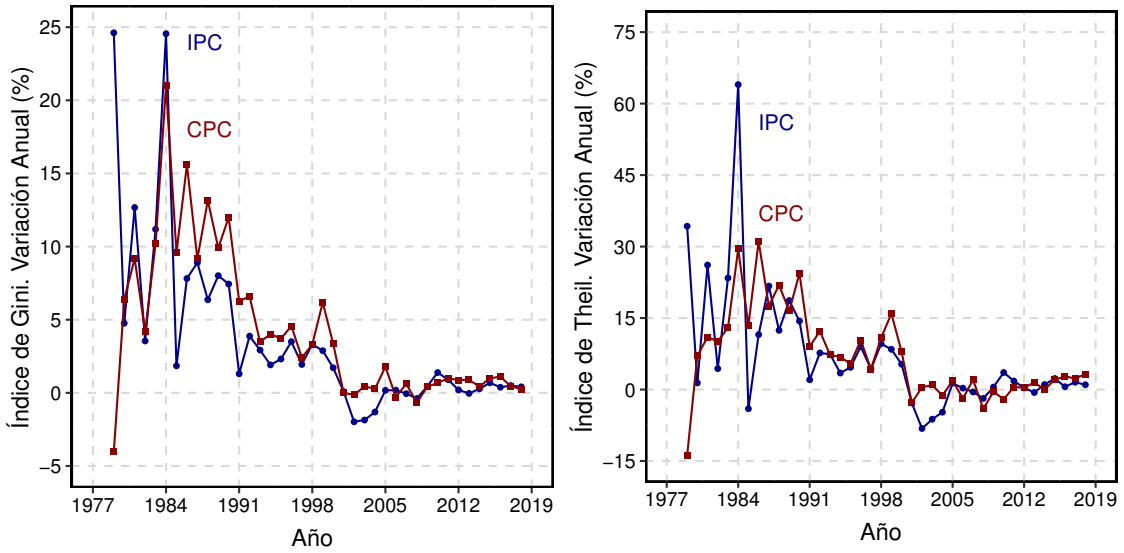

Figura 3.6: Crecimiento (%) de la desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, estimada a partir del coeficiente de Gini (izquierda y el índice de Theil (derecha). Periodo:1979-2018.

En las figuras 3.7 y 3.8 se muestra la evolución de la curva de Lorenz para la desigualdad total en la frecuencia de uso de códigos tecnológicos IPC y CPC, respectivamente. En estas figuras, se grafica la curva de Lorenz para 6 años seleccionados: 1977, 1980, 1990, 2000, 2010 y 2018. La visualización de los cambios (aumento) en el área contenida entre la línea de perfecta igualdad (en color rojo) y la curva de Lorenz (en color azul) nos permite inferir un considerable incremento de la desigualdad en la frecuencia de uso de códigos tecnologicos desde 1977 hasta el año 2000 (correspondiente a la primera fase descrita previamente), y el ligero incremento de la desigualdad desde el año 2000 hasta el año 2018 (en concordancia con la segunda fase de relativa estabilización).

La figura 3.9 muestra la proporción de la frecuencia de uso de códigos tecnológicos correspondiente al primer decil $P90-100$ para cada año entre 1977 y 2018. Dentro del primer decil, además se presenta el porcentaje del total de la frecuencia de uso para los fractiles $P99-100$ $P95-99$ y $P90-95$. Se puede observar que los patrones encontrados en la evolución del porcentaje de frecuencia de uso para el primer decil, Top 10 %, refleja la evolución de la desigualdad total en la frecuencia de uso (ver figura 3.4).

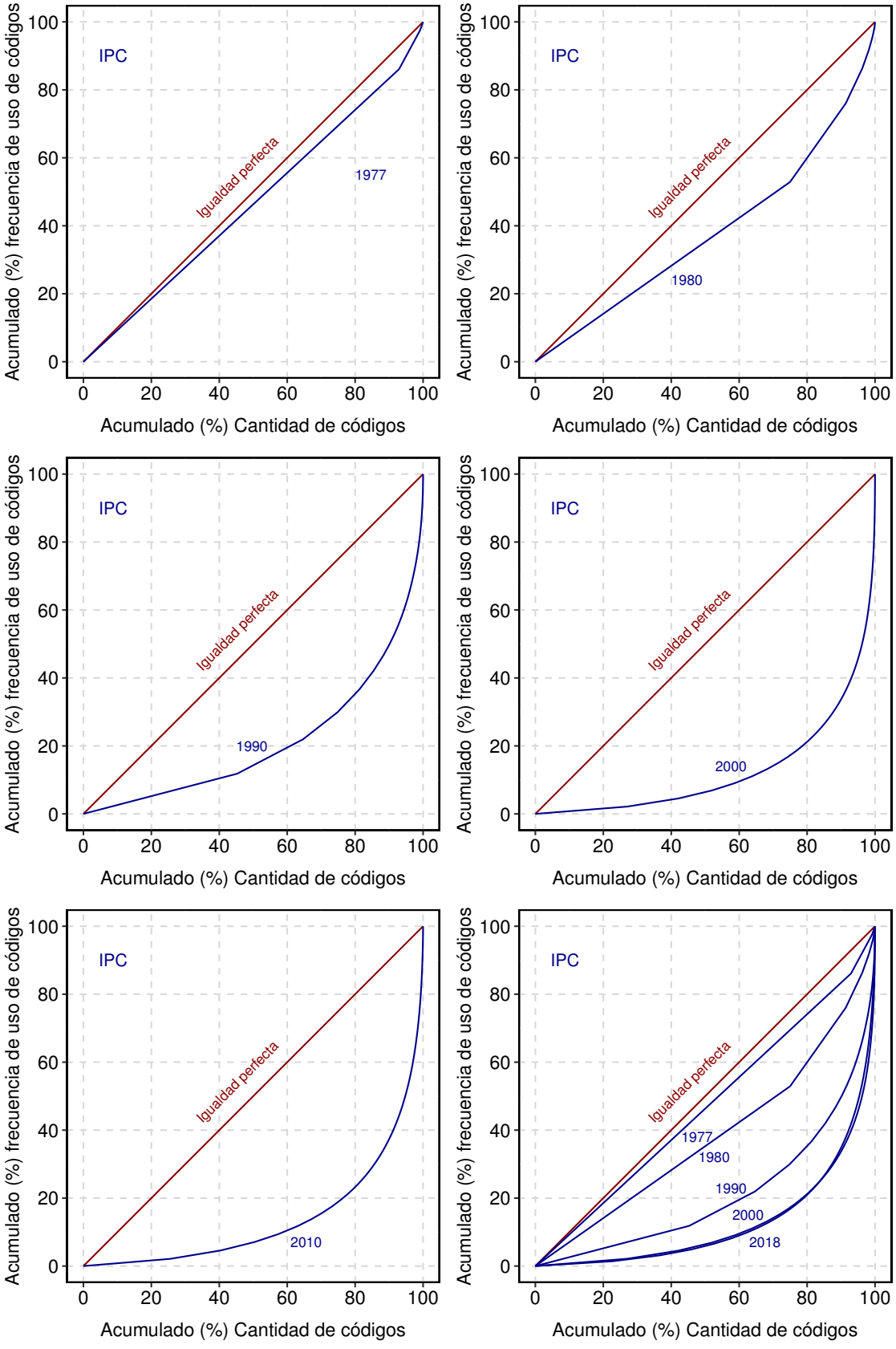

Figura 3.7: Curva de Lorenz, correspondiente a los códigos tecnológicos IPC para los años seleccionados: 1977, 1980, 1990, 2000, 2010 y 2018.

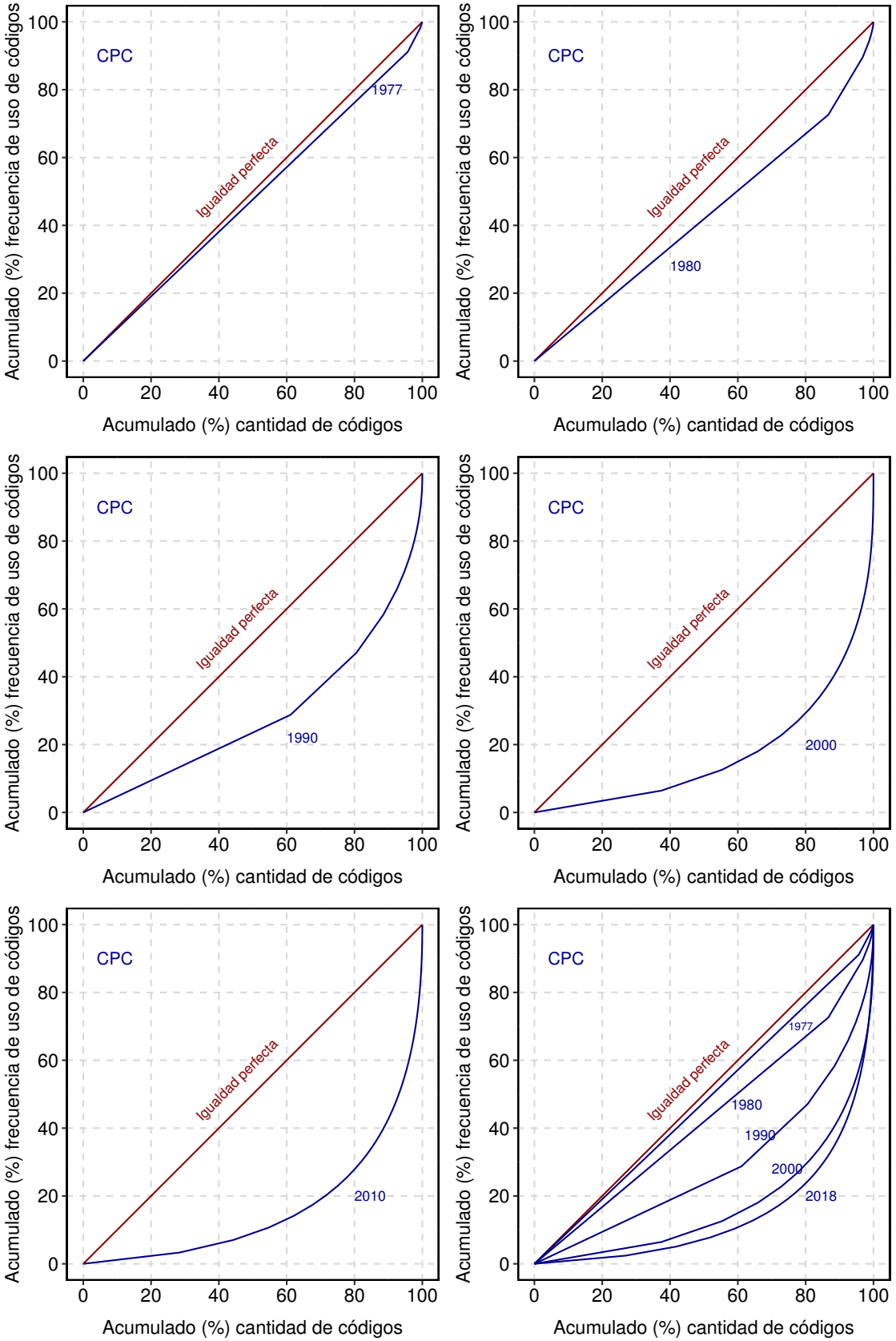

Figura 3.8: Curva de Lorenz, correspondiente a los códigos tecnológicos CPC para los años seleccionados: 1977, 1980, 1990, 2000, 2010 y 2018.

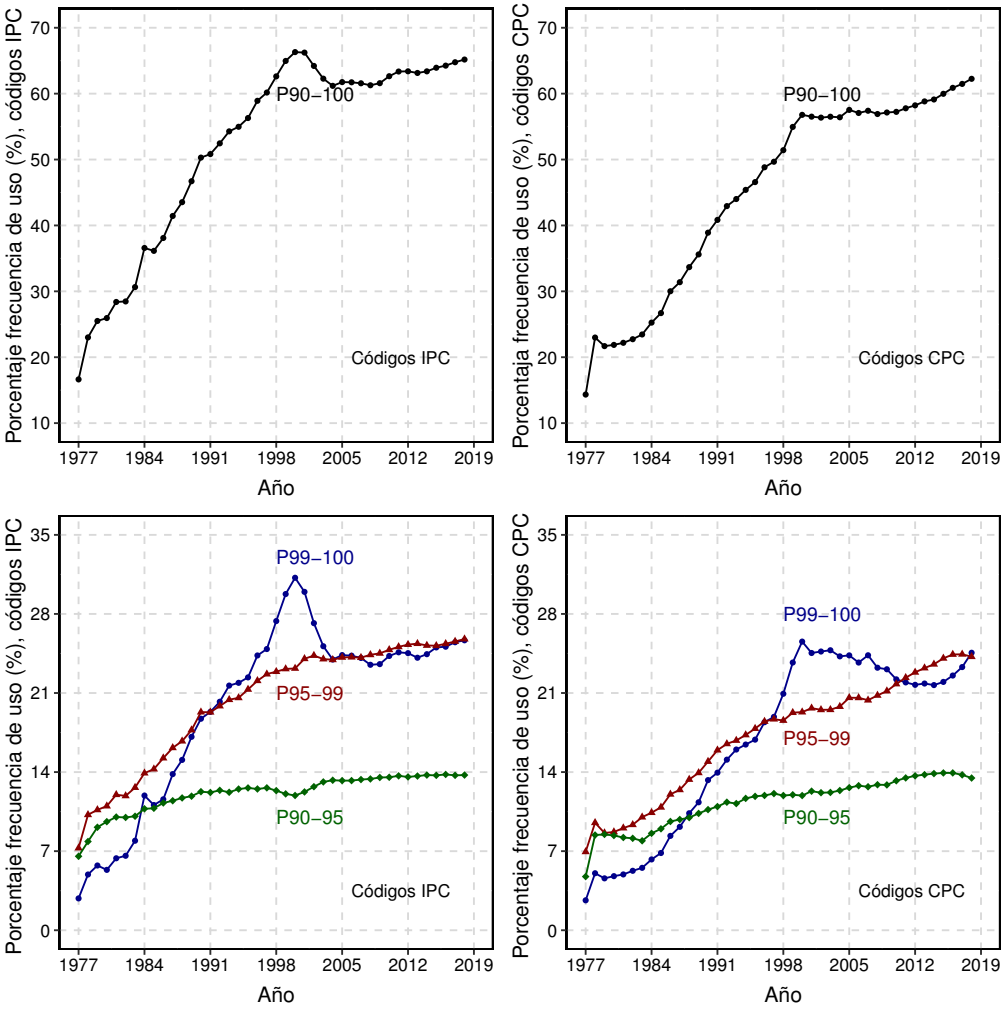

Figura 3.9: Porcentaje (%) de frecuencia de uso para el primer decil, P90-100 (arriba) y los fractiles P99 - 100, P95 - 99 y P90 - 95 (abajo); para códigos IPC (izquierda) y CPC (derecha). Periodo: 1977 - 2018.

El análisis del primer decil $90-100$ arroja resultados que concuerdan con la existencia de las dos fases descritas en la evolución de la desigualdad en la frecuencia de uso, y un patrón de crecimiento en los últimos años. Sin embargo, el porcentaje de la frecuencia de uso para los fractiles $P99-100$, $P95-99$ y $P90-95$ sigue diferentes caminos a lo largo de los años:

1. El grupo $P99-100$ tiene unos valores bajos durante los primeros años, tanto para el conjunto de datos de códigos IPC como CPC, luego estos valores crecen hasta un 31,2 % (IPC) y 25,5 % (CPC) en el año 2000, y finalmente, se mantienen relativamente estables desde el año 2000 (en el caso de los códigos

IPC despues de 4 años de declive) entre 23,4 % y 25,7 %, y con valores que oscilan entre el 21,7 % y 25,6 % para códigos CPC;

2. El grupo $P95-99$ siguió un patrón de crecimiento durante todos los años considerados, desde un 7,3 % a un 25,8 % para los códigos IPC, y desde 6,9 % hasta un 24,2 % para los códigos CPC, y alcanzando valores similares al grupo $P99-100$ en los últimos años del periodo considerado;

3. El grupo $P90-95$ muestra un suave incremento en comparación con los otros dos grupos, pasando de un 6,5 % a un 13,7 % en códigos IPC, y de un 4,7 % a un 13,5 % en el porcentaje de la frecuencia de uso para los códigos CPC.

Desigualdad en la frecuencia de uso de códigos tecnológicos y las fuentes de la novedad tecnológica.

La figura 3.10 muestra, de acuerdo a la clasificación de las diferentes fuentes de novedad tecnológica (ver apartado 3.2.5) y en base anualizada (periodo de ventana de análisis de 1 año), el porcentaje de patentes para cada una de las cuatros categorías: origen, combinación novedosa, combinación y refinamiento, para códigos IPC en la izquierda y para códigos CPC a la derecha. La figura 3.10 además expande la evidencia empírica acerca de un proceso de invención basado (en mayor medida y despues de algunos años iniciales) en el proceso combinatorio y acumulativo de conocimiento tecnológico previo (Lobo y Strumsky, 2019). Del comportamiento anual en términos porcentuales de las 4 categorías o fuentes de novedad tecnológica, durante el periodo de estudio (1977 -2018), podemos destacar que:

Las patentes clasificadas como "Origen" fueron predominantes entre 1977 y 1978 para los códigos IPC, y entre 1977 y 1980 para los códigos CPC. Esta fuente de novedad presenta un comportamiento decreciente en términos relativos a lo largo del periodo de estudio. Por su parte, las "Combinaciones Novedosas", muestran tanto para los códigos IPC como CPC, un patrón de U invertida: una primera fase de crecimiento sostenido, alcanzado para los códigos IPC su mayor peso relativo anual en el año 1980, y para los códigos CPC en 1984.

Las combinaciones novedosas fueron la fuente de novedad tecnológica más importante entre 1979 y 1981 (para códigos IPC), y entre 1981 y 1987 (para códigos CPC). Luego, tras alcanzar el punto máximo o mayor peso relativo en el año 1979 (IPC) y 1984 (CPC), la proporción anual de patentes consideradas combinaciones novedosas se caracteriza por un decrecimiento sostenido en el tiempo hasta llegar a valores anuales cercanos a cero, y muy similares a la categoría origen.

Las patentes clasificadas como "Combinaciones" se convierten en la fuente de novedad más importante a partir del año 1982 para los códigos IPC, y en 1988 para los códigos CPC. Esta fuente de novedad se mantiene como predominante para los restantes años del periodo de estudio (hasta 2018) para los códigos CPC. En el caso de los códigos IPC, las combinaciones son sobrepasadas (en porcentaje) en el año 2008 por las patentes clasificadas como "Refinamientos", tras una convergencia en el año 2007. En el año 2018 estas fuentes convergen nuevamente, llegando a representar cada una de ellas, casi la mitad de las aplicaciones de patente.

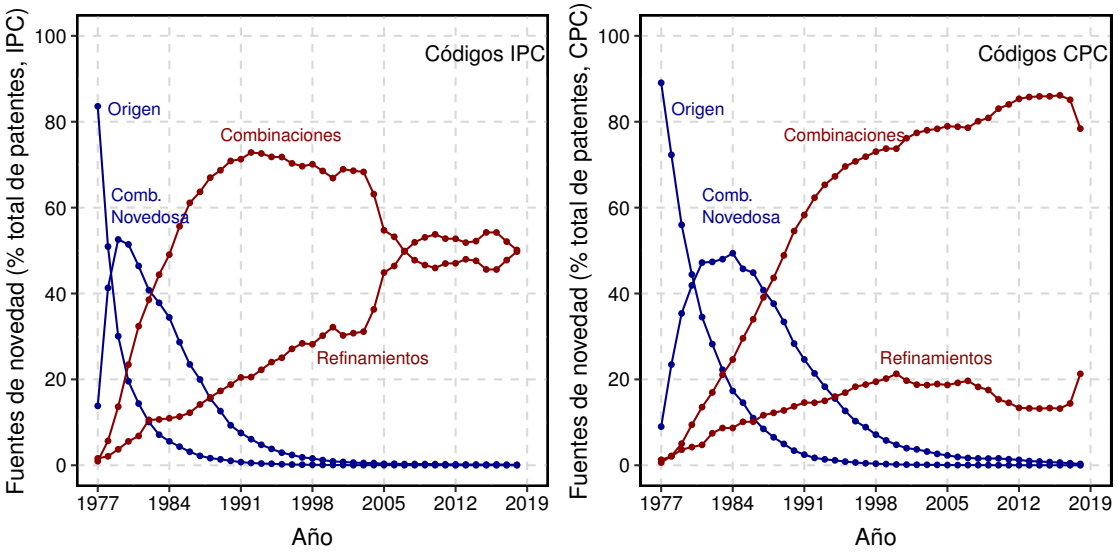

Figura 3.10: Porcentaje de aplicaciones de patentes por año (en base anual), clasificadas por las fuentes de novedad tecnológica (origen, combinación novedosa, combinación, y refinamiento), para códigos IPC (izquierda.) y códigos CPC (derecha.). Periodo: 1977-2018.

La figura 3.11 muestra el porcentaje de patentes con códigos tecnológicos nuevos (origen y combinaciones novedosas) y el porcentaje de patentes sin códigos nuevos (combinaciones y refinamientos). Se puede ver que la primera fase en la evolución anual de la desigualdad en la frecuencia de uso de códigos, en la cual esta desigualdad seguia un patrón de crecimiento (ver figura 3.4), parece coincidir con el periodo de cambio en las fuentes de la novedad inventiva, en donde el porcentaje de patentes con nuevos códigos decrece y las patentes conformadas por tecnologías conocidas incrementa su prominencia. Además, se nota que la segunda fase en la evolución anual de la desigualdad total en la frecuencia de uso, de relativa estabilización, parece estar en línea con la prevalencia de las combinaciones y refinamientos como principales fuentes de la novedad inventiva.

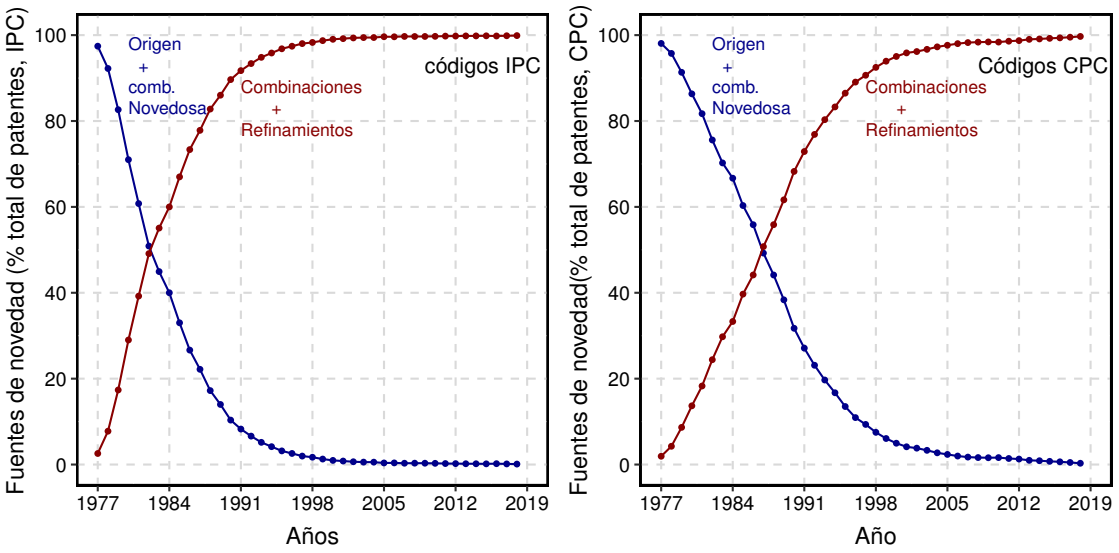

Figura 3.11: Porcentaje de aplicaciones de patentes por año (en base anual), agrupadas por fuentes de novedad tecnológica: (1) origen + combinaciones novedosas; (2) combinaciones + refinamientos;, para códigos IPC (izquierda.) y códigos CPC (derecha.). Periodo: 1977-2018.

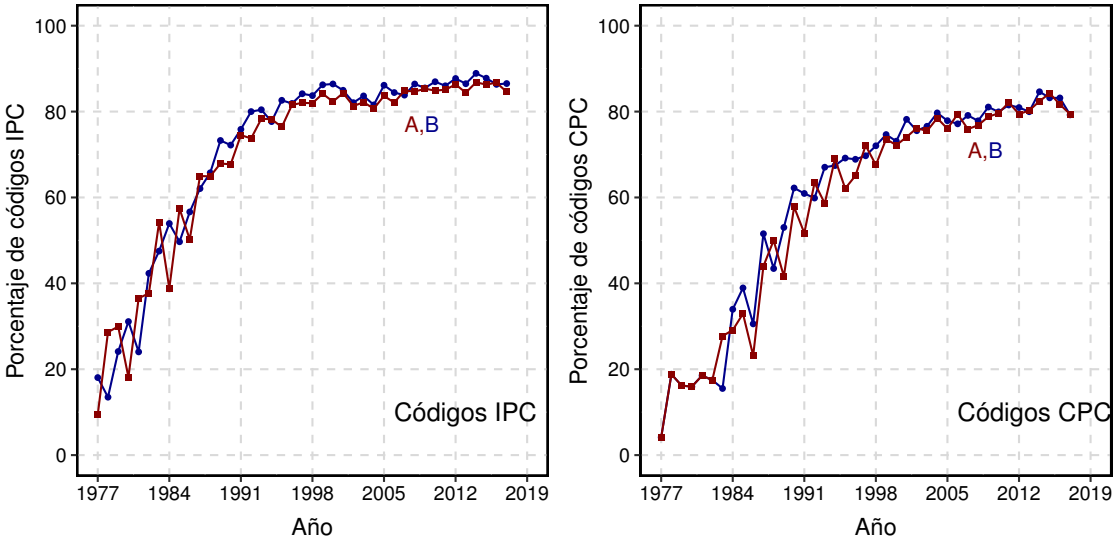

Figura 3.12: Porcentaje de códigos tecnológicos más usados en un año que mantienen este estatus de un año a otro: considerando códigos con frecuencia de uso en los fractiles $P95 - 100$ (línea A en color rojo) y en el fractil $P90 - 100$ (línea B en color azul); para códigos IPC (izquierda) y CPC (derecha). Periodo: 1977-2018.

Por otro lado, la figura 3.12 muestra la proporción de códigos usados más a menudo en un año dado, y que mantienen ese estatus en el siguiente año. Especificamente,

son considerados los códigos con frecuencia de uso en el fractil $P95 - 100$ (línea A en color rojo), y los códigos con frecuencia de uso en el fractil $P90 - 100$ (línea B en color azul) para los códigos IPC en la izquierda y códigos CPC en la derecha. En esta figura se evidencia que la proporción de códigos IPC y CPC que mantienen su estatus de un año a otro, crece considerablemente hasta valores cercanos al 80 %.

El comportamiento que se visualiza en la figura 3.12 concuerda con las 2 fases de la evolución de la desigualdad total en la frecuencia de uso de códigos tecnológicos. Además, en un entorno de crecimiento en el número total de patentes y códigos tecnológicos, la evolución de la probabilidad (creciente) de un código "rico"de mantenerse en este estatus puede interpretarse en términos de un mecanismo de "el rico se hace más rico".

Descomposición de la desigualdad en la frecuencia de uso de códigos tecnológicos por actividades económicas productivas

La figura 3.13 muestra el ranking para los años 2000 y 2018 (último año analizado) en términos de la desigualdad en la frecuencia de uso de códigos IPC para cada división NACE, estimada con el coeficiente de Gini (izquierda) y el índice de Theil (derecha). Como resultado tenemos que en el año 2018, las primeras 3 posiciones fueron ocupadas por las divisiones NACE 12 (Manufacture of Tobacco Products), NACE 21 (Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations) y NACE 26 (Manufacture of Computer, Electronic and Optical Products) respectivamente, con valores para ambos indicadores, por encima del valor de la desigualdad para todo el conjunto de datos (representandos la fila denominada "All"). Con respecto a las dos últimas posiciones del ranking, estas fueron ocupadas por las divisiones NACE 42 (Civil Engineering) y NACE 16 (Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials).

En comparación con el ranking del año 2000, se puede observar que algunas divisiones NACE cambiaron su ubicación en el ranking significativamente. Por ejemplo, la división NACE 21 (Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations) se mantiene en una de las primeras posiciones del ranking, la división NACE 62 (Computer Programming, Consultancy and Related Activities) se desplazó desde las principales posiciones a la parte baja del ranking, y por último, la división NACE 12 cuya situación fue la opuesta, pasando de las últimas posiciones del ranking a la primera posición en 2018.

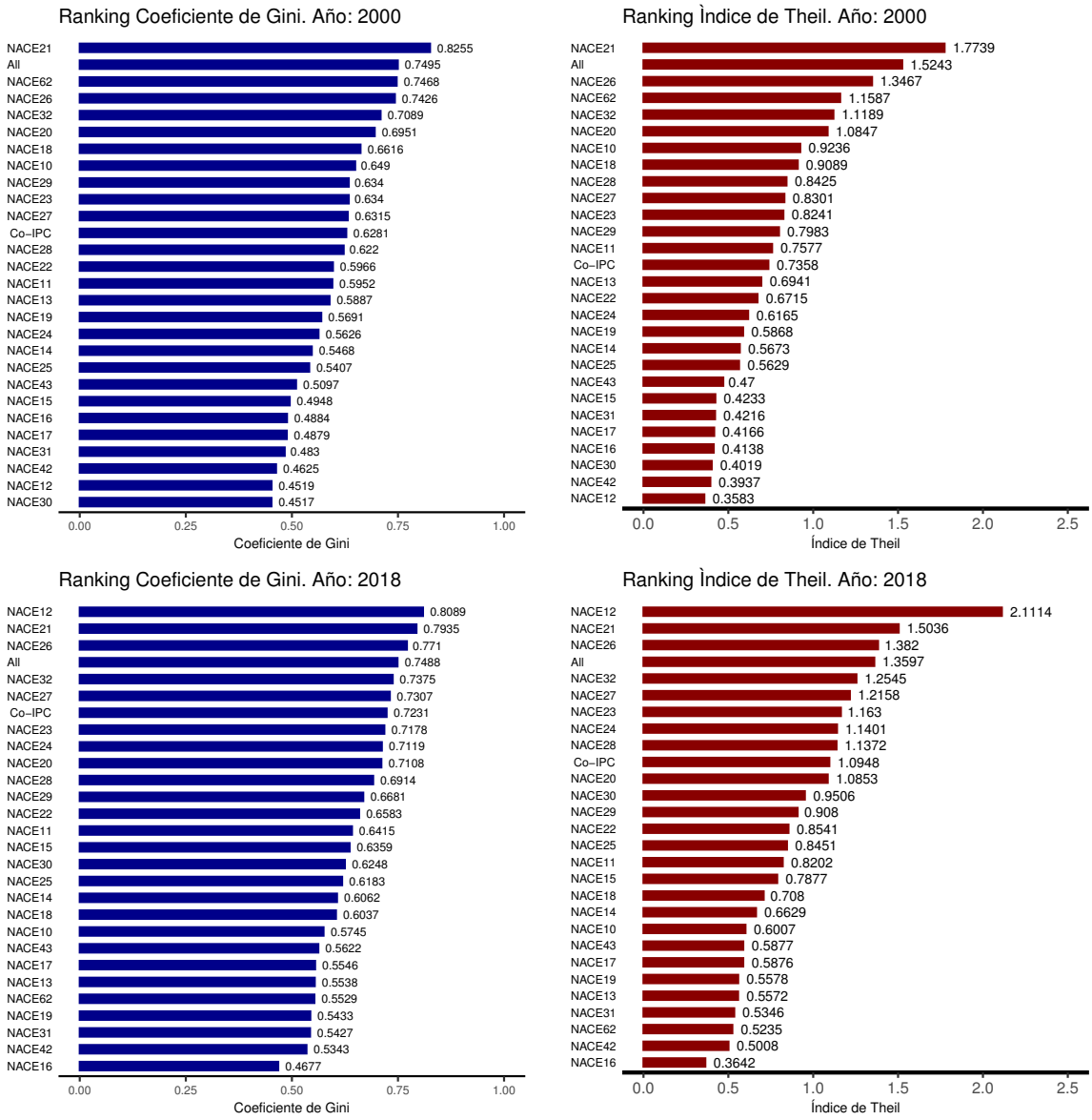

Figura 3.13: Ranking de divisiones NACE en los años 2000 y 2018 en términos de la desigualdad en la frecuencia de uso de códigos IPC, estimados por el coeficiente de Gini (izquierda) y el índice de Theil (derecha).

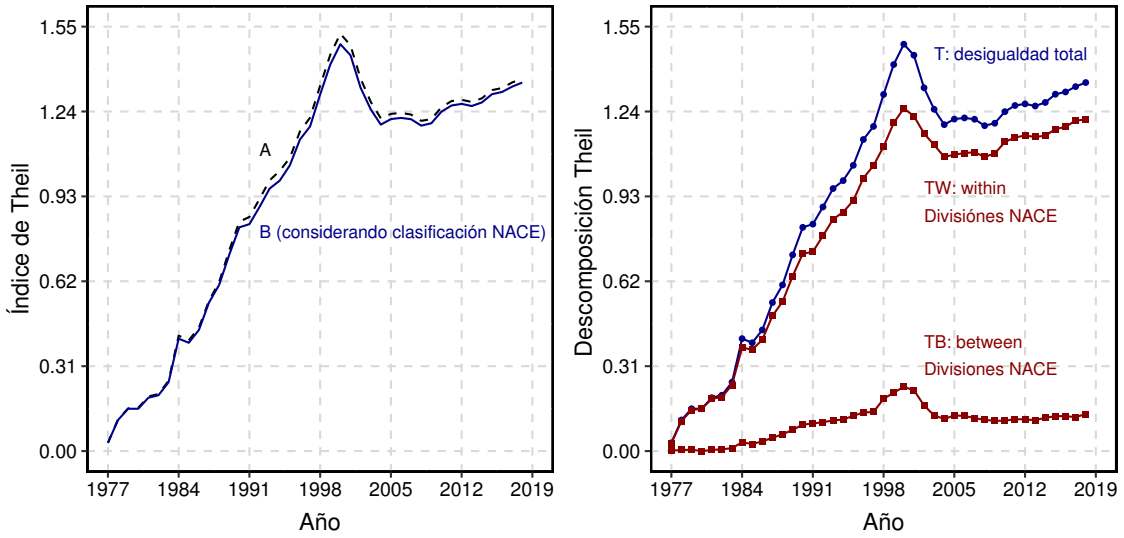

Figura 3.14: Desigualdad total en la frecuencia de uso de códigos tecnológicos IPC, estimada con el índice de Theil (Izquierda), y descomposición de la desigualdad en la frecuencia de uso de códigos IPC, estimada con el índice de Theil (dcha), periodo 1977 - 2018.

La figura 3.14 muestra, en la izquierda, la desigualdad total en la frecuencia de uso de códigos tecnológicos IPC de la base de datos (en la línea punteada, estimada por el índice de Theil), en comparación con el total de la desigualdad obtenido considerando la diferente clasificación de los códigos IPC (con símbolo de subclase C07B, C07C, C07F, C07G, C12M, C12S or C40B) en diferentes divisiones NACE, dependiendo de si están o no acompañados por códigos IPC del grupo A61K008/ (línea B). Consultar apartado 3.2.5 para más detalles. Se puede observar que las líneas A y B son muy similares, y además, el efecto de la división de los códigos mencionados es muy bajo en la desigualdad total en la frecuencia de uso anual obtenida.

La figura 3.14 muestra, en la derecha, la descomposición del total de la desigualdad en dos componentes, dados por la expresión (3.5): el término de la desigualdad within-group ($TW = T(\boldsymbol{y}, n)_W$) y el término de la desigualdad, between-group ($TB = T(\boldsymbol{y}, n)_B$). Se puede evidenciar que gran parte de la desigualdad, en todos los años estudiados, corresponde al componente within-group. Además, la evidencia sugiere que la desigualdad total en la frecuencia de uso de códigos IPC es conducida principalmente por las diferencias en la frecuencia de uso de estos códigos dentro de cada división NACE. Entre tanto, la contribución de las diferencias en la frecuencia media de uso entre divisiones NACE es muy baja.

3.4. Conclusiones

En este capítulo se mide por primera vez, desde lo que conocemos, la desigualdad en la frecuencia de uso de códigos tecnológicos en patentes. Además, indaga acerca de la relación entre la desigualdad en la frecuencia de uso de códigos y las fuentes de la novedad tecnológica en la invención como un proceso combinatorio. Con este propósito, fueron seleccionadas las aplicaciones de patentes PCT en fase internacional con la EPO como oficina designada, y se consideró la información de códigos IPC y CPC de estas patentes.

En primer lugar, se estimó la desigualdad usando el coeficiente de Gini y el índice de Theil. Se encontró que la desigualdad total en la frecuencia de uso de códigos tecnológicos tiene dos fases claramente diferenciadas. Durante la primera fase (1977-2000), la desigualdad total muestra un patrón de crecimiento. En la segunda fase (2001-2018), la desigualdad total muestra una relativa estabilización. Además, se evidencia que la desigualdad total anual en la frecuencia de uso de códigos tecnológicos tiende a ser persistentemente alta, sin importar si son considerados los códigos IPC o CPC, o si la estimación es realizada usando el coeficiente de Gini o el índice de Theil.

La evolución anual de la desigualdad total en la frecuencia de uso de códigos tecnológicos parece coincidir con la evolución de las fuentes de la novedad inventiva. Los resultados indican que la evolución de la desigualdad total en la frecuencia de uso estaría relacionada con los cambios que se presentan en la evolución de las fuentes de novedad inventiva en donde las patentes, en los años iniciales, son en su mayoría constituidas por códigos tecnológicos nuevos, y luego, se da paso a una segunda etapa en la cual predominan las patentes descritas por la combinación de códigos tecnológicos previamente usados. Esta investigación ofrece evidencia empírica adicional de la invención basada principalmente en el proceso combinatorio de tecnologías existentes.

Igualmente, se analizó la evolución de la probabilidad que tienen los códigos usados más frecuentemente, de mantener este estatus de un año a otro, a lo largo del periodo considerado. La proporción de la frecuencia de uso total que acaparan los códigos comprendidos en el primer decil $P90 - 100$ en el año 2018, fue superior al 60 %, y además, la probabilidad que tienen estos códigos de mantenerse en este decil, año tras año, se situó en alrededor de un 80 %. De este modo, en un contexto de crecimiento en el número total de patentes y de la frecuencia total de uso de códigos, el aumento de la probabilidad de un código de mantenerse en el estatus de "rico", haciendo la analogía de códigos ricos a aquellos más usados, puede interpretarse en términos de un mecanismo de "el rico se hace más rico" o "rich-get-richer".

Finalmente, se estudió la descomposición de esta desigualdad por actividades económicas productivas tomando como referencia la clasificación NACE. Tras agrupar los códigos IPC por actividades económicas productivas, se concluye que la desigualdad total esta principalmente conducida por las diferencias en la frecuencia de uso entre los códigos que conforman cada división NACE, el término de *within - group* o intragrupo del índice de Theil. Una pequeña contribución a la desigualdad total proviene de las diferencias entre las divisiones NACE, *between - group* o término intergrupo.

En futuras investigaciones, las estimaciones llevadas a cabo en este capítulo pueden ser extendidas a otras bases de datos de patentes, como por ejemplo, USPTO (US Patent and Trademark Office) o JPO (Japanese Patent Office). Al respecto es importante indicar que la reclasificación de patentes y los cambios en los sistemas de clasificación de patentes pueden brindarnos información clave acerca de la evolución tecnológica (Lafond y Kim, 2019). Otros retos abiertos por esta investigación son expandir el conocimento acerca del por qué algunas tecnológias son usadas más frecuentemente que otras, o analizar por qué la desigualdad dentro de las industrias difiere.

Capítulo 4

Modelización estadística de la frecuencia de uso de códigos tecnológicos

4.1. Introducción

En este capítulo se presenta una modelación estadística de las variables aleatorias discretas: frecuencia de uso de códigos tecnológicos IPC y CPC en aplicaciones de patentes PCT. Un análisis anualizado para todo el rango de valores que toman las variables consideradas, fue llevado a cabo para el periodo comprendido entre 1977 y 2018. Para ello, fueron especificados 13 modelos de distribución parámetrica (de 1, 2 y 3 parámetros), siendo los valores de los parámetros, estimados mediante el Método de Máxima Verosimilitud (MLE por sus sigles en inglés). Además, como medida de contraste para la comparación de los modelos de distribución ajustados, se utilizó el Criterio de Información de Akaike, más conocido por sus siglas en inglés, AIC. Por último, una vez fueron seleccionados aquellos modelos de distribución que mejor ajustan (modelizan) las variables, se efectúo una prueba de hipótesis estadística usando el test de bondad de ajuste Chi-Cuadrado (para cada año) con un nivel de significancia de $\alpha = 0{,}05$.

Desde lo que conocemos, un esfuerzo previo en modelizar la variable frecuencia de uso de códigos tecnológicos en patentes fue realizado por Youn et al., (2015) quienes usando un modelo de distribución continua de dos parámetros como lo es la distribución de probabilidad clásica de Pareto, aportan evidencia empírica que sugiere que la frecuencia de uso de códigos tecnológicos (acumulada) en patentes - a partir de los datos de patentes de la USPTO hasta el año 2013 - muestra un relación de ley de potencia (*power- law*) en su cola pesada (grandes valores). Este hallazgo según

los autores es discordante con la idea de comprender la búsqueda tecnológica como un proceso aleatorio; y se aproximan a la existencia de un mecanismo del "rico se hace más rico"que sugiere el uso intensivo de ciertos códigos tecnológicos frente a otros rara vez usados.

