

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE GRADUADOS



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

TÍTULO DE LA TESIS
MODELOS CHAIN LADDER ESTOCÁSTICOS Y
APLICACIONES AL CÁLCULO DE RESERVAS EN
COMPAÑÍAS DE SEGUROS

Tesis para optar el grado de Magister en Estadística

AUTOR

Gisella Gabriela Mazuelos Vizcarra

ASESOR

Dr. Luis Hilmar Valdivieso Serrano

JURADO

Dr. Cristian Luis Bayes Rodríguez

Dr. Luis Hilmar Valdivieso Serrano

Dr. Oscar Millones Destefano

LIMA-PERÚ

2013

Dedicatoria

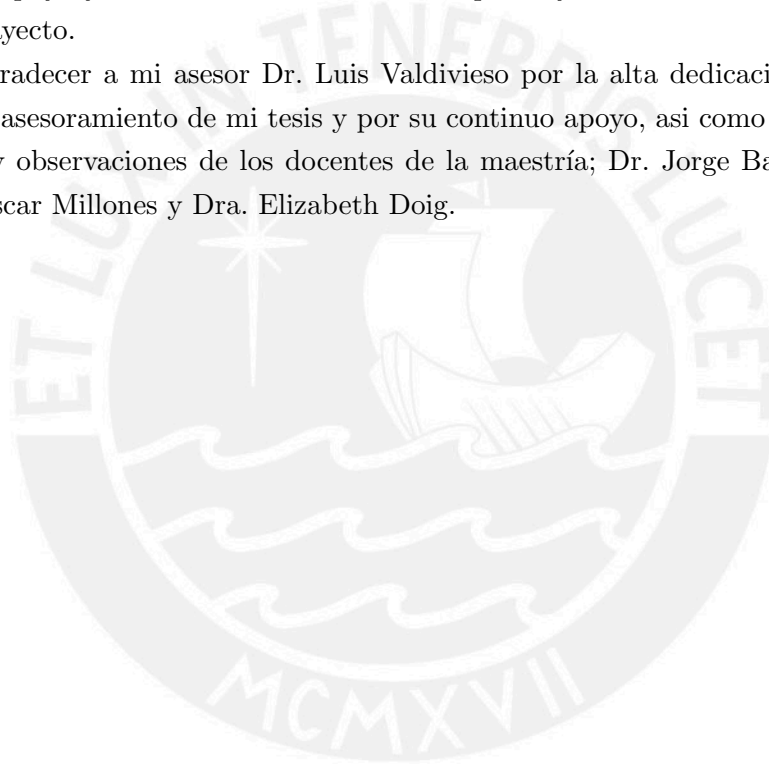
Con mucho amor, a mi hijo Sebastian, a mi esposo Eddy y a mis padres Doris y Julio, sin su apoyo incondicional, paciencia y cariño no hubiera podido lograr alcanzar esta meta.



Agradecimientos

Quiero agradecer de manera muy especial a Dios y a mis angeles que siempre estan iluminando mi camino y mi vida. A mis dos amores; mi hijo Sebastián y mi esposo Eddy, quienes fueron mi fuerza, empuje y aliento en la realización y culminación de este documento y quienes me prestaron horas de su vida para poder dedicarme a esta tesis. A mis adorados padres, que sin sus consejos, apoyo y amor incondicional de todo padre y madre hicieron que fuera menos difícil este trayecto.

Asimismo, agradecer a mi asesor Dr. Luis Valdivieso por la alta dedicación en la revisión, orientación y asesoramiento de mi tesis y por su continuo apoyo, así como a cada uno de los comentarios y observaciones de los docentes de la maestría; Dr. Jorge Bazán, Dr. Cristian Bayes, Dr. Oscar Millones y Dra. Elizabeth Doig.



Resumen

El presente documento tiene por objetivo profundizar en el estudio de los métodos univariados y multivariados del modelo Chain Ladder para la estimación de las reservas de una compañía de seguros. Se presenta de manera teórica y aplicativa tanto los métodos univariados Chain Ladder determinístico como Chain Ladder estocástico. Si bien el primero de estos métodos es el más utilizado por las compañías de seguro dada su simplicidad de cálculo y carencia de supuestos probabilísticos, el segundo, propuesto por Mack (1993), permite la construcción de intervalos de confianza para las reservas, lo cual es de invaluable ayuda para los investigadores.

Asimismo, desarrollamos el modelo General Multivariado Chain Ladder, el cual tiene como premisa básica analizar la posible relación que pueda existir entre diversos triángulos de desarrollo, dotando así de otra herramienta para mejorar las inferencias y predicciones de las reservas.

Estos métodos han sido desarrollados y aplicados sobre la base de datos de 3 tipos de seguros de salud, mostrando así las ventajas y desventajas de cada uno de ellos en diferentes escenarios y proporcionando distintos instrumentos para la toma de decisiones relacionados en el cumplimiento de las obligaciones futuras de las compañías de seguros.

Palabras-clave: Método Chain Ladder determinístico, Método Chain Ladder estocástico, Regresiones Aparentemente no Relacionadas, Modelo General Multivariado Chain Ladder.

Abstract

This document is intended to deepen the study of univariate and multivariate Chain Ladder methods for estimating reserves in an insurance company. It presents from a theoretical and applicative perspective both the univariate deterministic and stochastic Chain Ladder methods. Although, the first is the most used method by insurance companies due to its simplicity and lack of probabilistic assumptions, the second, proposed by [Mack \(1993\)](#), allows the construction of confidence intervals for the estimated reserves, which is invaluable for researchers.

We also develop the General Multivariate Chain Ladder model, which has the basic premise to analyze the possible relationship that may exist between different development triangles, thus providing another tool to improve inferences and predictions of reserves.

