

REPARAMETRIZACIÓN DE LAS PRINCIPALES DISTRIBUCIONES DE PROBABILIDAD EN EL ESTUDIO DEL NÚMERO DE SINIESTROS DEBIDO A LAS ANOMALÍAS MUESTRALES EN LAS CARTERAS DEL SEGURO DE RESPONSABILIDAD CIVIL DE AUTOMÓVILES. DETERMINACIÓN DEL ÍNDICE DE DISPERSIÓN.

Dr. José A. Álvarez Jareño¹
Dr. Prudencio Muñoz Rodríguez²

Resumen.- En el presente trabajo se identifican y analizan las principales anomalías detectadas en las carteras de seguros de responsabilidad civil de automóviles que son: el contagio, la *sobre-dispersión* y el inflado de ceros. Posteriormente, se redefinen los parámetros de las distribuciones de probabilidad del número de siniestros más utilizadas en la estadística actuarial en función de una de estas anomalías, la *sobre-dispersión*. Por último, se analiza la dispersión a través del coeficiente de simetría de las distribuciones de probabilidad reparametrizadas.

Palabras Clave: Sobre-dispersión, inflado de ceros, distribución Poisson Pascal Generalizada, reparametrización, coeficiente de asimetría, seguro de automóviles.

Abstract.- This paper identifies and discusses major anomalies detected in the portfolios of liability insurance of cars that are: the spread, the over-dispersion and the zero inflated. Subsequently, the parameters of probability distributions over the number of claims used in the actuarial statistics in terms of one of these anomalies, the over-dispersion, are redefined. Finally, we analyze the dispersion through the coefficient of symmetry of the reparametrized probability distributions.

¹ Profesor Asociado del Departamento de Economía de la Universitat de València. Avda. de los Naranjos, s/n, 46022, Valencia. Correo electrónico: Jose.A.Alvarez@uv.es

² Profesor Titular de Universidad del Departamento de Economía de la Universitat de València. Avda. de los Naranjos, s/n, 46022, Valencia. Correo electrónico: Prudencio.Muniz@uv.es

Este artículo ha sido recibido en versión revisada el 1 de junio de 2010.

Keywords: Over-dispersion, zero inflated, Generalized Poisson Pascal distribution, reparametrization, coefficient of skewness, automobile insurance.

1. Introducción

El tratamiento de los modelos estocásticos en los seguros no vida es derivado de la conocida distribución de Poisson, considerando la pregunta de cuantos siniestros pueden ocurrir durante un período definido de tiempo. Según Kupper (1963) las propiedades subyacentes a esta distribución son:

- a) La población estudiada es homogénea.
- b) La ocurrencia de un siniestro es un evento raro.
- c) La ocurrencia de un siniestro posterior no está influenciada por los anteriores siniestros, es decir, no existe contagio.

La distribución de Poisson tiene la siguiente función de probabilidad o de cuantía:

$$\Pr(N = k) = p_k = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (1)$$

Se puede comprobar que para la distribución de Poisson la varianza es igual a la media. El estimador máximo verosímil del parámetro λ es la media muestral.

Si el número de siniestros se distribuye como una Poisson con parámetro λ , los siniestros serán variables aleatorias independientes e idénticamente distribuidas con una distribución exponencial común con media $(1/\lambda)$. La distribución exponencial también tiene la propiedad de tener “ausencia de memoria”. Un proceso de Poisson es un proceso no contagioso, donde los siniestros ocurren totalmente al azar.

El incumplimiento de algunas de las principales propiedades de la distribución de Poisson de las carteras de seguro, producirá diferentes anomalías muestrales: el contagio, la *sobre-dispersión* y el *inflado de ceros*. En el segundo apartado se exponen las anomalías muestrales, y a continuación en el apartado 3 se presentan las 15 carteras de seguro de automóvil que se analizarán. En el apartado 4 se comprueba la existencia de las anomalías en todas las muestras analizadas. Las distribuciones del número de siniestros más utilizadas en la estadística actuarial se presentan en

el apartado 5 con la finalidad de tenerlas como referencia para en el apartado 6 redefinir los parámetros de estas distribuciones en función de las anomalías detectadas en las muestras, obteniendo de esta forma una reparametrización más acorde a la estadística actuarial. En el apartado 7 se efectúa un análisis del coeficiente de asimetría en las funciones de distribución del número de siniestros y como se ve afectado por las anomalías muestrales. Finalmente, en el apartado 8 se expondrán las conclusiones. La bibliografía y un Anexo gráfico cierran este trabajo.

2. Anomalías muestrales

Las anomalías surgen a consecuencia de comparar la distribución muestral del número de siniestros en una cartera de pólizas de seguro con la distribución de Poisson. Se han identificado tres anomalías que son el contagio, la *sobre-dispersión* y el inflado de ceros.

a. Contagio verdadero o aparente

Denuit *et al.* (2007) exponen que el contagio aparente se plantea desde el reconocimiento que los individuos de la muestra provienen de poblaciones heterogéneas en las cuales los individuos tienen una propensión a experimentar accidentes que es constante aunque diferente de unos a otros. Un individuo dado puede tener una alta (o baja) propensión a los accidentes pero la ocurrencia de un accidente no lo hace más (o menos) propenso a que le ocurra otro accidente. Sin embargo, la agregación a través de individuos heterogéneos puede generar un error estadístico que supone que la ocurrencia de un accidente incrementa la probabilidad de otro accidente; la observada pero persistente heterogeneidad puede ser malinterpretada como una dependencia serial fuerte.

El contagio verdadero se refiere a la dependencia entre la ocurrencia de eventos sucesivos. La ocurrencia de un evento, como un accidente o una enfermedad, puede cambiar la probabilidad de ocurrencia de eventos similares. El contagio verdadero positivo implica que la ocurrencia de un suceso reduce el tiempo de espera para la ocurrencia del siguiente suceso. Al igual que una persona con malos hábitos alimenticios es más propensa a sufrir enfermedades, un asegurado con malas costumbres a la hora de conducir tendrá una mayor probabilidad de provocar accidentes.

El presunto fenómeno de la propensión a los accidentes puede ser interpretado en términos de contagio positivo como sugiere que un individuo

que ha experimentado un accidente tiene mayor probabilidad de sufrir otro. En un marco longitudinal, los resultados actuales y futuros están directamente influenciados por los valores pasados, y esto causa un cambio sustancial a lo largo del tiempo en la correspondiente distribución.