La motivación de este capítulo nos conduce al estudio de la frecuencia de uso de palabras en textos, y específicamente a la conocida Ley de Zipf (1949) que ha sido a lo largo de las últimas décadas un tema de especial interés en el ámbito de la linguística, siendo además una materia que suscita interesantes debates que se mantienen aún vigentes (i Cancho y Solé, 2003; Corominas-Murtra et al., 2011; Perc, 2012; Piantadosi, 2014). En esta investigación hemos tomado el estudio de la frecuencia de uso de las palabras como inspiración; estableciendo una similitud entre los códigos tecnológicos y las palabras entendemos que, desde una perspectiva tecnológica, los códigos hacen parte del cuerpo de un "lenguaje" utilizado para describir e informar acerca de los componentes o tecnologías incorporadas en una invención.

El capítulo está estructurado de la siguiente manera: en el apartado 4.2 Datos y Metodología se describen, en primer lugar, las bases de datos utilizadas y se definen las variables aleatorias estudiadas. Posteriormente, son especificadas las expresiones matemáticas usadas para la estimación de los parámetros en los modelos de distribución seleccionados ajustados a las variables aleatorias, así como el método seguido para realizar dicho ajuste. Así mismo, se detalla la medida de contraste y la prueba de hipótesis estadística empleada. En el apartado 4.3 se presentan los resultados del ajuste de las variables aleatorias discretas frecuencia de uso de códigos tecnológicos IPC y CPC, en los modelos de distribución de 1 y 2 parámetros. En una extensión del apartado, son introducidos en el análisis los modelos de distribución de 3 parámetros considerados y sus resultados. Por último, en el apartado 4.4 son presentadas las conclusiones finales.

Destacar las principales aportaciones del presente capítulo:

- Se analiza, por primera vez en la literatura hasta lo que se conoce, la distribución de frecuencia de uso de códigos tecnológicos de patentes en todo el rango de su distribución, teniendo en cuenta la naturaleza discreta de dicha variable aleatoria.

- Se extiende el esfuerzo realizado en la literatura en el análisis de la frecuencia de aparición de las palabras en el lenguaje, de forma interdisciplinar, al ámbito de la frecuencia de aparición de los códigos en el "lenguaje"de las combinaciones de tecnologías en la invención.

- Se incorporan al análisis anterior, modelos de distribución de la jerarquía de distribuciones discretas de Pareto: tipo III y IV, y modelos de distribución de 3 parámetros como el modelo Pareto tipo IV anterior y la distribución de Waring generalizada desplazada, novedosos hasta lo que se conoce en este ámbito de investigación.

4.2. Datos y metodología

4.2.1. Datos

En este capítulo se utilizó información de patentes de la base de datos REGPAT, versión de julio de 2021, publicada por la Organización para el Desarrollo y la Cooperación Económica OCDE. Las variables aleatorias discretas estudiadas son la frecuencia de uso de códigos tecnológicos IPC y CPC en aplicaciones de patente PCT entre los años 1977 y 2018. Una descripción de las variables se presenta en el apartado 3.2.2 del capítulo previo que analiza la desigualdad en la frecuencia de uso de códigos tecnológicos.

Para la construcción de las variables aleatorias se usó información de códigos tecnológicos IPC del fichero (`202107_PCT_IPC.txt`), e información de códigos tecnológicos CPC extraída del fichero de códigos CPC (`202107_PCT_App_reg.txt`) y el fichero de aplicantes de patentes PCT (`202107_CPC_Classes.txt`). Tras el procesamiento de los ficheros de datos que se detalla en el apartado 2.1.4, fueron obtenidas dos bases de datos de formato largo con información de códigos tecnológicos IPC (`BaseLarga_IPC_Final.txt`) y CPC (`BaseLarga_CPC_Final.txt`) para un total de 3786795 aplicaciones de patentes PCT.

A continuación, se presenta una visualización de las primeras observaciones de las bases de datos de formato largo correspondientes a los códigos tecnológicos IPC y CPC, respectivamente:

```
> head(DataIPC)
       pct_nbr prio_year         IPC
1 WO1978000001      1977  C03C013/04
2 WO1978000001      1977 C03B037/016
3 WO1978000001      1977  G21F009/12
4 WO1978000001      1977  B01J035/06
5 WO1978000001      1977  C03C023/00
6 WO1978000001      1977  C03C011/00
> dim(DataIPC)
[1] 13615554        3

> head(DataCPC)
       pct_nbr prio_year    CPC_Class
```

```
1 WO1978000001      1977  B01D17/0202
2 WO1978000001      1977  B01D39/2003
3 WO1978000001      1977  B01D67/0088
4 WO1978000001      1977    B01D71/04
5 WO1978000001      1977 B01J20/28033
6 WO1978000001      1977 B01J20/28042
> dim(DataCPC)
[1] 25227358        3
```

En estas bases de datos de formato largo, cada fila u observación corresponde a un código tecnológico asignado a una aplicación de patente PCT. La base de datos de códigos tecnológicos IPC contiene 13615554 observaciones y 72607 códigos. Por su parte, la base de datos de códigos tecnológicos CPC comprende 25227358 observaciones y 234782 códigos diferentes.

Mediante un bucle con el comando `for` se obtuvieron los valores de X_i: frecuencia de uso, y O_i: número de veces usados (número de códigos correspondientes a cada uno de los valores que toma la frecuencia de uso X_i) para cada año i del periodo en estudio, y que posteriormente, son empleados para el ajuste de los modelos de distribución. Se puede observar a continuación el bucle `for` mencionado:

```
Data_year<-rep(list(data.frame()),numyears)
for (i in 1:numyears){
  year<-i+first.year-1
  message(year: ,year)
  Codes<-Data[Data$prio_year==year,3]
  Codes_xi<-table(Codes)
  Data_year[[i]]<-data.frame(
    Xi=as.numeric(rownames(table(Codes_xi))),
    Oi=as.numeric(table(Codes_xi))) }
```

A modo de ilustración, se presentan como ejemplo las tablas de frecuencia de uso obtenidas para los códigos tecnológicos IPC en los años 1978 (*i=2*) y 1980 (*i=4*):

```
> Data_year[[2]]
    xi     Oi
1    1   3260
2    2    521
3    3    116
4    4     48
5    5     19
6    6      4
7    7      2
8    8      3
9    9      3
10  12      2
11  18      1
```

```
> Data_year[[4]]
       xi     Oi
1       1   5454
2       2   1192
3       3    355
4       4    138
5       5     72
6       6     24
7       7     18
8       8      6
9       9      2
10     10      1
11     11      6
12     12      1
13     15      1
14     17      1
15     18      1
16     22      1
```

4.2.2. Metodología

A continuación, se detallan los modelos de distribución utilizados en este capítulo, para la modelización estadística de las variables aleatorias discretas: frecuencia de uso de los códigos tecnológicos IPC y CPC de patentes.

En una primera fase, se realizó un análisis con los modelos de distribuciones discretas con 1 y 2 parámetros, para lo cual se seleccionaron aquellos más conocidos en la literatura. Teniendo en cuenta la naturaleza discreta de la variable de interés, así como los valores que dicha variable aleatoria puede tomar (valores enteros mayores o iguales a uno, $x = 1, 2, 3, \dots$). Dichos modelos fueron los siguientes:

- Modelos de distribución de tipo discreto con 1 parámetro:
  - Distribución Geométrica
  - Distribución de Poisson desplazada
  - Distribución Zeta
  - Distribución de Yule-Simon
  - Distribución Logarítmica
- Modelos de distribución de tipo discreto con 2 parámetros:
  - Distribución Binomial Negativa desplazada
  - Distribución Poisson generalizada desplazada
  - Distribución Zipf-Mandelbrot
  - Distribución de Beta-Geométrica
  - Distribución de Pareto discreta generalizada tipo II
  - Distribución de Pareto discreta generalizada tipo III

Tras los resultados obtenidos en dicha primera fase, el análisis se amplió a una selección de dos modelos de distribuciones de tipo discreto con 3 parámetros:

- Modelos de distribución de tipo discreto con 3 parámetros:
  - Distribución de Pareto discreta generalizada tipo IV
  - Distribución de Waring generalizada desplazada

En el cuadro 4.1 se muestran las diferentes funciones de densidad de los modelos de distribución anteriormente indicados, así como la notación utilizada en el presente capítulo para dichos modelos. Más adelante, en este mismo apartado, se proporciona un mayor detalle sobre dichas distribuciones, detalle que puede ser ampliado por el lector en conocidos trabajos sobre distribuciones de probabilidad (Arnold, 2015; Johnson, Kotz y Kemp, 2005; Klugman, Panjer y Willmot, 2012), entre otros.

Cuadro 4.1: Funciones de densidad y notación de los modelos de distribución de probabilidad considerados.

Distribución	Notación	Función de densidad $f(x) = \Pr(X = x)$	
Geométrica	$G(p)$	$p(1-p)^{x-1}$	$x = 1, 2, \ldots;\ 0 < p < 1$
Poisson desplazada	$Po(\lambda)$	$e^{-\lambda}\lambda^{x-1}/(x-1)!$	$x = 1, 2, \ldots;\ \lambda > 0$
Zeta	$Zeta(\alpha)$	$[\zeta(\alpha+1)]^{-1}\, x^{-(\alpha+1)}$	$x = 1, 2, \ldots;\ \alpha > 0$
Yule-Simon	$Yule(\rho)$	$\rho\beta(x, \rho+1)$	$x = 1, 2, \ldots;\ \rho > 0$
Logarítmica	$Log(\theta)$	$-\theta^x / [x \log(1-\theta)]$	$x = 1, 2, \ldots;\ 0 < \theta < 1$
Binomial Negativa desplazada	$BN(r, p)$	$\binom{r+x-2}{x-1} p^r (1-p)^{x-1}$	$x = 1, 2, \ldots;\ 0 < p < 1;\ r > 0$
Poisson generalizada desplazada	$Po(\lambda_1, \lambda_2)$	$\frac{\lambda_1(\lambda_1 + (x-1)\lambda_2)^{x-2} \exp[-(\lambda_1 + (x-1)\lambda_2]}{(x-1)!}$	$x = 1, 2, \ldots;\ \lambda_1 > 0;\ \lvert\lambda_2\rvert < 1$
Zipf-Mandelbrot	$ZM(\alpha, \beta)$	$[\zeta(\alpha+1, \beta+1)]^{-1}\, (x+\beta)^{-(\alpha+1)}$	$x = 1, 2 \ldots;\ \alpha > 0;\ \beta \geq 0$
Beta Geométrica	$BG(\alpha, \beta)$	$\beta(\alpha+1, x+\beta-1)/\beta(\alpha, \beta)$	$x = 1, 2, \ldots;\ \alpha, \beta > 0$
Pareto discreta generalizada II	$DGPII(\sigma, \alpha)$	$\left[1 + \frac{x-1}{\sigma}\right]^{-\alpha} - \left[1 + \frac{x}{\sigma}\right]^{-\alpha}$	$x = 1, 2, \ldots;\ \sigma, \alpha > 0$
Pareto discreta generalizada III	$DGPIII(\sigma, \gamma)$	$\left[1 + \left(\frac{x-1}{\sigma}\right)^{(1/\gamma)}\right]^{-1} - \left[1 + \left(\frac{x}{\sigma}\right)^{(1/\gamma)}\right]^{-1}$	$x = 1, 2, \ldots;\ \sigma, \gamma > 0$
Pareto discreta generalizada IV	$DGPIV(\sigma, \gamma, \alpha)$	$\left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha}$	$x = 1, 2, \ldots;\ \sigma, \gamma, \alpha > 0$
Waring generalizada desplazada	$SWG(\alpha, \beta, \nu)$	$\frac{B(\nu+\alpha, x+\beta-1)\Gamma(x+\nu-1)}{B(\alpha, \beta)\Gamma(\nu)(x-1)!}.$	$x = 1, 2, \ldots;\ \alpha, \beta, \nu > 0$

A continuación, los modelos de probabilidad fueron ajustados usando el método de máxima verosimilitud. Este método permite estimar el vector de parámetros $\boldsymbol{\theta}$ que especifíca la función de probabilidad $P(X = x|\boldsymbol{\theta})$ en variables discretas, a partir de las observaciones o conjunto de datos, $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$. La estimación de máxima verosimilitud es el valor del vector de parámetros $\hat{\boldsymbol{\theta}}$ que máximiza la función de verosimilitud, y que resulta ser el valor más probable para generar los datos observados. La función de máxima verosimilitud es definida por la siguiente expresión:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta}) \tag{4.1}$$

En la expresión 4.1 de máxima verosimilitud, $\boldsymbol{\theta}$ es el vector de parámetros desconocido del modelo; $x_i,\ i = 1, \ldots, n$ es el conjunto de datos de la muestra; y el término $f(x)$ es la función de densidad de los modelos de distribución de probabilidad. En el cuadro 4.1, se muestran las funciones de densidad y los valores posibles de los parámetros de los modelos de distribución de probabilidad de 1 y 2 parámetros considerados.

Considerando que la función del logaritmo natural $\log(x)$ es una función creciente, el máximo de la función $L(x_i \mid \theta)$ coincide con el máximo de la función, $\log L(x_i \mid \theta)$ o función de log-verosimilitud que se define de la siguiente manera:

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i \mid \boldsymbol{\theta}) \tag{4.2}$$

Indicarque en varias de las distribuciones seleccionadas se dispuso de fórmulas cerradas para los estimadores de los parámetros. En aquellas en las que no fue así, la estimación de dichos parámetros se llevó a cabo utilizando el software `R`, mediante el paquete `optimx`, version 2021-10.12, (Nash y Varadhan, 2011; Nash, 2014).

La comparación de la bondad de ajuste de los diferentes modelos seleccionados se llevó a cabo utilizando el Criterio de Información de Akaike (AIC: Akaike Information Criterion), dado por la siguiente expresión:

$$AIC = -2 \log \ell + 2K \tag{4.3}$$

donde $\log \ell = \log L(\hat{\boldsymbol{\theta}})$ es la logverosimilitud de modelo evaluado en la estimación de máxima verosimilitud, K es el número de parámetros del modelo, y donde un valor del Criterio de Información de Akaike AIC más bajo indica una mejor bondad de ajuste del modelo.

Tras obtener los valores de AIC, se calculó para cada año el correspondiente valor de $\Delta = AIC - AIC_{min}$ (Burnham y Anderson, 2004; Burnham, Anderson y Huyvaert, 2011), siendo AIC_{min} el valor de AIC del modelo con mejor bondad de ajuste en dicho año, y se clasificaron los modelos analizados con base a las siguientes tres categorías:

- modelos con el mejor ajuste, dados por $\Delta \leq 2$;
- modelos con soporte relativamente pequeño, dados por $2 < \Delta \leq 20$;
- modelos sin soporte empírico, dados por $\Delta > 20$.

Para finalizar el análisis, se utilizó el test de bondad de ajuste de chi-cuadrado, propuesto por Pearson (1900). Este test proporciona una estimación de la cercanía de los valores observados con respecto a los valores esperados, siendo en nuestro caso los valores esperados aquellos obtenidos del ajuste de los modelos de distribución de probabilidad para variables discretas seleccionados. El test de bondad de ajuste o contraste de hipótesis chi-cuadrado se define por la siguiente expresión:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{4.4}$$

En la expresión 4.73, *Oi* representa los datos observados y *Ei*, los datos esperados para la variable frecuencia de uso de códigos tecnológicos. Expresamos la hipótesis nula de la investigación como:

Ho: los datos de la variable frecuencia de uso de códigos tecnológicos siguen el modelo de distribución discreta seleccionado.

Por su parte, la hipótesis alternativa se plantea como:

H_1: los datos de la variable frecuencia de uso de códigos tecnológicos no se ajustan al modelo de distribución discreta seleccionados.

Para las distribuciones de probabilidad consideradas, la hipótesis nula puede ser rechazada con un nivel de significancia $\alpha = 0{,}05$, si se cumple que $X^2 > X^2_{0,95;k-r-1}$; donde r es el número de parámetros del modelo, y k corresponde al número de observaciones.Así mismo, se obtuvo el correspondiente p-valor, a partir del cual la hipótesis nula puede rechazarse con el nivel de significancia elegido si p-valor < 0.05.

Distribución Geométrica

La distribución geométrica o de Pascal (Sarabia, Prieto y Jordá, 2020), que denotaremos como $X \sim G(p)$, depende de un único parámetro, p, que toma valores $0 < p < 1$. Esta distribución modela el número de intentos fallidos antes de obtener un éxito, x, en una serie de intentos independientes entre sí. El resultado posible en cada intento realizado es el éxito o el fracaso, siendo la probabilidad de éxito constante, y el número de intentos indefinido. Una variable aleatoria, X, con distribución geométrica y parámetro p, tiene como función de densidad:

$$f(x) = \Pr(X = x) = p(1-p)^x,\ x = 0, 1, 2, 3, \ldots \tag{4.5}$$

Ahora bien, cuando el soporte de la variable aleatoria discreta inicia en 1, como es el caso de las variables estudiadas, x puede interpretarse como el número de intentos necesarios para alcanzar el exito. Para esta parametrización de la distribución geométrica, la función de densidad se define de la siguiente forma:

$$f(x) = \Pr(X = x) = p(1-p)^{x-1},\ x = 1, 2, 3, ... \tag{4.6}$$

A partir de la función de densidad especificada en la expresión 4.6, definimos la función de máxima verosimilitud del modelo de distribución geométrica como:

$$L(p) = \prod_{j=1}^{\infty} f(j)^{N_j} = \prod_{j=1}^{\infty} \left[p(1-p)^{j-1}\right]^{N_j} \tag{4.7}$$

En la expresión 4.7, el término x (valores de la variable aleatoria en los datos observados) es sustituido por j, que ahora toma valores de números enteros, que van desde 1 hasta el infinito; y el término N_j es el número de veces que j aparece en el conjunto de datos de la muestra. con base en ello, la función de máxima verosimilitud de la distribución geométrica se define en la expresión 4.7 como el producto de las funciones de densidad para cada valor entre 1 e ∞ elevado al número de veces que dicho valor aparece en la muestra. Luego, al aplicar el logaritmo natural a la expresión 4.7, obtenemos la función de log-verosimilitud:

$$\begin{aligned} \log L(p) = \log \left[\prod_{j=1}^{\infty} f(j)^{N_j}\right] &= \sum_{j=1}^{\infty} \log \left[f(j)^{N_j}\right] \\ &= \sum_{j=1}^{\infty} N_j \log \left[f(j)\right] \end{aligned} \tag{4.8}$$

Ahora, teniendo en cuenta la propiedad de los logaritmos que señala que, el logaritmo de un producto es igual a la suma de los logaritmos de los factores, la función de log-verosimilitud puede considerarse como la sumatoria del logaritmo de las funciones de densidad para cada j (entre 1 e ∞), multiplicado el término N_j (ver expresión (4.8)). Al reemplazar en la expresión 4.8, el término $f(j)$ por la función de densidad de la distribución geométrica, obtenemos la siguiente expresión para la función de log-verosimilitud:

$$\log L(p) = \sum_{j=1}^{\infty} Nj \log \left[p(1-p)^{j-1}\right] \tag{4.9}$$

A continuación, se presenta el desarrollo o solución de la expresión 4.9, hasta obtener una expresión más sencilla de dicha función de log-verosimilitud:

$$\begin{aligned} \log L(p) &= \sum_{j=1}^{\infty} N_j \log(p) + \sum_{j=1}^{\infty} N_j (j-1) \log(1-p) \\ &= n \left[\log(p) + (\bar{x} - 1) \log(1-p)\right] \end{aligned} \tag{4.10}$$

Si establecemos el valor del estimador $\hat{p}$, que maximiza la función de log-verosimilitud:

$$\hat{p} = \arg \max \log L(p) \tag{4.11}$$

Para obtener dicho estimador $\hat{p}$ de máxima verosimilitud derivamos la expresión 4.10 en función del parámetro p, e igualamos a cero:

$$\frac{d \log L(p)}{dp} = \frac{n}{p} - \frac{n(\hat{x} - 1)}{1-p} = 0 \tag{4.12}$$

Despejando el término del parámetro p en la ecuación 4.12, tenemos como resultado que el valor del estimador $\hat{p}$ es igual a la inversa de la media del conjunto de datos:

$$\hat{p} = \frac{1}{\bar{x}} \tag{4.13}$$

siendo, el error estándar del estimador $\hat{p}$, tal como sigue:

$$s.e_{\,\hat{p}} = \left[\frac{\hat{p}^2(1-p)}{n}\right]^{1/2} = \hat{p} \left[\frac{1-\hat{p}}{n}\right]^{1/2} \tag{4.14}$$

Distribución Poisson Desplazada

La distribución de Poisson desplazada (Gómez-Déniz et al., 2020; Rothschild, 1986), $X \sim Po(\lambda)$, es un modelo de probabilidad aplicado a variables discretas positivas, donde el valor de la variable aleatoria, x, son números enteros positivos diferentes de cero. La distribución depende de un único parámetro que suele representarse con la letra griega lambda, λ, cuyo valor es mayor que cero ($\lambda > 0$).

La función de densidad de la distribución de Poisson desplazada se define por la siguiente expresión:

$$\Pr(X = x) = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\ , \quad x = 1, 2, 3, \ldots \tag{4.15}$$

Por otra parte, la estimación de $\hat{\lambda}$ por el método de los momentos viene dada por la expresión:

$$\hat{\lambda} = \bar{x} - 1 \tag{4.16}$$

siendo, la correspondiente función de máxima verosimilitud:

$$L(\lambda) = \prod_{j=1}^{\infty} \left[\frac{e^{-\lambda}\lambda^{j-1}}{(j-1)!}\right]^{N_j} \tag{4.17}$$

Luego, aplicamos el logaritmo natural a la expresión 4.17 de máxima verosimilitud. Se obtiene la función de log-verosimilitud:

$$\log L(\lambda) = \log\left\{\prod_{j=1}^{\infty} \left[\frac{e^{-\lambda}\lambda^{j-1}}{(j-1)!}\right]^{N_j}\right\} \tag{4.18}$$

La expresión 4.18 se soluciona hasta llegar a una expresión más sencilla de la función de log-verosimilitud:

$$\begin{aligned}\log L(\lambda) &= -\lambda \sum_{j=1}^{\infty} N_j + \sum_{j=1}^{\infty} N_j (j-1)\log(\lambda) - \sum_{j=1}^{\infty} N_j \log\left[(j-1)!\right] \\ &= -n\lambda + n\log(\lambda)(\bar{x} - 1) - \sum_{j=1}^{\infty} N_j \log\left[(j-1)!\right]\end{aligned} \tag{4.19}$$

Para establecer el valor del estimador $\hat{\lambda}$ que maximiza la función de logverosimilitud, derivamos la ecuación 4.19 en función del parámetro λ:

$$\frac{d \log L(\lambda)}{d\lambda} = -n + \frac{n(\bar{x} - 1)}{\lambda} = 0$$

$$\hat{\lambda} = \bar{x} - 1 \tag{4.20}$$

Puede observarse que la estimación de $\hat{\lambda}$ por el método de máxima verosimilitud y por el método de los momentos es equivalente en esta distribución. De tal forma que las expresiones 4.16 y 4.20 son idénticas.

El error estándar del estimador $\hat{\lambda}$ puede obtenerse a partir de la siguiente expresión:

$$s.e_{\,\hat{\lambda}} = \left[\frac{\hat{\lambda}(2 + \hat{\lambda}}{2n} \right]^{1/2} \tag{4.21}$$

Distribución Zeta

La distribución Zeta, tambien conocida como distribución de Zipf, (Arnold, 2015; Sichel, 1975; Zipf, 1932; Zipf, 1949), y que denotaremos como $X \sim Zeta(\alpha)$, es una distribución uniparamétrica, que depende del parámetro α, ($\alpha > 0$), y con soporte en el conjunto de números enteros positivos. La función de densidad para la distribución Zeta esta definida por la expresión:

$$\Pr(X = x) = [\zeta(\alpha + 1)]^{-1} \, x^{-(\alpha+1)} \, , x = 1, 2, 3, \ldots \tag{4.22}$$

donde $\zeta(s)$ es la función Zeta de Riemman, definida por:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{4.23}$$

A partir de la función de densidad especificada en la expresión 4.22, definimos la función de máxima verosimilitud del modelo de distribución:

$$L(\alpha) = \frac{\left[\prod_{j=1}^{\infty} j^{N_j}\right]^{-(\alpha+1)}}{[\zeta(\alpha+1)]^n} \tag{4.24}$$

Aplicando el logarítmo natural a la expresión 4.24, obtenemos la función de log-verosimilitud:

$$\log L(\alpha) = -(\alpha+1)\log\left[\prod_{j=1}^{\infty} j^{N_j}\right] - n\log\left[\zeta(\alpha+1)\right]$$

que puede expresarse como:

$$\log L(\alpha) = -(\alpha+1)\sum_{j=1}^{\infty} N_j \log(j) - n\log\left[\zeta(\alpha+1)\right] \tag{4.25}$$

Posteriormente, se deriva la expresión 4.25 en función de dicho parámetro α, y se iguala a cero:

$$\frac{d\log L(\alpha)}{d\alpha} = -\sum_{j=1}^{\infty} N_j \log(j) - n\frac{\zeta'(\alpha+1)}{\zeta(\alpha+1)} = 0 \tag{4.26}$$

Al despejar los términos de la ecuación 4.26, para se obtiene la siguiente expresión que nos permite obtener la estimación por el método de máxima verosimilitud del parámetro α:

$$-\frac{\zeta'(\alpha+1)}{\zeta(\alpha+1)} = \frac{1}{n}\sum_{j=1}^{\infty} N_j \log(j) \tag{4.27}$$

Distribución Logarítmica

La distribución logarítmica, también conocida como serie logarítmica, fue introducida por Fisher, Corbett y Williams (1943). Esta distribución de probabilidad discreta (Sarabia Alegría, Gómez Déniz y Vázquez Polo, 2007), que denotaremos como $X \sim Log(\theta)$, cuenta con un parámetro θ que puede tomar los valores $0 < \theta < 1$, y su función de densidad se define por la expresión:

$$\Pr(X = x) = \frac{a\theta^x}{x} = \frac{-\theta^x}{x \log(1-\theta)} \ , x = 1, 2, 3, \ldots \tag{4.28}$$

Donde el valor de a en la función de densidad (expresión 4.28) viene dado por:

$$a = -\left[\log[1-\theta)\right] \tag{4.29}$$

Para la estimación de $\hat{\theta}$, mediante el método de los momentos, se usa la siguiente expresión:

$$\hat{\theta} = 1 - \frac{\bar{x}}{s^2 + \bar{x}^2} \tag{4.30}$$

La función de máxima verosimilitud para la distribución logarítmica es:

$$\text{Ł}(\theta) = \prod_{j=1}^{\infty} \left[\frac{a\theta^j}{j}\right]^{N_j} = \prod_{j=1}^{\infty} \left[\frac{-\theta^j}{j \log(1-\theta)}\right]^{N_j} \tag{4.31}$$

Aplicando el logaritmo natural a la expresión 4.31, llegamos a la función de log-verosimilitud:

$$\log L(\theta) = -n \log\left[-\log(1-\theta)\right] + n\bar{x} \log(\theta) - \sum_{j=1}^{\infty} N_j \log(j) \tag{4.32}$$

Al aplicar una derivada a la expresión 4.32 en función del parámetro θ, e igualando a cero dicha derivada, se obtiene la ecuación requerida para estimar el valor del parámetro θ de máxima verosimilitud:

$$\bar{x} + \frac{\hat{\theta}}{(1-\hat{\theta}) \log(1-\hat{\theta})} = 0 \tag{4.33}$$

Distribución de Yule-Simon

La distribución de Yule-Simon, conocida también como distribución de Yule, (Simon, 1955; Xekalaki, 1983; Xekalaki, 1984; Yule, 1925), y que denotaremos como $X \sim Yule(\rho)$, presenta la siguiente función de densidad:

$$\Pr(X = x) = \rho B(x, \rho + 1) \tag{4.34}$$

El estimador de momentos de dicho parámetro $\hat{\rho}$ viene dado por la siguiente expresión:

$$\hat{\rho} = \frac{\bar{x}}{\bar{x} - 1} \tag{4.35}$$

La función de máxima verosimilitud para la distribución Yule - Simon es:

$$L(\rho) = \prod_{j=1}^{\infty} [\rho B(x, \rho + 1)]^{N_j} \tag{4.36}$$

y por último, la función de log-verosimilitud se expresa como:

$$\log L(\rho) = n \log(\rho) + \sum_{j=1}^{\infty} Nj \log [B(j, \rho + 1)] \tag{4.37}$$

Distribución Binomial Negativa Desplazada

La distribución Binomial Negativa desplazada (Ajiferuke, 1991; Vaidyanathan, Savariappan y Chandrasekhar, 2009; Wolfram, 1992), que denotaremos $X \sim BN(p, r)$, depende de dos parámetros: p y r, los cuales toman valores: $0 < p < 1$; $r > 0$. La función de densidad de la distribución se define por la siguiente expresión:

$$\Pr(X = x) = \binom{r + x - 2}{x - 1} p^r (1 - p)^{x-1} \ , \ x = 1, 2, 3, \ldots \tag{4.38}$$

Como caso particular, cuando $r = 1$, se obtiene la función de densidad de la distribución Geométrica vista anteriormente (ver expresión 4.6).

La estimación de los parámetros p y r, por el método de los momentos, puede realizarse a partir de las expresiones:

$$\hat{p} = \frac{\bar{x} - 1}{s^2} \tag{4.39}$$

$$\hat{r} = \frac{\hat{p}(\bar{x} - 1)}{1 - \hat{p}} \tag{4.40}$$

Partiendo de la función de densidad (expresión 4.38); sustituyendo los valores de la variable aleatoria en los datos observados por j (que toma valores de números enteros, que van desde 1 hasta el infinito) y siendo el término N_j el número de veces que j aparece en el conjunto de datos de la muestra; y finalmente, tras aplicar el logaritmo natural a la expresión 4.7, obtenemos la función de log-verosimilitud:

$$Ł(p, r) = \prod_{j=1}^{\infty} \left[\left(\frac{r + j + 2}{j - 1} \right) p^r (1 - p)^{j-1} \right]^{N_j} \tag{4.41}$$

Aplicando el logaritmo natural a la expresión 4.41 obtenemos la función de log-verosimilitud:

$$\begin{aligned} \log L(p, r) &= \sum_{j=1}^{\infty} N_j \log \left(\frac{r + j - 2}{j - 1} \right) + nr \log(p) + \sum_{j=1}^{\infty} N_j (j - 1) \log(p) \\ &= nr \log(p) + \log(1 - p) \sum_{j=1}^{\infty} N_j (j - 1) + \sum_{j=1}^{\infty} N_j \log \left(\frac{r + j - 2}{j - 1} \right) \end{aligned} \tag{4.42}$$

Distribución Poisson Generalizada Desplazada

La función de densidad de la distribución de Poisson generalizada desplazada (Ajiferuke, 1991; Consul y Jain, 1973; Consul y Shoukri, 1984), $X \sim Po(\lambda_1, \lambda_2)$, viene dada por la siguiente expresión:

$$\Pr(X = x) = \frac{\lambda_1 (\lambda_1 + (x - 1)\lambda_2)^{x-2} \exp\left[-(\lambda_1 + (x - 1)\lambda_2\right]}{(x - 1)!};\ x = 1, 2, 3, \ldots \tag{4.43}$$

La distribución depende de dos parámetros λ_1 y λ_2. El parámetro λ_1 toma valores positivos ($\lambda_1 > 0$) y λ_2 valores absolutos menores a 1, $|\lambda_2| < 1$. Como caso particular, cuando $\lambda_2 = 0$, se obtiene la función de densidad de la distribución de Poisson desplazada (ver expresión 4.15).

La estimación de $\hat{\lambda}_1$ y $\hat{\lambda}_2$ por el método de los momentos se efectúa con las expresiones:

$$\hat{\lambda}_2 = 1 - \left[\frac{(\bar{x}-1)}{s^2}\right]^{\frac{1}{2}} \tag{4.44}$$

$$\hat{\lambda}_1 = (\bar{x}-1)(1-\hat{\lambda}_2) \tag{4.45}$$

La función de máxima verosimilitud de la distribución puede expresarse de la siguiente forma:

$$Ł(\lambda_1, \lambda_2) = \prod_{j=1}^{\infty}\left[\frac{\lambda_1(\lambda_1+(j-1)\lambda_2)^{j-2}\exp\left[-\lambda_1+(j-1)\lambda_2\right]}{(j-1)!}\right]^{N_j} \tag{4.46}$$

y la función de log-verosimilitud viene dada por la siguiente expresión:

$$\begin{aligned}\log Ł(\lambda_1, \lambda_2) = n\log(\lambda_1) &+ \sum_{j=1}^{\infty} N_j(j-2)\log\left[\lambda_1+(j-1)\lambda_2\right] \\ &- \sum_{j=1}^{\infty} N_j\left[\lambda_1+(j-1)\lambda_2\right] - \sum_{j=1}^{\infty} N_j\log\left[(j-1)!\right]\end{aligned} \tag{4.47}$$

Distribución Zipf-Mandelbrot

La distribución Zipf-Mandelbrot es una distribución de probabilidad discreta (Mandelbrot, 1966; Mandelbrot, 1982; Manin, 2009), que denotaremos $X \sim ZM(\alpha, \beta)$. El nombre de esta distribución tiene su origen en el linguista George Kingsley Zipf, en la distribución de Zipf vista en el apartado anterior, y en el matemático Benoit Mandelbrot quien generalizó dicha distribución Zipf introduciendo un segundo parámetro (β). Entendiendose así que, la distribución de Zipf (Zeta) es un caso especial de la distribución Zipf-Mandelbrot, cuando β es igual a cero.

La distribución Zipf-Mandelbrot ha sido utilizada para modelar frecuencias clasificadas. En la literatura, su aplicación se ha realizado en mayor medida en el ámbito

de la linguística, pero tambien se ha extendido a otras áreas de la investigación científica como la medicina, economía, entre otras (Mačutek, 2021). La función de densidad para la distribución Zipf-Mandelbrot está definida por la expresión:

$$\Pr(X = x) = [\zeta(\alpha + 1, \beta + 1)]^{-1} (x + \beta)^{-(\alpha+1)} , x = 1, 2, 3, \ldots \tag{4.48}$$

donde $\alpha > 0$, $\beta \geq 0$, y donde $\zeta(s, a)$ es la función Zeta de Hurwitz, definida por:

$$\zeta(s, a) = \sum_{n=0}^{\infty} \frac{1}{(n + a)^s} \tag{4.49}$$

La función de máxima verosimilitud de la distribución Zipf-Mandelbrot viene dada por la siguiente expresión:

$$L(\alpha, \beta) = \frac{\left[\prod_{j=1}^{\infty} (j + \beta)^{N_j}\right]^{-(\alpha+1)}}{[\zeta(\alpha + 1, \beta + 1)]^n} \tag{4.50}$$

Tras aplicar el logaritmo natural a la expresión 4.50, llegamos a la función de log-verosimilitud:

$$\log L(\alpha, \beta) = -(\alpha + 1) \sum_{j=1}^{\infty} N_j \log(j + \beta) - n \log [\zeta(\alpha + 1, \beta + 1)] \tag{4.51}$$

Luego, se deriva la expresión 4.51 de log-verosimilitud en función de los parámetros α y β. Las derivadas parciales son igualadas a cero, obteniendo un sistema de dos ecuaciones para la estimación de los parámetros α y β de máxima verosimilitud:

$$\frac{\partial \log L(\alpha, \beta)}{\partial \alpha} = -\sum_{j=1}^{\infty} N_j \log(j + \beta) - n \frac{\dfrac{\partial \zeta(\alpha + 1, \beta + 1)}{\partial \alpha}}{\zeta(\alpha + 1, \beta + 1)} = 0 \tag{4.52}$$

$$\frac{\partial \log L(\alpha, \beta)}{\partial \beta} = -(\alpha + 1) \sum_{j=1}^{\infty} N_j \frac{1}{j + \beta} - n \frac{\dfrac{\partial \zeta(\alpha + 1, \beta + 1)}{\partial \beta}}{\zeta(\alpha + 1, \beta + 1)} = 0 \tag{4.53}$$

Distribución Beta-Geométrica

La distribución Beta-Geométrica, también conocida como distribución Geométrica tipo I o distribución Waring desplazada, (Irwin, 1968; Irwin, 1975; Paul, 2005), que denotaremos como $X \sim BG(\alpha, \beta)$, está caracterizada por dos parámetros α y β, tomando ambos parámetros valores positivos ($\alpha >0$, $\beta > 0$).

La función de densidad de la distribución Beta-Geométrica viene dada por la siguiente expresión:

$$\Pr(X = x) = \frac{B(\alpha + 1, x + \beta - 1)}{B(\alpha, \beta)},\ x = 1, 2, 3, \ldots \tag{4.54}$$

Como caso particular, cuando $\beta = 1$, se obtiene la función de densidad de la distribución de Yule-Simon vista anteriormente (expresión 4.34), tal como puede comprobarse a continuación:

$$\Pr(X = x) = \frac{B(\alpha + 1, x)}{B(\alpha, 1)} = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)\Gamma(1)} B(x, \alpha + 1) = \alpha B(x, \alpha + 1) \tag{4.55}$$

Utilizando el método de los momentos, pueden obtenerse los siguiente estimadores de los parámetros $\alpha, \beta)$, tal como sigue:

$$\hat{\alpha} = \frac{2s^2}{s^2 - \bar{x}(\bar{x} - 1)} \tag{4.56}$$

$$\hat{\beta} = (\hat{\alpha} - 1)(\bar{x} - 1) \tag{4.57}$$

La función de máxima verosimilitud para la distribución se define por la expresión a continuación:

$$Ł(\alpha, \beta) = \prod_{j=1}^{\infty} \left[\frac{B(\alpha + 1, j + \beta - 1)}{B(\alpha, \beta)} \right]^{N_j} \tag{4.58}$$

y la función de log-verosimilitud viene dada por la expresión:

$$\log Ł(\alpha, \beta) = -n \log \left[B(\alpha, \beta) \right] + \sum_{j=1}^{\infty} N_j \log \left[B(\alpha + 1, j + \beta - 1) \right] \tag{4.59}$$

Distribución Pareto Discreta Generalizada II

La distribución Pareto discreta generalizada (Prieto, Gómez-Déniz y Sarabia, 2014) es una distribución de tres parámetros (α, λ, μ), que si particularizamos para nuestra variable de interés (que toma valores enteros mayores o igual a 1, $\mu = 1$), y consideramos el parámetro de escala $\sigma = 1/\lambda$, denotaremos entonces como $X \sim DGPII(\sigma, \alpha)$ y su función de densidad vendrá expresada como:

$$\Pr(X = x) = \left[1 + \frac{x-1}{\sigma}\right]^{-\alpha} - \left[1 + \frac{x}{\sigma}\right]^{-\alpha}, x = 1, 2, 3, \ldots;\ \sigma, \alpha > 0 \tag{4.60}$$

La expresión que define la función de máxima verosimilitud de esta distribución es la siguiente:

$$Ł(\sigma, \alpha) = \prod_{j=1}^{\infty} \left(\left[1 + \frac{x-1}{\sigma}\right]^{-\alpha} - \left[1 + \frac{x}{\sigma}\right]^{-\alpha} \right)^{N_j} \tag{4.61}$$

y la correspondiente función de log-verosimilitud:

$$\log L(\sigma, \alpha) = \sum_{j=1}^{\infty} N_j \log \left(\left[(1 + \frac{x-1}{\sigma}\right]^{-\alpha} - \left[1 + \frac{x}{\sigma}\right]^{-\alpha} \right) \tag{4.62}$$

Distribución Discreta Generalizada de Pareto III

La distribución discreta generalizada de Pareto tipo III puede definirse en términos de su función de distribución acumulada. Si $F(x) = \Pr(X \leq x)$, se define:

$$F(x) = 1 - [1 + [(x - \mu + 1)/\sigma]^{(1/\gamma)}]^{-1},\ \ x = \mu, \mu + 1, \ldots \tag{4.63}$$

y $F(x) = 0$ si $x < \mu$, donde $\gamma, \sigma, \mu > 0$ son respectivamente los parámetros de forma, escala y localización (Arnold, 2015).

Si particularizamos dicho modelo de distribución para nuestra variable de interés (que toma valores enteros mayores o igual a 1, $\mu = 1$), la función de distribución anterior 4.64 vendrá expresada como:

$$F(x) = 1 - [1 + (x/\sigma)^{(1/\gamma)}]^{-1},\ \ x = 1, 2, 3, \ldots \tag{4.64}$$

Su función de supervivencia $\bar{F}(x) = \Pr(X \geq x) = 1 - F(x - 1)$ puede expresarse de la siguiente forma (Xekalaki, 1983b):

$$\bar{F}(x) = [1 + [(x-1)/\sigma]^{(1/\gamma)}]^{-1},\ \ x = 1, 2, 3, \ldots \tag{4.65}$$

Denotaremos a una variable aleatoria, con función de distribución acumulada 4.64, por $X \sim \mathcal{DGPIII}(\sigma, \gamma)$, y su función de densidad de la distribución discreta generalizada de Pareto III vendrá dada por:

$$\Pr(X = x) = \left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1}, \quad x = 1, 2, 3, \ldots \tag{4.66}$$

Indicar que para obtener la función de densidad anterior, se aplicó la siguiente expresión: $\Pr(X = x) = \bar{F}(x) - \bar{F}(x+1)$ y se llevó a cabo la discretización de la distribución continua Pareto generalizada (Kemp, 2008; Nakagawa, 1975; Gómez-Déniz, 2010; Gómez-Déniz y Calderín-Ojeda, 2011; Gómez-Déniz, Sarabia y Calderín-Ojeda, 2011).

A partir de la función de densidad 4.66, se define la función de máxima verosimilitud como:

$$L(\sigma, \gamma) = \prod_{j=1}^{\infty} \left(\left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1} \right)^{N_j} \tag{4.67}$$

$$\tag{4.68}$$

Y por último, la correspondiente función de log-verosimilitud viene dada como:

$$\log L(\sigma, \gamma) = \sum_{j=1}^{\infty} N_j \log \left(\left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-1} \right) \tag{4.69}$$

Distribución Discreta Generalizada de Pareto IV

La distribución discreta generalizada de Pareto tipo IV puede definirse en términos de su función de distribución acumulada $F(x) = \Pr(X \leq x)$ como sigue:

$$F(x) = 1 - [1 + [(x - \mu + 1)/\sigma]^{(1/\gamma)}]^{-\alpha}, \quad x = \mu, \mu + 1, \ldots \tag{4.70}$$

y $F(x) = 0$ si $x < \mu$, donde $\alpha, \gamma > 0$ son los parámetros de forma y $\sigma, \mu > 0$ son respectivamente los parámetros de escala y localización (Arnold, 2015).