These methods have been developed and applied to a database of 3 types of health insurance, thus showing the advantages and disadvantages of each of them in different scenarios and providing various tools for decision making in meeting the future obligations of insurance companies.

Keywords: Deterministic Chain Ladder Method, Stochastic Chain Ladder Method, Seemingly Unrelated Regressions, General Multivariate Chain Ladder.

Índice general

Lista de Abreviaturas	VIII
Índice de figuras	IX
Índice de cuadros	X
1. Introducción	1
1.1. Consideraciones Preliminares	1
1.1.1. Introducción	1
1.2. Objetivos	2
1.3. Organización del Trabajo	2
2. Modelos Chain Ladder univariados	3
2.1. El método Chain Ladder determinístico	3
2.2. El método Chain Ladder estocástico de Mack	8
2.2.1. Formulación del Modelo	8
2.2.2. El estimador Chain Ladder para los factores de desarrollo	9
2.2.3. Predicciones en el Modelo de Mack	15
3. Regresiones Aparentemente no Relacionadas (SUR)	19
3.1. Estimación de parámetros	21
3.2. El estimador iterativo de Mínimos Cuadrados Factible Generalizado (IFGLS)	24
4. El modelo General Multivariado Chain Ladder	25
4.1. Formulación del Modelo	26
4.2. El modelo GMCL con intercepto	30
5. Aplicación al Cálculo de Reservas en Compañías de Seguros	32
5.1. Aspectos generales	32
5.1.1. Descripción de las bases de datos	32
5.2. El Método Chain Ladder determinístico	34
5.3. El Método Chain Ladder Estocástico	40
5.4. El Modelo General Multivariado Chain Ladder	44
5.5. Comparación de los modelos multivariados con el Método Chain Ladder univariado	49

6. Conclusiones	52
6.1. Conclusiones	52
6.2. Sugerencias para investigaciones futuras	53
A. Programas en R	54
A.1. Código utilizado para el Método Chain Ladder Estocástico	54
A.2. Código utilizado para el Modelo General Multivariado Chain Ladder	55
B. Residuales de los distintos modelos	57
Bibliografía	60



Lista de Abreviaturas

IBNR	Siniestros incurridos pero no reportados(<i>Incurred but not reported</i>).
IBNER	Siniestros Incurridos pero aún no terminados de reportar (<i>Incurred but not ended reported</i>).
Run Off	Triángulo de desarrollo.
LFGN	Ley fuerte de los grandes números.
SUR	Regresiones Aparentemente no Relacionadas (<i>Seemingly Unrelated Regressions</i>).
MCO	Mínimos Cuadrados Ordinarios.
MELI	Mejor Estimador Lineal Insesgado.
GLS	Mínimos Cuadrados Generalizados (<i>Generalized Least Squares</i>).
FGLS	Mínimo Cuadrado Factible Generalizado (<i>Feasible Generalized Least Squares</i>).
IFGLS	Mínimo Cuadrado Iterativo Factible Generalizado (<i>Iterated Feasible Generalized Least Squares</i>).
MCL	Modelo Multivariado Chain Ladder (<i>Multivariate Chain Ladder</i>).
GMCL	Modelo General Multivariado Chain Ladder (<i>General Multivariate Chain Ladder</i>).

Índice de figuras

2.1. Esquema de montos de algunos reclamos entre 6 periodos distintos. En este ejemplo los reportes pueden hacerse sólo hasta finalizarse el año 2008 o periodo 6	4
2.2. Triángulo de desarrollo	4
2.3. Triángulo de desarrollo acumulado	5
5.1. Distribución del monto de reclamos por tipo de seguro de salud.	34
5.2. Reservas futuras y análisis de residuales para el seguro de tipo A	42
5.3. Reservas futuras y análisis de residuales para el seguro de tipo B.	43
5.4. Reservas futuras y análisis de residuales para el seguro de tipo C.	44
5.5. Boxplot de residuales para el seguro de tipo A, bajo los distintos modelos en análisis.	49
5.6. Boxplot de residuales para el seguro de tipo B, bajo los distintos modelos en análisis.	50
5.7. Boxplot de residuales para el seguro de tipo C, bajo los distintos modelos en análisis.	50
B.1. Gráfico de los residuales del modelo GMCL con restricciones y con intercepto, por tipo de seguro	57
B.2. Gráfico de los residuales del modelo GMCL con restricciones y sin intercepto, por tipo de seguro	58
B.3. Gráfico de los residuales del modelo GMCL sin restricciones y con intercepto, por tipo de seguro	58
B.4. Gráfico de los residuales del modelo GMCL sin restricciones y sin intercepto, por tipo de seguro	59

