

Reservas de Pérdidas Estocásticas utilizando modelos MCMC Bayesianos

Diego Gonzales
Seminario de Investigación
Universidad Nacional Autónoma de Honduras
e-mail: diego.gonzales@unah.hn

ÍNDICE

I.	Introducción	1
II.	Objetivos	2
III.	Definición del Problema	2
IV.	Preliminares y notación	2
IV-A.	La base de Datos de reservas de pérdidas de CAS	3
IV-B.	Validación del Modelo de Mack	4
IV-C.	Validación del modelo ODP de Bootstrap	4
IV-D.	Modelo Bayesiano para Datos de Pérdidas Incurridas	5
IV-D1.	Modelo de Chain-Ladder Correlacionado (CCL)	5
IV-E.	Modelo Bayesiano para Datos de Pérdidas Pagadas	5
IV-E1.	Modelo de Tasa de Asentamiento Cambiante (CSR)	5
V.	Resultados	5
V-A.	Modelo de Mack	5
V-B.	Modelo ODP de Bootstrap	6
V-C.	Modelo de Chain-Ladder Correlacionado (CCL)	8
V-D.	Modelo de Tasa de Asentamiento Cambiante (CSR)	8
VI.	Conclusiones	9
VII.	Trabajo futuro	9
VIII.	Repositorio de archivos del proyecto	9
	Referencias	9

ÍNDICE DE FIGURAS

1.	Gráficos p-p Prueba de Uniformidad Ilustrativo	3
2.	Gráficos p-p para el Modelo de Mack en Triángulos de Pérdidas Incurridas	6
3.	Gráficos p-p para el Modelo de Mack en Triángulos de Pérdidas Pagadas	6
4.	Gráficos p-p para el Modelo de ODP Bootstrap en Triángulos de Pérdidas Incurridas	7
5.	Gráficos p-p para el Modelo de ODP Bootstrap en Triángulos de Pérdidas Pagadas	7
6.	Gráficos p-p para el Modelo CCL en Triángulos de Pérdidas Incurridas	8
7.	Gráficos p-p para el Modelo CSR en Triángulos de Pérdidas Pagadas	8

ÍNDICE DE CUADROS

I.	Prima Neta Suscrita de la Aseguradora Ilustrativa	4
II.	Pérdidas Incurridas de la Aseguradora Ilustrativa Netas de Reaseguro	4
III.	Pérdidas Pagadas de la Aseguradora Ilustrativa Netas de Reaseguro	4
IV.	Resultados del Modelo Mack para las Pérdidas Incurridas del Asegurador Ilustrativo	6
V.	Resultados del Modelo Mack para las Pérdidas Pagadas del Asegurador Ilustrativo	6
VI.	Resultados del Modelo ODP para las Pérdidas Incurridas del Asegurador Ilustrativo	7
VII.	Resultados del Modelo ODP para las Pérdidas Pagadas del Asegurador Ilustrativo	7
VIII.	Resultados del Modelo CCL para las Pérdidas Incurridas del Asegurador Ilustrativo	8
IX.	Resultados del Modelo CSR para las Pérdidas Pagadas del Asegurador Ilustrativo	8

Reservas de Pérdidas Estocásticas utilizando modelos MCMC Bayesianos

Resumen—La estimación de reservas de pérdidas en seguros requiere métodos capaces de capturar adecuadamente la incertidumbre y la variabilidad de los resultados, utilizando modelos estocásticos basados en el enfoque Bayesiano con Cadenas de Markov Monte Carlo (MCMC). Se examinan las limitaciones de los modelos tradicionales más usados, el de Mack y el Bootstrap ODP, al demostrar que tienden a subestimar o sobreestimar la variabilidad de los resultados cuando se aplican a datos de pérdidas incurridas y pagadas. Para mejorar estas deficiencias, se desarrollan modelos bayesianos que incorporan características adicionales: correlación entre años de accidente, posibilidad de pérdidas incrementales negativas, tendencias en los años de pago y cambios en la velocidad de liquidación de siniestros. Con ayuda de una base de datos amplia de triángulos de desarrollo de pérdidas (CAS Loss Reserve Database), se realizan validaciones que muestran que estos modelos bayesianos, en particular el denominado Correlated Chain Ladder, ofrecen una mejor representación de la incertidumbre en comparación con los enfoques tradicionales. Se enfatiza la importancia de conocer la complejidad y el riesgo inherente en la estimación de reservas, y se propone que los actuarios consideren la implementación de estos modelos más flexibles para obtener estimaciones más realistas del rango esperado de resultados.

I. INTRODUCCIÓN

Para las aseguradoras generales, el pasivo más importante es la provisión para pérdidas no pagadas. A lo largo de los años, se han desarrollado una serie de modelos estocásticos para abordar este problema. Dos de los modelos no propietarios más prominentes son los de Mack (1993, 1994) y England y Verall (2002).

Con el permiso de la National Association of Insurance Commissioners (NAIC) y Shi (2011), se pudo reunir una base de datos que consta de una gran cantidad de triángulos del anexo P para seis líneas de seguro. Estos triángulos provenían de las Declaraciones Anuales de la NAIC reportadas en 1997.

Las secciones IV-B y IV-C intentan validar los modelos de Mack (1993,1994) y England y Verall (2002). Resulta que estos modelos no predicen con precisión la distribución de los resultados para los datos incluidos en la base de datos en cuestión. La explicación de estos resultados incluye lo siguiente:

- El entorno de pérdidas en seguros es demasiado dinámico para ser capturado en un único modelo estocástico de reservas para pérdidas.
- Podría haber otros modelos que se ajusten mejor a los datos existentes.
- Los datos utilizados para calibrar el modelo carecen de información crucial necesaria para realizar una predicción confiable. Ejemplos de tales cambios podrían incluir modificaciones en la forma en que se lleva a cabo el negocio subyacente, como cambios en los procesos de reclamaciones o cambios en la composición de reaseguro

directo/cedido/asumido de los valores de las reclamaciones en triángulos.

Las posibles formas de descartar el primer elemento mencionado anteriormente son (1) encontrar un mejor modelo; y/o (2) encontrar mejores datos.

Una característica de los modelos de reservas de pérdidas es que son complejos en el sentido de que tienen un número relativamente grande de parámetros. Una gran dificultad para cuantificar la incertidumbre en los parámetros de un modelo complejo ha sido que se requiere un considerable esfuerzo para derivar una fórmula para la distribución predictiva de los resultados. Hace un tiempo los modelos Bayesianos no eran prácticos para modelos de cualquier complejidad, pero con la reciente introducción de los modelos Bayesianos de Monte Carlo con cadenas de Markov (MCMC), los modelos complejos Bayesianos estocásticos de reservas de pérdidas son ahora prácticos en el entorno informático actual.

Aunque las cadenas de Markov han sido estudiadas durante mucho tiempo por los teóricos de la probabilidad, pasó un tiempo antes de que se reconociera su aplicación a la estadística bayesiana. A partir de la década de 1930, los físicos comenzaron a usar el muestreo estadístico de cadenas de Markov para resolver algunos de los problemas más complejos en la física nuclear.