Si particularizamos dicho modelo de distribución para nuestra variable de interés (que toma valores enteros mayores o igual a 1, $\mu = 1$), la función de distribución anterior 4.64 vendrá dada por:

$$F(x) = 1 - [1 + (x/\sigma)^{(1/\gamma)}]^{-\alpha}, \quad x = 1, 2, 3, \ldots \tag{4.71}$$

Su función de supervivencia podrá expresarse como:

$$\bar{F}(x) = [1 + [(x-1)/\sigma]^{(1/\gamma)}]^{-\alpha}, \quad x = 1, 2, 3, \ldots \tag{4.72}$$

Denotaremos a una variable aleatoria, con función de distribución acumulada 4.71, por $X \sim \mathcal{DGPIV}(\sigma, \gamma, \alpha)$, y su correspondiente función de densidad de la distribución discreta generalizada de Pareto IV vendrá dada por:

$$\Pr(X = x) = \left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha}, \quad x = 1, 2, 3, \ldots \tag{4.73}$$

Puede comprobarse que la distribución discreta generalizada de Pareto IV incluye como casos particulares la distribución discreta generalizada de Pareto tipo II (cuando $\gamma = 1$), y al modelo Pareto tipo III (cuando $\alpha = 1$).

La función de máxima verosimilitud para dicha distribución es:

$$L(\sigma, \gamma, \alpha) = \prod_{j=1}^{\infty} \left(\left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} \right)^{N_j} \tag{4.74}$$

y la función de log-verosimilitud:

$$\log L(\sigma, \gamma, \alpha) = \sum_{j=1}^{\infty} N_j \log \left(\left[1 + \left(\frac{x-1}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} - \left[1 + \left(\frac{x}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\alpha} \right) \tag{4.75}$$

Distribución Discreta Generalizada de Waring Desplazada

La función de densidad de la distribución discreta generalizada de Waring desplazada viene dada por la siguiente expresión (Burrell, 2005; Rodríguez-Avi et al., 2009; Sarabia y Castillo, 2003; Xekalaki, 1983c; Xekalaki, 2008):

$$\Pr(X = x) = \frac{\Gamma(\nu + \alpha)}{B(\alpha, \beta)\Gamma(\nu)} \frac{\Gamma(x + \beta - 1)\Gamma(x + \nu - 1)}{(x-1)!\,\Gamma(x + \alpha + \beta + \nu - 1)}, \quad x = 1, 2, 3, \ldots$$

Dicha función de densidad también puede expresarse como sigue:

$$\Pr(X = x) = \frac{B(\nu + \alpha, x + \beta - 1)\Gamma(x + \nu - 1)}{B(\alpha, \beta)\Gamma(\nu)(x-1)!}, \quad x = 1, 2, 3, \ldots \tag{4.76}$$

Puede observarse que cuando $x = 1$, la expresión 4.76 toma el siguiente valor:

$$\Pr(X = 1) = \frac{B(\nu + \alpha, \beta)}{B(\alpha, \beta)} \tag{4.77}$$

y cuando $x > 1$, tenemos que la función de densidad de la distribución es:

$$\Pr(X = x) = \frac{B(\nu + \alpha, x + \beta - 1)}{(x-1)B(\alpha, \beta)B(\nu, x-1)}, \quad x = 2, 3, \ldots \tag{4.78}$$

Indicar que la distribución discreta generalizada de Waring desplazada incluye como caso particular la distribución Beta Geométrica (Waring desplazada) cuando $\nu = 1$. Denotaremos a una variable aleatoria, con función de densidad 4.76, $X \sim \mathcal{SGW}(\alpha, \beta, \nu)$.

A partir de las expresiones 4.77 y 4.78, tenemos que la función de máxima verosimilitud para la distribución es:

$$L(\alpha, \beta, \nu) = \left[\frac{B(\nu + \alpha, \beta)}{B(\alpha, \beta)}\right]^{N_1} \prod_{j=2}^{\infty} \left[\frac{B(\nu + \alpha, j + \beta - 1)}{(j-1)B(\alpha, \beta)B(\nu, j-1)}\right]^{N_j} \tag{4.79}$$

siendo la función de log-verosimilitud:

$$\begin{aligned}\log L(\alpha, \beta, \nu) = {} & N_1 \left[\log\left[B(\nu + \alpha, \beta)\right] - \log\left[B(\alpha, \beta)\right]\right] + \sum_{j=2}^{\infty} N_j \log\left[B(\nu + \alpha, j + \beta - 1)\right] \\ & - \sum_{j=2}^{\infty} N_j \log(j-1) - \sum_{j=2}^{\infty} N_j \log\left[B(\alpha, \beta)\right] - \sum_{j=2}^{\infty} N_j \log\left[B(\nu, j-1)\right]\end{aligned} \tag{4.80}$$

4.3. Resultados

4.3.1. Distribución de probabilidad de la frecuencia de uso de códigos tecnológicos IPC

Estimación de los parámetros y errores estándar

Los cuadros 4.2 y 4.3 muestran los parámetros estimados y sus errores estándar en los 11 modelos de distribución de probabilidad de 1 y 2 parámetros, ajustados a la variable frecuencia de uso de códigos tecnológicos IPC en todo el rango por el método de máxima verosimilitud, para el periodo 1977 y 2018. En el cuadro 4.2 se encuentran los resultados mencionados para las distribuciónes: Geométrica (parámetro p), Negativa Binomial Desplazada (parámetros p y r), Poisson Desplazada (parámetro λ), Poisson Generalizada Desplazada (parámetros λ_1 y λ_2) y Logarítmica (parámetro θ).

Cuadro 4.2: Parámetros estimados y errores estándar para los modelos de distribución de probabilidad 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018.

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
1977	0.9264	0.8567	0.4754	0.0795	0.0736	0.0743	0.1402
	(0.0079)	(0.0670)	(0.2531)	(0.0090)	(0.0087)	(0.0373)	(0.0156)
1978	0.7851	0.5350	0.3149	0.2737	0.1971	0.2799	0.3724
	(0.0058)	(0.0228)	(0.0257)	(0.0088)	(0.0073)	(0.0177)	(0.0110)
1979	0.7238	0.4747	0.3449	0.3816	0.2575	0.3253	0.4594
	(0.0049)	(0.0157)	(0.0191)	(0.0087)	(0.0069)	(0.0132)	(0.0096)
1980	0.7054	0.4839	0.3916	0.4177	0.2850	0.3176	0.4840
	(0.0045)	(0.0140)	(0.0194)	(0.0083)	(0.0067)	(0.0116)	(0.0090)
1981	0.6675	0.4073	0.3425	0.4982	0.3089	0.3800	0.5326
	(0.0043)	(0.0115)	(0.0140)	(0.0087)	(0.0066)	(0.0107)	(0.0084)
1982	0.6519	0.4037	0.3615	0.5340	0.3296	0.3828	0.5517
	(0.0041)	(0.0108)	(0.0139)	(0.0088)	(0.0066)	(0.0101)	(0.0083)
1983	0.6117	0.3484	0.3394	0.6347	0.3597	0.4333	0.5990
	(0.0039)	(0.0089)	(0.0112)	(0.0093)	(0.0066)	(0.0093)	(0.0080)
1984	0.5155	0.2416	0.2994	0.9398	0.4303	0.5421	0.7004
	(0.0032)	(0.0055)	(0.0070)	(0.0105)	(0.0063)	(0.0073)	(0.0068)
1985	0.4959	0.2467	0.3330	1.0167	0.4724	0.5354	0.7192
	(0.0030)	(0.0051)	(0.0073)	(0.0104)	(0.0064)	(0.0066)	(0.0065)
1986	0.4532	0.2132	0.3269	1.2064	0.5142	0.5738	0.7578
	(0.0027)	(0.0042)	(0.0064)	(0.0112)	(0.0064)	(0.0060)	(0.0061)

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
1987	0.4041	0.1801	0.3239	1.4744	0.5667	0.6156	0.7986
	(0.0023)	(0.0033)	(0.0054)	(0.0119)	(0.0063)	(0.0052)	(0.0058)
1988	0.3628	0.1561	0.3249	1.7563	0.6194	0.6473	0.8301
	(0.0021)	(0.0027)	(0.0049)	(0.0130)	(0.0063)	(0.0046)	(0.0054)
1989	0.3090	0.1281	0.3284	2.2360	0.6982	0.6877	0.8673
	(0.0017)	(0.0020)	(0.0043)	(0.0143)	(0.0063)	(0.0039)	(0.0051)
1990	0.2613	0.1010	0.3176	2.8270	0.7603	0.7311	0.8968
	(0.0014)	(0.0015)	(0.0037)	(0.0165)	(0.0064)	(0.0034)	(0.0047)
1991	0.2475	0.0976	0.3288	3.0412	0.8013	0.7365	0.9048
	(0.0013)	(0.0014)	(0.0037)	(0.0170)	(0.0064)	(0.0032)	(0.0047)
1992	0.2173	0.0833	0.3271	3.6016	0.8594	0.7614	0.9212
	(0.0011)	(0.0011)	(0.0034)	(0.0189)	(0.0065)	(0.0029)	(0.0044)
1993	0.1984	0.0741	0.3235	4.0392	0.8946	0.7785	0.9309
	(0.0010)	(0.0010)	(0.0032)	(0.0202)	(0.0065)	(0.0027)	(0.0043)
1994	0.1790	0.0671	0.3297	4.5859	0.9543	0.7919	0.9403
	(0.0009)	(0.0009)	(0.0031)	(0.0218)	(0.0066)	(0.0025)	(0.0042)
1995	0.1598	0.0596	0.3328	5.2561	1.0134	0.8072	0.9490
	(0.0008)	(0.0007)	(0.0029)	(0.0239)	(0.0067)	(0.0023)	(0.0041)
1996	0.1404	0.0501	0.3230	6.1212	1.0532	0.8279	0.9574
	(0.0007)	(0.0006)	(0.0027)	(0.0265)	(0.0066)	(0.0020)	(0.0040)
1997	0.1258	0.0441	0.3208	6.9516	1.1012	0.8416	0.9634
	(0.0006)	(0.0005)	(0.0026)	(0.0291)	(0.0067)	(0.0019)	(0.0038)
1998	0.1075	0.0360	0.3095	8.2981	1.1469	0.8618	0.9703
	(0.0005)	(0.0004)	(0.0024)	(0.0332)	(0.0067)	(0.0017)	(0.0037)
1999	0.0904	0.0295	0.3055	10.0574	1.2185	0.8788	0.9764
	(0.0004)	(0.0003)	(0.0022)	(0.0388)	(0.0069)	(0.0015)	(0.0036)
2000	0.0796	0.0254	0.3014	11.5643	1.2689	0.8903	0.9801
	(0.0004)	(0.0003)	(0.0021)	(0.0433)	(0.0070)	(0.0014)	(0.0034)
2001	0.0766	0.0246	0.3038	12.0617	1.2936	0.8928	0.9811
	(0.0004)	(0.0003)	(0.0021)	(0.0447)	(0.0070)	(0.0014)	(0.0034)
2002	0.0820	0.0275	0.3159	11.1968	1.2979	0.8841	0.9793
	(0.0004)	(0.0003)	(0.0022)	(0.0414)	(0.0070)	(0.0014)	(0.0034)
2003	0.0868	0.0304	0.3295	10.5171	1.3076	0.8757	0.9777
	(0.0004)	(0.0003)	(0.0023)	(0.0390)	(0.0070)	(0.0014)	(0.0034)
2004	0.0919	0.0331	0.3386	9.8856	1.3038	0.8681	0.9760
	(0.0004)	(0.0003)	(0.0024)	(0.0364)	(0.0069)	(0.0015)	(0.0035)
2005	0.0985	0.0345	0.3267	9.1520	1.2334	0.8652	0.9736
	(0.0005)	(0.0004)	(0.0023)	(0.0347)	(0.0068)	(0.0016)	(0.0035)
2006	0.0948	0.0333	0.3288	9.5512	1.2602	0.8681	0.9749
	(0.0004)	(0.0004)	(0.0023)	(0.0358)	(0.0068)	(0.0015)	(0.0035)

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
2007	0.0935	0.0329	0.3297	9.6910	1.2702	0.8689	0.9754
	(0.0004)	(0.0004)	(0.0023)	(0.0362)	(0.0069)	(0.0015)	(0.0034)
2008	0.0955	0.0340	0.3331	9.4714	1.2670	0.8663	0.9747
	(0.0004)	(0.0004)	(0.0024)	(0.0354)	(0.0069)	(0.0015)	(0.0035)
2009	0.0910	0.0324	0.3347	9.9927	1.2969	0.8702	0.9762
	(0.0004)	(0.0003)	(0.0023)	(0.0368)	(0.0069)	(0.0015)	(0.0034)
2010	0.0830	0.0291	0.3308	11.0498	1.3333	0.8793	0.9789
	(0.0004)	(0.0003)	(0.0022)	(0.0398)	(0.0069)	(0.0014)	(0.0034)
2011	0.0770	0.0268	0.3296	11.9884	1.3691	0.8858	0.9809
	(0.0003)	(0.0003)	(0.0022)	(0.0426)	(0.0070)	(0.0013)	(0.0033)
2012	0.0743	0.0258	0.3305	12.4595	1.3917	0.8883	0.9818
	(0.0003)	(0.0003)	(0.0022)	(0.0440)	(0.0071)	(0.0013)	(0.0033)
2013	0.0716	0.0252	0.3344	12.9594	1.4237	0.8901	0.9826
	(0.0003)	(0.0003)	(0.0022)	(0.0453)	(0.0071)	(0.0013)	(0.0033)
2014	0.0702	0.0245	0.3329	13.2473	1.4300	0.8921	0.9831
	(0.0003)	(0.0002)	(0.0022)	(0.0462)	(0.0072)	(0.0012)	(0.0032)
2015	0.0661	0.0230	0.3317	14.1268	1.4607	0.8966	0.9843
	(0.0003)	(0.0002)	(0.0021)	(0.0482)	(0.0071)	(0.0012)	(0.0031)
2016	0.0630	0.0218	0.3317	14.8767	1.4876	0.9000	0.9853
	(0.0003)	(0.0002)	(0.0021)	(0.0502)	(0.0072)	(0.0011)	(0.0031)
2017	0.0628	0.0214	0.3271	14.9322	1.4722	0.9014	0.9854
	(0.0003)	(0.0002)	(0.0021)	(0.0508)	(0.0072)	(0.0012)	(0.0031)
2018	0.0624	0.0210	0.3224	15.0211	1.4573	0.9030	0.9855
	(0.0003)	(0.0002)	(0.0020)	(0.0514)	(0.0072)	(0.0012)	(0.0032)

A continuación, en el cuadro 4.3 se muestran los parámetros estimados y sus errores estándar para las distribuciónes de probabilidad: Zeta (parámetro α), Zipf- Mandelbrot (parámetros α y β), Yule-Simon (parámetro ρ), Beta - Geométrica (parámetros α y β), Discreta Generalizada de Pareto II (parámetros α y σ) y Discreta Generalizada de Pareto III (parámetros γ y σ).

Las figuras 4.1, 4.2, 4.3, 4.4, 4.5 muestran, a modo de validación gráfica del ajuste obtenido por cada uno de los modelos, el valor de las frecuencias observadas (en negro) y esperadas (en rojo) para cada uno de los valores de nuestra muestra, en escala log-log, para los años 1990, 2000 y 2018, de los diferentes modelos indicados: distribuciones geométrica y binomial negativa desplazada (figura 4.1); distribuciones zeta y Zipf-Mandelbrot (figura 4.2); distribuciones Yule-Simon y Beta Geométrica (figura 4.3); distribuciones logarítmica y Poisson desplazada (figura 4.4); distribuciones discreta generalizada de Pareto tipos II y III (figura 4.5).

Cuadro 4.3: Parámetros estimados y errores estándar para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018.

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
1977	3.2207	18.4346	5.7916	13.5366	28.4423	2.1819	6.6140	18.7475	0.4674	0.2954
	(0.1475)	(21.4068)	(8.1883)	(1.5663)	(33.0639)	(2.6120)	(8.1648)	(21.3653)	(0.0527)	(0.0412)
1978	2.0581	4.3195	1.3887	4.6517	5.0876	1.1134	2.1341	4.4667	0.5009	0.4585
	(0.0396)	(0.4889)	(0.3019)	(0.1625)	(0.5858)	(0.1457)	(0.3000)	(0.4912)	(0.0167)	(0.0172)
1979	1.7904	4.0273	1.6174	3.5734	4.5410	1.3444	2.3339	4.1379	0.5487	0.4985
	(0.0270)	(0.3277)	(0.2389)	(0.0904)	(0.3720)	(0.1274)	(0.2371)	(0.3285)	(0.0127)	(0.0131)
1980	1.7064	4.4393	2.0962	3.2994	4.9509	1.6481	2.7997	4.5386	0.5786	0.5025
	(0.0231)	(0.3487)	(0.2708)	(0.0734)	(0.3904)	(0.1494)	(0.2690)	(0.3491)	(0.0113)	(0.0114)
1981	1.6048	3.4976	1.5641	2.9313	3.8293	1.4066	2.2572	3.5800	0.5811	0.5475
	(0.0203)	(0.2136)	(0.1782)	(0.0584)	(0.2340)	(0.1017)	(0.1762)	(0.2136)	(0.0109)	(0.0112)
1982	1.5521	3.6190	1.7814	2.7785	3.9340	1.5599	2.4673	3.6960	0.6050	0.5502
	(0.0189)	(0.2156)	(0.1881)	(0.0526)	(0.2342)	(0.1098)	(0.1862)	(0.2156)	(0.0105)	(0.0105)
1983	1.4593	3.0849	1.5123	2.4914	3.3192	1.4598	2.1937	3.1551	0.6319	0.5737
	(0.0167)	(0.1507)	(0.1415)	(0.0430)	(0.1619)	(0.0860)	(0.1398)	(0.1507)	(0.0102)	(0.0099)
1984	1.2754	2.4450	1.2954	1.9942	2.5718	1.4262	1.9567	2.4946	0.6966	0.6395
	(0.0124)	(0.0851)	(0.0949)	(0.0275)	(0.0890)	(0.0620)	(0.0933)	(0.0849)	(0.0096)	(0.0087)
1985	1.2115	2.5604	1.6056	1.8545	2.6749	1.6684	2.2541	2.6030	0.7480	0.6468
	(0.0111)	(0.0861)	(0.1039)	(0.0236)	(0.0894)	(0.0700)	(0.1025)	(0.0859)	(0.0094)	(0.0080)
1986	1.1316	2.2928	1.5190	1.6688	2.3765	1.6625	2.1557	2.3262	0.7898	0.6887
	(0.0099)	(0.0687)	(0.0913)	(0.0196)	(0.0708)	(0.0636)	(0.0897)	(0.0685)	(0.0097)	(0.0078)
1987	1.0536	2.1323	1.5595	1.5048	2.1985	1.7490	2.1898	2.1621	0.8678	0.7030
	(0.0084)	(0.0545)	(0.0802)	(0.0158)	(0.0559)	(0.0579)	(0.0789)	(0.0544)	(0.0095)	(0.0070)
1988	0.9840	2.0099	1.6402	1.3661	2.0604	1.8586	2.2595	2.0338	0.9443	0.7311
	(0.0075)	(0.0471)	(0.0767)	(0.0134)	(0.0479)	(0.0572)	(0.0755)	(0.0469)	(0.0100)	(0.0067)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
1989	0.8986	1.8860	1.8120	1.2073	1.9225	2.0587	2.4192	1.9046	1.0772	0.7559
	(0.0062)	(0.0383)	(0.0723)	(0.0105)	(0.0389)	(0.0561)	(0.0712)	(0.0382)	(0.0104)	(0.0061)
1990	0.8303	1.6735	1.7467	1.0829	1.6990	2.0667	2.3452	1.6880	1.1926	0.7991
	(0.0055)	(0.0302)	(0.0643)	(0.0088)	(0.0305)	(0.0514)	(0.0633)	(0.0301)	(0.0114)	(0.0059)
1991	0.8033	1.6973	1.9572	1.0399	1.7206	2.2593	2.5511	1.7107	1.2750	0.7945
	(0.0051)	(0.0295)	(0.0669)	(0.0081)	(0.0298)	(0.0542)	(0.0660)	(0.0295)	(0.0116)	(0.0056)
1992	0.7553	1.5985	2.0445	0.9591	1.6158	2.3755	2.6286	1.6085	1.4009	0.8264
	(0.0047)	(0.0262)	(0.0660)	(0.0071)	(0.0263)	(0.0546)	(0.0651)	(0.0261)	(0.0127)	(0.0055)
1993	0.7293	1.5426	2.0894	0.9169	1.5573	2.4389	2.6705	1.5516	1.4867	0.8373
	(0.0044)	(0.0238)	(0.0638)	(0.0065)	(0.0239)	(0.0534)	(0.0630)	(0.0237)	(0.0132)	(0.0053)
1994	0.6931	1.5052	2.2833	0.8603	1.5172	2.6392	2.8573	1.5126	1.6319	0.8537
	(0.0040)	(0.0223)	(0.0658)	(0.0059)	(0.0224)	(0.0560)	(0.0650)	(0.0222)	(0.0142)	(0.0052)
1995	0.6602	1.4571	2.4435	0.8100	1.4669	2.8126	3.0122	1.4634	1.7897	0.8664
	(0.0037)	(0.0205)	(0.0666)	(0.0053)	(0.0206)	(0.0575)	(0.0658)	(0.0205)	(0.0153)	(0.0050)
1996	0.6348	1.3719	2.4167	0.7707	1.3802	2.8200	2.9833	1.3776	1.9171	0.8840
	(0.0035)	(0.0180)	(0.0625)	(0.0049)	(0.0181)	(0.0546)	(0.0618)	(0.0180)	(0.0162)	(0.0049)
1997	0.6095	1.3207	2.5135	0.7329	1.3274	2.9333	3.0751	1.3254	2.0723	0.9039
	(0.0033)	(0.0167)	(0.0629)	(0.0045)	(0.0168)	(0.0555)	(0.0623)	(0.0167)	(0.0175)	(0.0048)
1998	0.5833	1.2404	2.5129	0.6939	1.2457	2.9641	3.0705	1.2442	2.2479	0.9307
	(0.0030)	(0.0148)	(0.0604)	(0.0041)	(0.0148)	(0.0538)	(0.0597)	(0.0147)	(0.0189)	(0.0048)
1999	0.5536	1.1929	2.7034	0.6519	1.1971	3.1703	3.2568	1.1961	2.5155	0.9452
	(0.0028)	(0.0136)	(0.0617)	(0.0037)	(0.0136)	(0.0557)	(0.0611)	(0.0135)	(0.0210)	(0.0047)
2000	0.5329	1.1517	2.8237	0.6225	1.1551	3.3046	3.3728	1.1543	2.7319	0.9654
	(0.0027)	(0.0127)	(0.0626)	(0.0035)	(0.0127)	(0.0570)	(0.0621)	(0.0127)	(0.0228)	(0.0046)
2001	0.5239	1.1271	2.8418	0.6100	1.1301	3.3328	3.3899	1.1294	2.8314	0.9735
	(0.0026)	(0.0122)	(0.0622)	(0.0034)	(0.0122)	(0.0568)	(0.0616)	(0.0121)	(0.0236)	(0.0046)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
2002	0.5265	1.1577	2.9595	0.6140	1.1608	3.4359	3.5068	1.1600	2.8190	0.9656
	(0.0026)	(0.0126)	(0.0643)	(0.0034)	(0.0126)	(0.0587)	(0.0638)	(0.0126)	(0.0232)	(0.0046)
2003	0.5280	1.1952	3.1266	0.6169	1.1984	3.5848	3.6730	1.1975	2.8228	0.9542
	(0.0026)	(0.0132)	(0.0677)	(0.0034)	(0.0132)	(0.0618)	(0.0672)	(0.0132)	(0.0229)	(0.0045)
2004	0.5333	1.2338	3.2268	0.6250	1.2373	3.6682	3.7739	1.2363	2.7778	0.9395
	(0.0026)	(0.0137)	(0.0690)	(0.0034)	(0.0137)	(0.0629)	(0.0685)	(0.0137)	(0.0219)	(0.0044)
2005	0.5538	1.2186	2.8200	0.6527	1.2227	3.2740	3.3725	1.2217	2.5287	0.9379
	(0.0027)	(0.0137)	(0.0627)	(0.0036)	(0.0137)	(0.0566)	(0.0622)	(0.0137)	(0.0204)	(0.0045)
2006	0.5445	1.2142	2.9452	0.6397	1.2180	3.3990	3.4952	1.2170	2.6306	0.9443
	(0.0027)	(0.0136)	(0.0646)	(0.0035)	(0.0136)	(0.0586)	(0.0641)	(0.0135)	(0.0212)	(0.0045)
2007	0.5407	1.2102	2.9877	0.6344	1.2140	3.4424	3.5363	1.2128	2.6715	0.9491
	(0.0027)	(0.0135)	(0.0655)	(0.0035)	(0.0135)	(0.0595)	(0.0650)	(0.0135)	(0.0216)	(0.0045)
2008	0.5434	1.2189	2.9830	0.6385	1.2227	3.4345	3.5330	1.2217	2.6457	0.9413
	(0.0027)	(0.0136)	(0.0652)	(0.0035)	(0.0136)	(0.0591)	(0.0646)	(0.0136)	(0.0212)	(0.0045)
2009	0.5334	1.2085	3.0991	0.6245	1.2120	3.5529	3.6464	1.2110	2.7624	0.9464
	(0.0026)	(0.0133)	(0.0664)	(0.0034)	(0.0133)	(0.0604)	(0.0658)	(0.0132)	(0.0220)	(0.0044)
2010	0.5193	1.1707	3.1501	0.6048	1.1737	3.6189	3.6961	1.1730	2.9260	0.9591
	(0.0025)	(0.0124)	(0.0658)	(0.0032)	(0.0124)	(0.0602)	(0.0653)	(0.0124)	(0.0232)	(0.0044)
2011	0.5072	1.1441	3.2285	0.5882	1.1468	3.7076	3.7728	1.1462	3.0853	0.9678
	(0.0024)	(0.0119)	(0.0663)	(0.0031)	(0.0119)	(0.0609)	(0.0658)	(0.0119)	(0.0245)	(0.0043)
2012	0.5004	1.1376	3.3275	0.5790	1.1401	3.8078	3.8700	1.1395	3.1887	0.9735
	(0.0024)	(0.0118)	(0.0678)	(0.0030)	(0.0118)	(0.0626)	(0.0673)	(0.0118)	(0.0253)	(0.0043)
2013	0.4922	1.1334	3.4758	0.5679	1.1357	3.9559	4.0164	1.1351	3.3260	0.9767
	(0.0023)	(0.0117)	(0.0700)	(0.0029)	(0.0117)	(0.0648)	(0.0695)	(0.0116)	(0.0263)	(0.0043)
2014	0.4895	1.1257	3.4873	0.5642	1.1278	3.9702	4.0271	1.1273	3.3663	0.9816
	(0.0023)	(0.0116)	(0.0702)	(0.0029)	(0.0116)	(0.0650)	(0.0697)	(0.0115)	(0.0267)	(0.0043)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
2015	0.4809	1.1133	3.6019	0.5527	1.1152	4.0890	4.1404	1.1147	3.5169	0.9868
	(0.0022)	(0.0112)	(0.0707)	(0.0028)	(0.0112)	(0.0657)	(0.0702)	(0.0111)	(0.0276)	(0.0042)
2016	0.4734	1.1006	3.6976	0.5426	1.1024	4.1890	4.2345	1.1019	3.6561	0.9932
	(0.0022)	(0.0109)	(0.0716)	(0.0027)	(0.0109)	(0.0668)	(0.0712)	(0.0109)	(0.0286)	(0.0042)
2017	0.4758	1.0862	3.5432	0.5456	1.0880	4.0427	4.0816	1.0876	3.5936	0.9952
	(0.0022)	(0.0107)	(0.0693)	(0.0028)	(0.0107)	(0.0645)	(0.0689)	(0.0107)	(0.0284)	(0.0043)
2018	0.4776	1.0682	3.3863	0.5476	1.0701	3.8952	3.9257	1.0697	3.5416	1.0008
	(0.0022)	(0.0105)	(0.0669)	(0.0028)	(0.0105)	(0.0622)	(0.0665)	(0.0105)	(0.0283)	(0.0043)

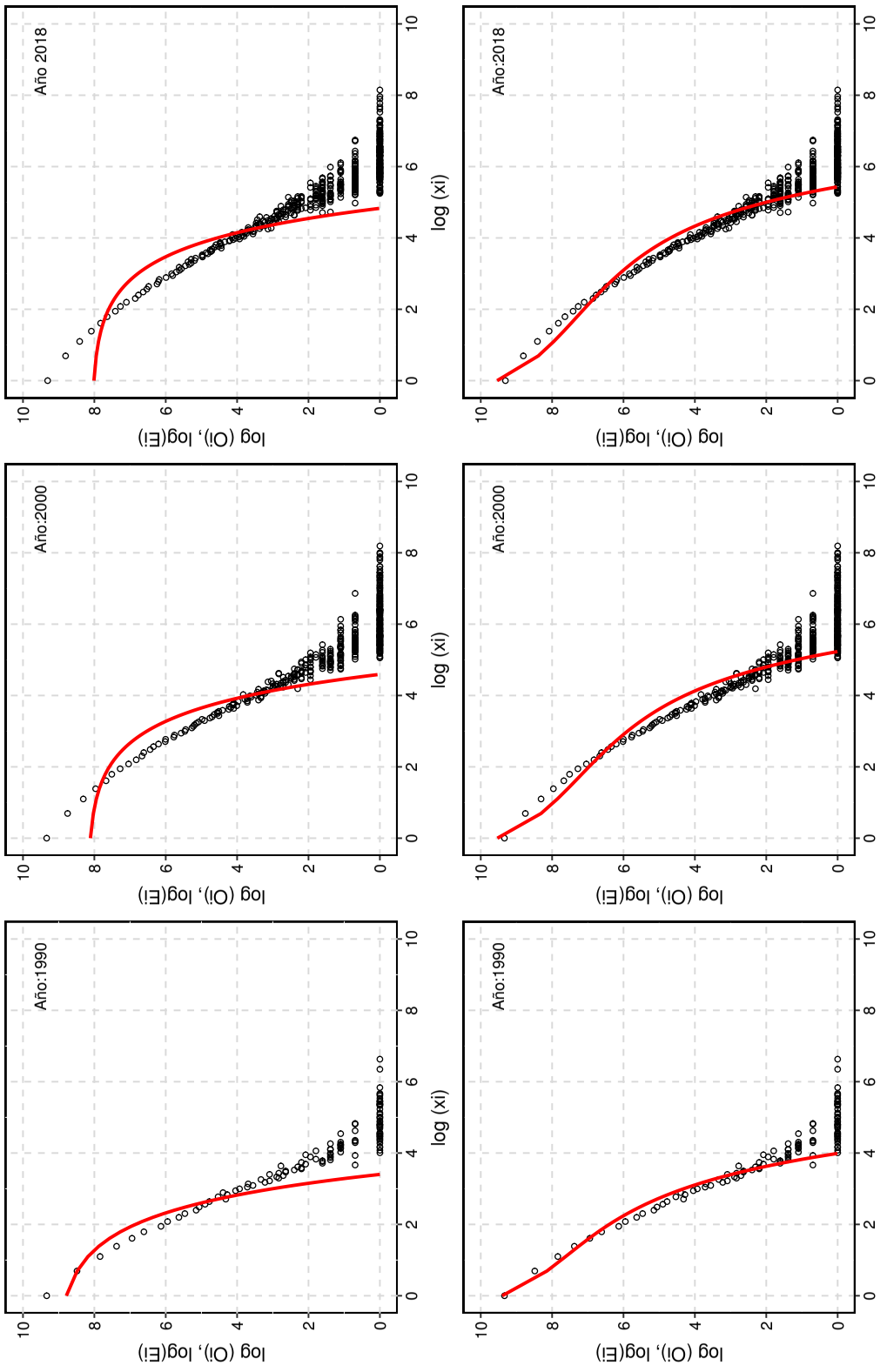

Figura 4.1: Ajuste de los modelos de distribución de probabilidad Geométrica (arriba) y Binomial Negativa Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

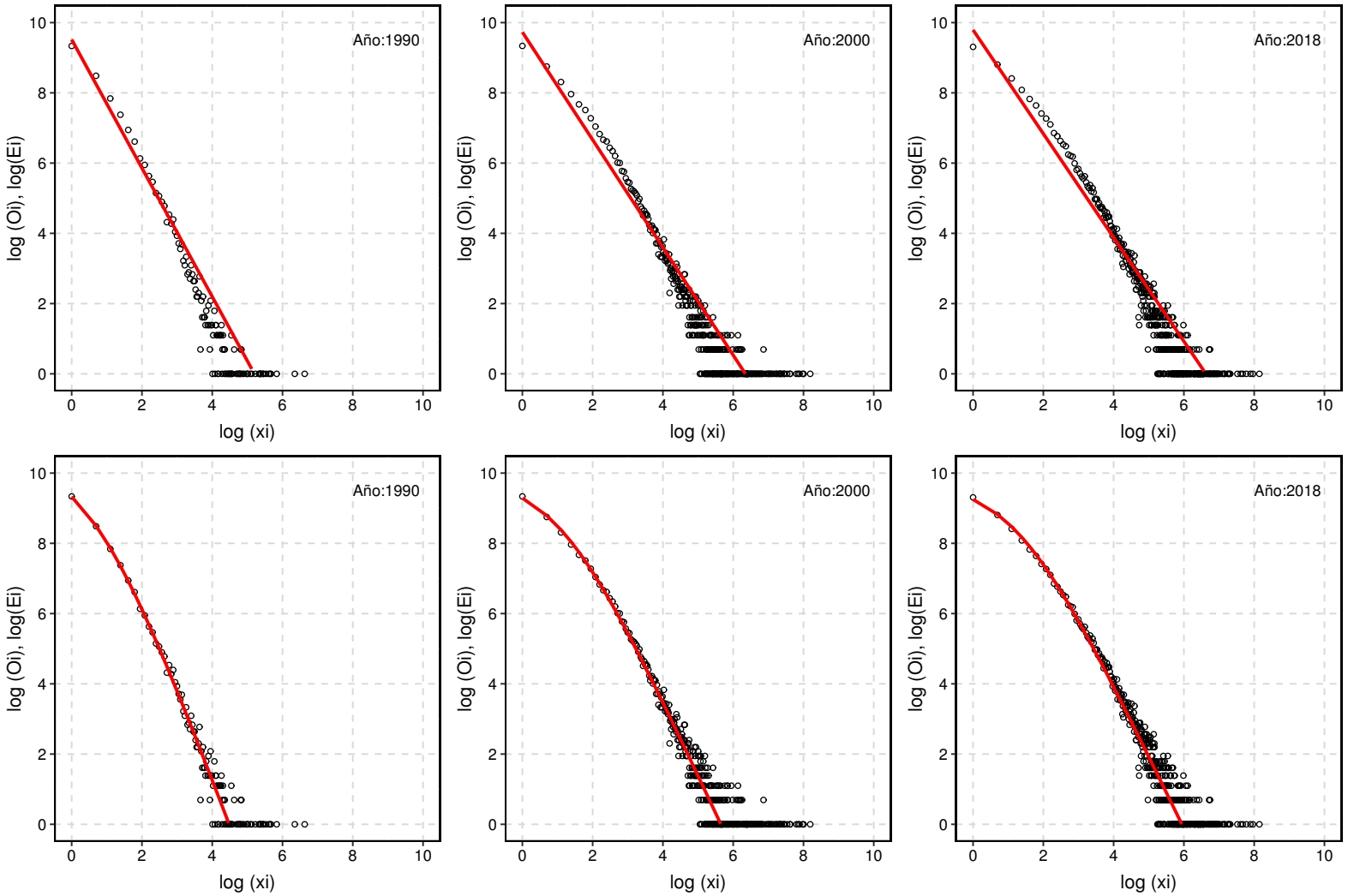

Figura 4.2: Ajuste de los modelos de distribución de probabilidad Zeta (arriba) y Zipf-Mandelbrot (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

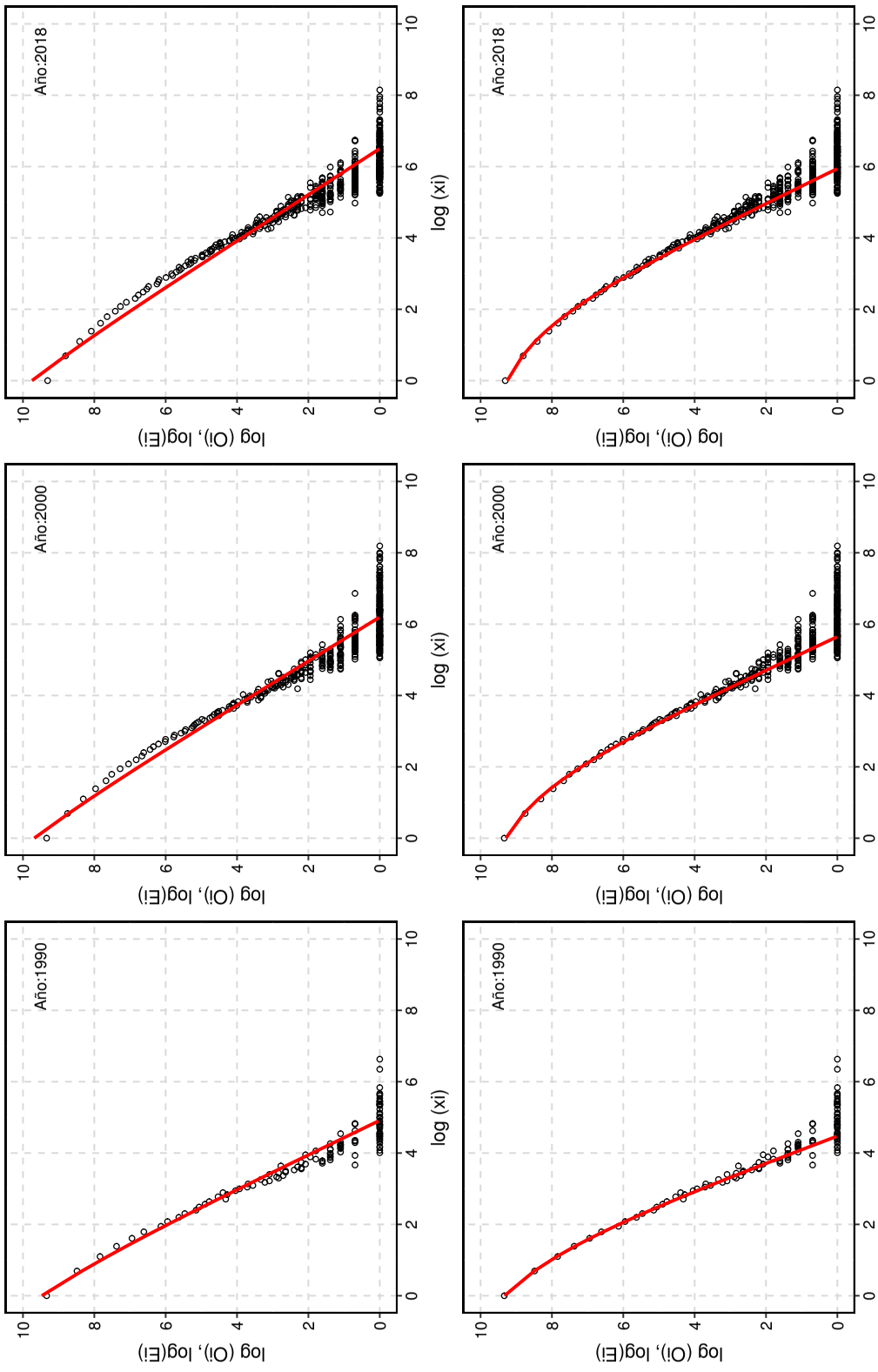

Figura 4.3: Ajuste de los modelos de distribución de probabilidad Yule-Simon (arriba) y Beta-Geométrica (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

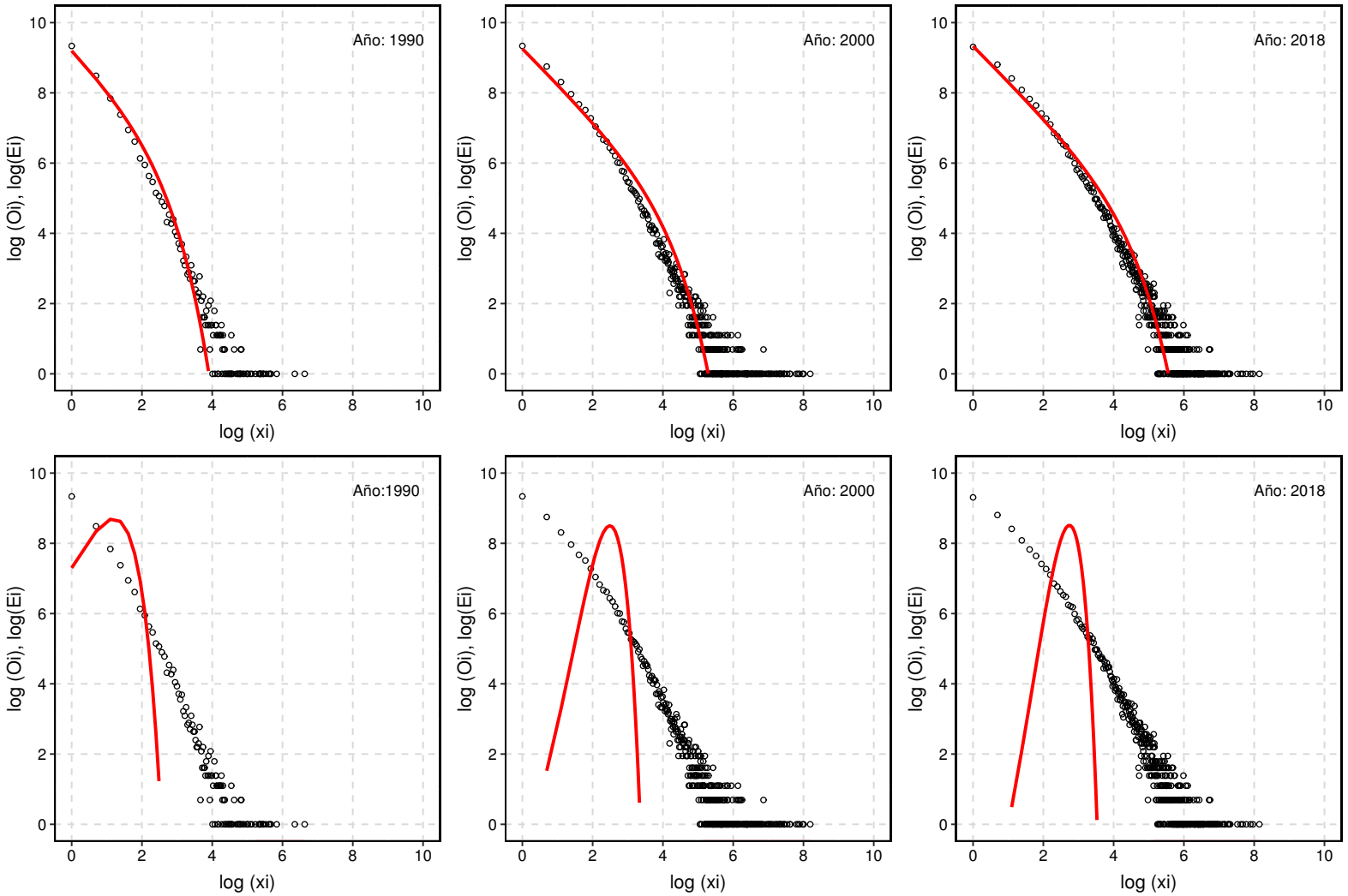

Figura 4.4: Ajuste de los modelos de distribución de probabilidad Logarítmica (arriba) y Poisson Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

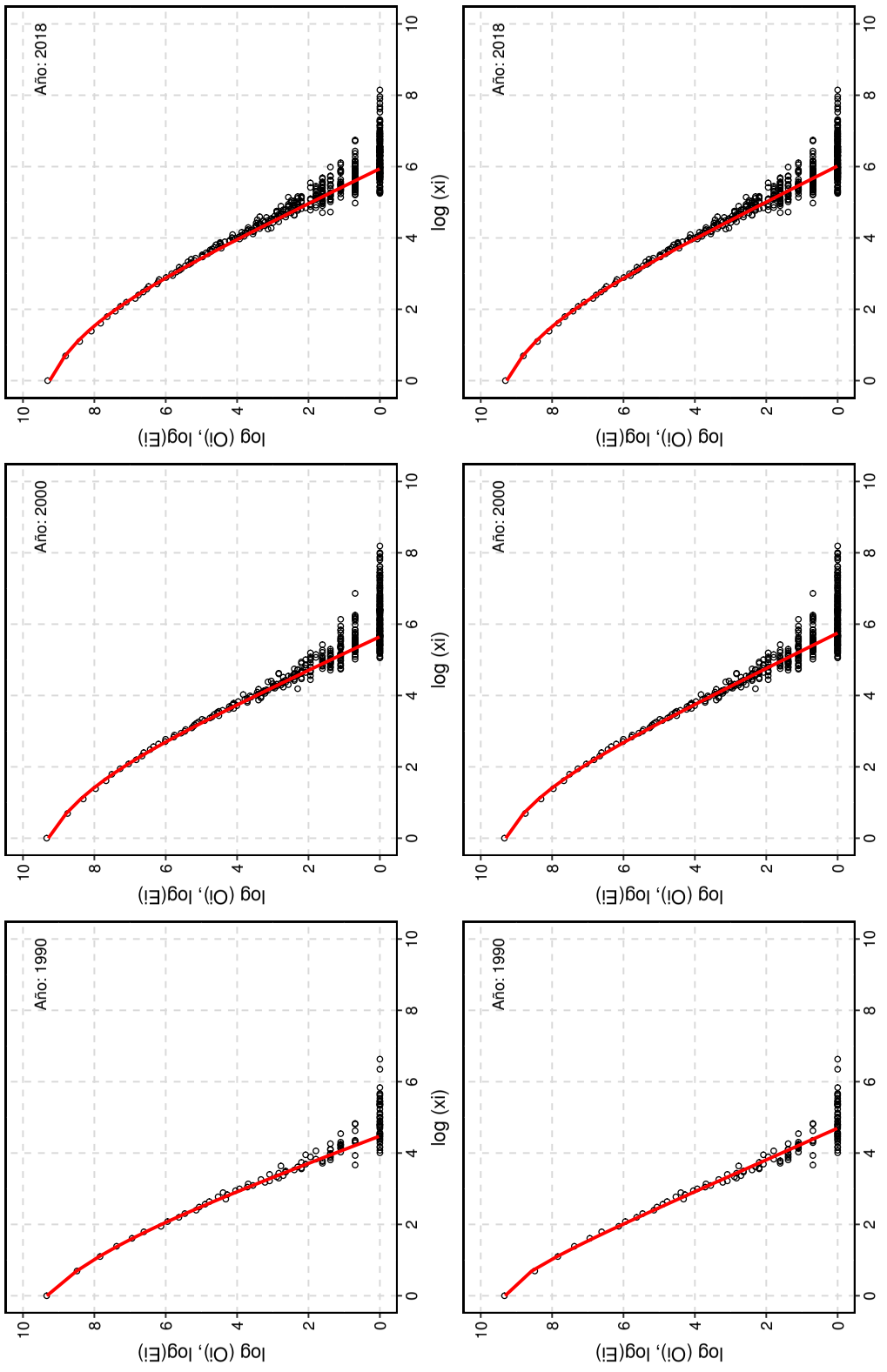

Figura 4.5: Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto II (arriba) y Discreta Generalizada de Pareto III (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

En el cuadro 4.4 se presentan los valores obtenidos mediante máxima verosimilitud, con el Criterio de Información de Akaike (AIC) para los modelos de distribución de probabilidad de 1 y 2 parámetros aplicados a la variable frecuencia de uso de códigos tecnológicos IPC entre 1977 y 2018.