Índice de cuadros

2.1. Triángulo de desarrollo	6
2.2. Triángulo de desarrollo de severidades acumuladas	6
2.3. Totales por columna	6
2.4. Factores de desarrollo	7
2.5. Triángulo de desarrollo completo	7
2.6. Estimación de reservas	7
5.1. Distribución de reclamos - Data seguro de salud de tipo A.	33
5.2. Distribución de reclamos - Data seguro de salud de tipo B.	33
5.3. Distribución de reclamos - Data seguro de salud de tipo C.	34
5.4. Triángulo de desarrollo (seguro de tipo A).	35
5.5. Triángulo de desarrollo de severidades acumuladas (seguro de tipo A).	35
5.6. Factores de desarrollo (seguro de tipo A).	35
5.7. Triángulo de desarrollo completo (seguro de tipo A).	36
5.8. Estimación de reservas (seguro de tipo A).	36
5.9. Triángulo de desarrollo (seguro de tipo B).	37
5.10. Triángulo de desarrollo de severidades acumuladas (seguro de tipo B).	37
5.11. Factores de desarrollo (seguro de tipo B).	37
5.12. Triángulo de desarrollo completo (seguro de tipo B).	38
5.13. Estimación de reservas (seguro de tipo B).	38
5.14. Triángulo de desarrollo (seguro de tipo C)	38
5.15. Triángulo de desarrollo de severidades acumuladas (seguro de tipo C).	39
5.16. Factores de desarrollo (seguro de tipo C).	39
5.17. Triángulo de desarrollo completo (seguro de tipo C).	39
5.18. Estimación de reservas (seguro de tipo C).	40
5.19. Factores de desarrollo (seguro de tipo A).	40
5.20. Triángulo de desarrollo completo (seguro de tipo A).	41
5.21. Reservas estimadas e intervalos de confianza (seguro de tipo A).	41
5.22. Reservas estimadas e intervalos de confianza (seguro de tipo B).	42
5.23. Reservas estimadas e intervalos de confianza (seguro de tipo C).	43
5.24. Parámetros estimados del modelo GMCL.	46
5.25. Triángulo de desarrollo completo (seguro de tipo A).	46
5.26. Triángulo de desarrollo completo (seguro de tipo B).	47
5.27. Triángulo de desarrollo completo (seguro de tipo C).	47
5.28. Reservas estimadas (seguro de tipo A).	47

5.29. Reservas estimadas (seguro de tipo B).	48
5.30. Reservas estimadas (seguro de tipo C).	48
5.31. Reservas estimadas del portafolio (TIPO A + TIPO B + TIPO C).	48



Capítulo 1

Introducción

1.1. Consideraciones Preliminares

1.1.1. Introducción

Uno de los objetivos centrales de toda compañía de seguros recae en la estimación correcta de sus reservas futuras. Para cada aseguradora es fundamental conocer sus pasivos y cómo estos pueden cubrirlos para evitar la insolvencia. Con el fin de lograr este propósito, se han desarrollado diversos métodos de estimación, en gran parte determinísticos, que permiten prever una reserva futura. Uno de éstos métodos es el llamado “Método Chain Ladder”.

El Método Chain Ladder es probablemente el método más utilizado en las compañías para la estimación de reservas futuras. Las razones de su popularidad recaen principalmente en que es fácil de usar, no requiere de un software especializado para implementarlo y carece de supuestos probabilísticos, con la excepción que la proporción de siniestros o reclamos que son reportados a la compañía de un periodo a otro (los cuales pueden medirse en semanas, meses, años, etc) se mantenga constante. Dado que en la práctica esta última suposición es bastante rebatible por la existencia de factores internos o externos que afectan tal comportamiento, se hace necesario el estudiar e introducir métodos con una base probabilística más formal que no solo nos permitan una estimación puntual de reservas más realista, sino también de su variabilidad e intervalos de confianza.

Es por ello, que a partir del modelo determinístico Chain Ladder, [Mack \(1993\)](#) introdujo una versión estocástica del mismo que permitió la estimación de la variabilidad de las reservas mediante el error cuadrático medio como una medida de incertidumbre contenida en la data de reclamos o siniestros. Es propósito de este trabajo el estudiar e investigar este modelo, así como también su más reciente extensión multivariada propuesta por [Zhang \(2010\)](#). El Modelo General Multivariado Chain Ladder o GMCL de sus siglas en inglés General Multivariate Chain Ladder, no sólo especifica correlaciones contemporáneas, sino que también permite conexiones estructurales entre los triángulos de desarrollo. Lo resaltante del GMCL es el empleo de la técnica de Regresión Aparentemente no Relacionada para la estimación de parámetros, el cual detallaremos más adelante. El uso de esta técnica es esencial para construir modelos flexibles y relacionados con las teorías estadísticas, las cuales se aplican para estudiar las propiedades de los estimadores existentes. Estimar reservas con múltiples triángulos es importante en tanto que entre los triángulos pueden existir conexiones estructurales; es decir, el desarrollo de un triángulo puede depender de información pasada de otros triángulos, o también debido a que el desarrollo conjunto tiene en cuenta las correlaciones

contemporáneas entre los triángulos podría dar lugar a estimaciones más eficientes.

1.2. Objetivos

El objetivo general de esta tesis es profundizar en el estudio de los modelos Chain Ladder estocásticos univariados y multivariados para la estimación de las reservas de una compañía de seguros.

Como objetivos específicos tenemos:

1. Revisar la literatura relacionada a los distintos modelos Chain Ladder.
2. Presentar una revisión de los modelos Chain Ladder determinístico y estocástico.
3. Introducir el modelo General Multivariado Chain Ladder.
4. Aplicar los modelos a una base de datos reales de una compañía de seguros.
5. Realizar un estudio comparativo entre las estimaciones de reservas obtenidas bajo los distintos modelos planteados.

1.3. Organización del Trabajo

En el Capítulo 1 presentamos los aspectos generales de esta investigación y sus objetivos.

En el Capítulo 2 se hace una presentación del método Chain Ladder determinístico y del método Chain Ladder estocástico propuesto por Mack (1993). Describiremos aquí no solo la metodología a emplearse en la previsión de reservas futuras, sino también el problema de estimación de los parámetros del modelo.

En el Capítulo 3 se explica la técnica econométrica SUR o de Regresiones Aparentemente no Relacionadas y sus métodos de estimación de parámetros, lo cual nos ayudará a entender más en profundidad el siguiente capítulo.

En el Capítulo 4 introduciremos el modelo General Multivariado Chain Ladder propuesto por Zhang (2010). Estudiaremos aquí el problema de la estimación de sus parámetros y su relación con otros modelos multivariados asociados.