Brouhns *et al.* (2003) hacen una interpretación del “contagio verdadero” en los siguientes términos. Un accidente de tráfico puede modificar la percepción del peligro al volante y reducir el riesgo de comunicar otro siniestro en el futuro. Los sistemas de tarificación de las compañías aseguradoras proporcionan incentivos para una conducción más cuidadosa y éstos pueden producir un contagio negativo. Sin embargo, la principal interpretación para el seguro del automóvil es exógena, si el contagio positivo se observa en el número de siniestros (los asegurados que comunicaron siniestros en el pasado es más probable que comuniquen siniestros en el futuro que aquellos que no los comunicaron), entonces el verdadero contagio debe ser negativo.

En la práctica se observa contagio positivo, lo que indica que los sucesos pasados no influyen verdaderamente en la probabilidad de comunicar un siniestro, y que proporcionan información sobre la verdadera naturaleza del conductor. Entonces, el término de heterogeneidad de los carteras de pólizas de seguros puede ser actualizado de acuerdo con el historial de siniestralidad del asegurado.

Si los datos sólo permiten observar el número total de sucesos al final del período, el contagio, como la heterogeneidad, es un proceso no observado. En el análisis de las carteras donde el contagio y la heterogeneidad son plausibles, ambos procesos no son distinguibles si los datos son agregados porque ambos conducen a la misma distribución de probabilidad para el número de sucesos.

b. Sobre-dispersión

Al hecho de que la varianza muestral sea superior a la media muestral en una cartera de seguros se le denomina “*sobre-dispersión*”, porque si la siniestralidad estuviera generada por una distribución de Poisson, la media sería igual a la varianza. Bermúdez *et al.* (2001) afirman que si la varianza es mayor que su media, implica que la aplicación del supuesto que el número de siniestros siguen una distribución de Poisson es incorrecto. Tremblay (1992) expone que es obvio que en las mixturas de Poisson (también en las distribuciones de Poisson compuestas) la varianza excede a la media, por lo que este tipo de distribuciones serán “más seguras” para las compañías aseguradoras.

Rolski *et al.* (1998) define el índice de dispersión como el cociente entre la varianza y la esperanza matemática, I_N , si este índice es superior a 1 la distribución presenta *sobre-dispersión*, y si es inferior a 1 la distribución muestra *bajo-dispersión*. La distribución Binomial Negativa es una distribución con *sobre-dispersión*.

$$I_N = \frac{\sigma_x^2}{\mu_x} \quad (2)$$

La interpretación que hacen Pitrebois *et al.* (2006) es que muchos factores no pueden ser tenidos en cuenta³ a la hora de efectuar una tarificación *a priori* para el seguro de responsabilidad civil frente a terceros. Por ejemplo, la rapidez de reflejos, la agresividad al volante o el conocimiento del código de circulación son difíciles de integrar en la clasificación del riesgo. Consecuentemente, las clases de tarificación son todavía heterogéneas después de la utilización de muchas variables de clasificación. Esta heterogeneidad⁴ residual produce *sobre-dispersión*: los datos relativos al número de siniestros exhiben una variabilidad que excede la explicada por los modelos de Poisson. Este fenómeno se puede modelar a través de un efecto aleatorio en un modelo estadístico.

Meng *et al.* (1999) exponen que este fenómeno es especialmente relevante en el seguro de automóviles, y proponen la utilización de la distribución Binomial Negativa mezclada con una distribución de Pareto.

En Denuit *et al.* (2007) se dispone de una demostración de la ocurrencia de *sobre-dispersión* en base a la heterogeneidad de los datos que componen una cartera de pólizas. Al mismo tiempo, proponen un contraste para la detección de la *sobre-dispersión*.

UNESPA ha elaborado durante décadas estadísticas comunes de siniestralidad en el seguro de automóviles. En el estudio editado en 1983, con datos del sector del automóvil de 1981, se pudo evidenciar que la frecuencia media de siniestralidad en la modalidad de responsabilidad civil suplementaria en vehículos turismos era de 0,10444 con una varianza de 0,11938, un 14,3% superior a la media. Este resultado, como exponen Boj *et*

³ Bermúdez *et al.* (2000) achacan este hecho a la imposibilidad de medir variables explicativas importantes, y por consiguiente, son excluidas incorrectamente del análisis de regresión.

⁴ En Denuit *et al.* (2007) se dispone de una demostración de la ocurrencia de *sobre-dispersión* en base a la heterogeneidad de los datos que componen una cartera de pólizas.

al. (2005), permitió cuestionar, en dicha clase de riesgo, el modelo de Poisson.

c. Inflado de ceros

El “inflado de ceros” se observa cuando los datos muestrales presentan una frecuencia más elevada para la ocurrencia de cero siniestros que la que se debería esperar si la muestra hubiera sido generada mediante una distribución de Poisson.

Se puede definir el índice del inflado de ceros como el cociente entre el número de asegurados con cero siniestros observados en una cartera y el número de asegurados con cero siniestros estimado por la distribución Poisson. Si el cociente es mayor que 1, la muestra presenta inflado de ceros, se observan más asegurados con cero siniestros que los esperados de acuerdo con la distribución de Poisson, y si es inferior a 1, los datos exhiben “desinflado de ceros”.

$$I_{zi}(0) = \frac{p_0(\text{observados})}{p_0(\text{Poisson})} \quad (3)$$

Una de las causas que produce el fenómeno del “inflado de ceros”, de acuerdo con Boucher *et al.* (2006) son los sistemas *bonus-malus*. Su argumentación es que muchas compañías aseguradoras, especialmente en Europa, tienen implementados mecanismos de tarificación por experiencia, y en aplicación de estos elementos, la comunicación de un siniestro implica un incremento en la prima del próximo año, esto induce el “hambre de bonus⁵”. Para el asegurado es óptimo retener todos aquellos siniestros que sean inferiores a una determinada cuantía, dependiendo del nivel ocupado en la escala *bonus-malus*. Este comportamiento generará una probabilidad inflada en el origen para el número de siniestros observados.