Se puede apreciar que durante los primeros años, los menores valores para el AIC corresponden, salvo ciertas excepciones, a la distribución de Zipf - Mandelbrot. A partir del año 1988, la distribución Beta - Geométrica presenta, de forma consistente, los menores valores AIC (exceptuando los años 1991 y 2018).

En el cuadro 4.5 se presentan los valores obtenidos de $\Delta = AIC - AIC_{min}$ (diferencia entre el valor que toma el estadístico AIC de una distribución específica, con respecto al menor valor de AIC entre las distribuciones consideradas, para cada año del periodo de estudio) para cada una de las distribuciones de probabilidad. Indicar que el valor obtenido de Δ para un modelo y año determinado, puede interpretarse en términos de si es un modelo de los que proporciona el mejor ajuste entre los considerados ($\Delta \leq 2$), si es un modelo con soporte relativamente pequeño ($2 < \Delta \leq 20$), o si es un modelo sin soporte empírico dicho año ($\Delta > 20$).

En el cuadro 4.6 se presenta una clasificación de los modelos analizados con base en el número de años en los que el valor de $\Delta = AIC - AIC_{min}$ obtenidos se encuentra dentro de los tres intervalos anteriormente indicados.

Los resultados obtenidos sugieren que las distribuciones de probabilidad con valores del criterio Δ comprendidos entre 0 y 20, para los 42 años de estudio, fueron los modelos de distribución Beta Geométrica y Pareto Discreta Generalizada tipo II, para todos los años salvo uno el modelo Zipf - Mandelbrot. En contraste, el valor de Δ obtenido en el resto de los modelos analizados, indica que dichos otros modelos no contaron con soporte empírico en la mayoría de los años. Finalmente, puede comprobarse que los mejores ajustes fueron proporcionados por las distribuciones Beta Geométrica y Pareto tipo II, con valores de Δ menores de 2 en la mayoría de dichos años.

Cuadro 4.4: Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977-2018.

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
1977	580	587	589	582	579	580	582	581	581	582	584
1978	5276	5772	5169	5108	5170	5108	5109	5134	5123	5108	5114
1979	9791	11015	9604	9456	9561	9466	9457	9510	9484	9456	9477
1980	12504	13949	12432	12185	12257	12220	12185	12226	12204	12185	12229
1981	15577	18173	15190	14949	15122	14972	14949	15028	14984	14949	15004
1982	17220	20180	16872	16561	16732	16606	16561	16658	16606	16561	16629
1983	20905	25761	20144	19806	20124	19852	19808	20004	19895	19806	19841
1984	33969	49331	31412	30933	31894	31006	30936	31571	31217	30934	30982
1985	39317	55827	37139	36425	37202	36592	36426	37012	36675	36425	36531
1986	46620	69321	43815	43022	43857	43216	43020	43592	43220	43021	43174
1987	60011	98577	55610	54526	55836	54844	54527	55610	54906	54526	54631
1988	70692	124879	65247	63904	65392	64358	63903	65214	64326	63903	64050
1989	92246	184776	84560	82603	84577	83402	82603	84629	83214	82603	82758
1990	109942	249915	99473	97269	99578	98224	97267	99683	97857	97268	97431
1991	120334	281472	109440	106691	109128	108002	106691	109566	107413	106691	106845
1992	136273	349591	123674	120515	122998	122126	120512	123595	121174	120513	120733
1993	150655	420329	135767	132218	135109	134097	132216	135951	132993	132217	132416
1994	166969	497162	151197	146976	149787	149365	146972	150958	147709	146974	147217
1995	184179	595512	166840	161918	164839	164854	161914	166432	162663	161915	162150
1996	204739	752030	183141	177799	181322	181054	177796	183330	178636	177797	177975
1997	220597	885049	196921	191097	194669	194760	191092	196966	191870	191093	191288
1998	245968	1156523	216712	210443	214597	214473	210437	217344	211235	210438	210607
1999	270975	1506090	236881	229664	234451	234506	229660	238043	230580	229660	229794
2000	292070	1833017	254270	246449	251443	251827	246443	255486	247320	246444	246568
2001	299499	1877048	261737	253689	258321	259270	253683	262495	254364	253684	253783
2002	297712	1661225	264389	255980	259833	261845	255974	263769	256486	255975	256114

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
2003	294324	1501737	265280	256423	259661	262655	256417	263472	256808	256418	256599
2004	295408	1399903	268498	259180	262305	265751	259175	266130	259594	259176	259387
2005	276987	1277177	248740	240873	244267	246197	240868	247677	241337	240868	241032
2006	285202	1346810	256923	248614	251872	254326	248609	255388	249028	248610	248795
2007	286952	1359310	259036	250639	253712	256440	250633	257173	250985	250634	250831
2008	285300	1314879	258082	249617	252682	255457	249611	256219	249966	249612	249794
2009	296698	1411404	268839	259821	262869	266134	259815	266637	260136	259816	260000
2010	314262	1624221	283600	274071	277292	280833	274066	281460	274342	274066	274219
2011	325711	1787175	293569	283592	286843	290766	283587	291328	283799	283588	283714
2012	331841	1865951	299600	289292	292433	296775	289287	296985	289451	289288	289416
2013	339747	1942362	308097	297225	300162	305217	297220	304849	297293	297220	297349
2014	343070	2004223	310765	299857	302785	307891	299851	307497	299908	299852	299975
2015	360814	2231835	326477	314817	317959	323499	314812	323045	314889	314812	314924
2016	371918	2383811	336700	324562	327666	333678	324557	332947	324577	324557	324654
2017	365634	2381072	329204	317641	320891	326262	317635	326185	317667	317636	317709
2018	361684	2385261	324070	313077	316333	321207	313072	321565	313055	313072	313119

Cuadro 4.5: $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977-2018.

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
1977	1.10	7.30	9.88	2.23	0.00	0.92	2.17	1.76	1.92	2.23	4.93
1978	168.42	664.36	61.15	0.40	62.25	0.00	1.30	26.24	14.70	0.55	6.33
1979	335.20	1559.21	148.56	0.00	104.76	9.84	1.03	54.04	27.92	0.19	20.86
1980	319.40	1764.48	247.44	0.00	72.31	34.94	0.17	41.78	18.85	0.01	43.85
1981	627.70	3223.85	241.20	0.28	173.31	22.84	0.00	79.28	35.00	0.17	54.63
1982	658.35	3618.63	310.49	0.01	170.29	44.25	0.11	96.55	44.94	0.00	67.79
1983	1099.27	5954.97	338.37	0.00	317.84	45.94	2.06	198.10	89.24	0.54	35.20
1984	3036.05	18397.95	479.14	0.00	960.70	73.41	3.28	637.66	283.98	1.15	49.38
1985	2892.21	19402.46	714.42	0.00	777.10	167.43	1.67	587.04	250.11	0.55	106.33
1986	3600.26	26301.03	794.92	2.13	836.94	195.99	0.00	571.89	200.11	1.19	154.22
1987	5485.78	44050.96	1084.35	0.00	1310.80	318.23	1.14	1083.99	380.14	0.42	105.37
1988	6789.22	60976.43	1343.65	0.55	1488.60	454.82	0.00	1311.19	423.54	0.21	147.52
1989	9643.13	102172.76	1956.50	0.24	1973.99	798.53	0.00	2025.69	611.01	0.05	155.30
1990	12674.44	152647.63	2205.32	2.08	2310.90	956.42	0.00	2415.93	590.09	0.81	163.47
1991	13642.94	174780.68	2748.51	0.10	2436.51	1310.63	0.02	2874.27	721.72	0.00	153.80
1992	15761.47	229079.28	3162.45	3.34	2486.53	1613.75	0.00	3083.32	662.51	1.25	221.24
1993	18439.60	288113.34	3551.14	2.65	2893.25	1880.96	0.00	3735.64	777.15	0.92	199.92
1994	19996.38	350190.27	4225.18	4.16	2815.12	2392.67	0.00	3985.87	737.03	1.41	245.20
1995	22264.94	433598.27	4925.49	4.01	2925.27	2939.55	0.00	4518.19	748.71	1.27	235.38
1996	26943.44	574234.68	5345.58	3.24	3526.16	3257.85	0.00	5533.92	840.00	0.88	179.53
1997	29504.76	693957.13	5828.65	4.40	3577.14	3667.59	0.00	5874.11	777.29	1.08	195.29
1998	35530.19	946085.45	6274.95	5.47	4159.68	4035.56	0.00	6906.60	797.72	1.08	169.38
1999	41315.40	1276430.29	7221.47	4.59	4791.22	4846.56	0.00	8383.34	920.64	0.75	134.66
2000	45627.22	1586574.20	7827.43	5.70	5000.49	5384.64	0.00	9043.65	877.11	0.77	124.76
2001	45816.14	1623365.12	8053.73	5.64	4638.13	5587.03	0.00	8811.80	680.73	0.65	100.49
2002	41737.44	1405250.26	8415.14	5.87	3858.74	5870.60	0.00	7794.41	511.19	0.82	140.12

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
2003	37906.63	1245319.93	8862.71	5.88	3243.79	6237.70	0.00	7054.91	391.20	0.98	181.84
2004	36233.17	1140728.34	9322.76	5.19	3129.81	6576.53	0.00	6954.61	419.21	1.00	212.19
2005	36119.21	1036309.21	7871.97	5.02	3399.17	5329.52	0.00	6809.31	469.27	0.91	164.38
2006	36593.10	1098201.04	8314.48	5.62	3263.09	5717.00	0.00	6779.38	419.82	1.01	186.10
2007	36318.93	1108677.27	8403.02	6.14	3078.64	5807.59	0.00	6540.37	351.64	1.08	197.82
2008	35688.10	1065267.62	8470.67	5.33	3070.89	5845.16	0.00	6607.68	354.89	0.97	182.66
2009	36882.43	1151588.79	9023.27	5.40	3053.45	6318.96	0.00	6821.92	320.79	0.95	184.91
2010	40196.11	1350155.57	9534.05	5.43	3226.40	6767.13	0.00	7394.24	276.90	0.80	153.88
2011	42124.12	1503588.06	9981.97	5.19	3256.25	7179.43	0.00	7741.19	212.01	0.67	126.72
2012	42554.47	1576664.02	10312.75	5.50	3145.77	7488.00	0.00	7698.02	164.13	0.68	129.42
2013	42526.81	1645142.54	10877.50	5.58	2942.07	7997.23	0.00	7628.86	73.23	0.67	129.16
2014	43219.17	1704372.14	10913.79	5.92	2934.20	8039.78	0.00	7645.33	57.15	0.67	123.97
2015	46002.53	1917023.27	11665.37	5.70	3147.02	8687.89	0.00	8233.12	77.09	0.59	112.30
2016	47361.53	2059254.52	12143.09	5.81	3109.30	9121.78	0.00	8389.97	20.22	0.54	96.95
2017	47998.83	2063436.08	11568.68	5.63	3255.14	8626.35	0.00	8550.01	31.12	0.46	73.11
2018	48629.55	2072206.82	11015.42	22.84	3278.43	8152.63	16.97	8510.27	0.00	17.35	64.94

Cuadro 4.6: Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1977 - 2018.

Distribución	$(\triangle \leq 2)$	$(2 < \triangle \leq 20)$	$(\triangle > 20)$
Geométrica	1	0	41
Poisson desplazada	1	0	41
Zeta	1	0	41
Yule-Simon	2	1	39
Logarítmica	1	0	41
Binomial Negativa desplazada	1	0	41
Poisson generalizada desplazada	2	2	38
Zipf-Mandelbrot	12	29	1
Beta Geométrica	38	4	0
Pareto discreta generalizada tipo II	40	2	0
Pareto discreta generalizada tipo III	0	2	40

4.3.2. Distribución de probabilidad de la frecuencia de uso de códigos tecnológicos CPC

Estimación de parámetros y errores estándar

Los cuadros 4.7 y 4.8 muestran los parámetros estimados y sus errores estándar en los modelos de distribución de probabilidad de 1 y 2 parámetros considerados, ajustados a la variable aleatoria: frecuencia de uso de códigos tecnológicos CPC en todo el rango por el método de máxima verosimilitud, para el periodo 1977 y 2018.

En el cuadro 4.7 se muestran los resultados mencionados para las distribuciones: Geométrica (parámetro p), Negativa Binomial Desplazada (parámetros p y r), Poisson Desplazada (parámetro λ), Poisson Generalizada Desplazada (parámetros λ_1 y λ_2) y Logarítmica (parámetro θ).

Cuadro 4.7: Parámetros estimados y errores estándar para los modelos de distribución de probabilidad considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018.

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
1977	0.9525	0.7898	0.1873	0.0499	0.0443	0.1125	0.0920
	(0.0064)	(0.0859)	(0.0924)	(0.0070)	(0.0066)	(0.0510)	(0.0121)
1978	0.8445	0.4350	0.1418	0.1841	0.1181	0.3586	0.2803
	(0.0051)	(0.0247)	(0.0120)	(0.0069)	(0.0054)	(0.0224)	(0.0084)
1979	0.8485	0.5469	0.2156	0.1786	0.1301	0.2714	0.2740
	(0.0040)	(0.0209)	(0.0161)	(0.0054)	(0.0045)	(0.0161)	(0.0073)
1980	0.8375	0.5464	0.2337	0.1940	0.1412	0.2721	0.2916
	(0.0038)	(0.0185)	(0.0155)	(0.0052)	(0.0043)	(0.0142)	(0.0069)
1981	0.8205	0.5428	0.2597	0.2188	0.1587	0.2745	0.3185
	(0.0035)	(0.0160)	(0.0149)	(0.0050)	(0.0042)	(0.0124)	(0.0066)
1982	0.8135	0.5065	0.2353	0.2293	0.1601	0.3018	0.3294
	(0.0034)	(0.0147)	(0.0121)	(0.0049)	(0.0040)	(0.0119)	(0.0062)
1983	0.7921	0.4887	0.2509	0.2625	0.1798	0.3149	0.3620
	(0.0033)	(0.0131)	(0.0115)	(0.0050)	(0.0040)	(0.0109)	(0.0061)
1984	0.7416	0.4538	0.2894	0.3484	0.2291	0.3425	0.4349
	(0.0029)	(0.0097)	(0.0098)	(0.0050)	(0.0039)	(0.0084)	(0.0056)
1985	0.7146	0.4184	0.2873	0.3993	0.2508	0.3719	0.4717
	(0.0028)	(0.0083)	(0.0084)	(0.0051)	(0.0039)	(0.0077)	(0.0053)
1986	0.6702	0.3394	0.2528	0.4920	0.2742	0.4427	0.5292
	(0.0026)	(0.0066)	(0.0061)	(0.0054)	(0.0038)	(0.0070)	(0.0049)
1987	0.6347	0.3208	0.2718	0.5756	0.3107	0.4602	0.5723
	(0.0024)	(0.0053)	(0.0054)	(0.0053)	(0.0036)	(0.0059)	(0.0045)
1988	0.5822	0.2752	0.2724	0.7175	0.3543	0.5062	0.6318
	(0.0021)	(0.0042)	(0.0045)	(0.0056)	(0.0036)	(0.0051)	(0.0041)
1989	0.5301	0.2560	0.3049	0.8865	0.4198	0.5264	0.6861
	(0.0018)	(0.0033)	(0.0042)	(0.0057)	(0.0035)	(0.0042)	(0.0039)
1990	0.4708	0.2103	0.2994	1.1242	0.4738	0.5786	0.7423
	(0.0016)	(0.0025)	(0.0035)	(0.0063)	(0.0036)	(0.0036)	(0.0036)
1991	0.4352	0.1875	0.2995	1.2980	0.5103	0.6069	0.7733
	(0.0015)	(0.0021)	(0.0031)	(0.0066)	(0.0035)	(0.0033)	(0.0034)
1992	0.3929	0.1655	0.3066	1.5455	0.5638	0.6352	0.8074
	(0.0013)	(0.0017)	(0.0029)	(0.0070)	(0.0036)	(0.0029)	(0.0032)
1993	0.3682	0.1547	0.3139	1.7159	0.6009	0.6498	0.8261
	(0.0012)	(0.0015)	(0.0027)	(0.0072)	(0.0035)	(0.0027)	(0.0031)
1994	0.3380	0.1407	0.3207	1.9586	0.6479	0.6692	0.8477
	(0.0011)	(0.0013)	(0.0025)	(0.0076)	(0.0035)	(0.0024)	(0.0030)
1995	0.3081	0.1277	0.3289	2.2459	0.7008	0.6880	0.8679
	(0.0009)	(0.0011)	(0.0024)	(0.0081)	(0.0035)	(0.0022)	(0.0028)

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
1996	0.2787	0.1114	0.3244	2.5880	0.7412	0.7135	0.8864
	(0.0008)	(0.0009)	(0.0022)	(0.0086)	(0.0035)	(0.0020)	(0.0027)
1997	0.2568	0.1028	0.3314	2.8933	0.7891	0.7273	0.8994
	(0.0008)	(0.0008)	(0.0021)	(0.0091)	(0.0035)	(0.0018)	(0.0026)
1998	0.2357	0.0917	0.3275	3.2434	0.8235	0.7461	0.9114
	(0.0007)	(0.0007)	(0.0020)	(0.0097)	(0.0035)	(0.0017)	(0.0025)
1999	0.1965	0.0717	0.3158	4.0888	0.8868	0.7831	0.9318
	(0.0006)	(0.0005)	(0.0017)	(0.0113)	(0.0036)	(0.0015)	(0.0024)
2000	0.1718	0.0610	0.3132	4.8196	0.9430	0.8043	0.9436
	(0.0005)	(0.0004)	(0.0016)	(0.0126)	(0.0036)	(0.0013)	(0.0023)
2001	0.1647	0.0603	0.3251	5.0699	0.9854	0.8056	0.9468
	(0.0005)	(0.0004)	(0.0016)	(0.0130)	(0.0037)	(0.0013)	(0.0023)
2002	0.1639	0.0606	0.3291	5.1022	0.9967	0.8046	0.9472
	(0.0005)	(0.0004)	(0.0016)	(0.0128)	(0.0036)	(0.0013)	(0.0022)
2003	0.1588	0.0588	0.3310	5.2972	1.0159	0.8082	0.9495
	(0.0004)	(0.0004)	(0.0016)	(0.0130)	(0.0036)	(0.0012)	(0.0022)
2004	0.1498	0.0570	0.3433	5.6758	1.0696	0.8116	0.9535
	(0.0004)	(0.0004)	(0.0016)	(0.0135)	(0.0037)	(0.0012)	(0.0022)
2005	0.1371	0.0510	0.3383	6.2915	1.1027	0.8247	0.9588
	(0.0004)	(0.0003)	(0.0015)	(0.0145)	(0.0037)	(0.0011)	(0.0021)
2006	0.1339	0.0509	0.3464	6.4659	1.1328	0.8248	0.9601
	(0.0004)	(0.0003)	(0.0015)	(0.0147)	(0.0037)	(0.0011)	(0.0021)
2007	0.1293	0.0489	0.3462	6.7366	1.1510	0.8291	0.9620
	(0.0003)	(0.0003)	(0.0015)	(0.0152)	(0.0037)	(0.0010)	(0.0021)
2008	0.1331	0.0508	0.3486	6.5108	1.1395	0.8250	0.9604
	(0.0003)	(0.0003)	(0.0015)	(0.0147)	(0.0037)	(0.0011)	(0.0021)
2009	0.1276	0.0490	0.3525	6.8381	1.1707	0.8288	0.9626
	(0.0003)	(0.0003)	(0.0015)	(0.0152)	(0.0037)	(0.0010)	(0.0021)
2010	0.1170	0.0453	0.3583	7.5501	1.2292	0.8372	0.9668
	(0.0003)	(0.0003)	(0.0015)	(0.0162)	(0.0038)	(0.0009)	(0.0020)
2011	0.1066	0.0414	0.3616	8.3782	1.2874	0.8463	0.9707
	(0.0003)	(0.0002)	(0.0014)	(0.0175)	(0.0038)	(0.0009)	(0.0020)
2012	0.0985	0.0380	0.3616	9.1545	1.3309	0.8546	0.9737
	(0.0002)	(0.0002)	(0.0014)	(0.0187)	(0.0039)	(0.0008)	(0.0019)
2013	0.0925	0.0353	0.3588	9.8096	1.3582	0.8615	0.9757
	(0.0002)	(0.0002)	(0.0014)	(0.0199)	(0.0039)	(0.0008)	(0.0019)
2014	0.0892	0.0339	0.3583	10.2092	1.3764	0.8652	0.9768
	(0.0002)	(0.0002)	(0.0014)	(0.0205)	(0.0039)	(0.0008)	(0.0019)
2015	0.0847	0.0317	0.3534	10.8035	1.3913	0.8712	0.9784
	(0.0002)	(0.0002)	(0.0013)	(0.0216)	(0.0040)	(0.0008)	(0.0018)

Año	Modelos de distribución de probabilidad						
	$G(p)$	$BN(r,p)$		$Po(\lambda)$	$Po(\lambda_1,\lambda_2)$		$Log(\theta)$
	$\hat{p}$	$\hat{p}$	$\hat{r}$	$\hat{\lambda}$	$\hat{\lambda_1}$	$\hat{\lambda_2}$	$\hat{\theta}$
2016	0.0792	0.0291	0.3486	11.6219	1.4138	0.8784	0.9802
	(0.0002)	(0.0002)	(0.0013)	(0.0230)	(0.0040)	(0.0007)	(0.0018)
2017	0.0799	0.0290	0.3443	11.5181	1.3968	0.8787	0.9800
	(0.0002)	(0.0002)	(0.0013)	(0.0231)	(0.0040)	(0.0008)	(0.0019)
2018	0.0896	0.0312	0.3273	10.1640	1.2848	0.8736	0.9767
	(0.0002)	(0.0002)	(0.0013)	(0.0214)	(0.0039)	(0.0008)	(0.0019)

A continuación, en el cuadro 4.8 se muestran los parámetros estimados y sus errores estándar para las distribuciónes de probabilidad: Zeta (parámetro α), Zipf- Mandelbrot (parámetros α y β), Yule-Simon (parámetro ρ), Beta - Geométrica (parámetros α y β), Discreta Generalizada de Pareto II (parámetros α y σ) y Discreta Generalizada de Pareto III (parámetros γ y σ).

Las figuras 4.6, 4.7, 4.8, 4.9, 4.10 muestran, a modo de validación gráfica del ajuste obtenido por cada uno de los modelos, el valor de las frecuencias observadas (en negro) y esperadas (en rojo) para cada uno de los valores de nuestra muestra, en escala log-log, para los años 1990, 2000 y 2018, de los diferentes modelos indicados: distribuciones geométrica y binomial negativa desplazada (figura 4.6); distribuciones zeta y Zipf-Mandelbrot (figura 4.7); distribuciones Yule-Simon y Beta Geométrica (figura 4.8); distribuciones logarítmica y Poisson desplazada (figura 4.9); distribuciones discreta generalizada de Pareto tipos II y III (figura 4.10).

Cuadro 4.8: Parámetros estimados y errores estándar para los modelos de distribución de probabilidad considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977 - 2018.

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
1977	3.7912	6.5347	0.8811	21.0838	10.8671	0.4929	1.7295	6.8677	0.3657	0.3247
	(0.1899)	(3.6800)	(1.1870)	(2.9993)	(6.6800)	(0.3212)	(1.1790)	(3.6874)	(0.0662)	(0.0559)
1978	2.5043	2.5043	0.0000	6.6434	3.0918	0.3898	0.7961	2.6895	0.3408	0.5202
	(0.0498)	(0.0000)	(0.0000)	(0.2625)	(0.2982)	(0.0442)	(0.1079)	(0.2349)	(0.0195)	(0.0242)
1979	2.4722	3.7546	0.6408	6.7135	4.6842	0.6535	1.4231	3.9448	0.4206	0.4413
	(0.0388)	(0.3370)	(0.1680)	(0.2126)	(0.4366)	(0.0691)	(0.1664)	(0.3400)	(0.0151)	(0.0158)
1980	2.3900	3.8591	0.7620	6.2526	4.7710	0.7257	1.5398	4.0458	0.4399	0.4385
	(0.0341)	(0.3132)	(0.1622)	(0.1762)	(0.4009)	(0.0692)	(0.1610)	(0.3160)	(0.0134)	(0.0139)
1981	2.2730	4.0738	0.9877	5.6488	4.9427	0.8531	1.7551	4.2481	0.4625	0.4419
	(0.0290)	(0.2996)	(0.1646)	(0.1384)	(0.3734)	(0.0731)	(0.1635)	(0.3020)	(0.0116)	(0.0120)
1982	2.2486	3.6299	0.7664	5.4744	4.3689	0.7618	1.5328	3.7972	0.4518	0.4580
	(0.0272)	(0.2332)	(0.1291)	(0.1253)	(0.2883)	(0.0576)	(0.1279)	(0.2351)	(0.0110)	(0.0117)
1983	2.1268	3.6402	0.8943	4.8986	4.2908	0.8512	1.6470	3.7885	0.4661	0.4749
	(0.0240)	(0.2162)	(0.1278)	(0.1013)	(0.2596)	(0.0593)	(0.1264)	(0.2175)	(0.0101)	(0.0108)
1984	1.8887	3.6644	1.2037	3.9115	4.1760	1.0845	1.9341	3.7871	0.5178	0.4940
	(0.0175)	(0.1716)	(0.1167)	(0.0623)	(0.1974)	(0.0596)	(0.1156)	(0.1723)	(0.0080)	(0.0084)
1985	1.7910	3.3327	1.1125	3.5302	3.7445	1.0773	1.8346	3.4432	0.5306	0.5162
	(0.0154)	(0.1335)	(0.0965)	(0.0504)	(0.1512)	(0.0512)	(0.0955)	(0.1340)	(0.0074)	(0.0078)
1986	1.6729	2.6838	0.7911	3.0817	2.9555	0.9465	1.5026	2.7746	0.5269	0.5703
	(0.0132)	(0.0857)	(0.0667)	(0.0388)	(0.0947)	(0.0368)	(0.0655)	(0.0859)	(0.0070)	(0.0076)
1987	1.5647	2.7465	1.0046	2.7637	2.9864	1.1082	1.7029	2.8269	0.5711	0.5764
	(0.0109)	(0.0764)	(0.0650)	(0.0297)	(0.0831)	(0.0375)	(0.0638)	(0.0765)	(0.0062)	(0.0064)
1988	1.4347	2.4784	0.9897	2.3866	2.6524	1.1550	1.6732	2.5436	0.6062	0.6159
	(0.0090)	(0.0578)	(0.0547)	(0.0222)	(0.0617)	(0.0333)	(0.0537)	(0.0578)	(0.0058)	(0.0059)
1989	1.3021	2.5465	1.3408	2.0630	2.6885	1.4421	2.0063	2.6001	0.6894	0.6242
	(0.0072)	(0.0516)	(0.0560)	(0.0163)	(0.0543)	(0.0363)	(0.0552)	(0.0516)	(0.0053)	(0.0049)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
1990	1.1913	2.2847	1.3281	1.7966	2.3849	1.4983	1.9796	2.3270	0.7478	0.6634
	(0.0061)	(0.0397)	(0.0486)	(0.0127)	(0.0412)	(0.0330)	(0.0478)	(0.0396)	(0.0053)	(0.0046)
1991	1.1260	2.1259	1.3126	1.6499	2.2062	1.5278	1.9560	2.1623	0.7903	0.6895
	(0.0054)	(0.0331)	(0.0439)	(0.0107)	(0.0342)	(0.0307)	(0.0431)	(0.0331)	(0.0054)	(0.0044)
1992	1.0485	2.0355	1.4348	1.4903	2.0983	1.6637	2.0663	2.0652	0.8635	0.7131
	(0.0048)	(0.0288)	(0.0424)	(0.0089)	(0.0295)	(0.0307)	(0.0417)	(0.0287)	(0.0055)	(0.0041)
1993	1.0042	2.0202	1.5741	1.4051	2.0752	1.7961	2.1987	2.0467	0.9211	0.7201
	(0.0043)	(0.0267)	(0.0421)	(0.0078)	(0.0273)	(0.0312)	(0.0415)	(0.0267)	(0.0055)	(0.0038)
1994	0.9497	1.9511	1.6880	1.3008	1.9958	1.9220	2.3000	1.9713	0.9927	0.7404
	(0.0039)	(0.0240)	(0.0414)	(0.0068)	(0.0244)	(0.0314)	(0.0407)	(0.0239)	(0.0056)	(0.0036)
1995	0.8960	1.8913	1.8358	1.2027	1.9271	2.0785	2.4416	1.9094	1.0808	0.7590
	(0.0035)	(0.0216)	(0.0410)	(0.0058)	(0.0219)	(0.0319)	(0.0404)	(0.0216)	(0.0059)	(0.0034)
1996	0.8544	1.7912	1.8594	1.1277	1.8207	2.1341	2.4601	1.8070	1.1569	0.7767
	(0.0032)	(0.0189)	(0.0386)	(0.0052)	(0.0191)	(0.0306)	(0.0380)	(0.0188)	(0.0061)	(0.0033)
1997	0.8151	1.7602	2.0261	1.0608	1.7847	2.3016	2.6188	1.7734	1.2486	0.7910
	(0.0029)	(0.0176)	(0.0392)	(0.0046)	(0.0178)	(0.0316)	(0.0387)	(0.0176)	(0.0063)	(0.0031)
1998	0.7868	1.7167	2.1124	1.0132	1.7379	2.3992	2.7009	1.7285	1.3242	0.8002
	(0.0027)	(0.0165)	(0.0389)	(0.0043)	(0.0166)	(0.0317)	(0.0384)	(0.0164)	(0.0066)	(0.0030)
1999	0.7333	1.5651	2.1204	0.9237	1.5803	2.4601	2.7011	1.5742	1.4750	0.8352
	(0.0024)	(0.0136)	(0.0362)	(0.0037)	(0.0137)	(0.0302)	(0.0357)	(0.0136)	(0.0073)	(0.0030)
2000	0.6946	1.4922	2.2333	0.8620	1.5041	2.5949	2.8071	1.4995	1.6192	0.8592
	(0.0022)	(0.0122)	(0.0359)	(0.0032)	(0.0122)	(0.0305)	(0.0354)	(0.0122)	(0.0079)	(0.0029)
2001	0.6747	1.4950	2.4230	0.8323	1.5057	2.7781	2.9932	1.5016	1.7210	0.8593
	(0.0021)	(0.0120)	(0.0375)	(0.0031)	(0.0121)	(0.0322)	(0.0371)	(0.0120)	(0.0082)	(0.0028)
2002	0.6710	1.5208	2.5398	0.8276	1.5314	2.8816	3.1085	1.5273	1.7483	0.8551
	(0.0021)	(0.0121)	(0.0383)	(0.0030)	(0.0121)	(0.0330)	(0.0379)	(0.0121)	(0.0081)	(0.0027)
2003	0.6610	1.5129	2.6185	0.8123	1.5228	2.9611	3.1853	1.5189	1.8003	0.8602
	(0.0020)	(0.0118)	(0.0385)	(0.0029)	(0.0118)	(0.0333)	(0.0381)	(0.0118)	(0.0082)	(0.0027)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
2004	0.6386	1.5226	2.9023	0.7797	1.5312	3.2338	3.4643	1.5278	1.9406	0.8596
	(0.0019)	(0.0116)	(0.0409)	(0.0027)	(0.0117)	(0.0357)	(0.0406)	(0.0116)	(0.0087)	(0.0026)
2005	0.6184	1.4435	2.8691	0.7483	1.4506	3.2316	3.4280	1.4479	2.0513	0.8818
	(0.0018)	(0.0105)	(0.0394)	(0.0025)	(0.0106)	(0.0347)	(0.0390)	(0.0105)	(0.0092)	(0.0026)
2006	0.6078	1.4549	3.0523	0.7332	1.4616	3.4063	3.6086	1.4590	2.1350	0.8818
	(0.0018)	(0.0106)	(0.0411)	(0.0024)	(0.0106)	(0.0364)	(0.0408)	(0.0106)	(0.0095)	(0.0025)
2007	0.5998	1.4470	3.1330	0.7215	1.4532	3.4883	3.6876	1.4508	2.1952	0.8856
	(0.0017)	(0.0104)	(0.0417)	(0.0024)	(0.0104)	(0.0370)	(0.0414)	(0.0104)	(0.0097)	(0.0025)
2008	0.6051	1.4447	3.0490	0.7292	1.4513	3.4075	3.6052	1.4488	2.1529	0.8828
	(0.0017)	(0.0104)	(0.0407)	(0.0024)	(0.0104)	(0.0361)	(0.0404)	(0.0104)	(0.0095)	(0.0025)
2009	0.5931	1.4387	3.1943	0.7120	1.4446	3.5522	3.7483	1.4424	2.2494	0.8857
	(0.0017)	(0.0102)	(0.0417)	(0.0023)	(0.0102)	(0.0371)	(0.0414)	(0.0102)	(0.0098)	(0.0025)
2010	0.5699	1.3970	3.3793	0.6783	1.4018	3.7505	3.9290	1.4000	2.4452	0.9002
	(0.0016)	(0.0095)	(0.0423)	(0.0021)	(0.0095)	(0.0380)	(0.0420)	(0.0095)	(0.0105)	(0.0024)
2011	0.5489	1.3622	3.5832	0.6485	1.3663	3.9655	4.1294	1.3648	2.6549	0.9108
	(0.0015)	(0.0090)	(0.0435)	(0.0020)	(0.0090)	(0.0395)	(0.0433)	(0.0090)	(0.0114)	(0.0024)
2012	0.5328	1.3228	3.6965	0.6256	1.3262	4.0935	4.2399	1.3250	2.8314	0.9249
	(0.0014)	(0.0085)	(0.0438)	(0.0019)	(0.0085)	(0.0400)	(0.0435)	(0.0085)	(0.0121)	(0.0024)
2013	0.5221	1.2877	3.7287	0.6104	1.2907	4.1396	4.2703	1.2896	2.9558	0.9362
	(0.0014)	(0.0081)	(0.0437)	(0.0018)	(0.0081)	(0.0401)	(0.0435)	(0.0081)	(0.0128)	(0.0024)
2014	0.5157	1.2679	3.7550	0.6014	1.2707	4.1740	4.2958	1.2697	3.0367	0.9422
	(0.0014)	(0.0079)	(0.0437)	(0.0018)	(0.0079)	(0.0402)	(0.0435)	(0.0079)	(0.0131)	(0.0024)
2015	0.5091	1.2339	3.7054	0.5920	1.2366	4.1395	4.2459	1.2357	3.1168	0.9507
	(0.0013)	(0.0075)	(0.0428)	(0.0017)	(0.0075)	(0.0394)	(0.0425)	(0.0075)	(0.0136)	(0.0024)
2016	0.5001	1.1984	3.6957	0.5793	1.2008	4.1449	4.2352	1.2000	3.2373	0.9622
	(0.0013)	(0.0072)	(0.0423)	(0.0017)	(0.0072)	(0.0390)	(0.0420)	(0.0072)	(0.0142)	(0.0024)
2017	0.5044	1.1975	3.5950	0.5852	1.2000	4.0460	4.1359	1.1993	3.1676	0.9588
	(0.0013)	(0.0072)	(0.0415)	(0.0017)	(0.0072)	(0.0382)	(0.0412)	(0.0072)	(0.0140)	(0.0024)

Año	Modelos de distribución de probabilidad									
	$Zeta(\alpha)$	$ZM(\alpha,\beta)$		$Yule(\rho)$	$BG(\alpha,\beta)$		$DGPII(\sigma,\alpha)$		$DGPIII(\sigma,\gamma)$	
	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\gamma}$
2018	0.5345	1.1894	2.9867	0.6255	1.1928	3.4497	3.5349	1.1919	2.7360	0.9545
	(0.0015)	(0.0074)	(0.0368)	(0.0019)	(0.0074)	(0.0335)	(0.0365)	(0.0074)	(0.0126)	(0.0025)