En el Capítulo 5 mostraremos la aplicación de los modelos planteados a una data real proveniente de una compañía de seguros del país. A través de esta aplicación realizaremos también un estudio comparativo de los distintos modelos.

Además, en el Capítulo 6 discutimos algunas conclusiones obtenidas en este trabajo, analizamos la ventajas y desventajas de los métodos propuestos y sugeriremos futuras investigaciones.

Finalmente, en los apéndices A y B presentamos los códigos del programa utilizado en las aplicaciones, así como gráficos residuales.

Capítulo 2

Modelos Chain Ladder univariados

2.1. El método Chain Ladder determinístico

Toda compañía de seguros tiene la obligación de mantener reservas técnicas para reclamos o siniestros que no han sido pagados o que están pendientes de pago pero que han ocurrido en periodos anteriores, éstos son los referidos siniestros IBNR (incurridos pero no reportados) e IBNER (incurridos pero no terminados de reportar), respectivamente. Según [Ordoñez \(2010\)](#) esta demora en el pago del reclamo es debido a que éstos normalmente se reportan a terceros que facilitan o contactan con la compañía aseguradora, por lo cual hay que esperar un periodo de tiempo hasta que el reclamo llegue a la misma aseguradora y a su base de datos. Así se traza una fecha de corte para estimar las reservas, son estos reclamos los que no son tomados en cuenta a pesar que ya ocurrieron dentro de la fecha de corte.

El método Chain Ladder es catalogado como el más popular para el establecimiento de reservas técnicas en una compañía de seguros. Este método se basa en la suposición de que la proporción de las cantidades o montos de reclamos es constante independientemente del periodo de ocurrencia u origen i , pero dependiente del periodo de desarrollo j , suponiendo así un proceso constante en el tiempo [Ordoñez \(2010\)](#). Una importante deficiencia de este método es que al no tener una base estadística, no es posible calcular intervalos de confianza para las reservas resultantes, aspecto que es esencial para el manejo de los riesgos derivados.

Sea cual sea el método que se utilice para la estimación de reservas, los datos generalmente se presentan en la forma de un triángulo de desarrollo (Run Off), como el presentado en la figura 2.2.

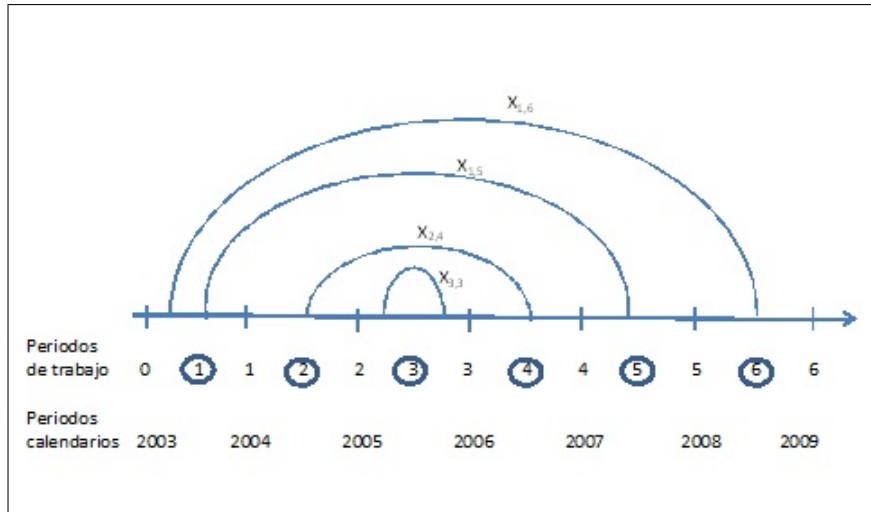


Figura 2.1: Esquema de montos de algunos reclamos entre 6 periodos distintos. En este ejemplo los reportes pueden hacerse sólo hasta finalizarse el año 2008 o periodo 6

Periodo de Ocurrencia		Periodo de desarrollo						
		1	2	...	j	n-1 n
1		$X_{1,1}$	$X_{1,2}$...	$X_{1,j}$	$X_{1,n-1} X_{1,n}$
2		$X_{2,1}$	$X_{2,2}$...	$X_{2,j}$	$X_{2,n-1}$
⋮		⋮	⋮	⋮	⋮			
i		$X_{i,1}$	$X_{i,2}$...	$X_{i,j}$...	$X_{i,n-i+1}$	
⋮		⋮	⋮					
n-1		$X_{n-1,1}$	$X_{n-1,2}$					
n		$X_{n,1}$						

Figura 2.2: Triángulo de desarrollo

En las filas se registran los períodos de origen u ocurrencia del reclamo o siniestro, en las columnas los periodos de desarrollo del reclamo o siniestro y $X_{i,j}$ denota a las severidades históricas (montos o cantidades) de los reclamos que ocurrieron en el periodo i y que fueron reportados en el periodo $i + j - 1$; es decir, después de haber transcurrido $j - 1$ periodos (véase la figura 2.1). El objetivo de los métodos de estimación de reservas es esencialmente completar la parte inferior derecha de este triángulo.

El método Chain Ladder permite la estimación de las reservas futuras mediante la introducción de las severidades acumuladas de los siniestros que ocurrieron en el periodo i y que fueron reportados hasta antes de finalizar el periodo $i + j - 1$. Estas se denotan por:

$$Y_{i,j} = \sum_{k=1}^j X_{i,k} ,$$

donde $i = 1, \dots, n$ y $j = 1, \dots, n + 1 - i$; y pueden resumirse también en un triángulo de desarrollo de severidades acumuladas. Tales severidades como sus contrapartes no acumuladas se muestran en las figuras 2.2 y 2.3.