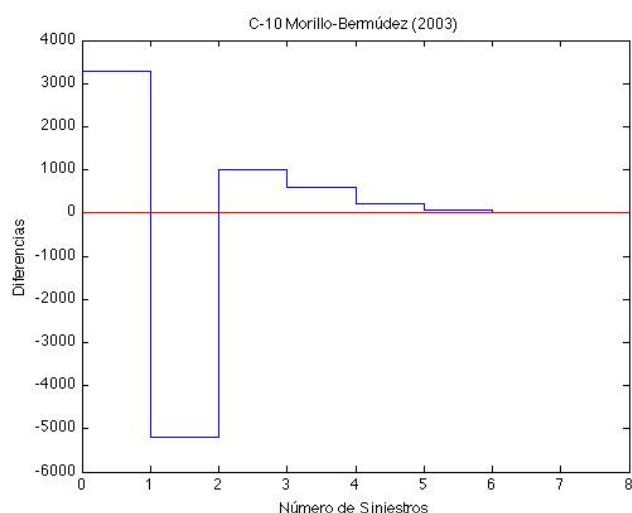
Walhin y Paris (2000) explican que los asegurados deberían costear los siniestros de pequeña cuantía para evitar las penalizaciones inducidas por el sistema *bonus-malus*. Consecuentemente, hay más accidentes que el número de siniestros en los registros de los aseguradores: los datos de siniestralidad

⁵ Traducción del término inglés “hunger of bonus” que expresa que los asegurados tienen un incentivo a no comunicar todos los siniestros ocurridos si el incremento de las primas futuras puede ser superior al beneficio que se obtiene del seguro.

de los aseguradores no están completos, ya que no disponen de los siniestros no comunicados por los asegurados (los datos están truncados).

De acuerdo con este planteamiento, se deberían encontrar más anomalías muestrales en la distribución de Poisson, ya que la suma de probabilidades de la cartera siempre será 1, y el inflado de ceros se deberá compensar en el resto de la distribución del número de siniestros.

Gráfico 1. Diferencias entre la muestra y la distribución de Poisson



La representación gráfica de la diferencia entre el número de asegurados observados para cada número de siniestros y el estimado por la distribución de Poisson, permitirá analizar dónde se producen las compensaciones de la probabilidad. Para todas las carteras de pólizas de automóviles analizadas, siempre se ha obtenido la misma representación gráfica: diferencias positivas para todos los valores excepto para el valor 1 que la diferencia siempre es negativa.

Se pueden identificar tres zonas (véase Gráfico 1):

- Cero siniestros: presenta inflado de ceros. Hay más asegurados con cero siniestros que los estimados por la distribución de Poisson. $I_{ZI}(0) > 1$

- Un siniestro: presenta desinflado de unos. Hay menos asegurados con un siniestro que los obtenidos con una distribución de Poisson. $I_{Z_i}(1) < 1$
- Resto de siniestros: muestra inflado de siniestros. Hay más asegurados con 2, 3, 4, ... que los determinados por un ajuste de Poisson. $I_{Z_i}(2,3,\dots) > 1$

3. Datos del análisis

A continuación se exponen 15 carteras⁶ de automóviles sobre las que se analizarán las anomalías expuestas. Las carteras son identificadas por el autor de la publicación en la que aparecieron por primera vez, así como el año al que pertenecen los datos.

Tabla 1. Carteras de automóviles

(k) Siniestros	C-1	C-2	C-3	C-4	C-5	C-6	C-7
0	7.840	20.592	103.704	370.412	96.978	881.705	57.178
1	1.317	2.651	14.075	46.545	9.240	142.217	5.617
2	239	297	1.766	3.935	704	18.088	446
3	42	41	255	317	43	2.118	50
4	14	7	45	28	9	273	8
5	4	0	6	3		53	
6	4	1	2				
7 ó más	1						
Total pólizas	9.461	23.589	119.853	412.240	106.974	1.044.454	63.299

(k) Siniestros	C-8	C-9	C-10	C-11	C-12	C-13	C-14	C-15
0	118.700	565.664	122.618	2.196.808	513.814	434.698	378.289	371.481
1	11.468	68.714	21.686	161.913	32.296	39.914	30.518	26.784
2	930	5.177	4.014	10.976	2.493	3.970	2.629	2.118
3	70	365	832	882	203	435	240	174
4	14	24	224	90	24	74	27	18
5		6	68	11		8	5	2
6			17	2		6	0	2
7			7	1		2	0	0
8 ó más			7			0	0	0
Total pólizas	131.182	639.950	149.483	2.370.683	548.830	479.107	411.708	400.579

Las carteras son C-1: Thyron (1960), C-2: Tröblinger (1961), C-3: Bühlmann (1970), C-4: Johson y Hey (1971), C-5: Lemaire (1985), C-6: Besson y Partrat (1992), C-7: Denuit (1993), C-8: Denuit (1994), C-9:

⁶ Las carteras analizadas pertenecen a diferentes países y diferentes períodos de tiempo, con la intención de que los resultados del estudio sean generalizables. No obstante, de las 15 carteras 6 son de compañías aseguradoras que operan en España.

Hossack (1999), C-10: Morillo y Bermúdez (2003), C-11: Vilar, Gil y Heras (2004), C-12: Boucher, Denuit y Guillén (2006), C-13: Álvarez (1999), C-14: Álvarez (2000) y C-15: Álvarez (2001).

4. Resultados del análisis

a. *Sobre-dispersión*

Todas las carteras analizadas presentan *sobre-dispersión*, tal como se muestra en la Tabla 2.

Tabla 2. “Sobre-dispersión” en las carteras analizadas

<i>Cartera</i>	<i>Media</i>	<i>Varianza</i>	<i>Sobre-dispersión</i>
Thyrión	0,2144	0,2889	1,3475
Troblinger	0,1442	0,1639	1,1366
Buhlmann	0,1551	0,1793	1,156
Johson-Hey	0,1317	0,1385	1,0516
Lemaire	0,1011	0,1074	1,0623
Besson-Partrat	0,1782	0,1974	1,1077
Denuit (Bélgica 93)	0,1057	0,1149	1,087
Denuit (Bélgica 94)	0,1036	0,1115	1,0763
Hossack	0,1255	0,13	1,0359
Morillo-Bermúdez	0,2251	0,2966	1,3176
Boucher-Denuit-Guillén	0,0788	0,0847	1,0749
Vilar-Gil-Heras	0,0692	0,0763	1,1026
Nueva cartera 1999	0,1034	0,1175	1,1364
Nueva cartera 2000	0,0889	0,0984	1,1069
Nueva cartera 2001	0,0789	0,0867	1,0989

La *sobre-dispersión* observada oscila entre el 34,75% de la cartera de Thyrión, que es la más antigua, y el 3,59% de la cartera de Hossack. La presencia de *sobre-dispersión* implica que la distribución de Poisson no será la más adecuada para ajustar el número de siniestros de las carteras de automóviles. Se precisarán distribuciones con colas más pesadas para obtener mejores ajustes, tales como la Binomial Negativa o la Poisson Pascal Generalizada. Aunque todo apunta a que una mayor media siniestral debería implicar una mayor *sobre-dispersión*, no se han hallado evidencias de una relación entre la *sobre-dispersión* y la siniestralidad. En el apartado 7 se analizará el coeficiente de asimetría, teniendo en cuenta que las carteras de seguro de automóvil presentan *sobre-dispersión*.

b. *Inflado de ceros*

Todas las carteras analizadas, independientemente del país e incluso del tiempo, presentan las tres zonas descritas en el apartado 2, y en todas ellas se

observa, como la frecuencia del valor un siniestro se encarga de compensar el inflado del resto de valores.