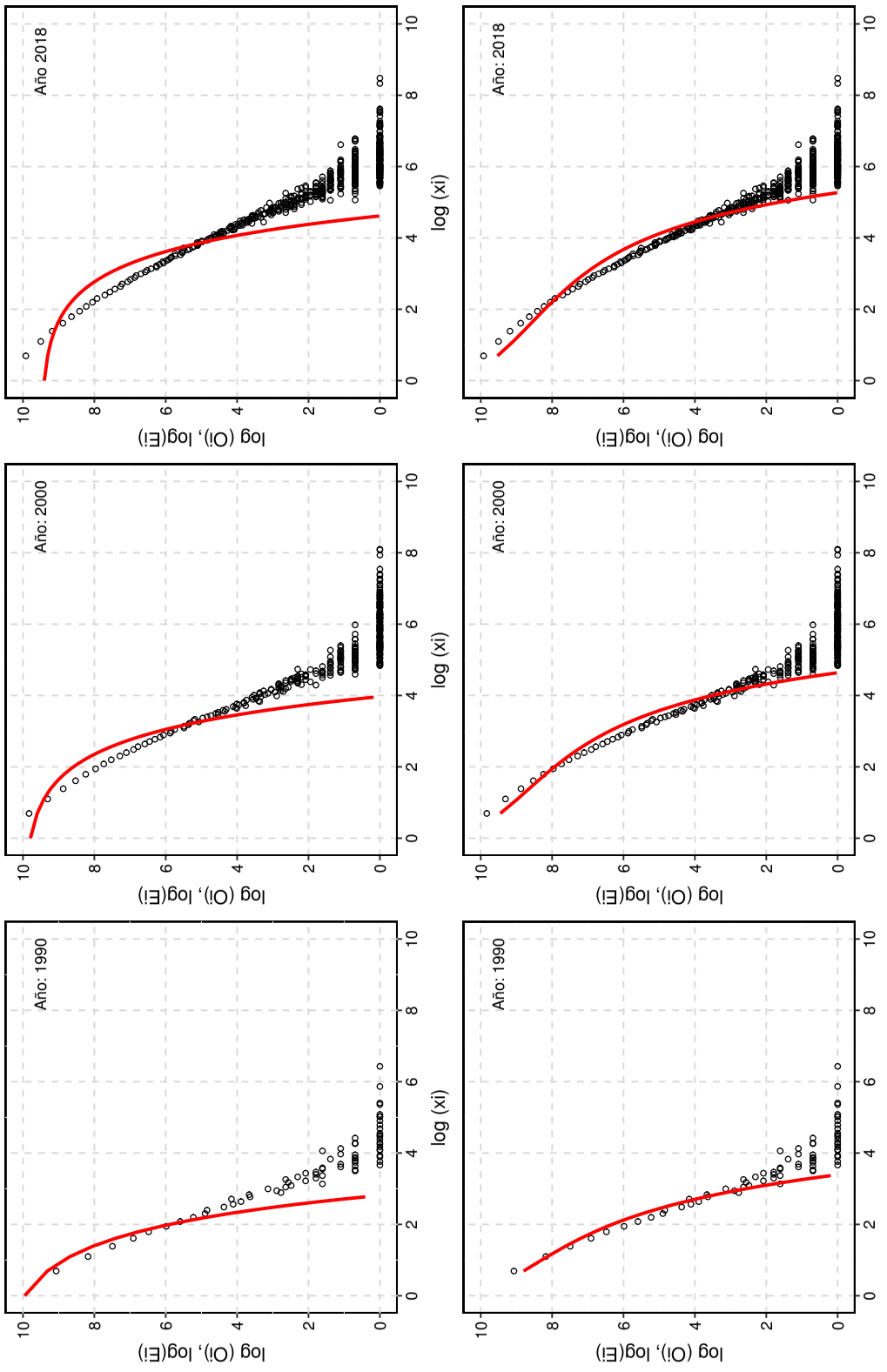

Figura 4.6: Ajuste de los modelos de distribución de probabilidad Geométrica (arriba) y Binomial Negativa Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018

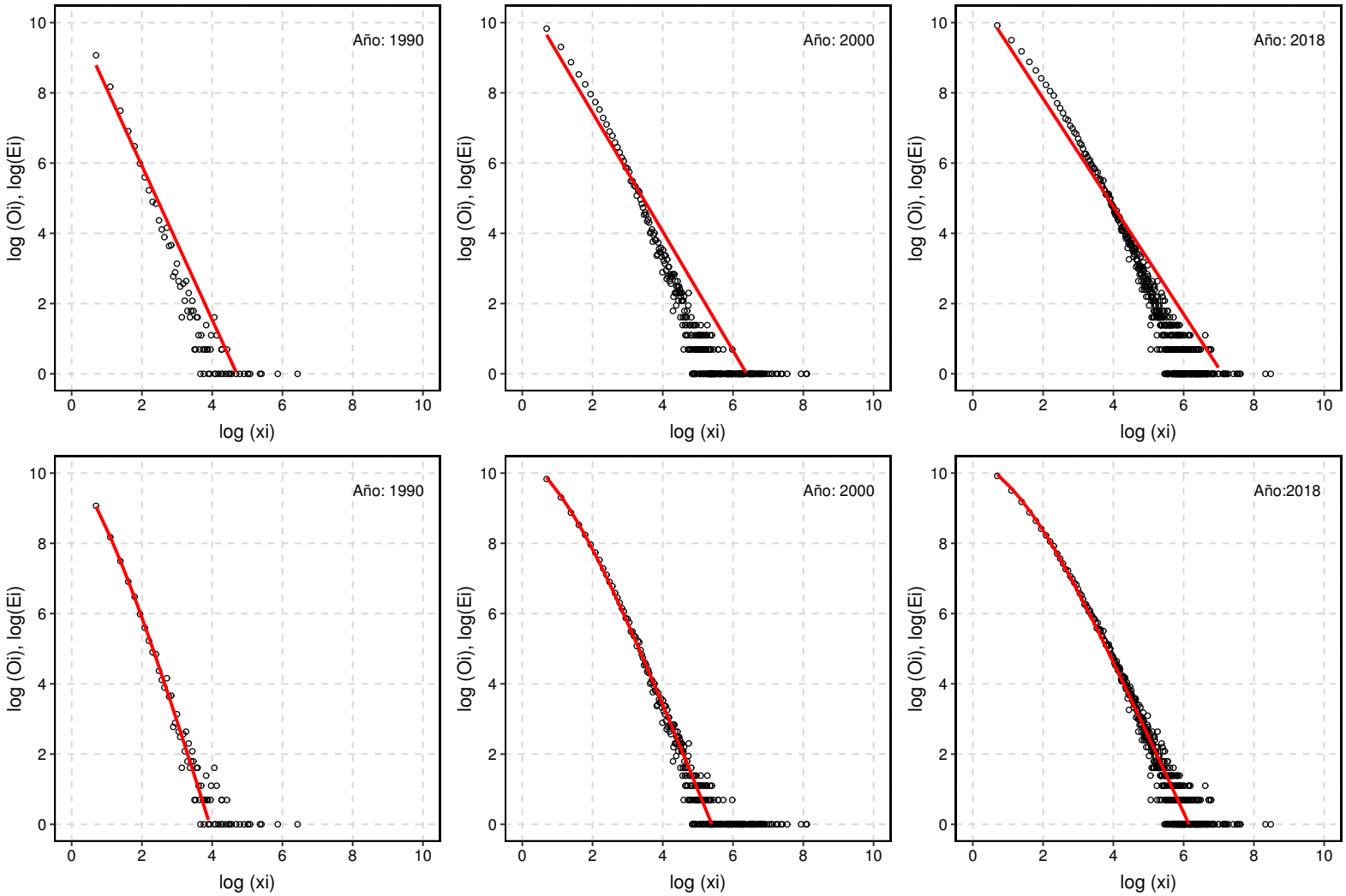

Figura 4.7: Ajuste de los modelos de distribución de probabilidad Zeta (arriba) y Zipf-Mandelbrot (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018.

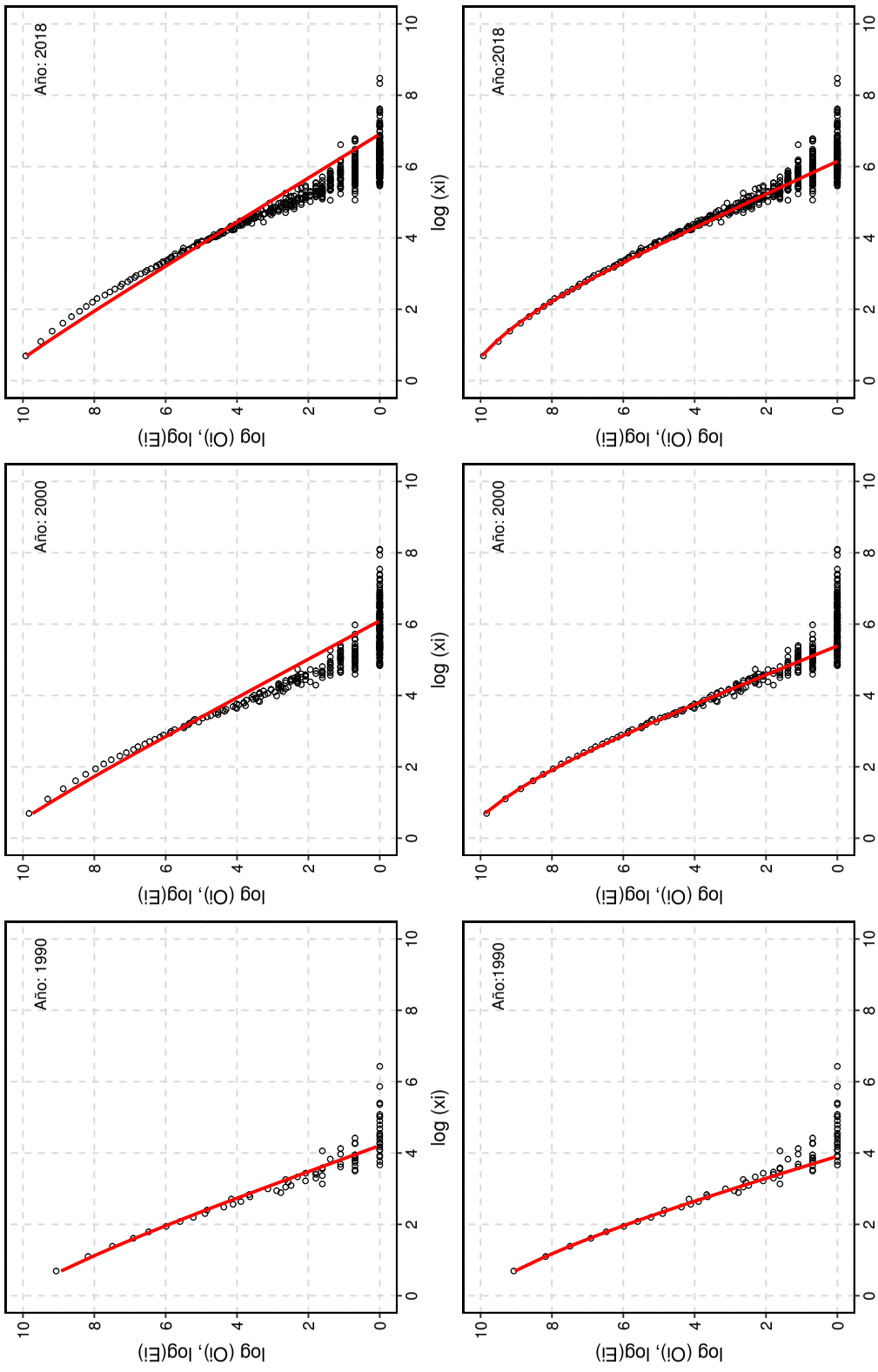

Figura 4.8: Ajuste de los modelos de distribución de probabilidad Yule-Simon (arriba) y Beta-Geométrica (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018.

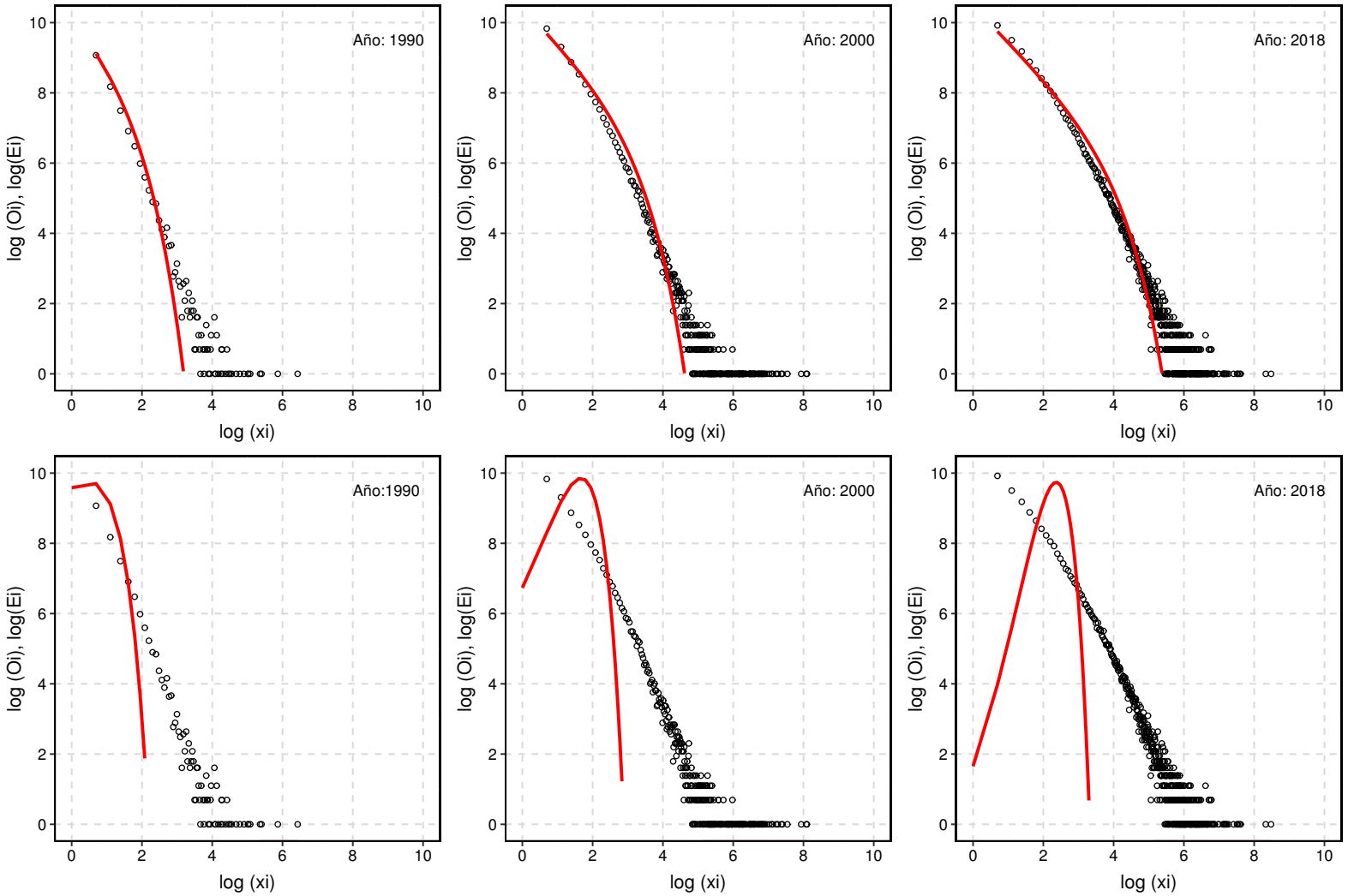

Figura 4.9: Ajuste de los modelos de distribución de probabilidad Logarítmica (arriba) y Poisson Desplazada (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018.

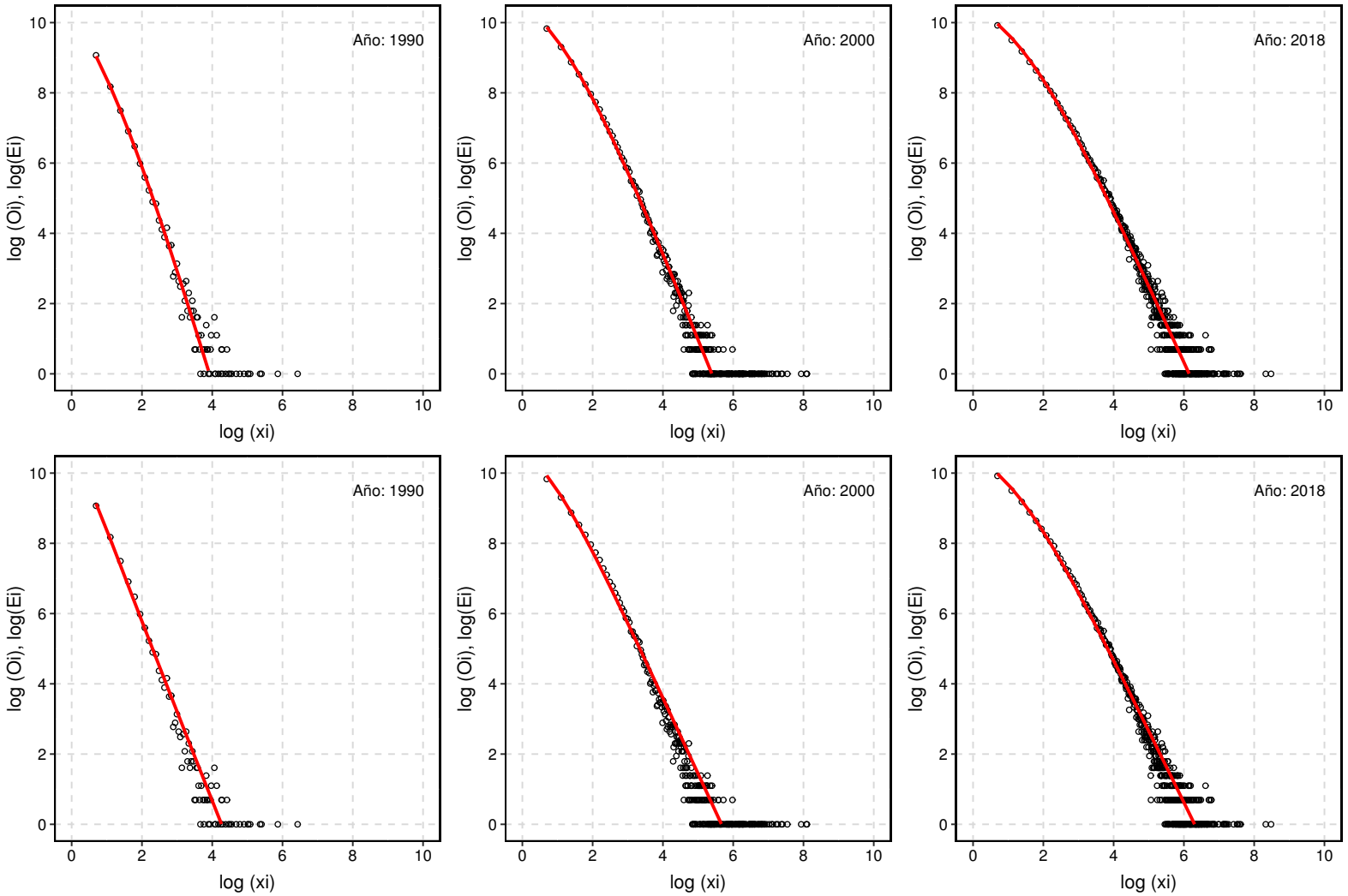

Figura 4.10: Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto II (arriba) y Discreta Generalizada de Pareto III (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018.

En el cuadro 4.9 se presentan los valores obtenidos mediante máxima verosimilitud, con el Criterio de Información de Akaike (AIC) para los modelos de distribución de probabilidad de 1 y 2 parámetros aplicados a la variable frecuencia de uso de códigos tecnológicos CPC entre 1977 y 2018.

En el cuadro 4.10 se presentan los valores obtenidos de $\Delta = AIC - AIC_{min}$ (diferencia entre el valor que toma el estadístico AIC de una distribución específica, con respecto al menor valor de AIC entre las distribuciones consideradas, para cada año del periodo de estudio) para cada una de las distribuciones de probabilidad.

En el cuadro 4.11 se presenta una clasificación de los modelos analizados con base en el número de años en los que el valor de $\Delta = AIC - AIC_{min}$ obtenidos se encuentra dentro de los tres intervalos anteriormente indicados.

En relación a la variable aleatoria, frecuencia de uso de códigos tecnológicos CPC, y con respecto a cuáles de los modelos de 1 y 2 parámetros considerados son los más adecuados, puede comprobarse que los resultados son similares a los obtenidos para la base de datos de códigos tecnológicos IPC anteriomente analizada: los mejores ajustes son proporcionados por los modelos Zipf-Mandelbrot, Beta-Geométrica y Discreta Generalizada de Pareto tipo II. En este caso, así mismo, puede verse que son los modelos Zipf-Mandelbrot y Discreta Generalizada Pareto II aquellos en que un mayor número de años proporcionan un valor de Δ menor de 2.

Cuadro 4.9: Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018.

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
1977	420	428	417	417	418	416	417	416	417	417	419
1978	4371	4900	4009	4011	4221	4063	4010	4021	4012	4010	4014
1979	6800	7390	6512	6487	6648	6501	6489	6525	6509	6487	6490
1980	8491	9248	8157	8115	8301	8126	8118	8169	8147	8115	8112
1981	11056	12104	10675	10594	10811	10600	10598	10676	10644	10595	10595
1982	12658	14078	12065	11997	12322	12013	12003	12108	12063	11998	11993
1983	15290	17166	14612	14508	14871	14515	14512	14638	14585	14509	14522
1984	25167	29015	24148	23874	24414	23881	23881	24148	24034	23876	23893
1985	31090	36492	29665	29343	30041	29352	29352	29679	29525	29345	29354
1986	40068	49923	37201	36927	38198	36935	36935	37413	37162	36929	36957
1987	55055	71168	51158	50624	52338	50643	50635	51491	51061	50627	50662
1988	73494	100309	67698	66979	69305	67012	66988	68148	67514	66982	67066
1989	102610	146773	95353	93837	96668	94078	93850	95783	94729	93841	93941
1990	131257	205910	120329	118388	122266	118763	118398	121141	119544	118392	118544
1991	156003	255649	142404	140112	144541	140603	140119	143282	141319	140115	140302
1992	188598	330315	171976	168894	173929	169735	168898	172862	170309	168896	169172
1993	218105	399772	198956	194974	200680	196253	194980	199960	196740	194977	195285
1994	253328	488276	231653	226687	232663	228497	226687	232275	228518	226687	227130
1995	294793	596602	270535	264238	270420	266832	264234	270556	266144	264236	264821
1996	337559	744206	307171	299885	307294	303070	299882	307810	302141	299883	300445
1997	380663	881614	347751	338866	346436	343100	338858	347640	341218	338861	339595
1998	415950	1060922	376928	366887	375936	371918	366881	377809	369838	366883	367609
1999	489798	1484748	437143	425611	436688	431720	425600	439440	428885	425604	426336
2000	555053	1885943	493124	479729	491837	487233	479712	495631	483157	479718	480564
2001	580255	1964048	521132	505877	516980	514829	505863	521837	509093	505868	506721
2002	602548	2051220	542781	526216	537750	536109	526202	543150	529771	526207	527156

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
2003	634342	2206158	572456	554646	566347	565463	554630	572238	558247	554635	555683
2004	672958	2379196	613348	592550	603745	605778	592537	610955	596144	592541	593635
2005	719933	2694420	654188	632612	643553	646511	632595	651156	635670	632600	633699
2006	741142	2762375	678576	655147	665387	670580	655130	673490	658032	655136	656331
2007	761127	2934797	696460	672046	682631	688314	672029	691218	675056	672034	673269
2008	750948	2761072	688920	665124	674968	680831	665107	683204	667765	665112	666254
2009	780862	2921131	718647	692992	702756	710253	692976	711642	695595	692981	694163
2010	848815	3263752	786149	757324	765960	777262	757306	775654	759190	757312	758545
2011	898785	3641269	835063	803425	811615	825873	803408	822213	804896	803412	804610
2012	953461	4045660	886832	853028	860660	877398	853009	871868	854001	853014	854197
2013	978731	4351373	908868	874406	881771	899462	874387	893361	875067	874391	875482
2014	995523	4517911	924302	889280	896371	914882	889261	908210	889695	889265	890290
2015	1016580	4833436	939869	904745	912090	930494	904726	924515	905032	904729	905602
2016	1045072	5289426	962148	926629	934155	952815	926610	947256	926742	926613	927348
2017	1021605	5247813	935659	901489	910025	926517	901472	923304	902014	901475	902154
2018	910616	4500546	819593	793095	802999	811495	793076	814337	794128	793080	793613

Cuadro 4.10: $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1 y 2 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977-2018.

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
1977	4.38	11.54	0.78	1.46	2.03	0.00	1.16	0.44	0.67	1.43	2.96
1978	361.90	891.24	0.00	2.00	212.07	53.93	0.96	11.90	3.45	1.54	5.56
1979	313.53	902.64	24.63	0.00	161.52	13.95	1.83	37.65	22.09	0.31	3.08
1980	378.99	1136.76	45.05	2.99	189.54	14.43	6.24	57.41	35.07	3.57	0.00
1981	462.17	1509.84	81.51	0.00	217.36	5.63	4.38	82.43	50.55	0.79	1.37
1982	664.91	2085.13	72.58	4.64	329.76	19.92	10.01	114.94	70.17	5.71	0.00
1983	781.94	2657.90	104.16	0.00	363.10	7.19	4.10	130.08	77.16	0.82	14.23
1984	1292.71	5140.72	273.65	0.00	540.40	7.48	7.25	273.66	160.09	1.56	19.33
1985	1746.97	7149.22	321.69	0.00	697.57	9.20	8.69	335.89	182.10	2.07	11.50
1986	3141.36	12995.85	274.30	0.00	1270.75	7.56	7.59	486.35	234.75	2.19	30.14
1987	4430.81	20543.34	534.27	0.00	1713.83	18.44	11.00	867.29	436.39	3.26	38.11
1988	6514.43	33329.30	718.41	0.00	2325.72	32.10	8.35	1168.64	534.97	2.71	86.76
1989	8773.13	52936.38	1515.85	0.00	2831.39	241.20	13.04	1945.82	892.16	4.42	103.86
1990	12869.60	87522.60	1941.00	0.00	3877.92	375.53	10.39	2753.46	1155.89	3.93	156.67
1991	15891.52	115537.59	2292.28	0.00	4429.11	491.54	7.66	3170.51	1207.02	3.05	189.90
1992	19704.27	161420.59	3081.97	0.00	5034.75	841.26	4.09	3968.22	1415.16	1.60	277.98
1993	23130.53	204798.13	3982.14	0.00	5706.27	1278.59	6.20	4986.25	1765.99	2.66	310.71
1994	26641.20	261589.52	4966.84	0.33	5976.36	1810.38	0.15	5588.35	1831.83	0.00	443.47
1995	30558.83	332367.62	6300.91	4.24	6185.46	2597.55	0.00	6321.99	1910.23	1.87	587.21
1996	37676.39	444323.37	7289.16	3.06	7411.83	3187.55	0.00	7928.10	2258.77	1.21	563.22
1997	41804.40	542755.74	8893.30	7.74	7578.15	4241.66	0.00	8782.07	2359.64	3.30	737.32
1998	49069.16	694041.69	10047.00	6.64	9054.98	5037.48	0.00	10928.29	2957.69	2.72	727.98
1999	64198.75	1059147.99	11543.69	11.27	11088.01	6120.24	0.00	13840.04	3285.70	4.08	736.31
2000	75340.67	1406231.44	13412.39	16.63	12124.72	7520.91	0.00	15919.11	3444.70	5.56	851.71
2001	74392.39	1458185.07	15268.72	14.60	11117.32	8965.89	0.00	15973.75	3229.83	4.88	858.20
2002	76346.06	1525018.11	16579.02	13.86	11547.96	9906.25	0.00	16947.31	3568.97	4.79	954.07

Año	$G(p)$	$Po(\lambda)$	$Zeta(\alpha)$	$ZM(\alpha,\beta)$	$Log(\theta)$	$Yule(\rho)$	$BG(\alpha,\beta)$	$BN(r,p)$	$Po(\lambda_1,\lambda_2)$	$DGPII(\sigma,\alpha)$	$DGPIII(\sigma,\gamma)$
2003	79712.55	1651528.13	17826.12	15.90	11717.37	10832.96	0.00	17608.30	3616.89	5.45	1052.83
2004	80421.69	1786659.64	20811.45	13.27	11208.81	13241.00	0.00	18418.50	3607.84	4.59	1098.56
2005	87338.35	2061824.84	21593.13	17.17	10957.70	13916.10	0.00	18560.74	3075.18	5.34	1104.46
2006	86012.07	2107244.47	23445.47	17.15	10257.06	15449.44	0.00	18359.83	2901.38	5.43	1200.70
2007	89098.77	2262768.31	24431.22	17.76	10602.38	16284.95	0.00	19189.56	3026.89	5.55	1240.05
2008	85840.94	2095965.30	23813.39	17.00	9861.22	15724.27	0.00	18097.48	2657.97	5.29	1147.03
2009	87885.80	2228155.52	25671.28	16.05	9780.48	17277.49	0.00	18666.61	2619.31	4.95	1187.68
2010	91508.92	2506446.01	28842.27	17.94	8653.57	19955.28	0.00	18347.54	1883.68	5.16	1238.50
2011	95376.97	2837861.52	31654.99	17.20	8207.29	22465.66	0.00	18804.88	1488.46	4.62	1202.44
2012	100451.92	3192650.12	33822.87	18.69	7650.78	24388.28	0.00	18858.63	991.19	4.61	1187.11
2013	104344.26	3476986.53	34481.74	19.28	7384.75	25075.57	0.00	18974.16	680.26	4.36	1095.57
2014	106262.18	3628649.86	35041.60	19.35	7110.22	25621.20	0.00	18949.57	434.47	4.14	1028.72
2015	111854.46	3928710.13	35143.07	18.87	7364.38	25767.77	0.00	19788.77	305.70	3.62	876.65
2016	118462.42	4362816.75	35538.27	19.31	7545.40	26204.94	0.00	20646.58	132.27	3.24	738.52
2017	120132.88	4346341.46	34187.85	17.68	8553.03	25045.11	0.00	21832.63	542.66	2.96	681.96
2018	117539.01	3707469.82	26516.32	18.60	9922.07	18418.14	0.00	21260.05	1051.06	3.01	536.59

Cuadro 4.11: Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1977 - 2018.

Distribución	$(\triangle \leq 2)$	$(2 < \triangle \leq 20)$	$(\triangle > 20)$
Geométrica	1	0	41
Poisson desplazada	0	1	41
Zeta	2	0	40
Yule-Simon	1	9	32
Logarítmica	1	0	41
Binomial Negativa desplazada	1	1	40
Poisson generalizada desplazada	1	1	40
Zipf- Mandelbrot	15	27	0
Beta - Geométrica	28	14	0
Pareto discreta generalizada tipo II	10	32	0
Pareto discreta generalizada tipo III	3	6	33

4.3.3. Test de bondad de ajuste Chi-Cuadrado para los modelos con mejor ajuste de 1 y 2 parámetros

El cuadro 4.12 muestra los p-valores obtenidos para los 3 distribuciones de probabilidad de 1 y 2 parámetros que mejor ajuste proporcionaron para la variable frecuencia de uso de códigos tecnológicos IPC: Zipf-Mandelbrot, Beta-Geométrica y Discreta Generalizada de Patero II.

Entre 1978 y 1984, la hipótesis nula, (H_0*: los datos siguen el modelo de distribución seleccionado*) no puede ser rechazada al nivel de significancia de 5 por ciento para los 3 modelos seleccionados.

Así mismo, los p-valores indican que la hipótesis nula no puede ser rechazada, a dicho nivel de significancia: para la distribución Zipf - Mandelbrot en los años 1987, 1988, 1989 y 1991; para la distribución Beta-Geométrica entre los años 1986 y 1993; y por último, la distribución discreta generalizada de Pareto II en los años 1988, 1989, 1991 y 1993. Para años posteriores a 1993, ninguno de los modelos supera el test Chi- Cuadrado, siendo los p-valores < 0.05, y por tanto, rechazándose la hipótesis nula al 5 por ciento de significancia.

Cuadro 4.12: Resultados del test de ajuste Chi- Cuadrado para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados.Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 -2018.

	Modelos de distribución de probabilidad								
	ZM			BG			DPII		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1978	**1.67**	**3.84**	**0.20**	**2.12**	**3.84**	**0.14**	**1.74**	**3.84**	**0.19**
1979	**0.63**	**5.99**	**0.73**	**0.86**	**5.99**	**0.65**	**0.66**	**5.99**	**0.72**
1980	**0.33**	**5.99**	**0.85**	**0.43**	**5.99**	**0.81**	**0.34**	**5.99**	**0.84**
1981	**1.23**	**7.81**	**0.75**	**0.70**	**7.81**	**0.87**	**1.08**	**7.81**	**0.78**
1982	**3.21**	**7.81**	**0.36**	**2.45**	**7.81**	**0.49**	**3.01**	**7.81**	**0.39**
1983	**4.30**	**9.49**	**0.37**	**5.59**	**9.49**	**0.23**	**4.62**	**9.49**	**0.33**
1984	**4.04**	**11.07**	**0.54**	**4.89**	**11.07**	**0.43**	**4.25**	**11.07**	**0.51**
1985	12.53	11.07	0.03	12.38	11.07	0.03	12.43	11.07	0.03
1986	14.72	12.59	0.02	**12.34**	**12.59**	**0.05**	13.70	12.59	0.03
1987	**1.29**	**15.51**	**1.00**	**1.68**	**15.51**	**0.99**			
1988	**10.51**	**15.51**	**0.23**	**9.19**	**15.51**	**0.33**	**9.83**	**15.51**	**0.28**
1989	**12.53**	**18.31**	**0.25**	**11.88**	**18.31**	**0.29**	**12.14**	**18.31**	**0.28**
1990	22.97	21.03	0.03	**21.30**	**21.03**	**0.05**	21.94	21.03	0.04
1991	**8.09**	**21.03**	**0.78**	**8.12**	**21.03**	**0.78**	**8.06**	**21.03**	**0.78**
1992	27.79	25.00	0.02	**24.82**	**25.00**	**0.05**	25.92	25.00	0.04
1993	26.68	25.00	0.03	**24.43**	**25.00**	**0.06**	**25.20**	**25.00**	**0.05**
1994	49.00	27.59	0.00	45.14	27.59	0.00	46.43	27.59	0.00
1995	51.76	28.87	0.00	48.39	28.87	0.00	49.43	28.87	0.00
1996	39.07	31.41	0.01	36.44	31.41	0.01	37.12	31.41	0.01
1997	66.19	32.67	0.00	62.53	32.67	0.00	63.38	32.67	0.00
1998	89.71	35.17	0.00	85.00	35.17	0.00	85.86	35.17	0.00
1999	90.64	38.89	0.00	86.44	38.89	0.00	87.09	38.89	0.00
2000	157.62	40.11	0.00	152.52	40.11	0.00	153.13	40.11	0.00
2001	147.67	41.34	0.00	142.67	41.34	0.00	143.16	41.34	0.00
2002	192.54	41.34	0.00	187.16	41.34	0.00	187.83	41.34	0.00
2003	189.08	41.34	0.00	183.87	41.34	0.00	184.64	41.34	0.00
2004	142.63	40.11	0.00	138.18	40.11	0.00	138.94	40.11	0.00
2005	125.80	40.11	0.00	121.47	40.11	0.00	122.20	40.11	0.00
2006	151.03	40.11	0.00	146.22	40.11	0.00	146.98	40.11	0.00
2007	218.23	40.11	0.00	212.74	40.11	0.00	213.56	40.11	0.00
2008	157.04	40.11	0.00	152.41	40.11	0.00	153.16	40.11	0.00
2009	159.80	41.34	0.00	155.18	41.34	0.00	155.88	41.34	0.00

	Modelos de distribución de probabilidad								
	ZM			BG			DPII		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.	χ^2	Umb.	p val.
2010	186.45	42.56	0.00	181.73	42.56	0.00	182.31	42.56	0.00
2011	206.75	44.99	0.00	202.18	44.99	0.00	202.65	44.99	0.00
2012	224.76	44.99	0.00	220.00	44.99	0.00	220.45	44.99	0.00
2013	291.68	46.19	0.00	286.73	46.19	0.00	287.18	46.19	0.00
2014	301.87	47.40	0.00	296.53	47.40	0.00	296.99	47.40	0.00
2015	325.60	48.60	0.00	320.48	48.60	0.00	320.85	48.60	0.00
2016	357.59	48.60	0.00	352.30	48.60	0.00	352.64	48.60	0.00
2017	326.52	48.60	0.00	321.48	48.60	0.00	321.74	48.60	0.00
2018	344.36	48.60	0.00	339.14	48.60	0.00	339.31	48.60	0.00

Por otro lado, considerando ahora la variable frecuencia de uso de códigos tecnológicos CPC, el cuadro 4.13 muestra los p-valores obtenidos para los 3 distribuciones de probabilidad de 1 y 2 parámetros que mejor ajuste proporcionaron para la variable Frecuencia de uso de códigos tecnológicos CPC: Zipf-Mandelbrot, Beta-Geométrica y Discreta Generalizada de Patero II.

En este caso, la hipótesis nula, H_0 no puede ser rechazada al nivel de significancia de 5 por ciento, para las 3 distribuciones seleccionadas en los años 1978, 1979, 1981, 1983, 1984, y entre 1992 y 1995.

Así mismo, los p-valores indican que la hipótesis nula no puede ser rechazada, a dicho nivel de significancia: para la distribución Zipf - Mandelbrot en los años 1986, 1990 y 1991; para la distribución Beta-Geométrica en el año 1996; y por último, la distribución discreta generalizada de Pareto II en los años 1986 y 1991. partir del año 1997, ninguno de los modelos supera el test Chi- Cuadrado, siendo los p-valores < 0.05, y por tanto, rechazándose la hipótesis nula al 5 por ciento de significancia.

Finalmente, el cuadro 4.14 resume los años, tanto para códigos IPC como códigos CPC, en los que la hipótesis nula (H_0*: los datos siguen el modelo de distribución seleccionado*) no puede ser rechazada al 5 por ciento de significancia, para los tres modelos de distribución de probabilidad de 1 y 2 parámetros que mejor ajuste proporcionaron.

Cuadro 4.13: Resultados del test de bondad de ajuste Chi- Cuadrado para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados. Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 -2018

	Modelos de distribución de probabilidad								
	ZM			BG			DPII		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1978	**0.39**	**3.84**	**0.53**	**0.70**	**3.84**	**0.40**	**0.41**	**3.84**	**0.52**
1979	**1.09**	**3.84**	**0.30**	**1.98**	**3.84**	**0.16**	**1.24**	**3.84**	**0.27**
1980	4.69	3.84	0.03	6.44	3.84	0.01	4.98	3.84	0.03
1981	**2.54**	**3.84**	**0.11**	**3.69**	**3.84**	**0.05**	**2.67**	**3.84**	**0.10**
1982	6.20	3.84	0.01	8.71	3.84	0.00	6.65	3.84	0.01
1983	**1.48**	**3.84**	**0.22**	**2.41**	**3.84**	**0.12**	**1.61**	**3.84**	**0.21**
1984	**3.57**	**5.99**	**0.17**	**5.58**	**5.99**	**0.06**	**3.88**	**5.99**	**0.14**
1985	11.31	7.81	0.01	15.55	7.81	0.00	12.20	7.81	0.01
1986	**6.11**	**7.81**	**0.11**	10.29	7.81	0.02	**7.26**	**7.81**	**0.06**
1987	14.46	9.49	0.01	20.00	9.49	0.00	15.95	9.49	0.00
1988	11.69	11.07	0.04	18.41	11.07	0.00	13.87	11.07	0.02
1989	16.77	12.59	0.01	25.07	12.59	0.00	19.45	12.59	0.00
1990	**13.21**	**14.07**	**0.07**	20.01	14.07	0.01	15.70	14.07	0.03
1991	**13.93**	**15.51**	**0.08**	18.80	15.51	0.02	**15.78**	**15.51**	**0.05**
1992	**9.37**	**16.92**	**0.40**	**12.03**	**16.92**	**0.21**	**10.36**	**16.92**	**0.32**
1993	**11.95**	**18.31**	**0.29**	**16.22**	**18.31**	**0.09**	**13.73**	**18.31**	**0.19**
1994	**12.25**	**19.68**	**0.34**	**11.47**	**19.68**	**0.40**	**11.69**	**19.68**	**0.39**
1995	**21.48**	**22.36**	**0.06**	**16.73**	**22.36**	**0.21**	**18.88**	**22.36**	**0.13**
1996	26.23	23.68	0.02	**23.38**	**23.68**	**0.05**	24.52	23.68	0.04
1997	46.63	26.30	0.00	38.49	26.30	0.00	41.99	26.30	0.00
1998	44.24	27.59	0.00	37.01	27.59	0.00	40.00	27.59	0.00
1999	88.16	30.14	0.00	77.07	30.14	0.00	81.08	30.14	0.00
2000	194.85	32.67	0.00	178.38	32.67	0.00	183.87	32.67	0.00
2001	194.27	35.17	0.00	179.66	35.17	0.00	184.52	35.17	0.00
2002	174.41	33.92	0.00	160.64	33.92	0.00	165.39	33.92	0.00
2003	236.05	36.42	0.00	220.17	36.42	0.00	225.60	36.42	0.00
2004	197.91	36.42	0.00	184.97	36.42	0.00	189.41	36.42	0.00
2005	305.68	38.89	0.00	289.51	38.89	0.00	294.44	38.89	0.00
2006	378.60	40.11	0.00	362.20	40.11	0.00	367.27	40.11	0.00
2007	431.42	41.34	0.00	414.40	41.34	0.00	419.61	41.34	0.00
2008	372.10	41.34	0.00	355.90	41.34	0.00	360.82	41.34	0.00

	Modelos de distribución de probabilidad								
	ZM			BG			DPII		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.	χ^2	Umb.	p val.
2009	335.27	41.34	0.00	320.28	41.34	0.00	324.79	41.34	0.00
2010	503.73	44.99	0.00	486.96	44.99	0.00	491.59	44.99	0.00
2011	554.01	47.40	0.00	538.30	47.40	0.00	542.27	47.40	0.00
2012	761.66	49.80	0.00	744.49	49.80	0.00	748.39	49.80	0.00
2013	907.88	52.19	0.00	890.07	52.19	0.00	893.73	52.19	0.00
2014	935.56	53.38	0.00	917.83	53.38	0.00	921.23	53.38	0.00
2015	909.04	53.38	0.00	891.84	53.38	0.00	894.71	53.38	0.00
2016	996.60	55.76	0.00	978.95	55.76	0.00	981.46	55.76	0.00
2017	777.36	54.57	0.00	761.50	54.57	0.00	763.77	54.57	0.00
2018	548.02	49.80	0.00	531.54	49.80	0.00	533.85	49.80	0.00

Cuadro 4.14: Años en los que no se puede rechazar la hipótesis nula H_0*: los datos siguen el modelo de distribución seleccionado* al 5 por ciento de significancia, para los modelos de distribución de probabilidad de 1 y 2 parámetros seleccionados.