CAPÍTULO 2. MODELOS CHAIN LADDER UNIVARIADOS

Periodo de Ocurrencia	Periodo de desarrollo								
	1	2	...	j	j+1	n-1	n
1	$Y_{1,1}$	$Y_{1,2}$...	$Y_{1,j}$	$Y_{1,j+1}$	$Y_{1,n-1}$	$Y_{1,n}$
2	$Y_{2,1}$	$Y_{2,2}$...	$Y_{2,j}$	$Y_{2,j+1}$	$Y_{2,n-1}$	
...				
i	$Y_{i,1}$	$Y_{i,2}$...	$Y_{i,j}$	$Y_{i,j+1}$...	$Y_{i,n-i+1}$		
...					
n-j	$Y_{n-j,1}$	$Y_{n-j,2}$...	$Y_{n-j,j}$	$Y_{n-j,j+1}$				
...							
n-1	$Y_{n-1,1}$	$Y_{n-1,2}$							
n	$Y_{n,1}$								

Figura 2.3: Triángulo de desarrollo acumulado

En base a este último triángulo de desarrollo se definen los factores de desarrollo:

$$\hat{f}_j = \frac{\sum_{i=1}^{n-j} Y_{i,j+1}}{\sum_{i=1}^{n-j} Y_{i,j}},$$

donde $j = 1, 2, \dots, n-1$. Nótese que \hat{f}_j denota al cociente entre la suma (sobre los periodos de ocurrencia comunes) de todas las severidades acumuladas, que son reportadas luego de j periodos, y éstas mismas cantidades luego de $j-1$ periodos.

El método Chain Ladder permite, con los factores de desarrollo anteriores y las severidades acumuladas, estimar las reservas futuras correspondientes mediante:

$$\hat{Y}_{i,j+1} = \hat{f}_j \hat{Y}_{i,j},$$

donde $i = 2, 3, \dots, n$, $1 \leq j \leq n-1$ y algunas de las severidades acumuladas en el lado derecho de esta ecuación se pueden obtener directamente del triángulo de desarrollo.

Como ilustración del método Chain Ladder consideremos el ejemplo siguiente (Ordoñez, 2010), en donde se utiliza como base de datos los siniestros reportados entre los años 2003 y 2008 del seguro SOAT de una compañía de seguros peruana. Como primera medida se construye el triángulo de desarrollo entre los años 2003 al 2008, los cuales identificaremos con los números del 1 al 6, respectivamente:

CAPÍTULO 2. MODELOS CHAIN LADDER UNIVARIADOS

Año de Ocurrencia	Año de desarrollo					
	1	2	3	4	5	6
1 (2003)	32,633.4	5,096.9	1,095.2	1,212.5	421.9	72.2
2 (2004)	40,595.5	7,990.6	1,133.0	454.3	136.8	
3 (2005)	56,528.5	10,066.7	1,814.2	398.0		
4 (2006)	63,915.7	16,300.5	1,039.6			
5 (2007)	77,575.5	12,363.9				
6 (2008)	77,233.6					

Cuadro 2.1: Triángulo de desarrollo

Luego, se construye el triángulo de severidades acumuladas ($Y_{i,j}$).

Año de Ocurrencia	Año de desarrollo					
	1	2	3	4	5	6
1 (2003)	32,633.4	37,730.3	38,825.5	40,038.0	40,459.9	40,532.2
2 (2004)	40,595.5	48,586.1	49,719.0	50,173.3	50,310.1	
3 (2005)	56,528.5	66,595.2	68,409.4	68,807.4		
4 (2006)	63,915.7	80,216.1	81,255.7			
5 (2007)	77,575.5	89,939.4				
6 (2008)	77,233.6					

Cuadro 2.2: Triángulo de desarrollo de severidades acumuladas

El siguiente paso en el método Chain Ladder consiste en obtener los totales por columna, como se ilustra en la tabla siguiente:

Suma de severidades acumuladas hasta el periodo:	Año de desarrollo (j)					
	1	2	3	4	5	6
7-j (numerador)		323,067.1	238,209.6	159,018.7	90,770.0	40,532.2
6-j (denominador)	271,248.6	233,127.7	156,953.9	90,211.3	40,459.9	

Cuadro 2.3: Totales por columna

Así, se obtienen los factores de desarrollo siguientes:

CAPÍTULO 2. MODELOS CHAIN LADDER UNIVARIADOS

Año de desarrollo	1	2	3	4	5
Fact.Des.	1.191036	1.021798	1.013155	1.006193	1.001786

Cuadro 2.4: Factores de desarrollo

Una vez obtenidos los factores de desarrollo, y a partir del triángulo de desarrollo acumulado, se puede completar ahora las celdas con los montos estimados en negrita.

Año Ocurrencia	Año de desarrollo					
	1	2	3	4	5	6
1 (2003)	32,633.4	37,730.3	38,825.5	40,038.0	40,459.9	40,532.2
2 (2004)	40,595.5	48,586.1	49,719.0	50,173.3	50,310.1	50,399.9
3 (2005)	56,528.5	66,595.2	68,409.4	68,807.4	69,233.5	69,357.2
4 (2006)	63,915.7	80,216.1	81,255.7	82,324.7	82,834.5	82,982.5
5 (2007)	77,575.5	89,939.4	91,900.0	93,109.0	93,685.6	93,852.9
6 (2008)	77,233.6	91,988.0	93,993.3	95,229.8	95,819.6	95,990.7

Cuadro 2.5: Triángulo de desarrollo completo

Por último, las provisiones de este producto se calculan de la diferencia entre los montos acumulados estimados y su anterior.