En términos relativos es mucho más importante el desinflado de unos, o el inflado de dos o más siniestros, que el inflado de ceros, que en todos los casos es inferior al 3% de la estimación de la distribución de Poisson, tal como se observa en la Tabla 3.

De acuerdo con la argumentación expuesta por Walhin y Paris (2000) o Boucher *et al.* (2006), los asegurados con 1 ó 2 siniestros tienden a no comunicar alguno de los siniestros a la compañía aseguradora, sin embargo, una vez se han tenido 2 ó más siniestros en una misma anualidad de seguro, los asegurados comunican cualquier incidente por mínimo que sea. El asegurado intenta preservar su nivel de bonificación o mejorarlo, no comunicando siniestros a la compañía aseguradora, pero cuando se ve obligado a comunicar un siniestro, sabiendo que perderá buena parte de la bonificación, no tiene ningún inconveniente en comunicarlos todos. No obstante, el contagio positivo, también podría ser la explicación a este fenómeno.

Tabla 3. “Inflado de ceros” en las carteras analizadas

Cartera	$I_{zi}(0) > 1$	$I_{zi}(1) < 1$	$I_{zi}(2, 3, \dots) > 1$
Thyrion	1,0268	0,8047	1,6114
Troblinger	1,0084	0,9001	1,5518
Buhlmann	1,0105	0,8840	1,5935
Johson-Hey	1,0032	0,9569	1,2787
Lemaire	1,0030	0,9454	1,4794
Besson-Partrat	1,0088	0,9132	1,3933
Belgica 93	1,0040	0,9331	1,5288
Belgica 94	1,0036	0,9357	1,5421
Hossack	1,0021	0,9702	1,2022
Morillo-Bermúdez	1,0275	0,8071	1,5831
Boucher-Denuit-Guillén	1,0033	0,9111	2,1665
Vilar-Gil-Heras	1,0027	0,9372	1,7102
Nueva cartera 1999	1,0062	0,8934	1,8793
Nueva cartera 2000	1,0043	0,9107	1,8888
Nueva cartera 2001	1,0036	0,9162	1,9521

Suponiendo que las carteras están compuestas sólo por dos categorías⁷ de conductores, los “buenos conductores” los que tienen 0 ó 1 siniestros, y los “malos conductores” los que tienen 2 ó más siniestros, la distribución de

⁷ Lemaire (1985) utiliza este argumento para proponer una mixtura de distribuciones, donde las dos categorías de conductores tienen distribución de Poisson con distintos parámetros, λ_1 (buenos conductores) menor que λ_2 (malos conductores).

Poisson sobreestima los “buenos conductores” y subestima los “malos” (serán necesarias distribuciones con una cola más pesada). Si se deben utilizar distribuciones con una cola más pesada implicará una mayor dispersión de los datos, existiendo una relación entre el “desinflado de unos” y la *sobre-dispersión*. Yip y Yuan (2005) relacionan el inflado de ceros, que también denominan “exceso de ceros”, con la *sobre-dispersión* muestral respecto a la distribución de Poisson.

5. Funciones de distribución para modelizar el número de siniestros en una cartera de automóviles

Las funciones de distribución más utilizadas en la estadística actuarial para modelizar el número de siniestros de una cartera [Hogg y Klugman (1984), Lemaire (1985), Willmot (1988), Klugman (1998), Boucher *et al.* (2006), Denuit *et al.* (2007)] son las que se exponen a continuación y que se pueden obtener como distribuciones de Poisson compuestas. Se exponen estas funciones de distribución para estudiar la posibilidad de expresar los parámetros de las mismas desde una perspectiva actuarial que las explique en función de las características de las carteras.

La estimación de los parámetros de las distribuciones de probabilidad se calcula en base a las medidas de posición y dispersión de las carteras, si estas medidas reflejan las anomalías muestrales, los parámetros se podrán expresar en función de las mismas.

d. *Distribución Binomial Negativa (BN)*

La distribución binomial negativa⁸ (BN) se ha utilizado extensamente como una alternativa a la distribución de Poisson. La distribución binomial negativa se puede obtener por dos procedimientos diferentes, bien como una distribución compuesta, donde la distribución primaria es una Poisson y la distribución secundaria es una logarítmica, o bien como una mixtura⁹ de distribuciones, Poisson y Gamma.

La función de probabilidad de la distribución binomial negativa viene dada por la siguiente función de cuantía:

⁸ La distribución binomial negativa fue formulada inicialmente por Montmort en 1714, como la distribución del número de pruebas requeridas en un experimento para obtener un número dado de sucesos. También es conocida como Poisson-Gamma.

⁹ El número de accidentes de una cartera sigue una distribución de Poisson con media λ , y la media λ muestra la variabilidad de la cartera representada con una función de densidad continua de tipo Gamma.

$$\Pr(S = k) = p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k, \quad k = 0, 1, 2, \dots, \quad r > 0, \beta > 0 \quad (4)$$

La función generatriz de probabilidad para la distribución binomial negativa es

$$P(z) = [1 - \beta(z-1)]^{-r} \quad (5)$$

De aquí se sigue que la media y la varianza de una distribución binomial negativa son

$$E(S) = r\beta = \mu_s$$

$$\text{Var}(S) = r\beta \cdot (1 + \beta) = \sigma_s^2$$

e. Distribución Polya-Aeppli (P-A)

Si la distribución secundaria es una distribución Geométrica¹⁰ truncada en cero se obtendrá la distribución Polya-Aeppli.

$$P_2(z) = \frac{[1 - \beta(z-1)]^{-1} - (1 + \beta)^{-1}}{1 - (1 + \beta)^{-1}} \quad (6)$$

La función generatriz de probabilidad para la distribución Polya-Aeppli es

$$P(z) = e^{\lambda \left\{ \frac{[1 - \beta(z-1)]^{-1} - 1}{1 - (1 + \beta)^{-1}} \right\}} \quad (7)$$

¹⁰ Una variable aleatoria X (número de fracasos antes del primer éxito) sigue una distribución geométrica si la función de probabilidad viene dada por,

$$\Pr(X = k) = \frac{1}{1 + \beta} \cdot \left(\frac{\beta}{1 + \beta}\right)^k \quad k = 0, 1, 2, 3, \dots$$

donde $\beta > 0$, siendo además la media de la distribución. Lemaire (1985) utiliza una distribución geométrica generalizada para ajustar la función del número de siniestros.