Distribución	Años	
	Códigos IPC	Códigos CPC
Zipf - Mandelbrot	1978 - 1984 1987 -1989 1991	1978 - 1981 1983, 1984, 1986, 1990 1995
Beta- Geométrica	1978 - 1984 1986-1993	1978, 1979, 1981 1983, 1984, 1990 1995
Discreta de pareto II	1979 - 1984 1987 - 1989 1991, 1993	1978, 1979, 1981 1983, 1984, 1986 1991 - 1995

4.3.4. Análisis de modelos de distribución de probabilidad de 3 parámetros

En el apartado previo, el análisis de los ajustes de los modelos de distribución de probabilidad de 1 y 2 parámetos para las variables aleatorias frecuencia de uso de códigos tecnológicos IPC y CPC, evidencia que no existe una distribución que pueda ser considerada de forma clara como la mejor opción para modelizar las variables estudiadas. El número de años en que las distribuciones con mejor comportamiento aceptan la hipotesis nula es muy limitado. Por esta razón, se procedió a extender el análisis a distribuciones de 3 parámetros.

De este modo, se seleccionaron aquellas distribuciones de probabilidad de 1 y 2 parámetros con mejores resultados en la primera parte, y que cuentan con modelos generalizados con un tercer parámetro. En concreto, se seleccionaron las siguiente distribuciones de 3 parámetros: la distribución Discreta Generalizada de Pareto IV (que incluye como caso particular a las distribuciones discreta generalizada de Pareto tipo II y III) y Discreta Generalizada de Waring (que incluye como caso particular la distribución Beta Geométrica).

En los cuadros 4.15 (correspondiente a los códigos tecnológicos IPC) y 4.16 (códigos CPC), se muestran los parámetros estimados y sus errores estándar para las distribuciónes Pareto tipo IV (parámetros σ, γ, α) y Waring generalizada desplazada (parámetros α, β, ν) ajustadas a la frecuencia de uso de dichos códigos tecnológicos IPC y CPC, entre 1978 y 2018 (se excluyen los resultados correspondientes al año 1977, puesto que el número de valores que toma la variable aleatoria es muy reducido frente al número de parámetros en las distribuciones seleccionadas).

Las figuras 4.11,4.12 muestran, a modo de validación gráfica del ajuste obtenido por cada uno de los modelos, el valor de las frecuencias observadas (en negro) y esperadas (en rojo) para cada uno de los valores de nuestra muestra, en escala log-log, para los años 1990, 2000 y 2018, de los dos modelos de 3 parámetros indicados: distribución Pareto discreta generalizada de tipo IV y distribución de Waring discreta generalizada. Dichas figuras muestran los ajustes con datos de códigos tecnológicos IPC (figura 4.11) y de códigos tecnológicos CPC (figura 4.12).

En el cuadro 4.17 se presentan los valores obtenidos mediante máxima verosimilitud del estadístico AIC, y en el cuadro 4.18 se muestran los valrores obtenidos de $\Delta = AIC - AIC_{min}$, para los modelos de distribución de probabilidad de 1, 2 y 3 parámetros aplicados a la variable frecuencia de uso de códigos tecnológicos IPC entre 1978 y 2018. Por otro lado, en los cuadros 4.19 y 4.20, se muestran los valores obtenidos del estadístico AIC y los valores obtenidos de $\Delta = AIC - AIC_{min}$,

respectivamente, para el caso de los códigos tecnológicos CPC.

Cuadro 4.15: Parámetros estimados y errores estándar para los modelos de distribución de probabilidad Pareto discreta generalizada IV y Waring generalizada desplazada. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 - 2018.

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
1978	1.2108	0.8574	2.9109	5.0902	1.0557	1.0553
	(0.5652)	(0.1351)	(1.1254)	(0.5782)	(0.5948)	(0.5945)
1979	1.6492	0.9199	3.2141	4.5181	1.1718	1.1398
	(0.5616)	(0.0834)	(0.8352)	(0.3514)	(0.5463)	(0.5250)
1980	2.7325	0.9954	4.4616	4.8688	1.4268	1.1312
	(1.0260)	(0.0688)	(1.1945)	(0.4942)	(1.2759)	(0.9090)
1981	2.6528	1.0340	4.0000	3.9214	1.6581	0.8754
	(0.8416)	(0.0619)	(0.8665)	(0.5544)	(1.1520)	(0.4597)
1982	2.7442	1.0219	3.9777	3.9642	1.6450	0.9579
	(0.7681)	(0.0540)	(0.7694)	(0.4973)	(1.1481)	(0.5268)
1983	1.5290	0.9062	2.4246	3.2938	1.2016	1.2015
	(0.2902)	(0.0519)	(0.3553)	(0.1507)	(0.3249)	(0.3248)
1984	1.4943	0.9220	2.0516	2.5571	1.1887	1.1888
	(0.1963)	(0.0384)	(0.2044)	(0.0839)	(0.2584)	(0.2584)
1985	2.0914	0.9807	2.4705	2.6307	1.2749	1.2758
	(0.2756)	(0.0326)	(0.2304)	(0.0805)	(0.4911)	(0.4916)
1986	2.8913	1.0719	2.8280	2.6094	2.6473	0.7257
	(0.4123)	(0.0310)	(0.2617)	(0.1504)	(0.4761)	(0.0759)
1987	2.0073	0.9758	2.0347	2.1581	1.3024	1.3023
	(0.2050)	(0.0272)	(0.1469)	(0.0501)	(0.4395)	(0.4395)
1988	2.4060	1.0170	2.1228	2.0892	2.0423	0.9316
	(0.2350)	(0.0245)	(0.1403)	(0.0834)	(0.4099)	(0.1347)
1989	2.4442	1.0029	1.9181	1.9178	2.0229	1.0133
	(0.1981)	(0.0209)	(0.1061)	(0.0641)	(0.3915)	(0.1468)
1990	2.6751	1.0375	1.8440	1.7646	2.5956	0.8541
	(0.2052)	(0.0198)	(0.0930)	(0.0525)	(0.3160)	(0.0657)
1991	2.5575	1.0007	1.7137	1.7132	2.1933	1.0219
	(0.1719)	(0.0179)	(0.0783)	(0.0464)	(0.3305)	(0.1122)
1992	3.3353	1.0660	1.8815	1.7395	0.7896	3.4552
	(0.2320)	(0.0170)	(0.0835)	(0.0461)	(0.0381)	(0.3034)
1993	3.1805	1.0500	1.7424	1.6426	3.2550	0.8306
	(0.2012)	(0.0161)	(0.0712)	(0.0398)	(0.2831)	(0.0416)
1994	3.8468	1.0822	1.8355	1.6619	4.0922	0.7691
	(0.2436)	(0.0152)	(0.0728)	(0.0393)	(0.3009)	(0.0290)

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
1995	3.9911	1.0784	1.7573	1.5974	0.7840	4.2448
	(0.2376)	(0.0145)	(0.0655)	(0.0354)	(0.0279)	(0.2962)
1996	3.6684	1.0605	1.5779	1.4640	3.8581	0.8259
	(0.1983)	(0.0139)	(0.0541)	(0.0292)	(0.2692)	(0.0312)
1997	4.2160	1.0913	1.6248	1.4536	4.6189	0.7682
	(0.2277)	(0.0135)	(0.0541)	(0.0282)	(0.2842)	(0.0235)
1998	4.5514	1.1151	1.5991	1.3919	0.7319	5.1501
	(0.2392)	(0.0131)	(0.0506)	(0.0255)	(0.0195)	(0.2905)
1999	4.6045	1.1037	1.4932	1.3170	0.7571	5.1847
	(0.2231)	(0.0124)	(0.0438)	(0.0223)	(0.0197)	(0.2802)
2000	5.3817	1.1380	1.5485	1.3135	0.7109	6.2153
	(0.2623)	(0.0121)	(0.0443)	(0.0217)	(0.0156)	(0.3070)
2001	5.4062	1.1387	1.5143	1.2812	6.2261	0.7138
	(0.2616)	(0.0121)	(0.0429)	(0.0208)	(0.3068)	(0.0157)
2002	5.8944	1.1500	1.6039	1.3343	6.6690	0.7038
	(0.2931)	(0.0120)	(0.0464)	(0.0222)	(0.3224)	(0.0147)
2003	6.2857	1.1503	1.6732	1.3874	6.9717	0.7075
	(0.3213)	(0.0119)	(0.0498)	(0.0239)	(0.3405)	(0.0147)
2004	5.9571	1.1277	1.6477	1.4026	6.5052	0.7386
	(0.2939)	(0.0116)	(0.0482)	(0.0240)	(0.3212)	(0.0160)
2005	4.9283	1.1101	1.5554	1.3541	5.4150	0.7563
	(0.2450)	(0.0123)	(0.0466)	(0.0233)	(0.2904)	(0.0191)
2006	5.5872	1.1333	1.6346	1.3832	6.1788	0.7254
	(0.2825)	(0.0121)	(0.0488)	(0.0238)	(0.3127)	(0.0163)
2007	6.1043	1.1528	1.7038	1.4070	6.7487	0.7020
	(0.3175)	(0.0121)	(0.0516)	(0.0246)	(0.3327)	(0.0148)
2008	5.4735	1.1245	1.6105	1.3757	5.9986	0.7400
	(0.2751)	(0.0121)	(0.0481)	(0.0237)	(0.3087)	(0.0172)
2009	5.7844	1.1306	1.6185	1.3725	6.3513	0.7339
	(0.2864)	(0.0118)	(0.0474)	(0.0232)	(0.3152)	(0.0161)
2010	5.9477	1.1361	1.5804	1.3304	6.5491	0.7305
	(0.2878)	(0.0116)	(0.0449)	(0.0216)	(0.3159)	(0.0155)
2011	6.0331	1.1354	1.5380	1.2944	6.6300	0.7356
	(0.2862)	(0.0115)	(0.0428)	(0.0205)	(0.3158)	(0.0154)
2012	6.6087	1.1524	1.5885	1.3096	7.2748	0.7157
	(0.3178)	(0.0114)	(0.0442)	(0.0208)	(0.3353)	(0.0140)
2013	7.1131	1.1606	1.6154	1.3169	7.7779	0.7097
	(0.3449)	(0.0113)	(0.0450)	(0.0210)	(0.3533)	(0.0133)
2014	7.4862	1.1731	1.6492	1.3255	8.2175	0.6946
	(0.3673)	(0.0113)	(0.0461)	(0.0212)	(0.3668)	(0.0125)
2015	7.7531	1.1758	1.6371	1.3087	8.4432	0.6962
	(0.3711)	(0.0110)	(0.0445)	(0.0202)	(0.3658)	(0.0122)

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma, \gamma, \alpha)$			$SGW(\alpha, \beta, \nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
2016	8.1955	1.1844	1.6482	1.3044	8.9285	0.6884
	(0.3903)	(0.0109)	(0.0442)	(0.0199)	(0.3780)	(0.0116)
2017	7.4838	1.1726	1.5776	1.2674	8.2208	0.6999
	(0.3535)	(0.0111)	(0.0423)	(0.0192)	(0.3583)	(0.0125)
2018	7.1560	1.1728	1.5463	1.2438	7.9443	0.6971
	(0.3407)	(0.0113)	(0.0417)	(0.0189)	(0.3522)	(0.0127)

Cuadro 4.16: Parámetros estimados y errores estándar para los modelos de distribución de probabilidad Pareto discreta generalizada IV y Waring generalizada desplazada. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 - 2018.

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
1978	1.7116	1.2222	4.4057	3.4396	0.3006	1.5044
	(2.0611)	(0.2812)	(3.0784)	(1.1979)	(0.1760)	(1.5576)
1979	0.8022	0.8183	2.5041	4.6018	0.7994	0.7995
	(0.3405)	(0.1604)	(0.9690)	(0.4574)	(0.4089)	(0.4089)
1980	0.6150	0.6741	1.8051	4.7104	0.8453	0.8450
	(0.1821)	(0.1510)	(0.6513)	(0.4179)	(0.3133)	(0.3132)
1981	0.7859	0.7500	2.2011	4.9202	0.9212	0.9210
	(0.2036)	(0.1036)	(0.5734)	(0.3845)	(0.2858)	(0.2857)
1982	0.6375	0.6877	1.7691	4.3255	0.8675	0.8671
	(0.1441)	(0.1117)	(0.4706)	(0.2993)	(0.2416)	(0.2415)
1983	0.9971	0.8533	2.5795	4.2744	0.9204	0.9203
	(0.2485)	(0.0824)	(0.5447)	(0.2666)	(0.2807)	(0.2807)
1984	1.0338	0.8211	2.3321	4.1790	1.0420	1.0417
	(0.1730)	(0.0556)	(0.3423)	(0.1965)	(0.2101)	(0.2100)
1985	0.9406	0.7963	2.0320	3.7472	1.0385	1.0383
	(0.1324)	(0.0510)	(0.2624)	(0.1507)	(0.1797)	(0.1797)
1986	0.9920	0.8662	2.0194	2.9534	0.9724	0.9724
	(0.1314)	(0.0465)	(0.2239)	(0.0959)	(0.1737)	(0.1737)
1987	1.0559	0.8509	1.9608	2.9877	1.0529	1.0530
	(0.1110)	(0.0361)	(0.1743)	(0.0824)	(0.1462)	(0.1462)
1988	1.2568	0.9130	2.0604	2.6511	1.0742	1.0743
	(0.1203)	(0.0300)	(0.1530)	(0.0606)	(0.1551)	(0.1551)
1989	1.3429	0.8820	1.9303	2.6744	1.1953	1.1961
	(0.0979)	(0.0225)	(0.1122)	(0.0511)	(0.1258)	(0.1259)
1990	1.5012	0.9183	1.9036	2.3668	1.2162	1.2164
	(0.0984)	(0.0196)	(0.0948)	(0.0385)	(0.1328)	(0.1328)
1991	1.6001	0.9404	1.8723	2.1864	1.2263	1.2269
	(0.0983)	(0.0181)	(0.0851)	(0.0318)	(0.1444)	(0.1445)
1992	1.8689	0.9710	1.9246	2.0679	1.2753	1.2725
	(0.1065)	(0.0158)	(0.0781)	(0.0269)	(0.2002)	(0.1996)
1993	1.9245	0.9616	1.8637	2.0365	1.3189	1.3184
	(0.0975)	(0.0141)	(0.0677)	(0.0244)	(0.1773)	(0.1772)
1994	2.3104	1.0009	1.9774	1.9825	1.0427	1.8223
	(0.1153)	(0.0129)	(0.0677)	(0.0409)	(0.1175)	(0.2595)
1995	2.7054	1.0277	2.0462	1.9851	2.4679	0.8865
	(0.1285)	(0.0117)	(0.0648)	(0.0381)	(0.1954)	(0.0458)
1996	2.6362	1.0193	1.8937	1.8536	2.3836	0.9244

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
	(0.1125)	(0.0109)	(0.0542)	(0.0318)	(0.1843)	(0.0482)
1997	3.1241	1.0476	1.9943	1.8871	0.8291	3.0674
	(0.1301)	(0.0101)	(0.0541)	(0.0311)	(0.0277)	(0.1759)
1998	3.1210	1.0396	1.9036	1.8197	3.0542	0.8541
	(0.1207)	(0.0095)	(0.0482)	(0.0280)	(0.1692)	(0.0280)
1999	3.4138	1.0653	1.8364	1.7008	3.5676	0.7930
	(0.1244)	(0.0090)	(0.0426)	(0.0237)	(0.1640)	(0.0199)
2000	4.0252	1.0991	1.8923	1.6848	4.4082	0.7327
	(0.1428)	(0.0084)	(0.0415)	(0.0222)	(0.1713)	(0.0140)
2001	4.1467	1.0890	1.8544	1.6662	0.7605	4.4429
	(0.1426)	(0.0081)	(0.0397)	(0.0215)	(0.0145)	(0.1714)
2002	4.2623	1.0854	1.8739	1.6917	4.5355	0.7688
	(0.1420)	(0.0078)	(0.0390)	(0.0214)	(0.1694)	(0.0142)
2003	4.5983	1.0980	1.9237	1.7103	4.9346	0.7491
	(0.1522)	(0.0076)	(0.0393)	(0.0213)	(0.1742)	(0.0126)
2004	4.7875	1.0858	1.8828	1.6951	5.0366	0.7798
	(0.1507)	(0.0072)	(0.0369)	(0.0204)	(0.1731)	(0.0128)
2005	5.3166	1.1165	1.9141	1.6577	5.7202	0.7330
	(0.1688)	(0.0071)	(0.0369)	(0.0193)	(0.1819)	(0.0105)
2006	5.7450	1.1212	1.9591	1.6837	6.1209	0.7318
	(0.1823)	(0.0070)	(0.0376)	(0.0196)	(0.1899)	(0.0100)
2007	6.0240	1.1271	1.9781	1.6888	6.4384	0.7246
	(0.1899)	(0.0069)	(0.0375)	(0.0195)	(0.1948)	(0.0094)
2008	5.6125	1.1159	1.9191	1.6600	5.9802	0.7400
	(0.1761)	(0.0070)	(0.0365)	(0.0191)	(0.1867)	(0.0102)
2009	5.7944	1.1142	1.9022	1.6486	6.1642	0.7458
	(0.1760)	(0.0067)	(0.0351)	(0.0184)	(0.1860)	(0.0099)
2010	6.6195	1.1349	1.9426	1.6363	7.0206	0.7245
	(0.2003)	(0.0065)	(0.0351)	(0.0178)	(0.1990)	(0.0086)
2011	7.1162	1.1408	1.9179	1.5993	7.4996	0.7242
	(0.2110)	(0.0064)	(0.0337)	(0.0168)	(0.2050)	(0.0082)
2012	7.9434	1.1613	1.9537	1.5861	8.3504	0.7032
	(0.2362)	(0.0063)	(0.0338)	(0.0163)	(0.2184)	(0.0073)
2013	8.4034	1.1742	1.9547	1.5604	8.8240	0.6911
	(0.2513)	(0.0063)	(0.0337)	(0.0159)	(0.2266)	(0.0069)
2014	8.5607	1.1781	1.9375	1.5394	9.0049	0.6880
	(0.2548)	(0.0063)	(0.0331)	(0.0155)	(0.2290)	(0.0068)
2015	8.3026	1.1765	1.8641	1.4841	8.7886	0.6912
	(0.2424)	(0.0063)	(0.0312)	(0.0145)	(0.2222)	(0.0069)
2016	8.4696	1.1841	1.8321	1.4467	9.0261	0.6839
	(0.2455)	(0.0063)	(0.0302)	(0.0138)	(0.2250)	(0.0066)

Año	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
	$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$
2017	7.5857	1.1650	1.7431	1.4117	8.1786	0.7028
	(0.2137)	(0.0063)	(0.0283)	(0.0133)	(0.2082)	(0.0073)
2018	5.9692	1.1483	1.6527	1.3728	6.6253	0.7086
	(0.1730)	(0.0069)	(0.0279)	(0.0133)	(0.1853)	(0.0086)

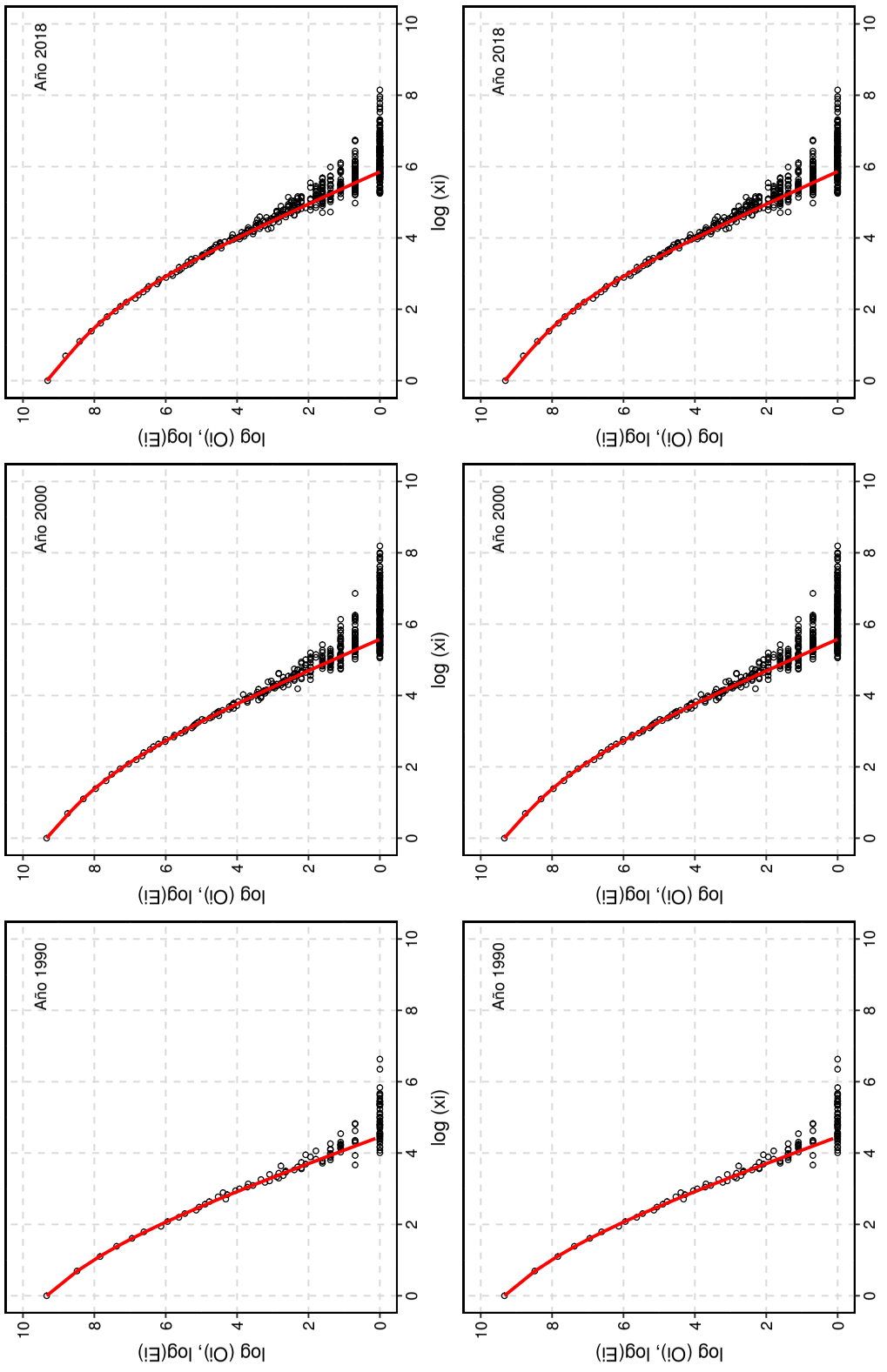

Figura 4.11: Ajuste de los modelos de distribución de probabilidad Discreta Generalizada de Pareto IV (arriba) y Discreta Generalizada de Waring (abajo) para la frecuencia de uso de códigos tecnológicos IPC, para los años 1990, 2000 y 2018.

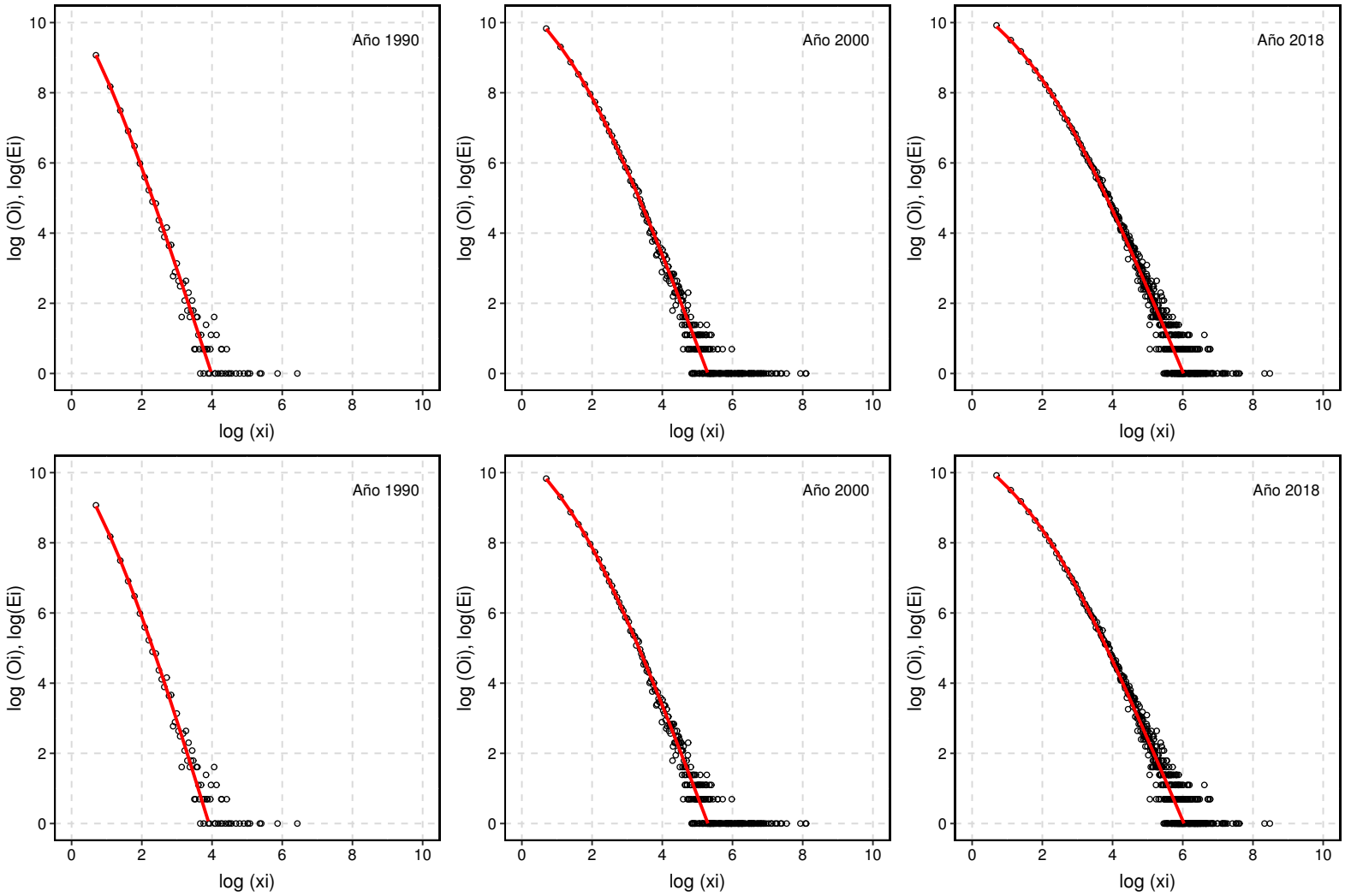

Figura 4.12: Ajuste de los modelos de distribución de probabilidad Discreta Gneralizada de Pareto IV (arriba) y Discreta Generalizada de Waring (abajo) para la frecuencia de uso de códigos tecnológicos CPC, para los años 1990, 2000 y 2018.

Cuadro 4.17: Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978-2018.

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1,\lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
1978	5276	5772	5169	5108	5170	5108	5109	5134	5123	5108	5114	5109	5111
1979	9791	11015	9604	9456	9561	9466	9457	9510	9484	9456	9477	9457	9459
1980	12504	13949	12432	12185	12257	12220	12185	12226	12204	12185	12229	12187	12187
1981	15577	18173	15190	14949	15122	14972	14949	15028	14984	14949	15004	14951	14951
1982	17220	20180	16872	16561	16732	16606	16561	16658	16606	16561	16629	16563	16563
1983	20905	25761	20144	19806	20124	19852	19808	20004	19895	19806	19841	19805	19809
1984	33969	49331	31412	30933	31894	31006	30936	31571	31217	30934	30982	30932	30937
1985	39317	55827	37139	36425	37202	36592	36426	37012	36675	36425	36531	36427	36428
1986	46620	69321	43815	43022	43857	43216	43020	43592	43220	43021	43174	43018	43018
1987	60011	98577	55610	54526	55836	54844	54527	55610	54906	54526	54631	54527	54528
1988	70692	124879	65247	63904	65392	64358	63903	65214	64326	63903	64050	63905	63905
1989	92246	184776	84560	82603	84577	83402	82603	84629	83214	82603	82758	82605	82605
1990	109942	249915	99473	97269	99578	98224	97267	99683	97857	97268	97431	97267	97267
1991	120334	281472	109440	106691	109128	108002	106691	109566	107413	106691	106845	106693	106693
1992	136273	349591	123674	120515	122998	122126	120512	123595	121174	120513	120733	120500	120500
1993	150655	420329	135767	132218	135109	134097	132216	135951	132993	132217	132416	132209	132209
1994	166969	497162	151197	146976	149787	149365	146972	150958	147709	146974	147217	146946	146946
1995	184179	595512	166840	161918	164839	164854	161914	166432	162663	161915	162150	161887	161888
1996	204739	752030	183141	177799	181322	181054	177796	183330	178636	177797	177975	177779	177781
1997	220597	885049	196921	191097	194669	194760	191092	196966	191870	191093	191288	191048	191050
1998	245968	1156523	216712	210443	214597	214473	210437	217344	211235	210438	210607	210360	210363
1999	270975	1506090	236881	229664	234451	234506	229660	238043	230580	229660	229794	229589	229593
2000	292070	1833017	254270	246449	251443	251827	246443	255486	247320	246444	246568	246307	246310
2001	299499	1877048	261737	253689	258321	259270	253683	262495	254364	253684	253783	253546	253552

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
2002	297712	1661225	264389	255980	259833	261845	255974	263769	256486	255975	256114	255810	255819
2003	294324	1501737	265280	256423	259661	262655	256417	263472	256808	256418	256599	256251	256261
2004	295408	1399903	268498	259180	262305	265751	259175	266130	259594	259176	259387	259050	259060
2005	276987	1277177	248740	240873	244267	246197	240868	247677	241337	240868	241032	240787	240796
2006	285202	1346810	256923	248614	251872	254326	248609	255388	249028	248610	248795	248483	248493
2007	286952	1359310	259036	250639	253712	256440	250633	257173	250985	250634	250831	250466	250477
2008	285300	1314879	258082	249617	252682	255457	249611	256219	249966	249612	249794	249503	249514
2009	296698	1411404	268839	259821	262869	266134	259815	266637	260136	259816	260000	259689	259700
2010	314262	1624221	283600	274071	277292	280833	274066	281460	274342	274066	274219	273923	273937
2011	325711	1787175	293569	283592	286843	290766	283587	291328	283799	283588	283714	283442	283460
2012	331841	1865951	299600	289292	292433	296775	289287	296985	289451	289288	289416	289098	289117
2013	339747	1942362	308097	297225	300162	305217	297220	304849	297293	297220	297349	297005	297027
2014	343070	2004223	310765	299857	302785	307891	299851	307497	299908	299852	299975	299599	299620
2015	360814	2231835	326477	314817	317959	323499	314812	323045	314889	314812	314924	314537	314564
2016	371918	2383811	336700	324562	327666	333678	324557	332947	324577	324557	324654	324245	324272
2017	365634	2381072	329204	317641	320891	326262	317635	326185	317667	317636	317709	317375	317403
2018	361684	2385261	324070	313077	316333	321207	313072	321565	313055	313072	313119	312821	312849

Cuadro 4.18: $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978-2018.

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
1978	168.42	664.36	61.15	0.40	62.25	0.00	1.30	26.24	14.70	0.55	6.33	1.39	3.29
1979	335.20	1559.21	148.56	0.00	104.76	9.84	1.03	54.04	27.92	0.19	20.86	1.25	2.92
1980	319.40	1764.48	247.44	0.00	72.31	34.94	0.17	41.78	18.85	0.01	43.85	2.01	2.10
1981	627.70	3223.85	241.20	0.28	173.31	22.84	0.00	79.28	35.00	0.17	54.63	1.87	1.96
1982	658.35	3618.63	310.49	0.01	170.29	44.25	0.11	96.55	44.94	0.00	67.79	1.84	2.10
1983	1100.02	5955.72	339.11	0.75	318.58	46.68	2.81	198.85	89.99	1.29	35.95	0.00	4.31
1984	3037.06	18398.96	480.15	1.01	961.71	74.42	4.29	638.67	284.99	2.16	50.39	0.00	5.58
1985	2892.21	19402.46	714.42	0.00	777.10	167.43	1.67	587.04	250.11	0.55	106.33	2.20	3.13
1986	3602.47	26303.24	797.13	4.34	839.15	198.20	2.21	574.10	202.32	3.40	156.43	0.00	0.18
1987	5485.78	44050.96	1084.35	0.00	1310.80	318.23	1.14	1083.99	380.14	0.42	105.37	1.63	2.17
1988	6789.22	60976.43	1343.65	0.55	1488.60	454.82	0.00	1311.19	423.54	0.21	147.52	1.73	1.82
1989	9643.13	102172.76	1956.50	0.24	1973.99	798.53	0.00	2025.69	611.01	0.05	155.30	2.03	1.99
1990	12675.25	152648.44	2206.13	2.89	2311.72	957.23	0.81	2416.74	590.90	1.62	164.28	0.00	0.09
1991	13642.94	174780.68	2748.51	0.10	2436.51	1310.63	0.02	2874.27	721.72	0.00	153.80	2.00	1.98
1992	15773.52	229091.32	3174.50	15.39	2498.58	1625.79	12.05	3095.36	674.56	13.30	233.29	0.00	0.12
1993	18446.46	288120.21	3558.00	9.52	2900.12	1887.83	6.87	3742.51	784.02	7.79	206.78	0.00	0.04
1994	20022.83	350216.73	4251.63	30.61	2841.57	2419.12	26.45	4012.33	763.48	27.86	271.65	0.00	0.46
1995	22291.90	433625.23	4952.46	30.98	2952.24	2966.51	26.96	4545.15	775.67	28.23	262.34	0.00	0.67
1996	26959.90	574251.14	5362.04	19.70	3542.62	3274.31	16.46	5550.38	856.46	17.34	195.99	0.00	1.52
1997	29549.15	694001.53	5873.05	48.80	3621.54	3711.99	44.40	5918.50	821.69	45.48	239.69	0.00	1.94
1998	35607.78	946163.04	6352.53	83.05	4237.26	4113.14	77.58	6984.18	875.30	78.66	246.96	0.00	2.75
1999	41385.72	1276500.61	7291.79	74.91	4861.54	4916.88	70.32	8453.66	990.96	71.08	204.98	0.00	3.49
2000	45762.64	1586709.62	7962.85	141.12	5135.91	5520.06	135.42	9179.07	1012.53	136.19	260.18	0.00	2.47
2001	45953.28	1623502.26	8190.87	142.78	4775.28	5724.17	137.14	8948.94	817.88	137.79	237.64	0.00	5.95

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
2002	41901.97	1405414.78	8579.66	170.39	4023.26	6035.13	164.53	7958.94	675.72	165.34	304.65	0.00	8.85
2003	38073.23	1245486.54	9029.31	172.48	3410.40	6404.31	166.61	7221.52	557.81	167.58	348.45	0.00	10.41
2004	36357.80	1140852.97	9447.39	129.82	3254.44	6701.16	124.63	7079.24	543.84	125.63	336.82	0.00	9.54
2005	36199.95	1036389.95	7952.71	85.76	3479.91	5410.26	80.74	6890.05	550.01	81.65	245.12	0.00	8.71
2006	36718.74	1098326.69	8440.13	131.26	3388.74	5842.65	125.64	6905.03	545.46	126.65	311.75	0.00	10.03
2007	36485.77	1108844.11	8569.86	172.98	3245.48	5974.43	166.84	6707.21	518.49	167.93	364.66	0.00	11.44
2008	35796.93	1065376.45	8579.50	114.16	3179.72	5953.99	108.83	6716.51	463.72	109.80	291.49	0.00	10.94
2009	37009.10	1151715.47	9149.95	132.07	3180.13	6445.64	126.68	6948.60	447.47	127.62	311.58	0.00	11.09
2010	40338.93	1350298.39	9676.87	148.25	3369.22	6909.95	142.82	7537.06	419.72	143.62	296.70	0.00	14.52
2011	42268.95	1503732.89	10126.79	150.02	3401.07	7324.26	144.83	7886.01	356.83	145.49	271.55	0.00	17.36
2012	42743.52	1576853.07	10501.80	194.55	3334.82	7677.05	189.05	7887.07	353.18	189.73	318.47	0.00	19.02
2013	42741.89	1645357.62	11092.57	220.65	3157.14	8212.30	215.08	7843.93	288.31	215.74	344.24	0.00	22.01
2014	43471.60	1704624.58	11166.22	258.35	3186.63	8292.21	252.43	7897.77	309.58	253.11	376.40	0.00	20.90
2015	46276.78	1917297.52	11939.62	279.95	3421.27	8962.14	274.25	8507.37	351.33	274.84	386.55	0.00	26.89
2016	47673.23	2059566.23	12454.79	317.51	3421.01	9433.48	311.70	8701.67	331.92	312.25	408.66	0.00	27.37
2017	48259.36	2063696.62	11829.22	266.17	3515.68	8886.89	260.54	8810.55	291.66	260.99	333.65	0.00	28.01
2018	48862.60	2072439.87	11248.48	255.89	3511.49	8385.68	250.03	8743.33	233.06	250.41	297.99	0.00	27.05

Cuadro 4.19: Criterio de Información de Akaike: (AIC) para los modelos de distribución de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978-2018.

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1,\lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
1978	4371	4900	4009	4011	4221	4063	4010	4021	4012	4010	4014	4012	4012
1979	6800	7390	6512	6487	6648	6501	6489	6525	6509	6487	6490	6488	6491
1980	8491	9248	8157	8115	8301	8126	8118	8169	8147	8115	8112	8112	8120
1981	11056	12104	10675	10594	10811	10600	10598	10676	10644	10595	10595	10590	10600
1982	12658	14078	12065	11997	12322	12013	12003	12108	12063	11998	11993	11991	12004
1983	15290	17166	14612	14508	14871	14515	14512	14638	14585	14509	14522	14508	14514
1984	25167	29015	24148	23874	24414	23881	23881	24148	24034	23876	23893	23867	23883
1985	31090	36492	29665	29343	30041	29352	29352	29679	29525	29345	29354	29330	29354
1986	40068	49923	37201	36927	38198	36935	36935	37413	37162	36929	36957	36923	36937
1987	55055	71168	51158	50624	52338	50643	50635	51491	51061	50627	50662	50612	50637
1988	73494	100309	67698	66979	69305	67012	66988	68148	67514	66982	67066	66976	66990
1989	102610	146773	95353	93837	96668	94078	93850	95783	94729	93841	93941	93816	93849
1990	131257	205910	120329	118388	122266	118763	118398	121141	119544	118392	118544	118376	118396
1991	156003	255649	142404	140112	144541	140603	140119	143282	141319	140115	140302	140106	140118
1992	188598	330315	171976	168894	173929	169735	168898	172862	170309	168896	169172	168894	168897
1993	218105	399772	198956	194974	200680	196253	194980	199960	196740	194977	195285	194971	194976
1994	253328	488276	231653	226687	232663	228497	226687	232275	228518	226687	227130	226689	226689
1995	294793	596602	270535	264238	270420	266832	264234	270556	266144	264236	264821	264232	264232
1996	337559	744206	307171	299885	307294	303070	299882	307810	302141	299883	300445	299882	299883
1997	380663	881614	347751	338866	346436	343100	338858	347640	341218	338861	339595	338841	338840
1998	415950	1060922	376928	366887	375936	371918	366881	377809	369838	366883	367609	366868	366867
1999	489798	1484748	437143	425611	436688	431720	425600	439440	428885	425604	426336	425552	425551
2000	555053	1885943	493124	479729	491837	487233	479712	495631	483157	479718	480564	479576	479572
2001	580255	1964048	521132	505877	516980	514829	505863	521837	509093	505868	506721	505746	505747

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
2002	602548	2051220	542781	526216	537750	536109	526202	543150	529771	526207	527156	526086	526084
2003	634342	2206158	572456	554646	566347	565463	554630	572238	558247	554635	555683	554465	554461
2004	672958	2379196	613348	592550	603745	605778	592537	610955	596144	592541	593635	592397	592397
2005	719933	2694420	654188	632612	643553	646511	632595	651156	635670	632600	633699	632324	632327
2006	741142	2762375	678576	655147	665387	670580	655130	673490	658032	655136	656331	654821	654827
2007	761127	2934797	696460	672046	682631	688314	672029	691218	675056	672034	673269	671675	671678
2008	750948	2761072	688920	665124	674968	680831	665107	683204	667765	665112	666254	664824	664831
2009	780862	2921131	718647	692992	702756	710253	692976	711642	695595	692981	694163	692681	692688
2010	848815	3263752	786149	757324	765960	777262	757306	775654	759190	757312	758545	756863	756878
2011	898785	3641269	835063	803425	811615	825873	803408	822213	804896	803412	804610	802895	802922
2012	953461	4045660	886832	853028	860660	877398	853009	871868	854001	853014	854197	852305	852339
2013	978731	4351373	908868	874406	881771	899462	874387	893361	875067	874391	875482	873559	873602
2014	995523	4517911	924302	889280	896371	914882	889261	908210	889695	889265	890290	888388	888434
2015	1016580	4833436	939869	904745	912090	930494	904726	924515	905032	904729	905602	903871	903928
2016	1045072	5289426	962148	926629	934155	952815	926610	947256	926742	926613	927348	925675	925736
2017	1021605	5247813	935659	901489	910025	926517	901472	923304	902014	901475	902154	900738	900791
2018	910616	4500546	819593	793095	802999	811495	793076	814337	794128	793080	793613	792581	792619

Cuadro 4.20: $\Delta = AIC - AIC_{min}$ entre los modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978-2018.