Año de ocurrencia	Año de desarrollo						Reservas
	1	2	3	4	5	6	
1 (2003)	32,633.4	37,730.3	38,825.5	40,038.0	40,459.9	40,532.2	-
2 (2004)	40,595.5	48,586.1	49,719.0	50,173.3	50,310.1	89.8	89.8
3 (2005)	56,528.5	66,595.2	68,409.4	68,807.4	426.2	123.6	549.8
4 (2006)	63,915.7	80,216.1	81,255.7	1,068.9	509.9	147.9	1,726.7
5 (2007)	77,575.5	89,939.4	1,960.6	1,209.0	576.7	167.3	3,913.5
6 (2008)	7,233.6	14,754.5	2,005.3	1,236.5	589.8	171.1	18,757.1
Reserva estimada							25,037.0

Cuadro 2.6: Estimación de reservas

Por consiguiente, de acuerdo al ejemplo mostrado, se puede decir que al término del año 2008 la provisión final o monto que debería de reservarse será de 25,037 dólares hasta el año 2013. En particular el monto de reserva para el año 2009 será de 18,300 dólares.

2.2. El método Chain Ladder estocástico de Mack

Mack (1993) introduce una versión estocástica del modelo determinístico Chain Ladder, con la finalidad de poder estimar, además de las reservas futuras, las variaciones de las mismas a través del error cuadrático medio, y así obtener intervalos de confianza para dichas reservas.

Para su formulación, será conveniente introducir los siguientes periodos de tiempo.

$$C_i = (i - 1, i], \quad i = 1, 2, \dots, n$$

y sus periodos acumulados

$$C_{i,j} = C_i \cup \dots \cup C_{i+j-1} = (i - 1, i + j - 1],$$

donde $j = 1, 2, \dots, n + 1 - i$.

En este contexto, $Y_{i,j}$ denotará a la severidad de los siniestros que ocurrieron en el periodo C_i y que se reportaron o pagaron en el periodo $C_{i,j}$.

Si interpretamos a C_n como el último periodo con información completa (en nuestro ejemplo C_6), el triángulo de desarrollo de la figura 2.2 contiene la información completa agregada sobre el total de las severidades $Y_{i,j}$ en los periodos previos y último con información, es decir, de las severidades para reclamos que ocurrieron en el periodo C_i para algún $i \leq n$, y cuyos pagos fueron ejecutados en el periodo $C_{i,j}$ para $j = 1, 2, \dots, n + 1 - i$.

2.2.1. Formulación del Modelo

Mikosch (2009) realiza una formulación del modelo Chain Ladder estocástico de Mack, que es la que adoptaremos y adaptaremos en nuestro trabajo.

Un problema práctico es el siguiente:

¿Cómo se puede utilizar la información representada por observaciones pasadas y presentes $Y_{i,j}$, para poder predecir las severidades acumuladas futuras totales de los siniestros $Y_{i,n+2-i}$?

Con esta pregunta en mente, Mack (1993) introdujo algunas condiciones en el modelo Chain Ladder que detallaremos seguidamente:

CL1: Las severidades (número y/o monto total de pagos por siniestros) para reclamos con diferentes periodos de ocurrencia son independientes e idénticamente distribuidos; es decir, $\forall i \neq l, Y_{i,j}$ e $Y_{l,k}$ son variables aleatorias independientes e idénticamente distribuidas cualquiera sea el valor de j y k .

CL2: Las variables aleatorias $Y_{i,j}$ para $i = 1, 2, \dots, n$ y $j = 1, \dots, n + 1 - i$, son positivas casi en todas partes (c.p.) y tienen esperanza finita.

CL3: Existen números reales no negativos f_j , los cuales denominamos factores de desarrollo, tales que para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, n + 1 - i$:

$$E[Y_{i,j+1} | Y_{i,1}, \dots, Y_{i,j}] = f_j Y_{i,j}. \quad (2.1)$$

CAPÍTULO 2. MODELOS CHAIN LADDER UNIVARIADOS

Con el propósito de simplificar la notación escribiremos así mientras no exista confusión:

$$Y_j = Y_{i,j}$$

Esto se justifica por la condición **CL1**.

Así la ecuación (2.1) se podría leer como:

$$E[Y_{j+1}|Y_1, \dots, Y_j] = f_j Y_j. \quad (2.2)$$

El requisito de finitud en la condición **CL2** es necesario para asegurar que el esperado condicional en la ecuación (2.2) sea finito con probabilidad 1.

Regresemos ahora a la pregunta sobre la predicción de las severidades futuras dadas las observaciones pasadas y últimas con información. En este contexto es necesario, recordar que la esperanza condicional:

$$E[Y_{j+1}|\mathcal{F}_j] = E[Y_{j+1}|Y_1, \dots, Y_j].$$

es la v.a. \mathcal{F}_j -medible M que minimiza el error cuadrático medio $E[(Y_{j+1} - M)^2]$ entre la clase de todas las variables aleatorias con varianza finita y que son medibles con respecto a la σ -álgebra:

$$\mathcal{F}_j = \sigma \langle Y_1, \dots, Y_j \rangle$$

generada por Y_1, \dots, Y_j . Por lo tanto, si $Var(Y_j) < \infty$ para $j = 1, 2, \dots, n$, la ecuación (2.1) define a la v.a. $f_j Y_{i,j}$ como el mejor predictor de $Y_{i,j+1}$ conocidas las severidades $Y_{i,1}, \dots, Y_{i,j}$ que se reportan en los periodos pasados y presentes $C_{i,1}, \dots, C_{i,j}$.

La principal motivación de Mack (1993) para introducir la ecuación (2.2) fue explicar la forma de los llamados estimadores “Chain Ladder” los cuales son comunmente usados en la práctica para la estimación de reservas futuras. En este sentido, el modelo de Mack es dictado por un procedimiento estadístico aplicado.