Como ocurrió con otras ciencias sociales, el MCMC bayesiano debería tener eventualmente un efecto profundo en la ciencia actuarial. Y de hecho, su efecto ya ha comenzado. Scollnik) presentó a los actuarios los modelos MCMC bayesianos.

Aquí está la situación a la que nos enfrentamos ahora. Primero, podemos construir una amplia variedad de modelos propuestos y predecir su distribución de resultados utilizando la metodología bayesiana MCMC. Segundo, podemos validar un modelo de reservas de pérdidas estocástico propuesto usando un gran número de aseguradoras en la base de Datos de Reservas de Pérdidas de la CAS.

El público al que va dirigido éste artículo consiste en actuarios de seguros generales que estén familiarizados con los modelos de Mack (1993, 1994) y de England y Verall (2002). Se espera que la mayoría de secciones sean comprendidas para un actuario, aquellos que deseen un entendimiento más profundo.

Los scripts de computadora utilizados para implementar estos modelos están escritos en el lenguaje de programación R.

PALABRAS CLAVE

- Reservas Estocásticas

- Modelos Bayesianos
- MCMC
- Triángulos de desarrollo
- Incertidumbre actuarial

II. OBJETIVOS

- Evaluar el desempeño predictivo de los modelos estocásticos tradicionales (Mack y Bootstrap ODP) en la base de datos de triángulos de desarrollo de la CAS, mediante pruebas de validación retrospectiva para identificar sus limitaciones en la cuantificación de la incertidumbre.
- Desarrollar e implementar modelos Bayesianos MCMC que incorporen estructuras de dependencia entre años de accidente, con el fin de capturar de manera más realista la variabilidad en los datos de pérdidas incurridas y mejorar la precisión de las distribuciones predictivas.
- Diseñar modelos estocásticos avanzados para pérdidas pagadas que integren explícitamente tendencias en el tiempo y cambios en la velocidad de liquidación de siniestros, para corregir el sesgo de sobrestimación encontrado en los enfoques tradicionales.
- Validar y comparar el rendimiento de los nuevos modelos propuestos (como el CCL y el CSR) frente a los métodos establecidos, utilizando técnicas estadísticas como gráficos p-p y la prueba de Kolmogorov-Smirnov para determinar su capacidad de generar percentiles de resultados uniformemente distribuidos.
- Evaluar el impacto de incorporar información actuarial experta a través de distribuciones previas informativas en los modelos MCMC, y medir cómo esto afecta la precisión y la incertidumbre de las estimaciones de reservas finales.

III. DEFINICIÓN DEL PROBLEMA

El problema central consiste en determinar con precisión las reservas de pérdidas en seguros, es decir, los montos que deben apartar las aseguradoras para cubrir siniestros ocurridos pero aún no pagados, enfrentando la dificultad de que los modelos tradicionales suelen subestimar o sobreestimar la variabilidad real de los resultados. Esto genera un reto actuarial: desarrollar y validar modelos estocásticos que capten de manera más realista la incertidumbre, incorporando elementos como correlaciones entre años, tendencias en los pagos y cambios en los procesos de liquidación, con el fin de mejorar la confiabilidad en la estimación de reservas.

IV. PRELIMINARES Y NOTACIÓN

Comenzaré explicando lo que es una reserva de pérdidas incurridas y pérdidas pagadas, para entender mucho mejor la idea central del artículo.

■ Reserva para Pérdidas Incurridas:

Es el estimado del monto total que una aseguradora eventualmente pagará por los siniestros ocurridos en un período determinado (año de accidente), independientemente de si se han pagado o no. Incluye:

- Pérdidas ya pagadas: El dinero que ya ha salido de la compañía para pagar siniestros.

- Provisiones para pérdidas reportadas pero no pagadas: Siniestros que se han reportado oficialmente, pero cuyo pago aún está pendiente.
- Provisiones para pérdidas incurridas pero no reportadas: Siniestros que ya han ocurrido pero que la aseguradora aún no ha recibido notificación.

- **Reserva para Pérdidas Pagadas:** Es el estimado del monto futuro que la aseguradora deberá desembolsar para liquidar los siniestros ocurridos en un periodo determinado. Esta reserva se centra exclusivamente en los flujos de caja futuros.

- No incluye el dinero que ya ha sido pagado.
- Se calcula proyectando los pagos futuros basándose en el historial de desembolsos pasados.

- De manera general, definimos una notación que nos ayudará a comprender el significado de variables mencionadas en este artículo, dependiendo del contexto, varias cantidades, como $C_{w,d}$, representarán observaciones, estimaciones o variables aleatorias.

En situaciones donde no pueda estar claro, adoptemos la convención de que para una cantidad X , \tilde{X} indicará que X está siendo tratada como una variable aleatoria o simulada, \hat{X} denotará una estimación de X , y una X será tratada como una observación o parámetro fijo.

■ Prueba de Uniformidad:

Una prueba visual de uniformidad que permite comprobar la significancia estadística es el gráfico p-p combinado con la prueba de Kolmogorov-Smirnov (K-S). Así es como funciona, Supongamos que se tiene una muestra de n percentiles pronosticados que van de 0 a 100 y se ordenan en orden creciente. El valor esperado de estos percentiles viene dado por $\{e_i\} = 100 \cdot \{1/(n+1), 2/(n+1), \dots, n/(n+1)\}$. Luego, se grafican los percentiles esperados en el eje horizontal contra los percentiles pronosticados ordenados en el eje vertical. Si los percentiles pronosticados están distribuidos uniformemente, esperamos que este gráfico se sitúe a lo largo de una línea de 45° . Según la prueba de K-S descrita por Klugman, Panjer y Willmot (2012, p.331), se puede rechazar la hipótesis de un conjunto de percentiles $\{p_i\}$ es uniforme al 5 % si $D \equiv \max |p_i - f_i|$ es mayor que su valor crítico, $136/\sqrt{n}$ donde $\{f_i\} = 100 \cdot \{1/n, 2/n, \dots, n/n\}$. Esto se representa visualmente en un gráfico p-p dibujando líneas a una distancia $136/\sqrt{n}$ por encima y por debajo de la línea de 45° . Rechazamos la hipótesis de uniformidad si el gráfico p-p se sitúa fuera de la banda definida por esas líneas. Para este artículo, un modelo se considerará validado si pasa la prueba K-S al nivel del 5 %. Un análisis gráfico se logra visualizar en la **figura 1**.

- **Distribuciones:** A continuación se mencionan algunas distribuciones utilizadas en cada uno de los modelos.