De aquí se deduce que la media y la varianza de la distribución son

$$E(S) = \lambda \cdot (1 + \beta) \quad \text{Var}(S) = \lambda \cdot (1 + \beta) \cdot (1 + 2\beta)$$

f. Distribución Poisson Inversa Gaussiana (PIG)

Si se utiliza una distribución Inversa Gaussiana¹¹ como distribución secundaria, cuya función generatriz de probabilidad es:

$$P_2(z) = \frac{\left[1 - 2\beta(z-1)\right]^{1/2} - (1+2\beta)^{1/2}}{1 - (1+2\beta)^{1/2}} \quad (8)$$

La distribución propuesta tendrá una distribución primaria de Poisson, y una distribución secundaria Inversa Gaussiana. La función generatriz de probabilidad para la distribución es

$$P(z) = e^{-\frac{\lambda}{\beta} \left\{ [1 - 2\beta(z-1)]^{0.5} - 1 \right\}} \quad (9)$$

De aquí se sigue que la media y la varianza de la distribución son

$$E(S) = \lambda \quad \text{Var}(S) = \lambda \cdot (1 + \beta)$$

g. Distribución Poisson Pascal Generalizada¹² (PPG)

Si a una distribución de Poisson primaria la componemos con una distribución secundaria Binomial Negativa truncada y extendida (ETNB) se obtendrá la distribución Poisson Pascal Generalizada (PPG). Vilar y Vegas

¹¹ A la distribución Inversa Gaussiana también se le conoce como distribución de Wald, y su función de cuantía es la siguiente:

$$f(x) = \frac{\mu}{\sqrt{2\pi\beta x^3}} \cdot e^{-\left[\frac{(x-\mu)^2}{2\beta x}\right]} \quad x > 0 \quad \beta = \frac{\mu^2}{\theta}$$

¹² Esta distribución también es conocida en la estadística actuarial como distribución Hofmann [Walhin y Paris (2000)], aunque presenta una parametrización distinta que permite una interpretación más sencilla de los parámetros de la distribución.

(1998) utilizan esta distribución para estimar la provisión y el recargo técnico en una cartera de pólizas de automóviles.

La función generatriz de probabilidad para la distribución Poisson Pascal Generalizada es la siguiente:

$$P(z) = e^{\lambda \left\{ \frac{[1-\beta(z-1)]^r - (1+\beta)^r}{1-(1+\beta)^r} - 1 \right\}} \quad \text{cuando } r > -1, \beta > 0 \text{ y } \lambda > 0 \quad (10)$$

De la función generatriz de probabilidad se deducen la esperanza matemática y la varianza de la distribución, así como el coeficiente de asimetría de esta distribución:

$$E(S) = \frac{\lambda \cdot \beta \cdot r}{1 - (1 + \beta)^r}$$

$$Var(S) = E(S) \cdot [1 + (r + 1)\beta]$$

$$g_1(S) = \frac{3 \cdot \sigma_s^2 - 2 \cdot \mu_s + \frac{r + 2}{r + 1} \cdot \frac{(\sigma_s^2 - \mu_s)^2}{\mu_s}}{\sigma_s^3} \quad (11)$$

Dependiendo de los valores de r el coeficiente de asimetría variará, pudiéndose obtener una cola derecha más gruesa o menos que las distribuciones anteriormente analizadas.

Las distribuciones expuestas anteriormente resultan ser casos particulares de la presente distribución y que se exponen en la siguiente Tabla 4.

Tabla 4. Parámetro r de la distribución Poisson Pascal Generalizada

Distribución	Parámetro
Poisson	r=-1
Binomial Negativa	r=0
Polya-Aeppli	r=1
Neyman Type A	r → ∞
Poisson Inversa Gaussiana	r=-0,5

Gómez-Déniz *et al.* (2009) exponen la distribución Poisson-Beta para la modelización del número de reclamaciones como una mezcla o mixtura de distribuciones de Poisson. Esta distribución presenta una cola más pesada que la distribución de Poisson y consiguientemente también refleja la *sobre-dispersión* de las muestras analizadas.

6. Redefinición de los parámetros de las funciones de distribución de probabilidad

Los datos analizados presentan todas las anomalías que se habían identificado a nivel teórico. Es evidente que las carteras analizadas son carteras heterogéneas, en las que, en el actual análisis, no se han identificado colectivos en función de las variables de tarificación.

Todas las carteras presentan *sobre-dispersión*, sin embargo, se debería realizar un análisis en mayor profundidad. Cabría pensar que la *sobre-dispersión* es proporcional a la media siniestral, es decir, a mayor media muestral mayor *sobre-dispersión* y que la distribución Binomial Negativa sería un buen modelo para modelizar el número de siniestros. Aunque en líneas generales parece que existe una relación entre ambas, no se han encontrado evidencias de una relación entre media y *sobre-dispersión*, y hay carteras con baja siniestralidad y alta *sobre-dispersión*, y viceversa.

Tabla 5. Parámetros de las distribuciones de probabilidad en función de la media muestral y el índice de sobre-dispersión

Distribución	Parámetros	
Binomial Negativa	$\beta = \frac{\sigma_x^2}{\mu_x} - 1 = I_N - 1$	$r = \frac{\mu_x}{I_N - 1}$
Polya-Aeppli	$\beta = 0,5 \cdot (I_N - 1)$	$\lambda = \frac{2\mu_x}{1 + I_N}$
Poisson Inversa Gausiana	$\beta = 2 \cdot (I_N - 1)$	$\lambda = \mu_x \cdot \frac{1 - \sqrt{2 \cdot I_N - 1}}{1 - I_N}$
Poisson Pascal Generalizada	$\beta \cdot (r + 1) = I_N - 1$ $\beta = \frac{I_N - 1}{r + 1}$	

Los parámetros de las distribuciones de probabilidad es posible expresarlos en función de la *sobre-dispersión* que presenta la muestra. Esta reparametrización permitirá expresar las distribuciones y sus parámetros de una forma más adaptada a la estadística actuarial, indicando información sobre la cartera que se este analizando. El parámetro β en las distribuciones Binomial Negativa es la proporción de *sobre-dispersión* de la cartera expresada en tanto por uno, mientras que en la Poisson Inversa Gaussiana es 2 veces esta proporción.

La distribución Poisson Pascal Generalizada admite dos interpretaciones en función de las dos expresiones que puede tomar, la primera, que el parámetro β es $\frac{1}{r+1}$ la proporción de la varianza que supera a la media, la segunda, que el producto de los parámetros β y r es esa misma proporción (es decir el parámetro β de la distribución Binomial Negativa).