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
1978	361.90	891.24	0.00	2.00	212.07	53.93	0.96	11.90	3.45	1.54	5.56	2.94	2.85
1979	313.53	902.64	24.63	0.00	161.52	13.95	1.83	37.65	22.09	0.31	3.08	0.94	3.65
1980	378.99	1136.76	45.05	2.99	189.54	14.43	6.24	57.41	35.07	3.57	0.00	0.13	8.04
1981	465.73	1513.40	85.07	3.56	220.93	9.19	7.94	85.99	54.11	4.35	4.93	0.00	9.87
1982	666.27	2086.49	73.94	6.00	331.12	21.28	11.37	116.30	71.53	7.07	1.36	0.00	13.12
1983	782.45	2658.40	104.66	0.51	363.60	7.69	4.61	130.58	77.66	1.32	14.74	0.00	6.54
1984	1300.07	5148.08	281.01	7.36	547.76	14.84	14.61	281.02	167.45	8.92	26.69	0.00	16.57
1985	1759.61	7161.86	334.33	12.64	710.22	21.84	21.34	348.53	194.74	14.71	24.14	0.00	23.29
1986	3145.61	13000.10	278.56	4.25	1275.00	11.81	11.84	490.60	239.01	6.44	34.39	0.00	13.82
1987	4443.15	20555.68	546.60	12.33	1726.17	30.77	23.33	879.63	448.72	15.60	50.45	0.00	25.19
1988	6518.18	33333.05	722.16	3.75	2329.47	35.85	12.11	1172.39	538.72	6.47	90.51	0.00	13.86
1989	8794.34	52957.59	1537.06	21.20	2852.60	262.41	34.25	1967.02	913.36	25.62	125.07	0.00	33.05
1990	12881.15	87534.15	1952.55	11.55	3889.46	387.08	21.94	2765.01	1167.44	15.48	168.22	0.00	20.31
1991	15897.36	115543.43	2298.12	5.84	4434.95	497.38	13.50	3176.34	1212.85	8.88	195.74	0.00	12.01
1992	19704.27	161420.59	3081.97	0.00	5034.75	841.26	4.09	3968.22	1415.16	1.60	277.98	0.25	3.00
1993	23133.30	204800.89	3984.91	2.77	5709.03	1281.36	8.97	4989.02	1768.76	5.43	313.48	0.00	4.96
1994	26641.20	261589.52	4966.84	0.33	5976.36	1810.38	0.15	5588.35	1831.83	0.00	443.47	1.99	1.99
1995	30560.65	332369.43	6302.72	6.05	6187.28	2599.36	1.81	6323.80	1912.04	3.68	589.03	0.05	0.00
1996	37676.39	444323.37	7289.16	3.06	7411.83	3187.55	0.00	7928.10	2258.77	1.21	563.22	0.06	0.23
1997	41822.06	542773.40	8910.96	25.40	7595.81	4259.32	17.66	8799.73	2377.30	20.96	754.98	0.29	0.00
1998	49082.67	694055.20	10060.51	20.16	9068.50	5051.00	13.52	10941.80	2971.21	16.23	741.49	0.70	0.00
1999	64247.56	1059196.79	11592.50	60.07	11136.82	6169.04	48.80	13888.84	3334.50	52.88	785.12	0.76	0.00
2000	75480.66	1406371.43	13552.39	156.63	12264.71	7660.91	139.99	16059.10	3584.69	145.56	991.70	4.41	0.00
2001	74508.79	1458301.47	15385.12	131.00	11233.71	9082.29	116.40	16090.15	3346.23	121.28	974.60	0.00	0.21

Año	G	$Po(\lambda)$	$Zeta$	ZM	Log	$Yule$	BG	BN	$Po(\lambda_1, \lambda_2)$	$DGPII$	$DGPIII$	$DGPIV$	SGW
2002	76464.11	1525136.16	16697.06	131.91	11666.01	10024.29	118.05	17065.36	3687.02	122.84	1072.11	1.69	0.00
2003	79881.03	1651696.61	17994.60	184.38	11885.85	11001.44	168.48	17776.78	3785.37	173.93	1221.31	3.40	0.00
2004	80561.74	1786799.70	20951.50	153.32	11348.86	13381.05	140.05	18558.55	3747.89	144.65	1238.61	0.09	0.00
2005	87609.49	2062095.98	21864.26	288.30	11228.83	14187.24	271.13	18831.87	3346.31	276.48	1375.59	0.00	3.02
2006	86321.05	2107553.45	23754.44	326.13	10566.03	15758.41	308.97	18668.80	3210.36	314.40	1509.67	0.00	5.88
2007	89451.99	2263121.53	24784.44	370.98	10955.60	16638.17	353.22	19542.78	3380.11	358.77	1593.27	0.00	3.05
2008	86123.70	2096248.06	24096.14	299.76	10143.98	16007.03	282.76	18380.24	2940.73	288.05	1429.79	0.00	7.19
2009	88180.10	2228449.82	25965.58	310.35	10074.78	17571.79	294.30	18960.91	2913.61	299.25	1481.98	0.00	6.11
2010	91952.74	2506889.83	29286.09	461.76	9097.39	20399.10	443.82	18791.36	2327.50	448.97	1682.32	0.00	15.17
2011	95889.54	2838374.09	32167.56	529.77	8719.87	22978.23	512.57	19317.46	2001.03	517.20	1715.02	0.00	26.60
2012	101156.53	3193354.73	34527.48	723.30	8355.40	25092.90	704.61	19563.24	1695.80	709.23	1891.72	0.00	34.41
2013	105172.18	3477814.45	35309.66	847.20	8212.67	25903.48	827.92	19802.08	1508.18	832.28	1923.49	0.00	43.22
2014	107134.78	3629522.45	35914.19	891.94	7982.81	26493.79	872.59	19822.16	1307.06	876.73	1901.31	0.00	45.68
2015	112709.01	3929564.68	35997.62	873.42	8218.93	26622.32	854.55	20643.32	1160.25	858.17	1731.20	0.00	56.60
2016	119397.09	4363751.43	36472.94	953.99	8480.08	27139.62	934.67	21581.25	1066.94	937.92	1673.19	0.00	60.79
2017	120866.91	4347075.49	34921.88	751.71	9287.05	25779.14	734.03	22566.66	1276.69	736.98	1415.99	0.00	53.44
2018	118034.03	3707964.84	27011.34	513.62	10417.09	18913.16	495.02	21755.07	1546.08	498.04	1031.61	0.00	37.41

En los cuadros 4.21 y 4.21 se presenta una clasificación de los modelos analizados con base en el número de años en los cuales el valor de $\Delta = AIC - AIC_{min}$ obtenidos se encuentra dentro de los tres intervalos anteriormente indicados, para los códigos tecnológicos IPC y CPC respectivamente.

Puede comprobarse que la distribución discreta generalizada de Pareto tipo IV es el modelo más adecuado para la variable de interés, tanto considerando códigos tecnológicos IPC como CPC. Dicha distribución obtuvo un valor de $\Delta = AIC - AIC_{min}$ inferior a 2 en 37 de los 41 años en el caso de códigos IPC y en 38 de los 41 años considerados en el caso de códigos CPC (sin incluir el primer año 1977), y presentó un valor inferior a 20 de de $\Delta = AIC - AIC_{min}$ en el resto de los años.

Finalmente, los cuadros 4.23 y 4.24 muestran los p-valores obtenidos para las distribuciones de probabilidad de 3 parámetros consideradas, y ajustadas para la variable frecuencia de uso de códigos tecnológicos IPC y CPC respectivamente. Puede observarse, en el caso de códigos IPC, que la hipotesis nula (H_0*: los datos siguen el modelo de distribución seleccionado*) no puede rechazarse para un nivel de significancia de 5 por ciento en los periodos 1979 - 1984 y 1986 -2001, en ambas distribuciones.

Así mismo, puede observarse, en el caso de códigos CPC, la hipótesis nula H_0 no puede rechazarse en el caso de la distribución discreta generalizada de Pareto IV en los siguientes años: 1984 - 1986, 1988-2000, y 2005; y en el caso de la distribución de Waring generalizada desplazada en los años 1992 - 2000 , y en el año 2005 (a modo de resumen, el cuadro 4.25 muestra los años en lo que H_0 no puede rechazarse al 5 por ciento de significancia para los cinco modelos de 2 y 3 parámetros con mejor ajuste a los datos utilizados).

En resumen, hemos encontrado que la distribución discreta generalizada de Pareto IV proporciona los mejores ajustes, tanto para la frecuencia de uso de códigos tecnológicos IPC como CPC, en comparación con el resto de los 12 modelos considerados. Así mismo, hemos encontrado que dicha distribución supera el test Chi-Cuadrado en muchos de los años analizados.

Cuadro 4.21: Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 - 2018.

Distribución	$(\triangle \leq 2)$	$(2 < \triangle \leq 20)$	$(\triangle > 20)$
Geométrica	0	0	41
Poisson desplazada	0	0	41
Zeta	0	0	41
Yule-Simon	1	1	39
Logarítmica	0	0	41
Binomial Negativa desplazada	0	0	41
Poisson generalizada desplazada	0	2	39
Zipf-Mandelbrot	12	4	25
Beta-Geométrica	11	6	24
Pareto discreta generalizada tipo II	12	5	24
Pareto discreta generalizada tipo III	0	1	40
Pareto discreta generalizada tipo IV	37	4	0
Waring generalizada desplazada	12	23	6

Cuadro 4.22: Número de años en cada intervalo $\Delta = AIC - AIC_{min}$, de los diferentes modelos de distribución de probabilidad de 1, 2 y 3 parámetros considerados. Variable aleatoria discreta: frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 - 2018.

Distribución	$(\triangle \leq 2)$	$(2 < \triangle \leq 20)$	$(\triangle > 20)$
Geométrica	0	0	41
Poisson desplazada	0	0	41
Zeta	1	0	40
Yule-Simon	0	7	34
Logarítmica	0	0	41
Binomial Negativa desplazada	0	1	40
Poisson generalizada desplazada	0	1	40
Zipf-Mandelbrot	4	13	24
Beta-Geométrica	5	12	24
Pareto discreta generalizada tipo II	6	12	23
Pareto discreta generalizada tipo III	2	4	35
Pareto discreta generalizada tipo IV	38	3	0
Waring generalizada desplazada	11	18	12

Cuadro 4.23: Resultado de test de bondad de ajuste Chi- Cuadrado para modelos de distribución de probabilidad de 3 parámetros.Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos IPC. Periodo 1978 -2018.

	Modelos de distribución de probabilidad					
	$DGPIV(\sigma, \gamma, \alpha)$			$SGW(\alpha, \beta, \nu)$		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1978	0.73	0.00	0.00	2.11	0.00	0.00
1979	**0.78**	**3.84**	**0.38**	**0.79**	**3.84**	**0.37**
1980	**0.32**	**3.84**	**0.57**	**0.38**	**3.84**	**0.54**
1981	**0.50**	**5.99**	**0.78**	**0.54**	**5.99**	**0.76**
1982	**2.26**	**5.99**	**0.32**	**2.34**	**5.99**	**0.31**
1983	**2.59**	**7.81**	**0.46**	**5.18**	**7.81**	**0.16**
1984	**4.99**	**9.49**	**0.29**	**4.47**	**9.49**	**0.35**
1985	13.39	9.49	0.01	12.65	9.49	0.01
1986	**6.38**	**11.07**	**0.27**	**6.56**	**11.07**	**0.26**
1987	**1.38**	**14.07**	**0.99**	**1.20**	**14.07**	**0.99**
1988	**8.44**	**14.07**	**0.30**	**8.44**	**14.07**	**0.30**
1989	**12.03**	**16.92**	**0.21**	**11.93**	**16.92**	**0.22**
1990	**19.21**	**19.68**	**0.06**	**19.25**	**19.68**	**0.06**
1991	**8.07**	**19.68**	**0.71**	**8.06**	**19.68**	**0.71**
1992	**12.67**	**23.68**	**0.55**	**12.67**	**23.68**	**0.55**

	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1993	**17.17**	**23.68**	**0.25**	**17.05**	**23.68**	**0.25**
1994	**18.79**	**26.30**	**0.28**	**19.19**	**26.30**	**0.26**
1995	**26.32**	**27.59**	**0.07**	**26.41**	**27.59**	**0.07**
1996	**24.18**	**30.14**	**0.19**	**24.76**	**30.14**	**0.17**
1997	**29.42**	**31.41**	**0.08**	**29.63**	**31.41**	**0.08**
1998	**23.39**	**33.92**	**0.38**	**24.07**	**33.92**	**0.34**
1999	**23.50**	**37.65**	**0.55**	**25.74**	**37.65**	**0.42**
2000	**37.82**	**38.89**	**0.06**	**37.91**	**38.89**	**0.06**
2001	**31.22**	**40.11**	**0.26**	**33.23**	**40.11**	**0.19**
2002	43.36	40.11	0.02	48.79	40.11	0.01
2003	51.53	40.11	0.00	56.59	40.11	0.00
2004	45.25	38.89	0.01	49.18	38.89	0.00
2005	61.18	38.89	0.00	65.98	38.89	0.00
2006	52.68	38.89	0.00	56.86	38.89	0.00
2007	70.81	38.89	0.00	78.18	38.89	0.00
2008	68.81	38.89	0.00	75.35	38.89	0.00
2009	62.12	40.11	0.00	67.27	40.11	0.00
2010	72.57	41.34	0.00	81.11	41.34	0.00
2011	89.38	43.77	0.00	100.69	43.77	0.00
2012	80.20	43.77	0.00	90.05	43.77	0.00
2013	118.32	44.99	0.00	132.05	44.99	0.00
2014	90.88	46.19	0.00	104.47	46.19	0.00
2015	100.11	47.40	0.00	117.36	47.40	0.00
2016	94.85	47.40	0.00	113.23	47.40	0.00
2017	115.88	47.40	0.00	133.81	47.40	0.00
2018	144.01	47.40	0.00	161.32	47.40	0.00

Cuadro 4.24: Resultado de test de bondad de ajuste Chi- Cuadrado para modelos de distribución de probabilidad de 3 parámetros.Variable aleatoria discreta: Frecuencia de uso de códigos tecnológicos CPC. Periodo 1978 -2018.

	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1978	1.17	0.00	0.00	1.11	0.00	0.00
1979	0.24	0.00	0.00	1.92	0.00	0.00

	Modelos de distribución de probabilidad					
	$DGPIV(\sigma,\gamma,\alpha)$			$SGW(\alpha,\beta,\nu)$		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.
1980	0.56	0.00	0.00	6.42	0.00	0.00
1981	2.04	0.00	0.00	3.77	0.00	0.00
1982	0.54	0.00	0.00	8.73	0.00	0.00
1983	1.76	0.00	0.00	2.46	0.00	0.00
1984	**2.66**	**3.84**	**0.10**	5.47	3.84	0.02
1985	**4.74**	**5.99**	**0.09**	15.44	5.99	0.00
1986	**3.00**	**5.99**	**0.22**	10.31	5.99	0.01
1987	8.58	7.81	0.04	19.80	7.81	0.00
1988	**6.87**	**9.49**	**0.14**	18.18	9.49	0.00
1989	**2.96**	**11.07**	**0.71**	22.33	11.07	0.00
1990	**4.92**	**12.59**	**0.55**	17.08	12.59	0.01
1991	**9.60**	**14.07**	**0.21**	16.07	14.07	0.02
1992	**8.51**	**15.51**	**0.39**	**9.74**	**15.51**	**0.28**
1993	**9.20**	**16.92**	**0.42**	**11.68**	**16.92**	**0.23**
1994	**11.63**	**18.31**	**0.31**	**11.53**	**18.31**	**0.32**
1995	**11.63**	**21.03**	**0.48**	**11.64**	**21.03**	**0.48**
1996	**21.41**	**22.36**	**0.07**	**21.62**	**22.36**	**0.06**
1997	**15.33**	**25.00**	**0.43**	**15.45**	**25.00**	**0.42**
1998	**18.63**	**26.30**	**0.29**	**18.26**	**26.30**	**0.31**
1999	**23.52**	**28.87**	**0.17**	**23.61**	**28.87**	**0.17**
2000	**31.29**	**31.41**	**0.05**	**28.91**	**31.41**	**0.09**
2001	53.46	33.92	0.00	55.40	33.92	0.00
2002	34.33	32.67	0.03	34.58	32.67	0.03
2003	41.69	35.17	0.01	40.67	35.17	0.01
2004	38.42	35.17	0.02	40.08	35.17	0.02
2005	**27.73**	**37.65**	**0.32**	**31.77**	**37.65**	**0.16**
2006	59.21	38.89	0.00	66.96	38.89	0.00
2007	64.96	40.11	0.00	70.67	40.11	0.00
2008	82.39	40.11	0.00	90.06	40.11	0.00
2009	45.39	40.11	0.01	50.64	40.11	0.00
2010	79.05	43.77	0.00	91.69	43.77	0.00
2011	93.98	46.19	0.00	112.68	46.19	0.00
2012	129.55	48.60	0.00	153.76	48.60	0.00
2013	162.87	51.00	0.00	195.43	51.00	0.00
2014	171.11	52.19	0.00	200.82	52.19	0.00

	Modelos de distribución de probabilidad					
	$DGPIV(\sigma, \gamma, \alpha)$			$SGW(\alpha, \beta, \nu)$		
Año	χ^2	Umb.	p val.	χ^2	Umb.	p val.
2015	175.82	52.19	0.00	213.49	52.19	0.00
2016	196.22	54.57	0.00	235.61	54.57	0.00
2017	168.79	53.38	0.00	200.27	53.38	0.00
2018	136.86	48.60	0.00	158.74	48.60	0.00

Cuadro 4.25: Años en los que no se puede rechazar la hipótesis nula H_0*: los datos siguen el modelo de distribución seleccionado* al 5 por ciento de significancia, para los modelos de distribución de probabilidad de 2 y 3 parámetros con mejor ajuste a los datos.

Distribución	Años	
	Códigos IPC	Códigos CPC
Zipf - Mandelbrot	1978 - 1984 1987 -1989 1991	1978 - 1981 1983, 1984, 1986, 1990 1995
Beta- Geométrica	1978 - 1984 1986-1993	1978, 1979, 1981 1983, 1984, 1990 1995
Pareto discreta generalizada tipo II	1979 - 1984 1987 - 1989 1991, 1993	1978, 1979, 1981 1983, 1984, 1986 1991 - 1995
Pareto discreta generalizada tipo IV	1979 -2001 (excepto 1985)	1984 -1986 1988- 1995 2005
Waring generalizada desplazada	1979 -1984 1986-2001	1992 -2000 2005

4.4. Conclusiones

En el presente capítulo se analiza, por primera vez en la literatura hasta lo que se conoce, la distribución de frecuencia de uso de códigos tecnológicos de patentes, considerando todo el rango de su distribución, y teniendo en cuenta la naturaleza discreta de dicha variable aleatoria.

El análisis realizado extiende, de forma interdisciplinar, la literatura existente en el análisis de la frecuencia de aparición de las palabras en el lenguaje, al ámbito de la frecuencia de aparición de los códigos en el "lenguaje"de las combinaciones de tecnologías en la invención.

Para llevar a cabo el análisis, se incluyeron modelos de distribución de la jerarquía de distribuciones discretas de Pareto: tipo III y IV, y la distribución de tres parámetros Waring generalizada desplazada, novedosos hasta lo que se conoce en este ámbito de investigación.

Como resultado del análisis realizado, hemos encontrado que la distribución Pareto discreta generalizada de tipo IV ha sido el modelo de distribución que mejor ajuste ha proporcionado para la variable de interés (frecuencia de uso de códigos tecnológicos), en todo su rango, tanto para códigos IPC como para códigos CPC, durante el periodo 1978-2018, en comparación con otros 12 modelos de distribución discreta muy conocidos en la literatura.

Dichos resultados positivos de la distribución Pareto discreta generalizada tipo IV ha sido reforzados con la superación del test Chi-Cuadrado en muchos de los años considerados.

Como futuras líneas de investigación, en primer lugar, indicar que este estudio está abierto a su ampliación con nuevos modelos de distribución, que puedan mejorar los resultados obtenidos con la distribución Pareto IV. Así mismo, teniendo en cuenta que los años en los que dicha distribución superó el test Chi-cuadrado pertenecen a la primera mitad del periodo de análisis, en los que el tamaño muestral de la misma fue menor, un interesante estudio futuro aprovechando los datos utilizados en este capítulo es el análisis de la relación entre el tamaño muestral con los límites de dicho test - como referencia de ello, se citan algunos trabajos que abordan cómo los test de significancia son sensibles comunmente al tamaño muestral (Bergh, 2015; Ferguson, 2017; Spruill III, 1973).

Capítulo 5

Redes de códigos tecnológicos en patentes.

5.1. Introducción

El análisis de redes tiene como objetivo identificar patrones en la forma global de las redes e interpretar caracteristicas desconocidas. En el ámbito de la innovación, son numerosas las redes que emergen, y que por tanto, han sido objeto de esta perspectiva de análisis. Entre estas redes, podemos destacar las relaciones de colaboración entre inventores (Goossen y Paruhuri, 2022; Soda et al, 2021; Bauer et al.,2022), aplicantes de pantentes que a su vez son actores de los sistemas de innovación, por ejemplo, empresas y universidades (Xiao et al, 2021; Bianchi et al., 2021; Françoso y Vonortas, 2022), el estudio del el ámbito geográfico de los invenciones, a partir de la localización o procedencia de los inventoress y aplicantes (Yu et el., 2021; Ye, 2021; Li et al.,2021), las redes de citas entre patentes y publicaciones científicas (Kim et al., 2021; Malhotr et al., 2021) y las redes de códigos tecnológicos en patentes (Gao et al., 2018; Leng et al., 2021; Son y Cho, 2020; Xiao et al., 2021), que son comunmente usadas para caracterizar los procesos de innovación y fusión de tecnologías.

En este estudio se exploran las redes de códigos tecnológicos IPC en aplicaciones de patentes PCT desagregadas por actividades económicas productivas segun las divisiones NACE (nivel 2 - digítos) en el periodo 1977 -2018. Para ello, fueron estimados indicadores e interpretados sus significados aplicando el concepto de teoría de redes (Cohen y Havlin, 2010; Dorogovtsev, 2010; Easly y Kleinberg, 2012; Latora, Nicosia y Giovanni, 2017).

El documento esta organizado de la siguiente manera. En el apartado 5.2 se presentan los datos utilizados, se detalla la metodología seguida para la construcción de las redes de códigos tecnológicos, y las características de las mismas; igualmente, se

describen los indicadores de red estimados. El apartado 5.3 presenta los resultados del ejercicio de exploración de las redes de códigos tecnológicos. En el apartado 5.4 son presentadas las conclusiones.

5.2. Datos y metodología

En el análisis de las redes de códigos tecnológicos en patentes se utilizó las base de datos de formato largo "BaseLarga_IPC_Final"construida a partir del fichero de datos "202107_PCT_IPC.txt" de la base de datos de patentes REGPAT publicada por la OCDE, versión julio de 2021. La base de datos usada proporciona información acerca de los códigos tecnológicos IPC asignados a un total de 3786795 aplicaciones de patente PCT, con fecha de prioridad entre los años 1977 y 2018.

La base de datos incluye las variables Número de publicación PCT (*pct_nbr*), Año de prioridad (*prio_year*) y Códigos IPC (*IPC*) asignados; y contiene 13615554 observaciones correspondientes a 72607 códigos tecnológicos IPC diferentes, a nivel de subgrupo en la clasificación IPC. El proceso para obtener la base de datos de códigos tecnológicos IPC se detalla en el apartado 2.1.4 del capítulo de Datos.

En este estudio, las redes complejas analizadas se caracterizaron por los códigos tecnológicos IPC, que constituyen los nodos (o vértices) de dichas redes, y por el hecho de que dos nodos hayan formado parte de una misma patente, que constituye un enlace entre ambos.

Para realizar dicho análisis, se consideraron las divisiones NACE de los códigos tecnológicos, es decir, se usó la concordancia entre las mencionadas divisiones NACE (nivel 2-dígitos) y los códigos IPC (ver Van Looy, Vereyen, Schmoch 2015, pp. 8-11, table 2). Tal como se menciona en el apartado 3.2.6, esta concordancia permitió efectuar una partición de la base de datos de patentes en 27 subconjuntos, a partir de la cual se estudiaron la 27 subredes complejas correspondientes a cada una de dichas actividades económicas productivas. Por tanto, el análisis se llevó a cabo considerando los nodos (códigos tecnológicos) de una misma división NACE enlazados por una misma patente. Es preciso indicar que no se consideraron en la red aquellos códigos tecnológicos que formaron parte de forma individual (en solitario) en aplicaciones de patente PCT.

Tal como se detalló en el mencionado apartado 3.2.6 de esta tesis doctoral, es preciso indicar que hay códigos tecnológicos IPC que según su simbología en la clasificación de patentes IPC pueden corresponder a una división NACE u otra, si están acom-

pañados (o no) de un código tecnológico con símbolo de una subclase tecnológica específica. Por tanto, estos códigos IPC son considerados como códigos diferentes en términos de la correspondencia NACE, y por tanto, contabilizados en varias divisiones NACE, siendo la suma de los códigos IPC en las divisiones NACE (un total de 75202) superior al total de códigos IPC (un total de 72607) considerando la base de datos completa.

A partir de la base de datos anterior, se construyeron las listas de adyacencia para cada división NACE, es decir, para cada división NACE: en primer lugar, se creó una matriz cuadrada simétrica con tantas filas y columnas como códigos tecnológicos de dicha división NACE, códigos utilizados en alguna patente en el periodo considerado; a continuación, se completó dicha matriz con un valor de 0 en la fila/-columna correspondiente a cada pareja de códigos que no hubieran formado parte de una misma patente, y con un valor de 1 si dicha pareja de códigos sí formaron parte de una misma patente, identificando la existencia de un enlace en la red compleja con dicho valor 1 (indicar que la diagonal de dicha matriz se mantuvo con el valor cero en todos los casos, no considerando un código como enlazado consigo mismo); y finalmente, se eliminaron aquellas filas/columnas correspondientes a códigos aislados, que no compartieron patente alguna con los demás códigos en el periodo analizado. Para realizar dicha matriz de adyacencia, se consideraron los enlaces entre los códigos tecnológicos IPC pertenecientes a una misma división NACE, durante los 42 años del periodo de estudio (1977- 2018).

En términos matemáticos, denotamos la matriz de adyacencia como

$$\boldsymbol{A} = (a_{ij}),\ i, j \in \{1, ...N\}, \tag{5.1}$$

una matriz cuadrada simétrica $N \times N$, siendo N el número de códigos tecnológicos de una red NACE en particular, y teniendo que $a_{ij} = 1$ si existe un enlace que conecte el nodo i con el nodo j. De no ser así, si los nodos (códigos) i y j no se encuentran enlazados (no formaron parte de una misma patente), tendríamos que $a_{ij} = 0$.

Destacar que la red de códigos tecnológicos de patentes considerada es una red compleja no dirigida, dado que en nuestro caso, el sentido del enlace no es relevante (si un código ha compartido una patente con otro, el segundo código lógicamente también lo ha hecho con el primero), siendo por tanto la matriz de adyacencia simétrica, $A = A^T$, donde A^T expresa la transpuesta de la matriz A. Así mismo, para nuestro análisis, consideramos una red compleja no ponderada, dado que no se contabiliza el número de veces que dos códigos han compartido una misma patente, y sólo se contabiliza el hecho de que dichos códigos hayan compartido al menos una

patente.

Para el análisis de las redes de códigos tecnológicos anteriormente descritas, fueron realizadas dos medidas de red: por un lado, se analizó el Grado de Centralidad (*Degree Centrality*) de cada nodo (ver página 168, Newman, 2010), con el objetivo de identificar los códigos tecnológicos más importantes, en términos de combinación con el resto de códigos de su misma actividad económica productiva NACE; y por otro lado, se analizó la Densidad de Red (*Density of a Network*), definida como la la proporción de relaciones existentes (enlaces) entre los nodos que conforman una red con respecto al número de posibles conexiones (ver página 29, Jackson, 2008), como medida global de cada subred NACE, y como indicación de lo cerca o alejada que cada una de dichas subredes se encontraba de una red completa (red en la que todos los nodos están conectados, *Complete Network*).

Cuando hablamos del *Grado de Centralidad* de un nodo, o simplemente *Grado* de un nodo, hacemos referencia al número de enlaces o conexiones que tiene un nodo i perteneciente a una red. Entendemos que, un nodo i, en nuestro caso, un código tecnológico IPC tiene un mayor grado de centralidad en la medida en que está enlazado con un número mayor de nodos en la red. El Grado de Centralidad k_i de un nodo i en una red, puede obtenerse a partir de la información disponible en la matriz de adyacencia $N \times N$ (definida en 5.1), a partir de la siguiente expresión (Gómez, 2019):

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{5.2}$$

Con respecto a la *Densidad* de una red, tal como se indicó, mide la proporción de relaciones existentes (enlaces) entre los nodos que conforman una red con respecto al número de posibles conexiones. Teniendo en cuenta que el número de posibles conexiones P_N entre los N nodos de una red no dirigida puede calcularse a partir de la siguiente expresión:

$$P_N = \frac{N(N-1)}{2}, \tag{5.3}$$

la densidad D_m de una subred de m códigos NACE se puede obtener como resultado de dividir el número de enlaces existente en dicha subred de interés L entre el número de enlaces posibles P_m, a partir de la siguiente expresión:

$$D_m = \frac{L}{P_m} = \frac{L}{\frac{m(m-1)}{2}} = \frac{2L}{m(m-1)} \tag{5.4}$$

Destacar que dicha medida de densidad D_m también puede obtenerse dividiendo el grado medio de centralidad de dicha subred entre el valor de $(m-1)$.

Para finalizar, indicar que tanto los cálculos como las gráficas del presente trabajo, se realizaron utilizando el software `R` (Sarabia, Prieto y Jordá, 2018), en particular utilizando las librerías `igraph` (Csardi y Nepusz, 2006; Csardi y Nepusz, 2010) y `network` (Butts, 2008).

5.3. Resultados

En el cuadro 5.1 se presenta el número total de códigos tecnológicos IPC para cada división NACE, que correspondería en este estudio con el número de nodos de cada subred (se excluyen aquellos códigos tecnológicos IPC que no se enlazan con códigos de su misma divisón NACE), así como el número total de enlaces en la red. Cabe indicar que, en cuanto al número de códigos tecnológicos IPC, la división NACE 28 Manufacture of Machinery and Equipment N.E.C ocupa el primer lugar con 19391 códigos IPC diferentes y una red conformada por 19216 nodos que se enlazan un total de 662871 veces.

En cuanto al número de nodos en la red, la segunda posición es ocupada por la división NACE 26 Manufacture of Computer, Electronic and Optical Products con 10967 nodos, y en tercer lugar, la división NACE 20 Manufacture of Chemicals and Chemical Products con 10359 nodos. Las divisiones NACE con menor número de nodos son las divisiones NACE 62 Computer Programming, Consultancy and Related Activities con 58 nodos, y la división NACE 16 Manufacture of Wood and of Products of Wood and Cork,except Furniture; Manufacture of Articles of Straw and Plaiting Materials con 63 nodos.

Entre las redes analizadas aquella con un mayor número de enlaces fue la red de códigos tecnológicos IPC correspondiente a la división NACE 21 Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations con un total de 1505494 enlaces. Por otro lado, la red de códigos tecnológicos IPC con menor número de enlaces es la división NACE 16 Manufacture of Wood and of Products of Wood and Cork,except Furniture; Manufacture of Articles of Straw and Plaiting Materials

con 464 enlaces.

En el cuadro 5.2 se presenta la estimación (en porcentaje) de la intensidad de la red de códigos tecnológicos IPC para cada una de las divisiones NACE (nivel 2-dígitos) durante el periodo 1977 -2018. La subred de códigos tecnológicos IPC con menor densidad es la división NACE 28 Manufacture of Machinery and Equipment N.E.C. con un proporción de 0.0036, de los enlaces existentes con respecto a los enlaces posibles en dicha subred.

Tal como se puede observar en el cuadro 5.2 y en la figura 5.1, hay un grupo de redes de códigos con una estimación de dicha densidad de red inferior al 0.02. En este grupo se encuentran las siguientes divisiones: NACE 27 Manufacture of Electrical Equipment (0.0115), NACE 25 Manufacture of Fabricated Metal Products, except Machinery and Equipment (0.0124), NACE 26 Manufacture of Computer, Electronic and Optical Products (0.0134), NACE 32 Other Manufacturing (0.0147) y NACE 30 Manufacture of Other Transport Equipment (0.0151).

Por otro lado, la subred de códigos IPC con mayor densidad fue la correspondiente a la actividad económica productiva NACE 62, con un valor de densidad del 0.8482, extraordinariamente alto y cercano al valor máximo de 1 que correspondería a una red completa (en la que todos los códigos de dicha división NACE se habrían combinado con todos los demás en alguna de las patentes existentes).

Cuadro 5.1: Número de nodos y enlaces de cada subred NACE (nivel 2-dígitos) de códigos IPC. Periodo 1977-2018.

NACE	Descripción	Códigos	Nodos	Enlaces
10	Manufacture of Food Products	845	839	29234
11	Manufacture of Beverages	161	157	2475
12	Manufacture of Tobacco Products	184	175	1692
13	Manufacture of Textiles	673	668	29148
14	Manufacture of Wearing Apparel	202	200	2926
15	Manufacture of Leather and Related Products	242	238	5001
16	Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials	65	63	464
17	Manufacture of Paper and Paper Products	410	406	12751
18	Printing and Reproduction of Recorded Media	201	197	3659
19	Manufacture of Coke and Refined Petroleum Products	408	406	13273
20	Manufacture of Chemicals and Chemical Products	10401	10359	979922
21	Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations	6353	6342	1505494
22	Manufacture of Rubber and Plastic Products	1148	1141	49154
23	Manufacture of Other Non-Metallic Mineral Product	1400	1391	56543
24	Manufacture of Basic Metals	926	921	32345
25	Manufacture of Fabricated Metal Products, except Machinery and Equipment	2857	2811	49022
26	Manufacture of Computer, Electronic and Optical Products	11071	10967	809801
27	Manufacture of Electrical Equipment	5618	5553	178272
28	Manufacture of Machinery and Equipment N.E.C.	19391	19216	662871
29	Manufacture of Motor Vehicles,Trailers and Semi-Trailers	3076	3067	118970
30	Manufacture of Other Transport Equipment	2169	2114	33789
31	Manufacture of Furniture	628	623	11107
32	Other Manufacturing	4997	4947	180057
42	Civil Engineering	190	188	1489
43	Specialised Construction Activities	749	746	23558
62	Computer Programming, Consultancy and Related Activities	59	58	1402
Co_IPC	IPC codes not elsewhere classified	778	760	12820

Cuadro 5.2: Densidad de cada subred NACE (nivel 2-dígitos) de códigos IPC. Periodo 1977-2018.

NACE	Descripción	Densidad
10	Manufacture of Food Products	0.0832
11	Manufacture of Beverages	0.2021
12	Manufacture of Tobacco Products	0.1111
13	Manufacture of Textiles	0.1308
14	Manufacture of Wearing Apparel	0.1470
15	Manufacture of Leather and Related Products	0.1773
16	Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials	0.2376
17	Manufacture of Paper and Paper Products	0.1551
18	Printing and Reproduction of Recorded Media	0.1895
19	Manufacture of Coke and Refined Petroleum Products	0.1614
20	Manufacture of Chemicals and Chemical Products	0.0183
21	Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations	0.0749
22	Manufacture of Rubber and Plastic Products	0.0756
23	Manufacture of Other Non-Metallic Mineral Product	0.0585
24	Manufacture of Basic Metals	0.0763
25	Manufacture of Fabricated Metal Products, except Machinery and Equipment	0.0124
26	Manufacture of Computer, Electronic and Optical Products	0.0135
27	Manufacture of Electrical Equipment	0.0116
28	Manufacture of Machinery and Equipment N.E.C.	0.0036
29	Manufacture of Motor Vehicles,Trailers and Semi-Trailers	0.0253
30	Manufacture of Other Transport Equipment	0.0151
31	Manufacture of Furniture	0.0573
32	Other Manufacturing	0.0147
42	Civil Engineering	0.0847
43	Specialised Construction Activities	0.0848
62	Computer Programming, Consultancy and Related Activities	0.8482
Co_IPC	IPC codes not elsewhere classified	0.0444

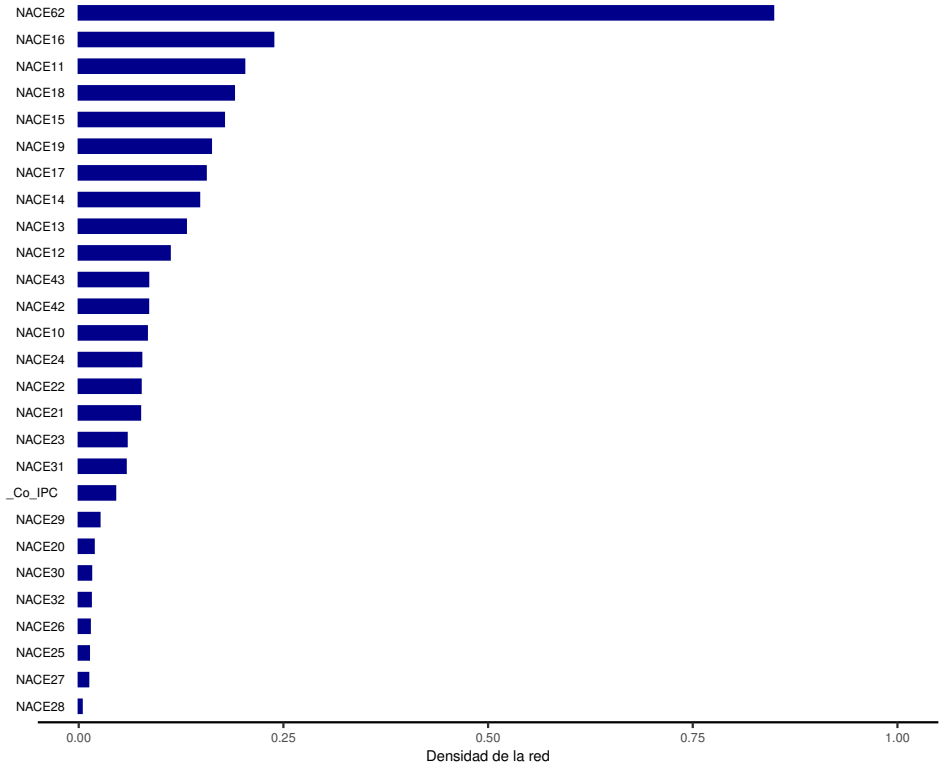

Figura 5.1: Ranking de divisiones NACE en términos de la densidad de la red de códigos tecnológicos IPC. Periodo 1977 -2018.

Con respecto a la medida de Grado de Centralidad de nodo, que se realizó con el objetivo de identificar a los nodos más centrales de cada subred, en el cuadro 5.3 se listan los 5 principales nodos de las redes de códigos tecnológicos IPC para actividades económicas productivos de acuerdo con la correspondencia IPC - divisiones NACE (nivel 2 - dígitos).

Cuadro 5.3: Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018.

Ranking	**Código IPC**	**Grado Centralidad**	**Código IPC**	**Grado Centralidad**	**Código IPC**	**Grado Centralidad**
	NACE 10		**NACE 11**		**NACE 12**	
1	A23L001/00	472	C12C011/00	92	A24F047/00	132
2	A23L001/30	421	A23L002/38	89	A24B015/16	66
3	A23G003/34	371	C12G001/00	89	A24B015/28	64
4	A23G003/00	344	C12G003/02	89	A24D003/04	63
5	A23K001/16	329	C12G003/00	85	A24B013/00	59

Cuadro 5.3: Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018.

Ranking	Código IPC	Grado Centralidad	Código IPC	Grado Centralidad	Código IPC	Grado Centralidad
	NACE 13		**NACE14**		**NACE 15**	
1	D06M015/263	358	A41D013/00	123	A43B013/14	147
2	D06M015/643	329	A41D031/00	101	A43B003/00	145
3	D06M013/02	325	A41D001/00	97	A43B003/24	145
4	D06M101/00	324	A41D027/00	96	A43B023/02	143
5	D04H001/42	321	A41D013/05	94	A43B005/00	142
	NACE 16		**NACE17**		**NACE 18**	
1	B27M003/00	37	D21H027/00	236	B42D015/00	140
2	B27N003/04	33	D21H021/16	222	B41M003/14	134
3	B27M001/02	30	D21H017/67	220	B42D015/10	121
4	B27N003/00	29	D21H021/18	205	B41M007/00	114
5	B27D001/04	29	D21H017/37	199	B41M005/00	107
	NACE 19		**NACE20**		**NACE 21**	
1	C10G045/02	242	C07B061/00	3999	A61P035/00	5079
2	C10G003/00	237	C08L101/00	2949	A61P043/00	5029
3	C10G009/00	215	C08K005/00	2775	A61P029/00	4739
4	C10G045/00	214	C09K003/00	2659	A61P009/10	4243
5	C10G002/00	211	A61Q019/00	2437	A61P025/28	4233
	NACE 22		**NACE23**		**NACE 24**	
1	B29C047/00	578	C04B028/02	456	C22C038/00	419
2	B29C035/02	535	C04B038/00	424	C22C001/02	344
3	B29C035/08	514	C04B111/00	411	C22C038/02	340
4	B29C045/00	484	B32B009/00	405	B22D011/00	333
5	B29C065/00	467	C04B041/50	386	C22F001/00	330
	NACE 25		**NACE26**		**NACE 27**	
1	E05B047/00	381	H04L029/06	2351	H02J007/00	1001
2	E05B017/00	346	H04L029/08	2177	H05B037/02	678
3	E05B065/00	291	G06F019/00	2076	H01M002/10	655
4	E05B063/00	278	H04N005/225	2057	F21Y115/10	632
5	E05B015/00	261	H04L012/28	1954	F21S002/00	591

Cuadro 5.3: Ranking de nodos en la red de códigos tecnológicos IPC según división NACE (Nivel - 2 dígitos) en términos del Grado de Centralidad. Periodo 1977 - 2018.