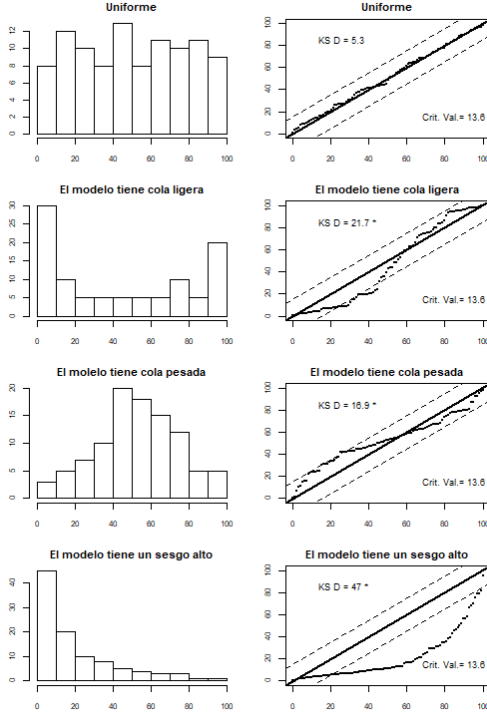


Figura 1. Gráficos p-p Prueba de Uniformidad Ilustrativo

- **Distribución log-normal**

$X \sim \log \text{ normal}(\mu, \sigma^2)$ con $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0$$

$$E(X) = \exp(\mu + \sigma^2/2).$$

$$\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$

- **Distribución Poisson**

$X \sim \text{Poisson}(\lambda)$ con $\lambda > 0$.

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \text{para } x = 0, 1, \dots$$

$$E(X) = \lambda.$$

$$\text{Var}(X) = \lambda.$$

- **Distribución Uniforme**

$X \sim \text{unif}(a, b)$ con $a < b$.

$$f(x) = \frac{1}{b-a}, \quad \text{para } x \in (a, b).$$

$$E(X) = \frac{a+b}{2}.$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}.$$

Ahora, daremos una introducción a los Modelos Bayesianos de MCMC:

Desde el reconocimiento de las Cadenas de Markov Monte Carlo (MCMC) como una herramienta poderosa para realizar análisis Bayesianos en 1990, ha habido muchos esfuerzos para crear software que ayude en estos análisis.

Así es como funcionan los análisis Bayesianos MCMC en la práctica.

1. El usuario especifica la distribución previa, $p(y)$, y la distribución condicional, $f(x|y)$.
2. El usuario selecciona un vector inicial, x_1 , y luego, usando una simulación por computadora, ejecuta la cadena de Markov a través de un número suficientemente grande, t_1 , de iteraciones. Esta primera fase de la simulación se llama fase adaptativa, donde el algoritmo se modifica automáticamente para aumentar su eficiencia.
3. Luego, el usuario ejecuta t_2 iteraciones adicionales. Esta fase se llama fase de burn-in o calentamiento. t_2 se selecciona para ser lo suficientemente alto como para que una muestra tomada de los t_3 periodos subsiguientes represente la distribución posterior.
4. El usuario luego ejecuta t_3 iteraciones adicionales y luego toma una muestra, $\{x_i\}$, desde el paso $(t_2 + 1)$ hasta el paso $(t_2 + t_3)$ para presentar la distribución posterior $f(y|x)$.
5. A partir de la muestra, se construyen varios estadísticos de interés que son relevantes para el problema abordado por el análisis.

Una cadena de Markov es un proceso aleatorio en el que la transición al siguiente estado depende solo de su estado actual, y no de estados anteriores. Formalmente, una cadena de Markov, X_t para $t = 1, 2, \dots$ es una secuencia de vectores que satisface la propiedad de que

$$\begin{aligned} P(X_{t+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) \\ = P(X_{t+1} = x | X_t = x_t). \end{aligned}$$

Hay una cierta clase de Cadenas de Markov, generalmente llamadas ergódicas, para las cuales los vectores, $\{X_t\}$, se aproximan a una distribución límite. Es decir, a medida que T aumenta, la distribución de $\{X_t\}$ para todo $t > T$ se aproxima a una distribución límite única.

Sea x un vector de observaciones y sea y un vector de parámetros en un modelo. En los análisis Bayesianos MCMC, la cadena de Markov se define en términos de la distribución previa, $p(y)$, y la distribución condicional $f(x|y)$. La distribución límite es la distribución posterior, $f(y|x)$. Es decir, si dejamos que la cadena se ejecute el tiempo suficiente, la cadena visitará aleatoriamente todos los estados con una frecuencia que es proporcional a sus probabilidades posteriores.

IV-A. La base de Datos de reservas de pérdidas de CAS

El Anexo P del Estado Anual de la NAIC (National Association of Insurance Commissioners) contiene triángulos de liquidación a nivel de aseguradora de pérdidas agregadas por línea de seguro. En el Anexo P se informan triángulos tanto de pérdidas pagadas como incurridas (netas de reaseguro).

Para ilustrar los ejemplos de este artículo, se seleccionaron triángulos de pérdidas incurridas y pagadas de una sola aseguradora en la base de datos, cuyos datos se encuentran en las tablas I, II y III. La información por encima de la diagonal en estos triángulos, corresponden a la Declaración Anual de 1997, sirvió para desarrollar los modelos de reservas. Por otro lado, los datos por debajo de la diagonal, obtenidos de declaraciones anuales posteriores (detalladas en la columna Fuente), se emplearon exclusivamente para validar la precisión predictiva de dichos modelos.

Tabla I
PRIMA NETA SUSCRITA DE LA ASEGURADORA ILUSTRATIVA

AY	1	2	3	4	5	6	7	8	9	10
Prima	5812	4908	5454	5165	5214	5230	4992	5466	5226	4962

Tabla II
PÉRDIDAS INCURRIDAS DE LA ASEGURADORA ILUSTRATIVA NETAS DE REASEGURO

AY/Lag	1	2	3	4	5	6	7	8	9	10	Fuente
1988	1722	3830	3603	3835	3873	3895	3918	3918	3917	3917	1997
1989	1581	2192	2528	2533	2528	2530	2534	2541	2538	2532	1998
1990	1834	3009	3488	4000	4105	4087	4112	4170	4271	4279	1999
1991	2305	3473	3713	4018	4295	4334	4343	4340	4342	4341	2000
1992	1832	2625	3086	3493	3521	3563	3542	3541	3541	3587	2001
1993	2289	3160	3154	3204	3190	3206	3351	3289	3267	3268	2002
1994	2881	4254	4841	5176	5551	5689	5683	5688	5684	5684	2003
1995	2489	2956	3382	3755	4148	4123	4126	4127	4128	4128	2004
1996	2541	3307	3789	3973	4031	4157	4143	4142	4144	4144	2005
1997	2203	2934	3608	3977	4040	4121	4147	4155	4183	4181	2006

Tabla III
PÉRDIDAS PAGADAS DE LA ASEGURADORA ILUSTRATIVA NETAS DE REASEGURO

AY/Lag	1	2	3	4	5	6	7	8	9	10	Source
1988	952	1529	2813	3647	3724	3832	3899	3907	3911	3912	1997
1989	849	1564	2202	2432	2468	2487	2513	2526	2531	2527	1998
1990	983	2211	2830	3832	4039	4065	4102	4155	4268	4274	1999
1991	1657	2685	3169	3600	3900	4320	4332	4338	4341	4341	2000
1992	932	1940	2626	3332	3368	3491	3531	3540	3540	3583	2001
1993	1162	2402	2799	2996	3034	3042	3230	3238	3241	3268	2002
1994	1478	2980	3945	4714	5462	5680	5682	5683	5684	5684	2003
1995	1240	2080	2607	3080	3678	2004	4117	4125	4128	4128	1997
1996	1326	2412	3367	3843	3965	4127	4133	4141	4142	4144	2005
1997	1413	2683	3173	3674	3805	4005	4020	4095	4132	4139	2006

IV-B. Validación del Modelo de Mack

Esta sección describe un intento de validar el modelo de Mack con los datos de pérdidas incurridas de varias aseguradoras, incluidos en la base de datos de la CAS. Comenzaremos con el modelo Chain-Ladder clásico.