En general, es imposible calcular las esperanzas condicionales del tipo $E[Y_{j+1}|Y_1, \dots, Y_j]$ sin supuestos específicos para modelos estocásticos subyacentes. La condición **CL3** del modelo de Mack define estas esperanzas condicionales en una forma manejable.

2.2.2. El estimador Chain Ladder para los factores de desarrollo

La condición **CL1** implica que las severidades acumuladas derivadas de distintos periodos C_i tienen la misma distribución y son mutuamente independientes. Esta propiedad abre la puerta a la teoría asintótica estándar cuando $n \rightarrow \infty$. Por el supuesto **CL1** $(Y_{i,j})_{i \geq 1}$ es una secuencia de variables aleatorias independientes e idénticamente distribuidas para cada valor fijo de $j \geq 1$ y por tanto la ley fuerte de los Grandes Números (LFGN) garantiza la consistencia fuerte de las siguientes versiones muestrales de $E[Y_j]$, $Var(Y_j)$ y $Cov(Y_j, Y_k)$:

$$\bar{Y}_n^{(j)} = \frac{1}{n-j+1} \sum_{i=1}^{n-j+1} Y_{i,j} \quad (2.3)$$

$$(S_n^{(j)})^2 = \frac{1}{n-j} \sum_{i=1}^{n-j+1} (Y_{i,j} - \bar{Y}_n^{(j)})^2 \quad (2.4)$$

y

$$\hat{\gamma}_n(j, k) = \frac{1}{n-k+1} \sum_{i=1}^{n-k+1} (Y_{i,j} - \bar{Y}_n^{(j)})(Y_{i,k} - \bar{Y}_n^{(k)}), \quad (2.5)$$

respectivamente; es decir,

$$\bar{Y}_n^{(j)} \xrightarrow{c.p.} E[Y_j],$$

$$[S_n^{(j)}]^2 \xrightarrow{c.p.} \text{Var}(Y_j)$$

y

$$\hat{\gamma}_n(j_1, j_2) \xrightarrow{c.p.} \text{Cov}(Y_{j_1}, Y_{j_2})$$

donde $j < k$. Aquí se asume que $E[Y_j] < \infty$ y $\text{Var}(Y_j) < \infty$.

En virtud de la ecuación (2.2), se observa que para $j = 1, \dots, n$

$$E[Y_{j+1}] = E[E[Y_{j+1}|Y_1, \dots, Y_j]] = f_j E[Y_j],$$

donde se asume por **CL2** que Y_j es positivo para cada j . Por tanto

$$f_j = \frac{E[Y_{j+1}]}{E[Y_j]}, \quad j = 1, 2, \dots, n-1. \quad (2.6)$$

Esta relación y la LFGN para $\bar{Y}_n^{(j)}$ sugieren que el esperado de la ecuación (2.6) debería ser remplazada por su correspondiente estimador muestral. En efecto, el estimador de momentos propuesto para f_j viene dado por:

$$\hat{f}_j^{(n)} = \frac{\bar{Y}_n^{(j+1)}}{\bar{Y}_n^{(j)}} = \frac{\sum_{i=1}^{n-j} Y_{i,j+1}}{\sum_{i=1}^{n-j} Y_{i,j}} \quad (2.7)$$

Este es llamado el *estimador Chain Ladder* de f_j . La estructura del ratio de $\hat{f}_j^{(n)}$ justifica el supuesto **CL2** en el modelo de Mack que las variables aleatorias $Y_{i,j}$ sean positivas. Si $P(Y_j = 0) > 0$, existe una probabilidad positiva que el denominador $\sum_{i=1}^{n-j} Y_{i,j}$ desaparezca. En efecto, la condición **CL1** garantiza que:

$$\begin{aligned} P(\sum_{i=1}^{n-j} Y_{i,j} = 0) &= P(Y_{i,j} = 0, i = 1, \dots, n-j) \\ &= (P(Y_j = 0))^{n-j} > 0 \end{aligned}$$

A continuación se resume algunas propiedades asintóticas de los estimadores Chain Ladder.

Proposición 2.2.1 *Asumiendo las condiciones del modelo de Mack, se cumple que:*

1. El estimador Chain Ladder $\hat{f}_j^{(n)}$ es un estimador fuertemente consistente de f_j ; es decir:

$$\hat{f}_j^{(n)} \xrightarrow{c.p.} f_j \text{ conforme } n \rightarrow \infty$$

2. Si $\text{Var}(Y_j) < \infty$, entonces $\hat{f}_j^{(n)}$ satisface la siguiente versión del Teorema de Límite Central:

$$\sqrt{n}(\hat{f}_j^{(n)} - f_j) \xrightarrow{\mathcal{D}} N(0, \frac{E[Y_{j+1}^2] - f_j^2 E[Y_j^2]}{E[Y_j]^2})$$

Demostración: Dado que $\bar{Y}_n^{(j+1)}$ es por la LFGN un estimador fuertemente consistente de $E[Y_{j+1}]$ e igualmente $\bar{Y}_{n-1}^{(j)}$ de $E[Y_j]$, donde $P(\bar{Y}_{n-1}^{(j)} = 0) > 0$, se sigue de (2.6) que $\hat{f}_j^{(n)}$ es un estimador fuertemente consistente de f_j . Más precisamente que:

$$\hat{f}_j^{(n)} = \frac{\bar{Y}_n^{(j+1)}}{\bar{Y}_{n-1}^{(j)}} \xrightarrow{c.p.} \frac{E[Y_{j+1}]}{E[Y_j]} = f_j.$$