Ranking	**Código IPC**	**Grado Centralidad**	**Código IPC**	**Grado Centralidad**	**Código IPC**	**Grado Centralidad**
	NACE 28		**NACE29**		**NACE 30**	
1	B23D061/00	1034	F02D045/00	732	B64C039/02	401
2	B23K026/00	974	F02D041/00	710	B64C029/00	264
3	B01D046/04	953	F02B075/02	666	B63B035/44	237
4	B08B003/02	884	B60R016/02	664	B64C027/08	231
5	B23P015/00	806	B60W010/06	618	B63B035/00	230
	NACE 31		**NACE32**		**NACE 42**	
1	A47C007/00	166	A61B005/00	1097	E03B001/04	48
2	A47C007/02	150	A61B019/00	937	E03B007/07	42
3	A47F005/00	140	A61B017/00	891	E02B003/10	42
4	A47C013/00	138	A61M001/00	730	E03C001/04	40
5	A47C031/00	130	A61F002/00	710	E02B007/20	40
	NACE 43		**NACE62**		**Co_IPC**	
1	E04B001/00	319	G06Q040/00	57	A23L029/00	162
2	E04F013/08	293	G06Q030/02	57	B29L031/00	158
3	E04B001/343	278	G06Q050/10	57	A23L033/00	156
4	E04B001/38	271	G06Q050/00	57	A23L019/00	149
5	E04C002/38	263	G06Q030/04	57	A23L027/00	148

A modo ilustrativo, en las figuras 5.2, 5.3 y 5.4 se muestran las subredes de códigos tecnológicos IPC formadas por los enlaces entre los 5 principales nodos (con mayor Grado de Centralidad) y los códigos con los que dichos nodos están enlazados (para mayor claridad no se han incluido los enlaces entre los nodos no principales), para las divisiones NACE 16 Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting Materials, NACE 42 Civil Engineering y NACE 18 Printing and Reproduction of Recorded Media, respectivamente.

La subred correspondiente a la división NACE 16 (figura 5.2), contiene 55 diferentes nodos. En dicha figura, los nodos principales son identificados con un número, que en este caso, representa su Grado de Centralidad. B27M0003/00= 37; B27N003/04=33; B27M001/02=30; B27N003/00 =29; y B27D001/04=29. Así mismo, dicho Grado de Centralidad también viene reflejado en el tamaño de los nodos.

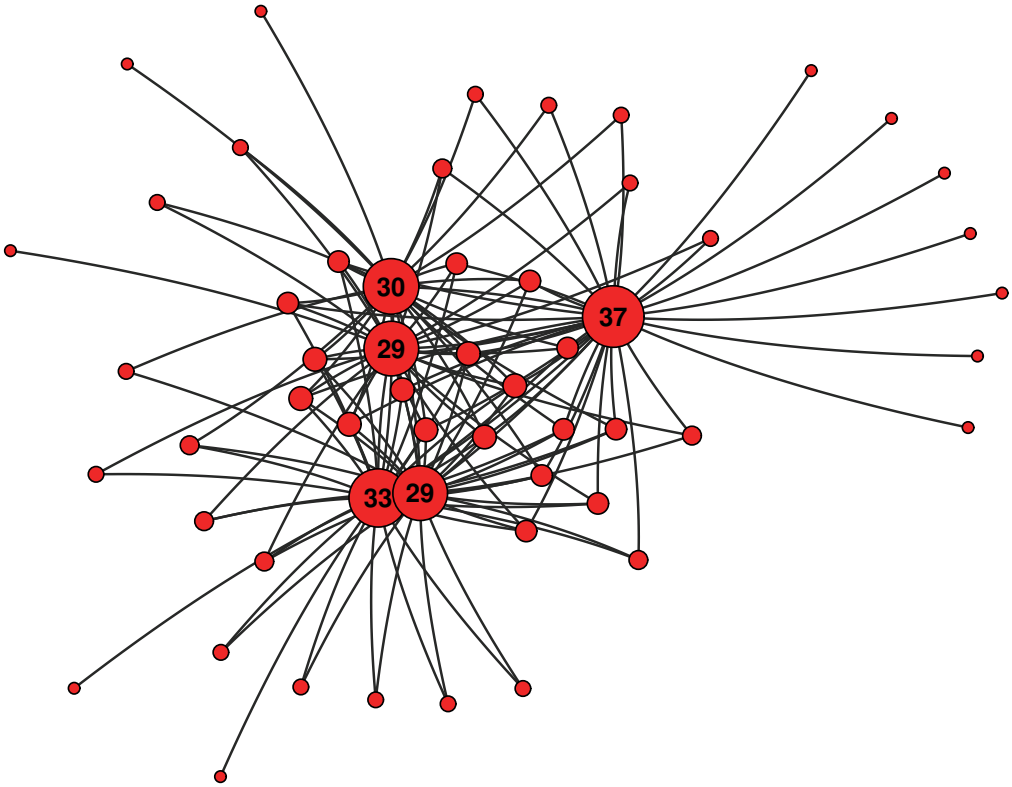

Figura 5.2: Subred correspondiente a la división NACE 16 Manufacture of Wood and of Products of Wood and Cork, except Furniture; Manufacture of Articles of Straw and Plaiting (nivel 2 - dígitos). Enlaces de los 5 principales nodos en términos del Grado de Centralidad. Periodo 1977 - 2018.

Por otro lado, en la figura 5.3 se muestra la subred correspondiente a la división NACE 42 Civil Engineering, incluyendo tan solo los enlaces a los nodos principales (más centrales) de la red. Cabe resaltar que la red representada está conformada por 123 nodos de los 188 nodos que comprende la red completa de esta división NACE. En la figura 5.3, los nodos principales son identificados con un número que representa una posición en la red, y que en este caso, corresponde a los siguientes códigos tecnológicos: **(1)**E03B007/07; **(8)** E03C001/04; **(12)** E02B007/20; **(16)** E03B001/04; y **(24)** E02B003/10.

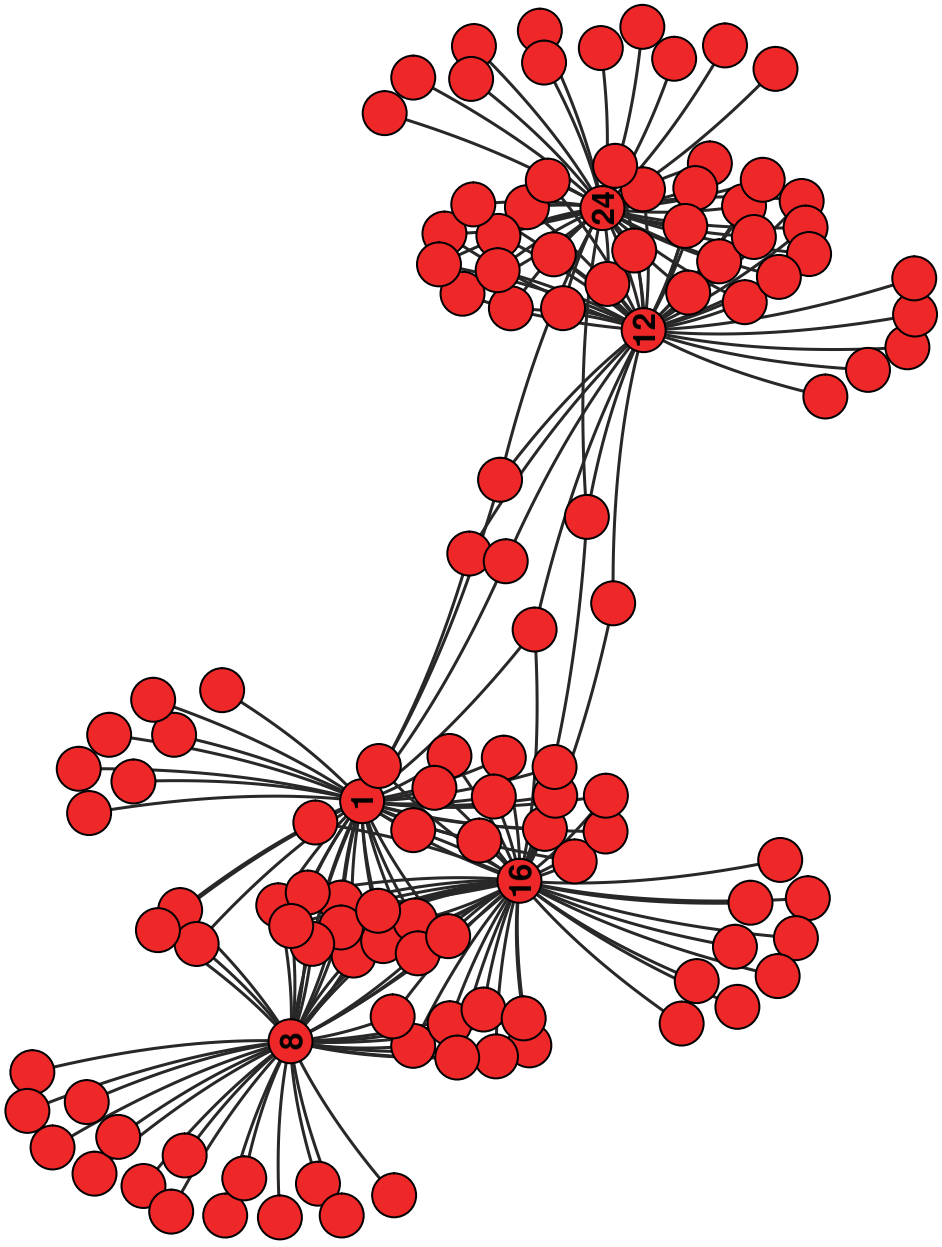

Figura 5.3: Subred correspondiente a la división NACE 42 Civil Engineering (nivel 2 - dígitos). Enlaces de los 5 principales nodos, en términos del Grado de Centralidad. Periodo 1977 - 2018.

En la figura 5.4 se muestra la subred correspondiente a la división NACE 18 Printing and Reproduction of Recorded Media, incluyendo solamente los enlaces a los nodos principales (más centrales) de la red. En este ejemplo, observamos una red con un número de nodos más amplio, siendo necesario para su correcta visualización un esquema circular y un tamaño de nodos (círculo rojo) más pequeño.

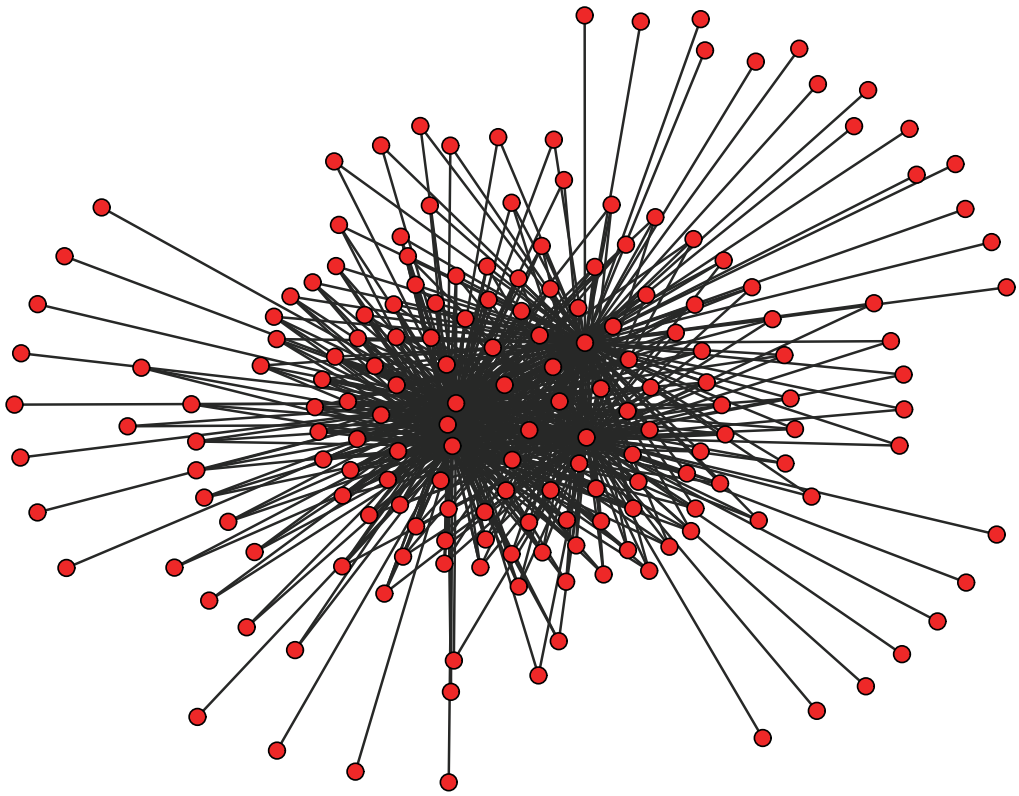

Figura 5.4: Subred correspondiente a la división NACE 18 Printing and Reproduction of Recorded Media (nivel 2 - dígitos). Enlaces a los 5 principales nodos, en términos del Grado de Centralidad. Periodo 1977 - 2018.

Para finalizar, en la figura 5.5 se muestra la red de códigos tecnológicos IPC correspondiente a la totalidad de los nodos principales (más centrales) del cuadro 5.3 para el periodo 1977 - 2018. La red construida tiene un total de 135 nodos y 2.344 enlaces.

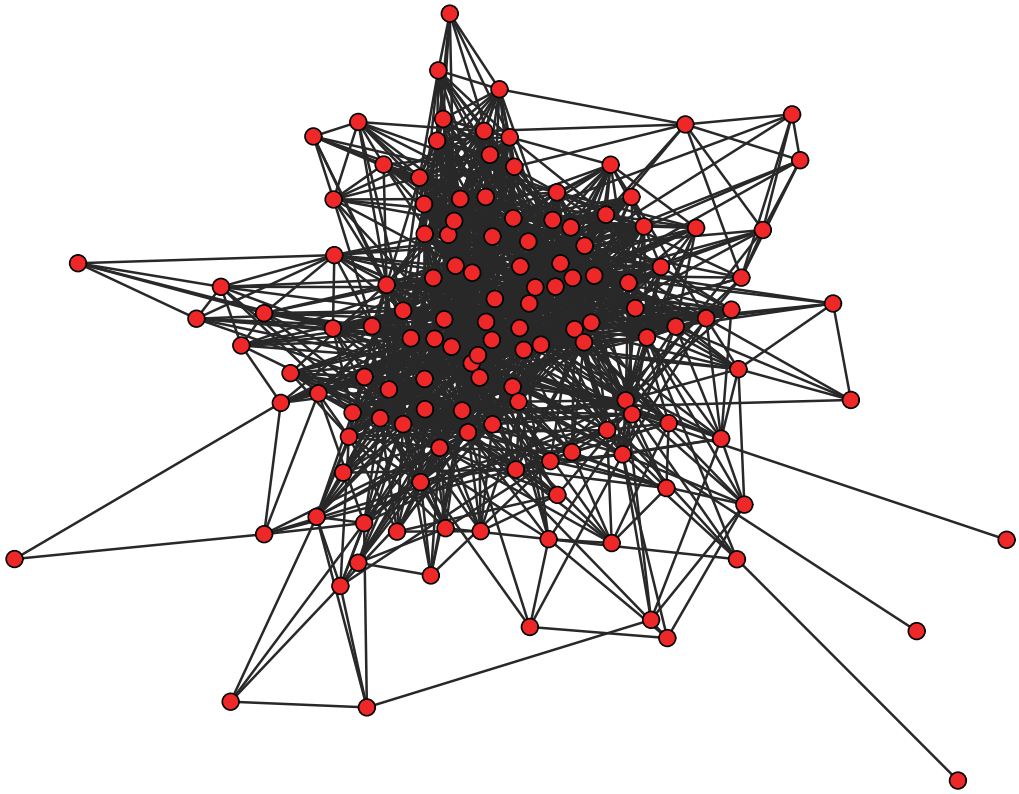

Figura 5.5: Red de códigos tecnológicos IPC conformada a partir de los enlaces entre los nodos principales (mayor grado de centralidad) de las divisiones NACE analizadas. Periodo 1977 - 2018.

5.4. Conclusiones

Dentro del marco general de las Ciencias de la Complejidad, y en particular dentro de la Teoría de la Redes SocioEconómicas Complejas, en este capítulo se exploran las características estructurales de las redes de códigos tecnológicos en patentes clasificadas por actividades económicas productivas NACE.

En este trabajo, se explora la red compleja constituida por dichos códigos tecnológicos como nodos, agrupados por actividades económicas productivas NACE, y se considera como enlace entre dos nodos (códigos) el hecho de haberse combinado para formar una patente.

Dicho análisis se realiza en base a dos medidas de red: la densidad de red, definida

en términos de la proporción de enlaces (combinaciones de códigos tecnológicos) existentes frente al total de enlaces posibles; y la centralidad de grado de cada nodo, definida por el número de enlaces de cada código tecnológico.

Como resultado del presente estudio, se muestra el ranking de divisiones NACE en términos de la densidad de la subred correspondiente, en el que se encuentra una gran disparidad en los valores obtenidos, con subredes como la correspondiente a la división NACE 62 que alcanzó un valor de 0.8482 (muy cercano al valor 1 correspondiente a una red completa) y otras con valores de densidad de red inferiores al 0.02.

Así mismo, como fruto del presente estudio, se identifican los códigos tecnológicos más centrales dentro de cada división NACE (en términos del número de combinaciones con otros códigos de su actividad económica).

Referencias

Abernathy, W. (s.f.). "the productivity dilemma: roadblock to innovation in the automobile industry". *John Hopkins University, Baltimore*.

Abernathy, W. J., y Utterback, J. M. (1978). Patterns of industrial innovation. *Technology review*, *80*(7), 40–47.

Adrián Risso, W., y Sánchez Carrera, E. J. (2019). On the impact of innovation and inequality in economic growth. *Economics of innovation and new technology*, *28*(1), 64–81.

Ajiferuke, I. (1991). A probabilistic model for the distribution of authorships. *Journal of the American Society for Information Science*, *42*(4), 279–289.

Allen, T. J. (1977). Managing the flow of technology cambridge. *Masachusetts: The MIT*.

Allison, P. D. (1978). Measures of inequality. *American sociological review*, 865–880.

Alvarez, I., y Molero, J. (2005). Technology and the generation of international knowledge spillovers: An application to spanish manufacturing firms. *Research Policy*, *34*(9), 1440–1452.

Archibugi, D., y Michie, J. (1995). The globalisation of technology: a new taxonomy. *Cambridge journal of Economics*, *19*(1), 121–140.

Arnold, B. (2015). *Pareto distributions 2nd ed.* CRC Press. Taylor & Francis Group.

Arnold, B. C., y Sarabia, J. M. (2018). *Majorization and the Lorenz order with applications in applied mathematics and economics* (Vol. 7). Springer.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. En *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton University Press.

Arthur, W. B. (2007). The structure of invention. *Research policy*, *36*(2), 274–287.

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Arts, S., y Veugelers, R. (2015). Technology familiarity, recombinant novelty, and breakthrough invention. *Industrial and Corporate Change*, *24*(6), 1215–1246.

Audretsch, D. B., y Keilbach, M. (2008). Resolving the knowledge paradox: knowledge-spillover entrepreneurship and economic growth. *Research Policy*, *37*(10), 1697–1705.

Baptista, R. (2001). Geographical clusters and innovation diffusion. *Technological Forecasting & Social Change*, *66*(1), 31–46.

Bauer, V., Harhoff, D., y Kauermann, G. (2022). A smooth dynamic network model for patent collaboration data. *AStA advances in statistical analysis*, *106*(1), 97–116.

Bergh, D. (2015). Sample size and chi-squared test of fit- a comparison between a random sample approach and a chi-square value adjustment method using swedish adolescent data. En *Pacific rim objective measurement symposium (proms) 2014 conference proceedings* (pp. 197–211).

Bianchi, C., Galaso, P., y Palomeque, S. (2021). Patent collaboration networks in latin america: Extra-regional orientation and core-periphery structure. *Journal of scientometric research*, *10*(1s), s59–s70.

Biurrun, A. (2020). New evidence toward solving the puzzle of innovation and inequality. the role of institutions. *Economics of Innovation and New Technology*, 1–22.

Bogliacino, F., y Pianta, M. (2010). Innovation and employment: a reinvestigation using revised pavitt classes. *Research Policy*, *39*(6), 799–809.

Brown, M. C. (1994). Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on alberta data. *Social science & medicine*, *38*(9), 1243–1256.

Burnham, K. P., y Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261–304.

Burnham, K. P., Anderson, D. R., y Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology*, *65*(1), 23–35.

Burrell, Q. L. (2005). The use of the generalized waring process in modelling informetric data. *Scientometrics*, *64*(3), 247–270.

Butts, C. T. (2008). network: a package for managing relational data in R. *Journal of statistical software*, *24*, 1–36.

Cagetti, M., y De Nardi, M. (2008). Wealth inequality: Data and models. *Macroeconomic Dynamics*, *12*(S2), 285–313.

Cantwell, J. (1995). The globalisation of technology: what remains of the product cycle model? *Cambridge journal of economics*, *19*, 155–155.

Chae, S., y Gim, J. (2019). A study on trend analysis of applicants based on patent classification systems. *Information*, *10*(12), 364.

Chang, S.-H. (2017). The technology networks and development trends of university-industry collaborative patents. *Technological Forecasting and Social Change*, *118*, 107–113.

Chen, Q., Wang, J., Ou, R., y Tsai, S.-B. (2019). An empirical study on the network model and the online knowledge production structure. *Journal of Information Technology Research (JITR)*, *12*(4), 171–182.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

Clancy, M. S. (2018). Combinations of technology in us patents, 1926–2009: a weakening base for future innovation? *Economics of Innovation and new Technology*, *27*(8), 770–785.

Coco, G., Lagravinese, R., y Resce, G. (2020). Beyond the weights: a multicriteria approach to evaluate inequality in education. *The Journal of Economic Inequality*, *18*(4), 469–489.

Cohen, R., y Havlin, S. (2010). *Complex networks: structure, robustness and function*. Cambridge university press.

Cohen, W. M., y Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, *99*(397), 569–596.

Consul, P., y Jain, G. (1973). On some interesting properties of the generalized poisson distribution. *Biometrische Zeitschrift*, *15*(7), 495–500.

Consul, P., y Shoukri, M. (1984). Maximum likelihood estimation for the generalized poisson distribution. *Communications in statistics-Theory and Methods*, *13*(12), 1533–1547.

Corominas-Murtra, B., Fortuny, J., y Solé, R. V. (2011). Emergence of zipf 's law in the evolution of communication. *Physical Review E*, *83*(3), 036115.

Crespi, F., y Pianta, M. (2008). Diversity in innovation and productivity in europe. *Journal of Evolutionary Economics*, *18*(3), 529–545.

Csardi, G., y Nepusz, T. (2010). igraph reference manual. *URL: http://igraph. sourceforge. net/documentation. html*.

Csardi, G., Nepusz, T., y cols. (2006). The igraph software package for complex network research. *InterJournal, complex systems*, *1695*(5), 1–9.

Dorogovtsev, S. (2010). *Complex networks*. Oxford University Press, Oxford, UK.

Dosi, G. (1991). The research on innovation diffusion: An assessment. En *Diffusion of technologies and social behavior* (pp. 179–208). Springer.

Easley, D., Kleinberg, J., y cols. (2012). Networks, crowds, and markets. *Cambridge Books*.

Ejermo, O., Kander, A., y Henning, M. S. (2011). The R & D-growth paradox arises in fast-growing sectors. *Research policy*, *40*(5), 664–672.

EPO, y USPTO. (2017). *Guide to the CPC (Cooperative Patent Classification.* Autor.

Etzkowitz, H., y Leydesdorff, L. (2000). The dynamics of innovation: from national systems and "mode 2"to a triple helix of university–industry–government relations. *Research policy*, *29*(2), 109–123.

Eurostat, N. (2008). Rev. 2–statistical classification of economic activities in the european community. *Office for Official Publications of the European Communities, Luxemburg*.

Ezcurra, R., Pascual, P., y Rapún, M. (2007). Spatial inequality in productivity in the european union: sectoral and regional factors. *International Regional Science Review*, *30*(4), 384–407.

Fagerberg, J., Srholec, M., y Knell, M. (2007). The competitiveness of nations: Why some countries prosper while others fall behind. *World development*, *35*(10), 1595–1620.

Fagerberg, J., y Verspagen, B. (2009). Innovation studies - the emerging structure of a new scientific field. *Research policy*, *38*(2), 218–233.

Ferguson, T. S. (2017). *A course in large sample theory*. Routledge.

Fisher, R. A., Corbet, A. S., y Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 42–58.

Fleming, L., y Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, *30*(7), 1019–1039.

Françoso, M. S., y Vonortas, N. S. (2022). Gatekeepers in regional innovation networks: Evidence from an emerging economy. *The Journal of Technology Transfer*, 1–21.

Gao, Y., Zhu, Z., Kali, R., y Riccaboni, M. (2018). Community evolution in patent networks: technological change and network dynamics. *Applied network science*, *3*(1), 1–23.

Gisbert, F. J. G., y Notario, A. V. (2009). *Desigualdad y bienestar social: de la teoría a la práctica*. Fundacion BBVA.

Goldschlag, N., Lybbert, T. J., y Zolas, N. J. (2020). Tracking the technological composition of industries with algorithmic patent concordances. *Economics of Innovation and New Technology*, *29*(6), 582–602.

Gómez, S. (2019). Centrality in networks: finding the most important nodes. En *Business and consumer analytics: New ideas* (pp. 401–433). Springer.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, *19*(2), 399–415.

Gómez-Déniz, E., y Calderín-Ojeda, E. (2011). The discrete lindley distribution: properties and applications. *Journal of statistical computation and simulation*, *81*(11), 1405–1416.

Gómez-Déniz, E., Pérez-Rodríguez, J. V., Reyes, J., y Gómez, H. W. (2020). A bimodal discrete shifted poisson distribution. a case study of tourists 'length of stay. *Symmetry*, *12*(3), 442.

Gómez-Déniz, E., Sarabia, J. M., y Calderín-Ojeda, E. (2011). A new discrete distribution with actuarial applications. *Insurance: Mathematics and Economics*, *48*(3), 406–412.

Goossen, M. C., y Paruchuri, S. (2022). Measurement errors and estimation biases with incomplete social networks: replication studies on intra-firm inventor network analysis. *Research Policy*, *51*(1), 104404.

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California management review*, *33*(3), 114–135.

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. En *R&d and productivity: the econometric evidence* (pp. 287–343). University of Chicago Press.

Grossman, G. M., y Helpman, E. (1991). Quality ladders in the theory of growth. *The review of economic studies*, *58*(1), 43–61.

Harrigan, K. R., y Fang, Y. (2019). Financial implications of technology-class code popularity and usage among industry competitors. *Scientometrics*, *121*(1), 25–51.

Huang, J.-Y. (2016). Patent portfolio analysis of the cloud computing industry. *Journal of Engineering and Technology Management*, *39*, 45–64.

i Cancho, R. F., y Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*(3), 788–791.

Irwin, J. O. (1968). The generalized waring distribution applied to accident theory. *Journal of the Royal Statistical Society: Series A (General)*, *131*(2), 205–225.

Irwin, J. O. (1975). The generalized waring distribution. Part I. *Journal of the Royal Statistical Society: Series A (General)*, *138*(1), 18–31.

Jackson, M. O. (2008). *Social and economic networks, princeton, nj: Princeton univ.* Press.

Jeon, J., Chae, S., y Gim, J. (2019). Technical invention trend analysis of applicants based on CPC classification. En *International conference on web engineering* (pp. 10–17).

Johnson, N. L., Kotz, S., y Kemp, A. W. (2005). *Univariate discrete distributions*. John Wiley & Sons.

Jones, C. I. (1995). R & D-based models of economic growth. *Journal of political Economy*, *103*(4), 759–784.

Jordá, V., y Alonso, J. M. (2017). New estimates on educational attainment using a continuous approach (1970–2010). *World Development*, *90*, 281–293.

Kay, L., Newman, N., Youtie, J., Porter, A. L., y Rafols, I. (2014). Patent overlay mapping: Visualizing technological distance. *Journal of the Association for Information Science and Technology*, *65*(12), 2432–2443.

Kemp, A. W. (2008). The discrete half-normal distribution. En *Advances in mathematical and statistical modeling* (pp. 353–360). Springer.

Kim, D., Cerigo, D. B., Jeong, H., y Youn, H. (2016). Technological novelty profile and invention s future impact. *EPJ Data Science*, *5*(1), 1–15.

Kim, G., y Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, *117*, 228–237.

Kim, S., Yoon, S., Raghavan, N., Le, N.-T., y Park, H. (2021). Developmental trajectories in blockchain technology using patent-based knowledge network analysis. *IEEE Access*, *9*, 44704–44717.

Kline, S. J., y Rosenberg, N. (1986). An overview of innovation. the positive sum strategy: Harnessing technology for economic growth. *The National Academy of Science, USA*, *35*, 36.

Klugman, S. A., Panjer, H. H., y Willmot, G. E. (2012). *Loss models: from data to decisions* (Vol. 715). John Wiley & Sons.

Lafond, F., y Kim, D. (2019). Long-run dynamics of the us patent classification system. *Journal of Evolutionary Economics*, *29*(2), 631–664.

Lang, G. (2009). Measuring the returns of R&D - an empirical study of the german manufacturing sector over 45 years. *Research Policy*, *38*(9), 1438–1445.

Latora, V., Nicosia, V., y Russo, G. (2017). *Complex networks: principles, methods and applications*. Cambridge University Press.

Laursen, K., y Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among uk manufacturing firms. *Strategic management journal*, *27*(2), 131–150.

Leng, Z., Sun, H., Cheng, J., Wang, H., y Yao, Z. (2021). China's rare earth industry technological innovation structure and driving factors: A social network analysis based on patents. *Resources Policy*, *73*, 102233.

Leydesdorff, L. (2008). Patent classifications as indicators of intellectual organization. *Journal of the American Society for Information Science and Technology*, *59*(10), 1582–1597.

Li, Y., Zhang, Y., Lee, C.-C., y Li, J. (2021). Structural characteristics and determinants of an international green technological collaboration network. *Journal of Cleaner Production*, *324*, 129258.

Lobo, J., y Strumsky, D. (2019). Sources of inventive novelty: two patent classification schemas, same story. *Scientometrics*, *120*(1), 19–37.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, *22*(1), 3–42.

Lucas Jr, R. E. (1993). Making a miracle. *Econometrica: Journal of the Econometric Society*, 251–272.

Lundvall, B.-A. (1992). National systems of innovation: towards a theory of innovation and interactive learning.

Mačutek, J. (2021). Why Do Parameter values in the Zipf-Mandelbrot Distribution Sometimes Explode? *Journal of Quantitative Linguistics*, 1–12.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, *31*(2), 247–264.

Malhotra, A., Zhang, H., Beuse, M., y Schmidt, T. (2021). How do new use environments influence a technology 's knowledge trajectory? a patent citation network analysis of lithium-ion battery technology. *Research Policy*, *50*(9), 104318.

Mancusi, M. L. (2003). Geographical concentration and the dynamics of countries 'specialization in technologies. *Economics of Innovation and New Technology*, *12*(3), 269–291.

Mandelbrot, B. (1966). *Information theory and psycholinguistics: A theory of word frequencies, readings in mathematical social sciences.* MIT Press, MA, USA.

Mandelbrot, B. B., y Mandelbrot, B. B. (1982). *The fractal geometry of nature* (Vol. 1). WH freeman New York.

Manin, D. Y. (2009). Mandelbrot's model for zipf's law: Can mandelbrot's model explain zipf's law for language? *Journal of Quantitative Linguistics*, *16*(3), 274–285.

Maraut, S., Dernis, H., Webb, C., Spiezia, V., y Guellec, D. (2008). The OECD REGPAT database: a presentation.

Marsili, O., y Salter, A. (2005). "inequality.ºf innovation: Skewed distributions and the returns to innovation in dutch manufacturing. *Economics of Innovation and New Technology*, *14*(1-2), 83–102.

Martin, B. R. (2012). The evolution of science policy and innovation studies. *Research policy*, *41*(7), 1219–1239.

Mun, C., Yoon, S., y Park, H. (2019). Structural decomposition of technological domain using patent co-classification and classification hierarchy. *Scientometrics*, *121*(2), 633–652.

Nakagawa, T., y Osaki, S. (1975). The discrete weibull distribution. *IEEE transactions on reliability*, *24*(5), 300–301.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*, 1–14.

Nash, J. C., y Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, *43*, 1–14.

Networks: An introduction.Oxford Univ, author=Newman, MEJ, year=2010, publisher=Oxford Univ Press. (s.f.).

Nonaka, l., Takeuchi, H., y Umemoto, K. (1996). A theory of organizational knowledge creation. *International Journal of Technology Management*, *11*(7-8), 833–845.

OCDE. (2021). *REGPAT database.* Autor. doi: https://www.oecd.org/sti/inno/intellectual-property-statistics-and-analysis.htm.

OECD. (2009). *OECD patent statistics manual*. Descargado de `https://www.oecd-ilibrary.org/content/publication/9789264056442-en` doi: https://doi.org/https://doi.org/10.1787/9789264056442-en

Ó Huallacháin, B., y Lee, D.-S. (2011). Technological specialization and variety in urban invention. *Regional studies*, *45*(1), 67–88.

Parraguez, P., Škec, S., e Carmo, D. O., y Maier, A. (2020). Quantifying technological change as a combinatorial process. *Technological Forecasting and Social Change*, *151*, 119803.

Paul, S. R. (2005). Testing goodness of fit of the geometric distribution: an application to human fecundability data. *Journal of Modern Applied Statistical Methods*, *4*(2), 425–33.

Pavitt, K. (1984). Sectoral patterns of technical change: towards a taxonomy and a theory. *Research policy*, *13*(6), 343–373.

Pearson, K. (1900). X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175.

Perc, M. (2012). Evolution of the most common english words and phrases over the centuries. *Journal of The Royal Society Interface*, *9*(77), 3323–3328.

Piantadosi, S. T. (2014). Zipf 's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, *21*(5), 1112–1130.

Piketty, T., y Saez, E. (2003). Income inequality in the united states, 1913–1998. *The Quarterly journal of economics*, *118*(1), 1–41.

Prieto, F., Gómez-Déniz, E., y Sarabia, J. M. (2014). Modelling road accident blackspots data with the discrete generalized pareto distribution. *Accident Analysis & Prevention*, *71*, 38–49.

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A., Olmo-Jiménez, M. J., y Martínez-Rodríguez, A. M. (2009). A generalized waring regression model for count data. *Computational statistics & data analysis*, *53*(10), 3717–3725.

Rogers, E. M. (1962). Diffusion of innovations. The free Press of Glencoe. *NY*, *32*, 891–937.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy*, *94*(5), 1002–1037.

Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, *98*(5, Part 2), S71–S102.

Rothschild, L. (1986). The distribution of english dictionary word lengths. *Journal of statistical planning and inference*, *14*(2-3), 311–322.

Sarabia, J., Prieto, F., y Jordá, V. (2018). *Prácticas de estadística con R.* Comercial Grupo Anaya, S.A.

Sarabia, J., Prieto, F., y Jordá, V. (2020). *Apuntes de Estadística I I.* Ediciones TGD.

Sarabia, J. M., y Castillo, E. (2003). Extensiones multivariantes de la distribución de waring generalizada. *Proc XVII Nat Cong Statistic and Operational Research, Lleida (Spain), Apr*, 8–11.

Sarabia, J. M., Jorda, V., y Prieto, F. (2019). On a new pareto-type distribution with applications in the study of income inequality and risk analysis. *Physica A: Statistical Mechanics and Its Applications*, *527*, 121277.

Sarabia Alegría, J. M., Gómez Déniz, E., y Vázquez Polo, F. J. (2007). *Estadística actuarial. teoría y aplicaciones.* Pearson Prentice Hall.

Schumpeter, J. (1934). The theory of economic development harvard university press. *Cambridge, MA*.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy.* Allen & Unwin London.

Sen, A. K. (1997). From income inequality to economic inequality. *Southern Economic Journal*, *64*(2), 384–401.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, 613–625.

Sichel, H. S. (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association*, *70*(351a), 542–547.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, *42*(3/4), 425–440.

Sitthiyot, T., y Holasut, K. (2021). A simple method for estimating the lorenz curve. *Humanities and Social Sciences Communications*, *8*(1), 1–9.

Smojver, V., Štorga, M., Potočki, E., y cols. (2016). An extended methodology for the assessment of technical invention evolution. En *Ds 84: Proceedings of the design 2016 14th international design conference* (pp. 1135–1144).

Soda, G., Zaheer, A., Sun, X., y Cui, W. (2021). Brokerage evolution in innovation contexts: Formal structure, network neighborhoods and knowledge. *Research Policy*, *50*(10), 104343.

Solé, R., Amor, D. R., y Valverde, S. (2016). On singularities and black holes in combination-driven models of technological innovation networks. *Plos one*, *11*(1), e0146180.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, *70*(1), 65–94.

Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312–320.

Son, S., y Cho, N.-W. (2020). Technology fusion characteristics in the solar photovoltaic industry of South Korea: A patent network analysis using IPC Co-occurrence. *Sustainability*, *12*(21), 9084.

SPRUILL III, M. C. (1973). *Large sample properties of chi-square goodness-of-fit tests* (Tesis Doctoral no publicada). Purdue University.

Strumsky, D., y Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, *44*(8), 1445–1461.

Strumsky, D., Lobo, J., y Tainter, J. A. (2010). Complexity and the productivity of innovation. *Systems research and behavioral science*, *27*(5), 496–509.

Strumsky, D., Lobo, J., y Van der Leeuw, S. (2012). Using patent technology codes to study technological change. *Economics of Innovation and New technology*, *21*(3), 267–286.

Taalbi, J. (2017). What drives innovation? evidence from economic history. *Research Policy*, *46*(8), 1437–1453.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research policy*, *15*(6), 285–305.

Tidd, J., Bessant, J., y Pavitt, K. (1997). Managing innovation. *John Wiley&Sons, Ltd*.

Tilak, J. B., y cols. (1987). *The economics of inequality in education.* (n.º 44). Sage Publications India (Pvt) Ltd.

Utterback, J. M., y Abernathy, W. J. (1975). A dynamic model of process and product innovation. *Omega*, *3*(6), 639–656.

Uzzi, B., Mukherjee, S., Stringer, M., y Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.

Vaidyanathan, V., Savariappan, P., y Chandrasekhar, P. (2009). Statistical inference for a bulk-arri val queue..

Van Looy, B., Vereyen, C., y Schmoch, U. (2015). Patent statistics: Concordance IPC V8-NACE Rev. 2 (version 2.0). *Final report, EUROSTAT, Luxembourg*.

Venugopalan, S., y Rai, V. (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change*, *94*, 236–250.

Von Hippel, E. (2006). *Democratizing innovation*. the MIT Press.

Wang, P., Van De Vrande, V., y Jansen, J. J. (2017). Balancing exploration and exploitation in inventions: Quality of inventions and team composition. *Research Policy*, *46*(10), 1836–1850.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics*, *113*(2), 331–360.

Wickham, H., François, R., Henry, L., y Müller, K. (s.f.). *Rstudio. dplyr: a grammar of data manipulation.* Version.

Winter, S. G., y Nelson, R. R. (1982). An evolutionary theory of economic change. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

WIPO. (2019). *Guide to the International Patent Classification.* Autor.

Wolfram, D. (1992). Applying informetric characteristics of databases to ir system file design, part I: informetric models. *Information Processing & Management*, *28*(1), 121–133.

Xekalaki, E. (1983a). Hazard functions and life distributions in discrete time. *Communications in Statistics-Theory and Methods*, *12*(21), 2503–2509.

Xekalaki, E. (1983b). A property of the yule distribution and its applications. *Communications in Statistics-Theory and Methods*, *12*(10), 1181–1189.

Xekalaki, E. (1983c). The univariate generalized waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics*, 887–895.

Xekalaki, E. (1984). Linear regression and the yule distribution. *Journal of Econometrics*, *24*(3), 397–403.

Xekalaki, E., y Zografi, M. (2008). The generalized waring process and its application. *Communications in Statistics - Theory and Methods*, *37*(12), 1835–1854.

Xiao, J., Bao, Y., y Wang, J. (2021). Which neighbor is more conducive to innovation? the moderating effect of partners 'innovation. *The Journal of Technology Transfer*, 1–35.

Ye, Q. (2021). The impact of knowledge depth and breadth on the geography of analytical industry technological networks: Evidence from china 's biotechnology industry. *Growth and Change*, *52*(4), 2238–2255.

Youn, H., Strumsky, D., Bettencourt, L. M., y Lobo, J. (2015). Invention as a combinatorial process: evidence from us patents. *Journal of the Royal Society interface*, *12*(106), 20150272.

Yu, P., Jiang, F., Cai, Z., y Sun, Y. (2021). The spatio-temporal evolution of china 's hydrogen fuel cell vehicle innovation network: Evidence from patent citation at provincial level. *FRONTIERS IN ENVIRONMENTAL SCIENCE*, *9*.

Yule, G. U. (1925). II.- a mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character*, *213*(402-410), 21–87.

Zipf, G. (1932). *Selected studies of the principle of relative frequency in language.* Cambridge, MA: Harvard University Press.

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human ecology.

Zucman, G. (2019). Global wealth inequality. *Anual Review of Economics*, *11*, 109-138.