Sea $C_{w,d}$ el monto acumulado de pérdidas, ya sean incurridas o pagadas, para el año de accidente w , y el retraso de desarrollo d , para $1 \leq w \leq K$ y $1 \leq d \leq K$. Se conocen los valores de $C_{w,d}$ para el "triángulo" de datos especificado por $w + d \leq K + 1$. El objetivo de este modelo es estimar los montos de pérdida en la última columna de datos, $C_{w,K}$ para $w = 2, \dots, K$. Para usar el modelo Chain-Ladder, primero se calculan los factores de edad a edad dados por

$$f_d = \frac{\sum_{w=1}^{K-d} C_{w,d+1}}{\sum_{w=1}^{K-d} C_{w,d}} \text{ para } d = 1, \dots, K-1.$$

La estimación Chain-Ladder de $C_{w,K}$ es el producto de la última pérdida reportada $C_{w,K+1-w}$ y los subsiguientes

factores de edad a edad $f_{K+1-w} \cdots f_{K-1}$. Juntando esto, tenemos

$$C_{w,K} = C_{w,K+1-w} \cdot f_{K+1-w} \cdots f_{K-1}$$

Mack (1993, 1994) convierte el modelo Chain-Ladder determinístico en un modelo estocástico tratando primero a $\tilde{C}_{w,d}$ como una variable aleatoria que representa el monto de pérdida acumulada en la celda (w,d) . Luego hace tres supuestos:

1. $E[\tilde{C}_{w,d+1}|C_{w,1}, \dots, C_{w,d}] = C_{w,d} \cdot f_d$
2. Para cualquier d dado, las variables aleatorias $\tilde{C}_{w,d}$ y $\tilde{C}_{v,d}$ son independientes para $v \neq w$.
3. $Var[\tilde{C}_{w,d+1}|C_{w,1}, \dots, C_{w,d}] = C_{w,d} \cdot \alpha_d^2$,

donde α_d^2 es la varianza del factor de desarrollo, es un parámetro que captura la incertidumbre o volatilidad en la evolución de las pérdidas entre los periodos de desarrollo d y $d+1$.

La estimación de Mack para $E[C_{w,K}]$ para $w = 2, \dots, K$ está dada por

$$\hat{C}_{w,K} = C_{w,K+1-w} \cdot \hat{f}_{K+1-w} \cdots \hat{f}_{K-1}$$

donde

$$\hat{f}_d = \frac{\sum_{w=1}^{K-d} C_{w,d+1}}{\sum_{w=1}^{K-d} C_{w,d}}$$

La tabla IV aplica las expresiones de Mack a los datos de la aseguradora ilustrativa de la tabla II utilizando el paquete Chainladder de R. Además de las estadísticas de pérdidas calculadas por las expresiones de Mack, la tabla IV contiene los resultados observados $C_{w,10}$ de la tabla II. Siguiendo la sugerencia de Mack, se calcula el percentil de $\sum_{w=1}^{10} C_{w,10}$ asumiendo una distribución lognormal con la misma media y desviación estándar.

Se calcula el percentil para cada triángulo, el criterio de aplicabilidad del modelo es que estos percentiles deberían estar distribuidos uniformemente, y para una muestra suficientemente grande la uniformidad es comprobable.

IV-C. Validación del modelo ODP de Bootstrap

Este modelo fué diseñado para trabajar con pérdidas incrementales, $I_{w,d}$, en lugar de las pérdidas acumuladas $C_{w,d}$, donde $I_{w,1} = C_{w,1}$ y $I_{w,d} = C_{w,d} - C_{w,d-1}$ para $d > 1$.

Una suposición clave de este modelo es que las pérdidas incrementales se describen mediante la distribución de Poisson sobredispersa con

$$E[\tilde{I}_{w,d}] = \alpha_w \cdot \beta_d \text{ y } Var[\tilde{I}_{w,d}] = \Phi \cdot \alpha_w \cdot \beta_d$$

Donde, α_w representa el efecto del año de accidente, β_d representa el patrón de desarrollo y Φ representa el parámetro de dispersión.

Los parámetros del modelo pueden ser estimados por un paquete estándar de modelo lineal generalizado (GLM). Luego, utilizan un procedimiento de remuestreo bootstrap para cuantificar la volatilidad de la estimación.

IV-D. Modelo Bayesiano para Datos de Pérdidas Incurridas

Se propone un modelo que considera la correlación entre los años de accidente.

IV-D1. Modelo de Chain-Ladder Correlacionado (CCL): Sea:

1. Cada $\alpha_w \sim normal(\log(Prima_w) + logelr, \sqrt{10})$ donde el parámetro $logelr \sim uniform(-1, 0.5)$.
2. $\mu_{1,d} = \alpha_1 + \beta_d$.
3. $\mu_{w,d} = \alpha_w + \beta_d + \rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})$, $w > 1$.
4. $\tilde{C}_{w,d}$ tiene una distribución lognormal con media logarítmica $\mu_{w,d}$ y desviación estándar logarítmica σ_d sujeta a la restricción de que $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$.

Nótese que el modelo CCL se reduce a un modelo Nivelado cuando $\rho = 0$.