De igual forma, en todas las carteras se ha encontrado la anomalía del “inflado de ceros”, el “desinflado de unos” y el “inflado de la cola”. El patrón de comportamiento de la siniestralidad de una cartera es común a todas las carteras analizadas sin excepción, tal como se puede comprobar en los gráficos del Anexo. No se han encontrado referencias bibliográficas sobre el “desinflado de unos” cuando es un fenómeno con mayores consecuencias prácticas que el “inflado de ceros”.

Se han hallado evidencias de una relación entre la *sobre-dispersión* y el “desinflado de unos”, tanto a nivel teórico, como a nivel práctico, a mayor *sobre-dispersión* mayor “desinflado de unos”.

La relación entre la *sobre-dispersión* y el “desinflado de unos” viene determinada por el hecho que la frecuencia del número de asegurados en cero siniestros no influye directamente en el cálculo de la media muestral (cero por cualquier frecuencia es cero, el primer término del sumatorio es siempre cero), sin embargo, el resto de valores de la variable tendrán un peso más importante en la determinación, tanto de la media muestral como de la varianza muestral. En el cálculo de la varianza muestral, la aportación al resultado final de cero siniestros, es prácticamente la misma, independientemente de la función de distribución utilizada, siendo el resto de los valores los que soportan variaciones más importantes. El valor un siniestro será el que mayores variaciones genere. En la distribución de Poisson, el valor un siniestro acumula gran parte de la variación de la distribución, dejando una participación menor para los valores mayores que uno. En el resto de distribuciones de probabilidad utilizadas, al tener una

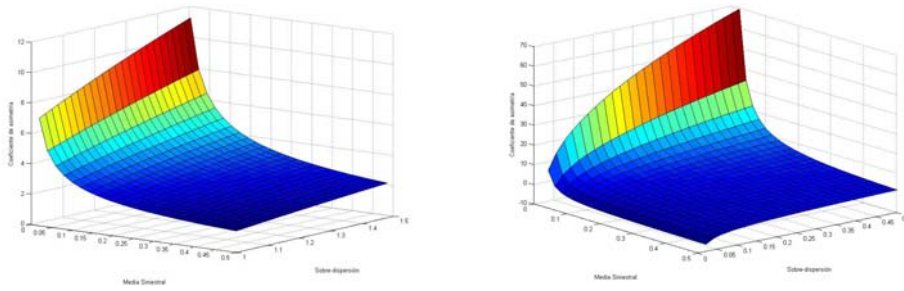
cola más pesada, el valor un siniestro reduce su peso y los valores mayores que uno lo incrementan en mayor medida que el valor uno. El resultado final es una mayor dispersión de las distribuciones compuestas.

Las distribuciones de probabilidad compuestas, o las mixturas, al calcularse en base a dos o más parámetros ajustan mejor la variabilidad de la muestra, y las diferencias entre la varianza muestral y la de la distribución es muy pequeña. El ajuste a una distribución de probabilidad u otra dependerá de otras medidas (coeficiente de asimetría o coeficiente de curtosis). A mayor número de parámetros de una distribución se incrementa la bondad del ajuste porque se incorpora una mayor información de los datos muestrales a la distribución teórica. Sin embargo, se deberán tener en cuenta dos aspectos, el primero de carácter teórico, un mayor número de parámetros implica un menor número de grados de libertad de la distribución estimada, y el segundo de carácter práctico, distribuciones más complejas (con mayor número de parámetros) no tienen porque aportar mejoras significativas en las características requeridas a un sistema o escala de primas. Se preferirá un modelo más sencillo a otro más complejo si las mejoras que introduce el segundo no mejoran sustancialmente el sistema (navaja de Occam o principio de economía o parsimonia).

7. Coeficiente de Asimetría

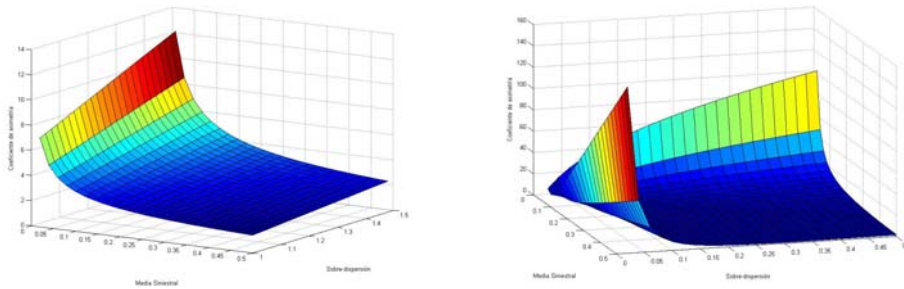
El coeficiente de asimetría en la distribución del número de siniestros será un valor positivo, es decir, la distribución es asimétrica por la derecha, hecho que es evidente, ya que el número de siniestros únicamente toma valores positivos, y su media aritmética es un valor inferior a 1. Se deberá añadir que todas las distribuciones del número de siniestros presentan *sobre-dispersión*, y que existe una relación entre la *sobre-dispersión* y el “desinflado de unos”, por consiguiente el coeficiente de asimetría (o el momento central de orden 3) estará influido por estas anomalías.

Gráfico 2. Coeficiente de Asimetría de la distribución Binomial Negativa
Coficiente de Asimetría con sobre-dispersión *Coficiente de Asimetría sin sobre-dispersión*



Partiendo de la distribución PPG, el parámetro “r” reflejará la *sobre-dispersión* de los datos. Si la media muestral es igual a la varianza muestral, se estaría en la distribución de Poisson, y el parámetro “r” sería igual a -1. A medida que haya mayor *sobre-dispersión*, también se incrementará el parámetro “r”, teniendo en cuenta, que no es el único elemento que influye en la determinación de este parámetro. El coeficiente de asimetría (o más concretamente, el momento central de orden 3) también se verá reflejado en el parámetro “r”, que a su vez viene determinado por la *sobre-dispersión* de los datos.