Si los parámetros $\{\alpha_w\}$, $\{\beta_d\}$, y ρ están dados, el parámetro ρ es igual al coeficiente de correlación entre $\log(\tilde{C}_{w-1,d})$ y $\log(\tilde{C}_{w,d})$. Para ver esto, primero notamos que incondicionalmente:

$$\begin{aligned} E(\log(\tilde{C}_{w,d})) &= \mu_{w,d} \\ &= \alpha_w + \beta_d + \rho \cdot (\log(\tilde{C}_{w-1,d}) - \mu_{w-1,d}) \\ &= \alpha_w + \beta_d \end{aligned}$$

Dado $C_{w-1,d}$ tenemos que:

$$\begin{aligned} &((\log(\tilde{C}_{w,d}) - \alpha_w + \beta_d) \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})) \\ &= (\mu_{w,d} - (\alpha_w + \beta_d)) \cdot (\log(C_{w-1,d}) - \mu_{w-1,d}) \\ &= \rho \cdot (\log(C_{w-1,d}) - \mu_{w-1,d})^2 \end{aligned}$$

Entonces el coeficiente de correlación entre $\tilde{C}_{w,d}$ y $\tilde{C}_{w-1,d}$ está dado por:

$$\begin{aligned} E_{C_{w-1,d}} \left(E_{C_{w,d}} \left(\frac{\log(\tilde{C}_{w,d}) - (\alpha_w + \beta_d)}{\sigma_d} \cdot \frac{\log(C_{w-1,d}) - \mu_{w-1,d}}{\sigma_d} \middle| C_{w-1,d} \right) \right) \\ = E_{C_{w-1,d}} \left(\frac{\rho \cdot (\log(\tilde{C}_{w-1,d}) - \mu_{w-1,d})^2}{\sigma_d^2} \right) = \rho \end{aligned}$$

Para evitar que el modelo esté sobredeterminado, se establece $\beta_{10} = 0$. A los parámetros $\{\alpha_w\}$, $\{\sigma_d\}$ y los $\{\beta_d\}$ restantes se les asignan distribuciones previas relativamente amplias de la siguiente manera:

1. Cada $\alpha_w \sim normal(\log(Prima_w) + logelr, \sqrt{10})$ donde el parámetro $logelr \sim uniform(-1, 0.5)$.
2. Cada $\beta_d \sim uniform(-5, 5)$ para $d < 10$.
3. $\rho \sim uniform(-1, 1)$ — El rango permisible completo para ρ .
4. Cada $\sigma_d = \sum_{i=d}^{10} a_i$ donde $a_i \sim uniform(0, 1)$.

IV-E. Modelo Bayesiano para Datos de Pérdidas Pagadas

Se propone el siguiente modelo con el objetivo principal de capturar de manera explícita las tendencias temporales en la velocidad de liquidación de los siniestros.

IV-E1. Modelo de Tasa de Asentamiento Cambiante (CSR): Sea:

1. Cada $\alpha_w \sim normal(\log(Prima_w) + logelr, \sqrt{10})$ donde el parámetro $logelr \sim uniform(-1, 0.5)$.
2. Cada $\beta_d \sim uniform(-5, 5)$ para $d = 1, \dots, 9$, $\beta_{10} = 0$.
3. $\mu_{w,d} = \alpha_w + \beta_d \cdot (1 - \gamma)^{w-1}$ $\gamma \sim normal(0, 0.025)$
4. Cada $\sigma_d = \sum_{i=d}^{10} a_i$ donde $a_i \sim uniform(0, 1)$.
5. $\tilde{C}_{w,d}$ tiene una distribución lognormal con media logarítmica $\mu_{w,d}$ y desviación estándar logarítmica σ_d sujeta a la restricción de que $\sigma_1 > \sigma_2 > \dots > \sigma_{10}$.

Dado que $\beta_{10} = 0$ y las pérdidas pagadas acumuladas generalmente aumentan con el año de desarrollo, d , β_d para $d < 10$ suele ser negativo. Entonces, para cada $d < 10$, un valor positivo de γ hará que $\beta_d \cdot (1 - \gamma)^{w-1}$ aumente con w y, por lo tanto, indique una aceleración en el arreglo de reclamaciones. De manera similar, un valor negativo de γ indicará una desaceleración en la liquidación de reclamaciones.

V. RESULTADOS

Para poder validar cada uno de los modelos, se hacen los cálculos para 200 triángulos de reservas seleccionados de cuatro líneas de seguro.

Líneas de seguro seleccionadas:

1. CA: Commercial auto.
2. PA: Personal auto.
3. WC: Workers' Compensation.
4. OL: Other Liability.

En los resultados, se muestran algunas tablas ilustrativas que contienen la siguiente información:

- w : año de accidente.
- $\hat{C}_{w,10}$: estimación del monto acumulado de pérdidas.
- DE: desviación estándar.
- CV: coeficiente de variación.
- $C_{w,10}$: monto acumulado de pérdidas.
- Percentil.

Para cada modelo, los gráficos se realizaron primero por separado para los percentiles de resultados en cada línea de seguro y luego se combinan las cuatro líneas de seguro para ver si el modelo subestima o sobreestima la variabilidad de los datos.

A continuación se presentan algunas tablas ilustrativas y gráficos para cada uno de los modelos, estas tablas corresponden a los resultados que se obtienen de cada triángulo de desarrollo y los gráficos corresponden a las líneas de seguro utilizadas.

V-A. Modelo de Mack

Aquí se presentan dos tablas ilustrativas, la tabla IV corresponde a las Pérdidas Incurridas y la tabla V a Pérdidas pagadas, y luego se muestran los gráficos para cada línea de seguro.

Tabla IV
RESULTADOS DEL MODELO MACK PARA LAS PÉRDIDAS INCURRIDAS DEL
ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3917	0	0.000	3917	
2	2538	0	0.000	2532	
3	4167	3	0.001	4279	
4	4367	37	0.009	4341	
5	3597	34	0.010	3587	
6	3236	40	0.012	3268	
7	5358	146	0.027	5684	
8	3765	225	0.060	4128	
9	4013	412	0.103	4144	
10	3955	878	0.222	4181	
Total	38914	1057	0.027	40061	86.03

Tabla V
RESULTADOS DEL MODELO MACK PARA LAS PÉRDIDAS PAGADAS DEL
ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3912	0	0.0000	3912	
2	2532	0	0.0000	2527	
3	4162	3	0.0007	4274	
4	4370	28	0.0064	4341	
5	3555	35	0.0098	3583	
6	3213	157	0.0489	3268	
7	5167	251	0.0486	5684	
8	3442	385	0.1119	4128	
9	4210	750	0.1781	4144	
10	4616	957	0.2073	4139	
Total	39177	1442	0.0368	40000	72.02

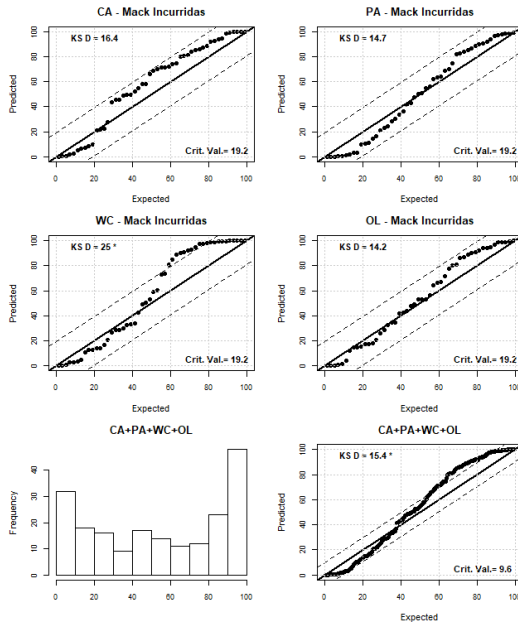


Figura 2. Gráficos p-p para el Modelo de Mack en Triángulos de Pérdidas Incurridas

Interpretación gráfica (figura 2):

El análisis gráfico revela que el modelo de Mack presenta un patrón sistemático de S inclinada en la gráfica p-p, característico de una distribución predictiva con colas ligeras. Si bien individualmente tres de las cuatro líneas (CA, PA, OL) no superan el valor crítico de K-S, todas exhiben la misma tendencia a subestimar la variabilidad de los resultados. Al combinar las 200 observaciones, se rechaza la hipótesis de uniformidad ($KS D = 15.04^* > 9.6$), confirmando que el

modelo no captura adecuadamente los eventos extremos en las colas de la distribución. Concluimos que la distribución predicha por el modelo de Mack es demasiado ligera en las colas para esos datos.

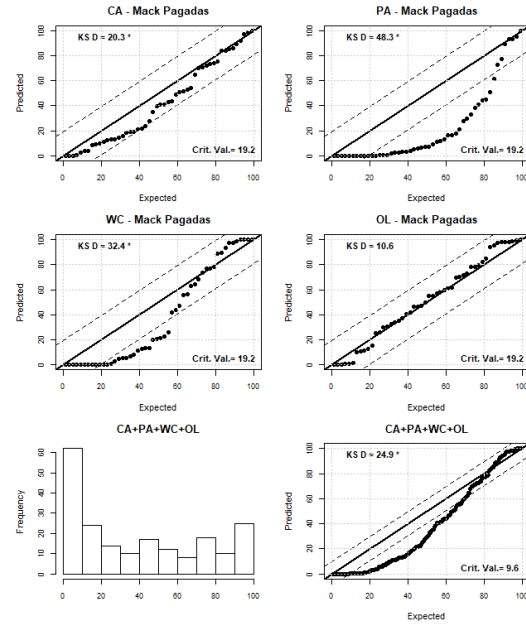


Figura 3. Gráficos p-p para el Modelo de Mack en Triángulos de Pérdidas Pagadas

Interpretación gráfica (figura 3):

El análisis gráfico revela que el modelo de Mack presenta un patrón sistemático de S invertida en los gráficos p-p, característico de una distribución que tiende a sobreestimar las pérdidas finales. Tres de las cuatro líneas (CA con $KS D = 20.3^*$, PA con $KS D = 48.3^*$ y WC con $KS D = 32.4^*$) exceden sus valores críticos de K-S de 19.2, siendo PA la que muestra el peor ajuste con una desviación extremadamente pronunciada de la línea de 45° en los percentiles medios. La línea de OL es la única que pasa la prueba con $KS D = 10.6$, manteniéndose dentro de las bandas de confianza. Al combinar las 200 observaciones, se rechaza la hipótesis de uniformidad ($KS D = 24.9^* > 9.6$) y el histograma muestra una concentración marcada de resultados observados en percentiles bajos (0-40), confirmando que el modelo produce estimaciones sistemáticamente más altas que los resultados reales. Este sesgo sugiere que el modelo de Mack no captura adecuadamente los cambios en el entorno de liquidación de siniestros que podrían estar presentes en los datos de pérdidas pagadas. Concluimos que la distribución predicha por el modelo de Mack tiende a ser sobreestimada para estos datos de pérdidas pagadas, lo que limita su aplicabilidad general y justifica la búsqueda de modelos alternativos más sofisticados.

V-B. Modelo ODP de Bootstrap

Aquí se presentan dos tablas ilustrativas, la tabla VI corresponde a las Pérdidas Incurridas y la tabla VII a Pérdidas pagadas, y luego se muestran los gráficos para cada línea de seguro.

Tabla VI
RESULTADOS DEL MODELO ODP PARA LAS PÉRDIDAS INCURRIDAS DEL
ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3917	0	0.0000	3917	
2	2538	0	0.0000	2532	
3	4167	46	0.0110	4279	
4	4367	95	0.0218	4341	
5	3596	101	0.0281	3587	
6	3235	118	0.0365	3268	
7	5351	206	0.0385	5684	
8	3758	277	0.0737	4128	
9	4014	379	0.1844	4144	
10	3954	729	0.1844	4181	
Total	38897	1004	0.0258	40061	87.62

Tabla VII
RESULTADOS DEL MODELO ODP PARA LAS PÉRDIDAS PAGADAS DEL
ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3912	0	0.0000	3912	
2	2532	21	0.0083	2527	
3	4161	49	0.0118	4274	
4	4369	86	0.0197	4341	
5	3556	98	0.0276	3583	
6	3212	148	0.0461	3268	
7	5166	245	0.0474	5684	
8	3440	338	0.0983	4128	
9	4216	578	0.1371	4144	
10	4634	1054	0.2274	4139	
Total	39197	1411	0.0360	40000	73.58

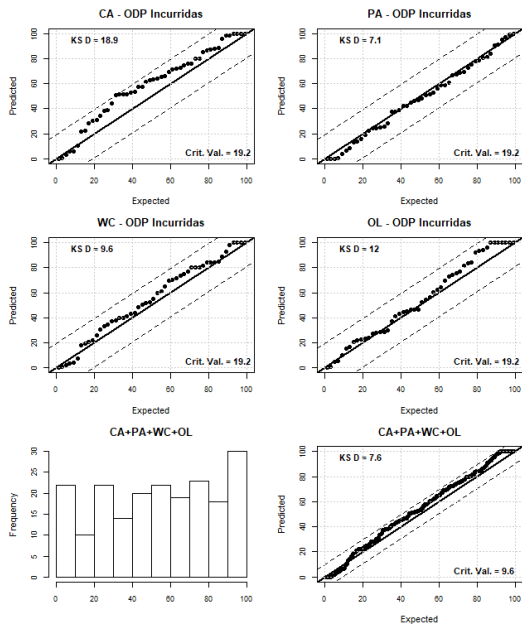


Figura 4. Gráficos p-p para el Modelo de ODP Bootstrap en Triángulos de Pérdidas Incurridas

Interpretación gráfica (figura 4):

El análisis gráfico revela que el modelo de Bootstrap para pérdidas incurridas presenta un desempeño notablemente superior al modelo de Mack, aunque persisten algunas deficiencias. Las cuatro líneas de seguro muestran valores de KS D por debajo de sus valores críticos de 19.2, lo que individualmente sugiere un ajuste aceptable. Sin embargo, al examinar los gráficos p-p se observa que todas las líneas mantienen un

patrón de S inclinada, característico de una distribución predictiva con colas ligeras indicando una tendencia residual a subestimar la variabilidad de los resultados. Al combinar las 200 observaciones de las cuatro líneas, se obtiene KS D = 7.6, que se mantiene dentro del valor crítico de 9.6, validando técnicamente el modelo al nivel del 5% de significancia. El histograma combinado muestra una distribución de percentiles más uniforme que la observada en el modelo de Mack, con una representación más equilibrada en todos los rangos de percentiles. La persistencia del patrón de S inclinada en los gráficos individuales sugiere que, aunque el modelo ODP captura mejor la incertidumbre que el modelo de Mack, aún existe una ligera subestimación de la variabilidad en las colas de la distribución. Concluimos que el modelo ODP de Bootstrap ofrece una mejora sobre el modelo de Mack para pérdidas incurridas y pasa la prueba de validación K-S.