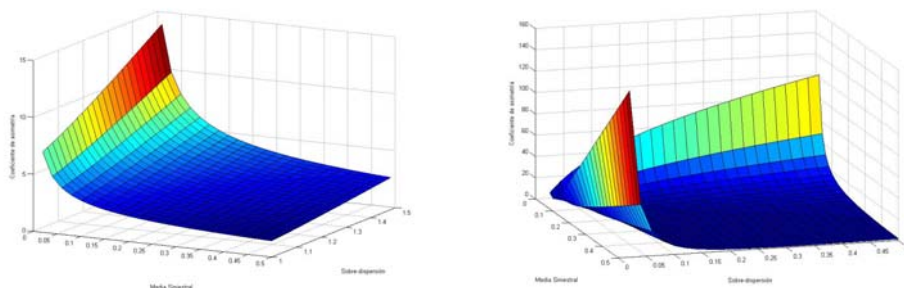
Gráfico 3. Coeficiente de Asimetría de la distribución Poisson Inversa Gausiana
Coficiente de Asimetría con sobre-dispersión *Coficiente de Asimetría sin sobre-dispersión*



Si no se tiene en cuenta que los datos presentan *sobre-dispersión* el comportamiento del coeficiente de asimetría para cada una de las distribuciones propuestas, tiene un comportamiento bastante diferenciado, tal como se muestra en los gráficos de la derecha de los Gráficos 2, 3 y 4. Sin embargo, al introducir el fenómeno de la *sobre-dispersión*, las tres distribuciones que mejores resultados obtienen en los ajustes tienen un

comportamiento muy similar, y únicamente se pueden observar pequeñas diferencias entre ellas.

Gráfico 4. Coeficiente de Asimetría de la distribución Poisson Pascal Generalizada
Coeficiente de Asimetría con sobre-dispersión *Coeficiente de Asimetría sin sobre-dispersión*



Todas las distribuciones empiezan con la inexistencia de *sobre-dispersión* ($I_N=1$), siendo los resultados coincidentes entre las diferentes distribuciones y a su vez con la distribución de Poisson. En función de la *sobre-dispersión* el coeficiente de variación se verá modificado en cada una de las distribuciones, aunque de forma poco significativa.

Si además se tiene en cuenta que el intervalo de la media siniestral no toma valores muy elevados en la realidad (el mínimo analizado es de 0,0691 y el máximo de 0,2251), los resultados para el coeficiente de asimetría de las tres distribuciones son muy similares para las sobre-dispersiones observadas.

8. Conclusiones

Del análisis realizado se deduce que todas las carteras analizadas presentan *sobre-dispersión*, inflado de ceros y desinflado de unos, es decir, la distribución de Poisson sobreestima a los “buenos conductores” (0 ó 1 siniestro) y subestima a los “malos conductores” (2 ó más siniestros). Son necesarias distribuciones con colas más pesadas, como la Binomial Negativa (BN), la Poisson Inversa Gaussiana (PIG) o la Poisson Pascal Generalizada (PPG) para poder ajustar mejor la distribución del número de siniestros a los datos reales.

Se han redefinido los parámetros de las funciones de distribución utilizadas en el análisis de siniestros para que éstos se expresen en función de la media muestral y el índice de *sobre-dispersión*, de esta forma la interpretación de los parámetros de la distribución se hace en función de las denominadas

anomalías muestrales que tiene un carácter más adaptado a la estadística actuarial. Esto significa que para la distribución Binomial Negativa, el parámetro β es el porcentaje de *sobre-dispersión* que presenta la cartera, mientras que el parámetro r es el cociente entre la media muestral y el porcentaje de *sobre-dispersión* de la cartera.

El parámetro β en las distribuciones Binomial Negativa, Polya-Aeppli y Poisson Inversa Gausiana está directamente relacionado con la *sobre-dispersión*.

El coeficiente de asimetría de las funciones de distribución utilizadas en este trabajo se ve condicionado por la existencia de *sobre-dispersión* en las muestras y presenta un comportamiento muy similar para todas las distribuciones, lo que implica que los parámetros de las distribuciones del número de siniestros están en un intervalo muy delimitado donde todas las distribuciones se asemejan bastante.

Bibliografía

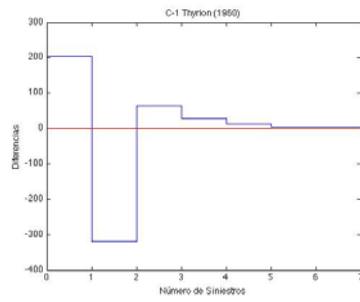
- Álvarez, J.A. (2009). Análisis de los sistemas *bonus-malus* en el seguro de automóviles español. Tesis Doctoral, Universitat de València.
- Bermúdez, Ll.; M. Denuit y J. Dhaene (2001). “*Exponential Bonus-Malus Systems integrating a priori Risk Classification*”. Journal of Actuarial Practice, Vol. 9, págs. 84-112.
- Boj, E.; M. Claramunt, J. Fortiana y Á. Vegas (2005). “*Bases de datos y estadísticas del seguro de automóviles en España: influencia en el cálculo de primas*”. Estadística Española Vol. 47, nº 160, págs. 539-566.
- Boucher, J.P.; M. Denuit y M. Guillén (2006). “*Risk classification for claims counts: Zero-inflated mixed Poisson and hurdle models*”. Working Paper 06-06, Institut des Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Bélgica.
- Brouhns, N.; M. Guillén, M. Denuit y J. Pinquet (2003). “*Bonus-Malus Scales in Segmented Tariffs with Stochastic Migration Between Segments*”. The Journal of Risk and Insurance Vol. 70, nº 4, págs. 577-599.
- Denuit, M.; X. Maréchal, S. Pitrebois y J.F. Walhin (2007). Actuarial Modeling of Claim Counts. Risk Classification, Credibility and Bonus-Malus Systems. John Wiley & Sons, Inc. Chichester.
- Gómez-Déniz, E.; J.M. Sarabia y J. Prieto (2009). “*La Distribución Poisson-Beta: Aplicaciones y Propiedades en la Teoría del Riesgo Colectivo*”. Anales del Instituto de Actuarios, 2009, págs. 141-160.

- Hogg, R.V. y S.A. Klugman (1984). *Loss Distributions*. John Wiley & Sons, Inc. New York.
- Klugman, S.A.; H. Panjer y G.E. Willmot (1998). *Loss Models: From Data to Decisions*. John Wiley & Sons, Inc. New York.
- Kupper, J. (1963). “*Some aspects of cumulative risk*”. *ASTIN Bulletin* Vol. 3, nº 1, págs. 85-102.
- Lemaire, J. (1985). *Automobile Insurance. Actuarial Models*. Kluwer Academic Publishers. Boston.
- Meng, W.Y. y G.A. Whitmore (1999). “*Accounting for Individual Over-Dispersion in a Bonus-Malus Automobile Insurance System*”. *ASTIN Bulletin* Vol. 29, nº 2, págs. 327-337.
- Pitrebois, S.; M. Denuit y J.F. Walhin (2006). “*An actuarial analysis of the French bonus-malus system*”. *Scandinavian Actuarial Journal* Volume 2006, Núm. 5, págs. 247-264.
- Rolski, T.; H. Schmidk, V. Schmidt y J. Teugels (1998). *Stochastic Processes for Insurance and Finance*. John Wiley & Sons, Inc. New York.
- Tremblay, L. (1992). “*Using the Poisson Inverse Gaussian in Bonus-Malus Systems*”. *ASTIN Bulletin* Vol. 22, nº 1, págs. 97-106.
- Vilar, J.L. y J. Vegas (1998). “*Estimación de la provisión de estabilización y del recargo técnico sobre primas a partir del ajuste de una distribución de Poisson compuesta para el número de siniestros*”. *Anales del Instituto de Actuarios*, 1998, págs. 69-91.
- Walhin, J.F. y J. Paris (2000). “*The True Claim Amount and Frequency Distribution of a Bonus-Malus System*”. *ASTIN Bulletin* Vol. 30, nº 2, págs. 391-403.
- Willmot, G. E. (1988). “*Sundt and Jewell’s family of discrete distributions*”. *ASTIN Bulletin* Vol. 18, págs. 17-29.