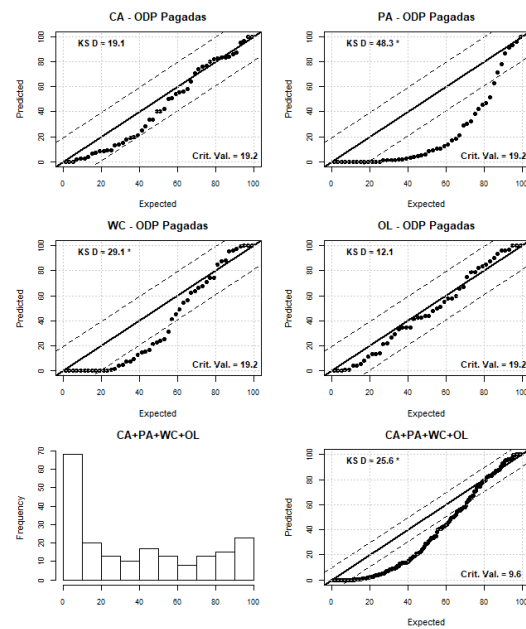


Figura 5. Gráficos p-p para el Modelo de ODP Bootstrap en Triángulos de Pérdidas Pagadas

Interpretación gráfica (figura 5):

El análisis gráfico revela ODP de Bootstrap para pérdidas pagadas presenta un comportamiento similar al modelo de Mack, exhibiendo un patrón sistemático de S invertida en los gráficos p-p que indica una tendencia a sobreestimar las pérdidas finales. Dos de las cuatro líneas exceden significativamente sus valores críticos de K-S de 19.2 (PA Y WC), CA con KS D = 19.1 apenas en el límite y OL con KS D = 12.1 se mantiene dentro de las líneas de confianza. Al combinar las 200 observaciones, se rechaza la hipótesis de uniformidad con KS D = 25.6* > 9.6, y el histograma muestra una marcada concentración de resultados en percentiles bajos (0-40), confirmando que el modelo produce estimaciones superiores a las pérdidas reales. Este patrón sugiere que tanto el modelo ODP como el de Mack no capturan adecuadamente los cambios en el entorno de liquidación de siniestros presentes en los datos de pérdidas pagadas. Concluimos que el modelo ODP de Bootstrap no valida para pérdidas pagadas y presenta

limitaciones similares al modelo de Mack, lo que justifica la necesidad de modelos alternativos más sofisticados que incorporen características adicionales específicas para datos de pérdidas pagadas.

V-C. Modelo de Chain-Ladder Correlacionado (CCL)

Tabla VIII
RESULTADOS DEL MODELO CCL PARA LAS PÉRDIDAS INCURRIDAS DEL ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3917	0	0.0000	3917	
2	2539	12	0.0047	2532	
3	4145	30	0.0072	4279	
4	4348	39	0.0090	4341	
5	3580	43	0.0120	3587	
6	3251	63	0.0194	3268	
7	5318	185	0.0348	5684	
8	3831	294	0.0767	4128	
9	4168	604	0.1449	4144	
10	4099	1084	0.2645	4181	
Total	39196	1292	0.0330	40061	77.56

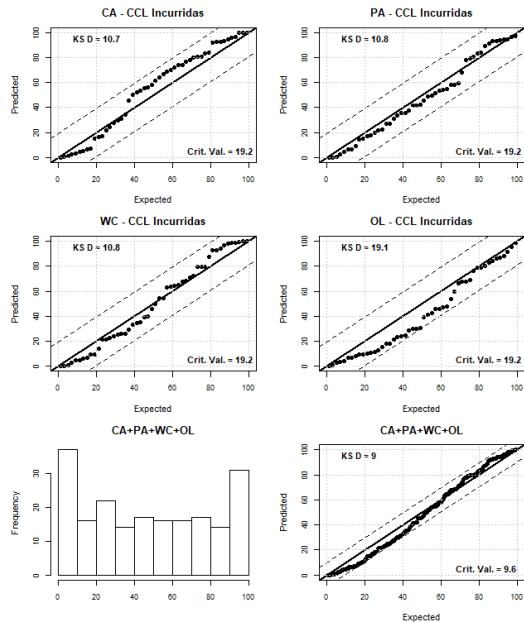


Figura 6. Gráficos p-p para el Modelo CCL en Triángulos de Pérdidas Incurridas

Interpretación gráfica (figura 6):

El análisis gráfico revela que el modelo CCL para pérdidas incurridas presenta el mejor desempeño entre todos los modelos evaluados. Las cuatro líneas de seguro individualmente muestran valores de KS D que se mantienen por debajo del valor crítico de 19.2, indicando un ajuste aceptable en cada línea. Los gráficos p-p individuales muestran una adherencia mejor a la línea de 45°, con las observaciones distribuyéndose de manera más uniforme dentro de las bandas de confianza.

Al combinar las 200 observaciones de las cuatro líneas de seguro, el modelo obtiene $KS D = 9$, que se mantiene por debajo del valor crítico de 9.6, validando exitosamente el modelo al nivel del 5% de significancia. El histograma

combinado muestra una distribución de percentiles más uniforme que la observada en los modelos de Mack y ODP, con una representación más equilibrada en todos los rangos de percentiles (0-100), lo que indica que el modelo captura adecuadamente tanto los eventos centrales como los extremos de la distribución.

La mejora en el desempeño del modelo CCL se atribuye directamente a la incorporación del parámetro ρ , que captura la dependencia entre años de accidente consecutivos. Esta característica permite al modelo reflejar de manera más realista la estructura de correlación presente en los datos de pérdidas incurridas, algo que los modelos tradicionales no consideran al asumir independencia entre años de accidente.

Concluimos que el modelo CCL valida exitosamente para pérdidas incurridas, ofreciendo una representación superior de la incertidumbre en comparación con los modelos tradicionales, y se recomienda su implementación para obtener estimaciones más confiables de las reservas de pérdidas incurridas.