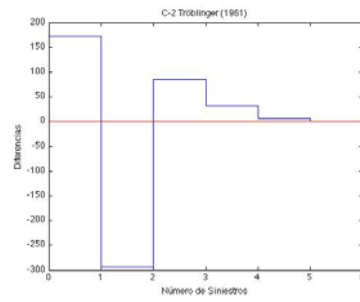
Anexo Gráficos

Los siguientes gráficos muestran las diferencias entre la cartera real y el ajuste mediante una distribución de Poisson para el número de siniestros. Todas las carteras tienen el mismo comportamiento, inflado de ceros, desinflado de unos y de nuevo inflado de las colas.

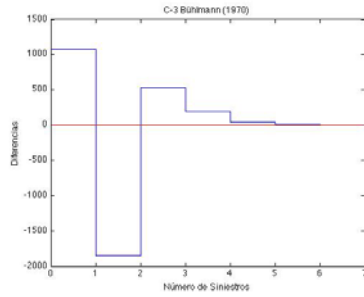
C-1 Cartera Thyron



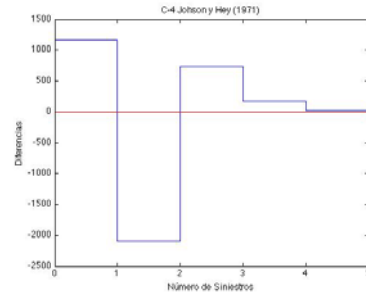
C-2 Cartera Tröblinger



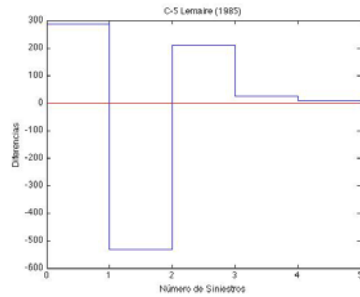
C-3 Cartera de Bühlmann



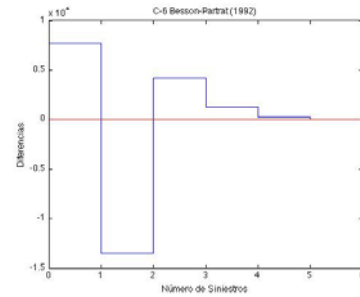
C-4 Cartera Johson-Hey



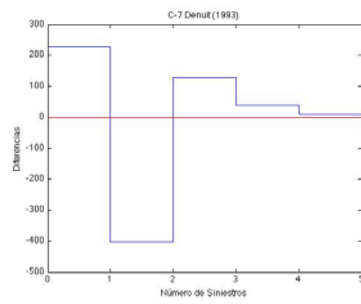
C-5 Cartera Lemaire



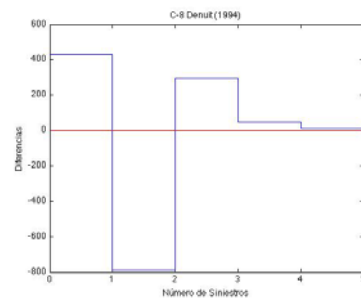
C-6 Cartera Besson-Partrat



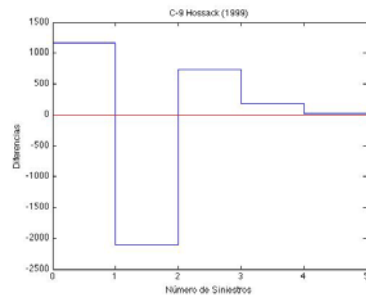
C-7 Cartera Bélgica 93



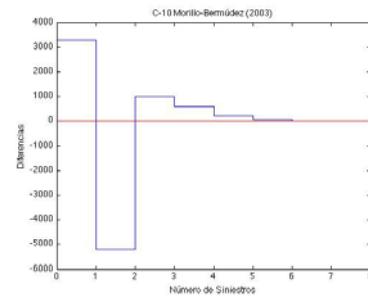
C-8 Cartera Bélgica 94



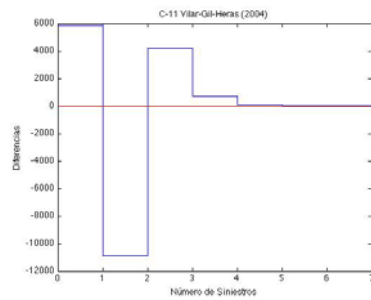
C-9 Cartera Hossack



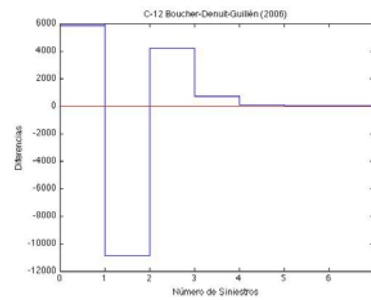
C-10 Cartera Morillo-Bermúdez



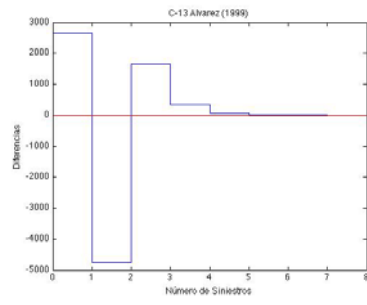
C-11 Cartera Boucher-Denuit-Guillén



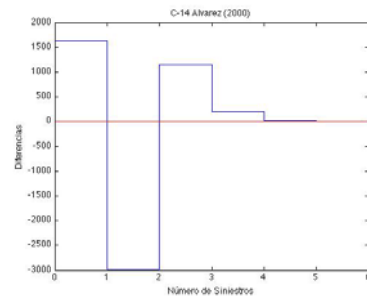
C-12 Cartera Vilar-Gil-Heras



C-13 Nueva Cartera 1999



C-14 Nueva Cartera 2000



C-15 Nueva Cartera 2001