V-D. Modelo de Tasa de Asentamiento Cambiante (CSR)

Tabla IX
RESULTADOS DEL MODELO CSR PARA LAS PÉRDIDAS PAGADAS DEL ASEGURADOR ILUSTRATIVO

w	$\hat{C}_{w,10}$	DE	CV	$C_{w,10}$	Percentil
1	3912	0	0.0000	3912	
2	2527	89	0.0349	2553	
3	4274	159	0.0385	4127	
4	4341	180	0.0421	4277	
5	3583	171	0.0487	3508	
6	3268	202	0.0611	3304	
7	5684	384	0.0776	4950	
8	4128	399	0.1211	3295	
9	4144	744	0.1996	3728	
10	4139	2062	0.5435	3794	
Total	40000	2735	0.0730	37449	87.45

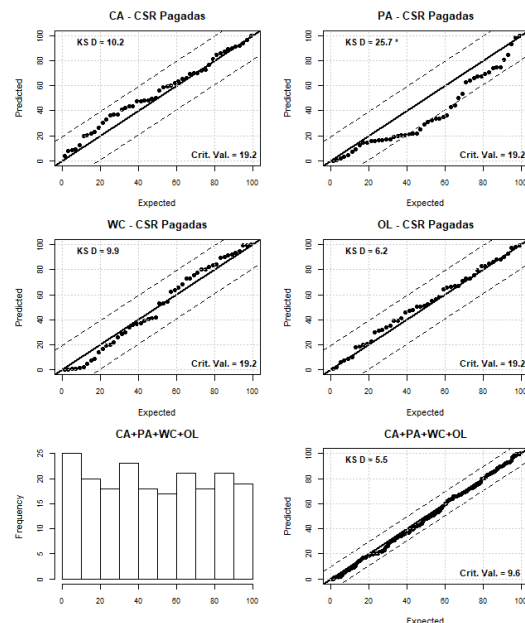


Figura 7. Gráficos p-p para el Modelo CSR en Triángulos de Pérdidas Pagadas

Interpretación gráfica (figura 7):

El análisis gráfico revela que el modelo CSR para pérdidas pagadas presenta un desempeño mixto entre las cuatro líneas de seguro evaluadas. Tres de las cuatro líneas de seguro individualmente muestran valores de KS D que se mantienen por debajo del valor crítico de 19.2, indicando un ajuste aceptable en éstas líneas. Sin embargo, la línea PA excede significativamente el valor crítico, mostrando un patrón de S invertida pronunciado que indica una tendencia a sobreestimar las pérdidas finales en esta línea específica.

Al combinar las 200 observaciones de las cuatro líneas de seguro, el modelo obtiene $KS D = 5.5$, que se mantiene muy por debajo del valor crítico de 9.6, validando exitosamente el modelo al nivel del 5% de significancia. El histograma muestra una distribución de percentiles notablemente uniforme, superior a la observada en los modelos de Mack y ODP de Bootstrap, con una representación equilibrada en todos los rangos de percentiles (0-100). Esto indica que el modelo captura adecuadamente los extremos de la distribución, a pesar de las dificultades observadas en la línea de PA.

Concluimos que el modelo CSR valida exitosamente para pérdidas pagadas, ofreciendo una representación superior de la incertidumbre en comparación con los modelos tradicionales de Mack y ODP (que ambos fallaron la validación para pérdidas pagadas con $KS D = 24.9^*$ y 25.6^* respectivamente), y se recomienda con su implementación para obtener estimaciones más confiables de las reservas de pérdidas pagadas.

VI. CONCLUSIONES

- Los modelos estocásticos tradicionales, como el de Mack y el de Bootstrap ODP, presentan limitaciones significativas para predecir de forma fiable el rango completo de resultados en reservas de pérdidas, ya que sistemáticamente subestiman la variabilidad en los datos incurridos y sobreestiman las estimaciones en los datos pagados.
- La incorporación de estructuras de dependencia, como la correlación entre años de accidente en el modelo CCL, demuestra ser un avance crucial, ya que produce distribuciones predictivas más realistas y valida exitosamente con los datos históricos observados, capturando mejor la incertidumbre inherente al desarrollo de las pérdidas incurridas.
- Para los datos de pérdidas pagadas, es esencial modelar explícitamente los cambios dinámicos en el entorno de liquidación de siniestros. el modelo CSR, que introduce un parámetro de aceleración o desaceleración de los pagos, logra corregir el sesgo de los modelos anteriores y ofrece una estimación de la incertidumbre más confiable.
- La validación retrospectiva a gran escala se confirma como una metodología indispensable para evaluar el rendimiento predictivo de los modelos de reservas, permitiendo identificar diferencias sistemáticas y confirmar la efectividad de las mejores propuestas en un contexto real.
- El enfoque de modelo Bayesiano MCMC se establece como una alternativa práctica y superior a los métodos

tradicionales, ofreciendo una flexibilidad necesaria para incorporar características complejas del datos, como correlaciones y tendencias y produciendo estimaciones de reservas que se alinean mejor con los resultados observados.

VII. TRABAJO FUTURO

Esta investigación abre varias líneas de estudio para profundizar en el desarrollo de modelos estocásticos de reservas:

- **Extensión temporal del análisis:** Aplicar la misma metodología de validación retrospectiva a datos de períodos más recientes (por ejemplo, usando estados financieros de 2010 en adelante) para verificar si las conclusiones obtenidas se mantienen en diferentes contextos temporales y regímenes de siniestralidad.
- **Análisis de líneas de seguro especializadas:** Extender la validación a líneas con patrones de desarrollo más complejos como responsabilidad profesional médica y garantías, donde los supuestos de los modelos tradicionales podrían ser aún más limitantes.

AGRADECIMIENTOS

La culminación de este proyecto de investigación fue posible gracias al invaluable acompañamiento de destacados profesionales que orientaron su desarrollo. Deseo expresar mi más sincero reconocimiento al Doctor Fredy Vides por su liderazgo académico y guía experto como docente del seminario de investigación. Sus aportes metodológicos y su visión crítica fueron fundamentales para la estructuración y rigor de este trabajo.

Asimismo, extiendo mi profundo agradecimiento a la licenciada Dariela Jissel Cerna Fonseca por su dedicación y asesoramiento especializado. Sus valiosas observaciones y su apoyo durante todo el proceso de investigación fueron decisivos para el éxito y la calidad final de este seminario.

VIII. REPOSITORIO DE ARCHIVOS DEL PROYECTO

El repositorio de github donde estará subida la base de datos y los scripts está en [6].

REFERENCIAS

- [1] <https://www.casact.org/sites/default/files/2021-02/01-Meyers.PDF>
- [2] <https://www.casact.org/publications-research/research/research-resources/loss-reserving-data-pulled-naic-schedule-p>
- [3] Mack, Thomas. 1993. "Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates." *ASTIN Bulletin* 23(2):213–225.
- [4] Mack, Thomas. 1994. "Measuring the Variability of Chain Ladder Reserve Estimates." *Casualty Actuarial Society Forum* (Spring):101–182.
- [5] Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot. 2012. *Loss Models, From Data to Decisions*. 4th ed., Hoboken, N.J.: Wiley.
- [6] <https://github.com/fernandg/Seminario-de-Investigacion-IIIPAC-2025>.
- [7] Verrall, Richard. 2007. "Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion." *Variance* 1:53–80.
- [8] Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Hoboken, N. J.: Wiley.
- [9] Meyers, Glenn G., and Peng Shi. 2011. "The Retrospective Testing of Stochastic Loss Reserve Models." *Casualty Actuarial Society E-Forum* (Summer):1–37.