Por otro lado,

$$\begin{aligned} \sqrt{n}[\hat{f}_j^{(n)} - f_j] &= \sqrt{n}[\frac{\bar{Y}_n^{(j+1)}}{\bar{Y}_{n-1}^{(j)}} - f_j] = \frac{\sqrt{n}}{\bar{Y}_{n-1}^{(j)}}[\bar{Y}_n^{(j+1)} - f_j \bar{Y}_{n-1}^{(j)}] \\ &= \frac{\sqrt{n/(n-j)}}{\bar{Y}_{n-1}^{(j)}}[\frac{1}{\sqrt{n-j}} \sum_{i=1}^{n-j} (Y_{i,j+1} - f_j Y_{i,j})] \end{aligned}$$

Puesto que por **CL3** la secuencia $(Y_{i,j+1} - f_j Y_{i,j})_{i \geq 1}$ es i.i.d con media común cero y varianza finita, una aplicación del Teorema del Límite Central en combinación con la Ley Fuerte de los Grandes Números $\bar{Y}_{n-1}^{(j)} \xrightarrow{c.p.} E[Y_j]$ implica, por el teorema de Slutsky, que

$$\sqrt{n}[\hat{f}_j^{(n)} - f_j] \xrightarrow{\mathcal{D}} N(0, \frac{\text{Var}(Y_{j+1} - f_j Y_j)}{E[Y_j]^2}) \quad (2.8)$$

Más aún, como:

$$\begin{aligned} \text{Cov}(Y_j, Y_{j+1}) &= E[Y_{j+1}Y_j] - E[Y_{j+1}]E[Y_j] \\ &= E[Y_j E[Y_{j+1}|Y_j]] - E[E[Y_{j+1}|Y_j]]E[Y_j] \\ &= f_j \text{Var}(Y_j), \end{aligned}$$

$$\text{Var}(Y_{j+1}) = E[Y_{j+1}^2] - f_j^2 E[Y_j]^2$$

y

$$\text{Var}(Y_{j+1} - f_j Y_j) = \text{Var}(Y_{j+1}) + f_j^2 \text{Var}(Y_j) - 2f_j \text{Cov}(Y_j, Y_{j+1}),$$

la varianza asintótica en la ecuación (2.8) viene dada por:

$$\frac{\text{Var}(Y_{j+1} - f_j Y_j)}{E[Y_j]^2} = \frac{E[Y_{j+1}^2] - f_j^2 E[Y_j^2]}{E[Y_j]^2}$$

□

La varianza asintótica de $\hat{f}_j^{(n)}$ en el Teorema de Límite Central, sugiere utilizar una aproximación por el método de momentos para estimar $E[Y_j^2]$ y $E[Y_{j+1}^2]$. Esto significa que la varianza asintótica $\frac{E[Y_{j+1}^2] - f_j^2 E[Y_j^2]}{E[Y_j]^2}$ podría aproximarse por el siguiente estimador fuertemente consistente:

$$\frac{\sum_{i=1}^{n-j} (Y_{i,j+1}^2 - (\hat{f}_j^{(n)})^2 Y_{i,j}^2)}{(n-j+1)^{-1} (\sum_{i=1}^{n-j+1} Y_{i,j})^2}.$$

Este estimador de la varianza asintótica permite construir intervalos de confianza asintóticos para f_j .

Vale destacar que el estimador $\hat{f}_j^{(n)}$ no es exactamente la versión muestral de $\frac{E[Y_{j+1}]}{E[Y_j]}$, la cual viene dada por:

$$\frac{\bar{Y}_n^{(j+1)}}{\bar{Y}_n^{(j)}} = \frac{n-j+1}{n-j} \frac{\sum_{i=1}^{n-j} Y_{i,j+1}}{\sum_{i=1}^{n-j+1} Y_{i,j}}.$$

Una de las razones para la elección de $\hat{f}_j^{(n)}$, es el hecho que este es un estimador insesgado de su contraparte determinística f_j . Esta propiedad, así como el hecho de que estos factores de desarrollo estimados son no correlacionados de contarse con una varianza finita se plantean en la siguiente proposición. Antes introduciremos para $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, n+1-i$ las σ -álgebras:

$$\mathcal{G}_j^{(n)} = \sigma \langle Y_{i,k} | 1 \leq i \leq n-j, 1 \leq k \leq j \rangle$$

y

$$\mathcal{F}_{i,j} = \sigma \langle Y_{i,1}, \dots, Y_{i,j} \rangle.$$

Proposición 2.2.2 *Asumiendo las condiciones del modelo de Mack, se cumple que:*

1. *Los estimadores Chain Ladder son insesgados:*

$$E[\hat{f}_j^{(n)}] = f_j,$$

2. *Si $\text{Var}(Y_j) < \infty$ para $j \geq 1$, entonces los estimadores Chain Ladder son no correlacionados:*

$$\text{Cov}(\hat{f}_j^{(n)}, \hat{f}_k^{(n)}) = 0, \quad \forall j \neq k \in 1, 2, \dots, n-1.$$

Demostración: Dado que $\sum_{i=1}^{n-j} Y_{i,j}$ es medible con respecto a $\mathcal{G}_j^{(n)}$,

$$E[\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}] = \frac{\sum_{i=1}^{n-j} E[Y_{i,j+1} | \mathcal{G}_j^{(n)}]}{\sum_{i=1}^{n-j} Y_{i,j}}.$$

De otro lado, por la condición **CL1**, la variable aleatoria $Y_{i,j+1}$ es independiente de $(Y_{l,k})_{k \geq 1}$ para $l \neq i$ y como $\mathcal{F}_{i,j} \subset \mathcal{G}_j^{(n)}$ para $i = 1, \dots, n-j$ y $1 \leq j \leq n-1$, se obtiene que:

$$E[Y_{i,j+1} | \mathcal{G}_j^{(n)}] = E[Y_{i,j+1} | \mathcal{F}_{i,j}] = f_j Y_{i,j}. \quad (2.9)$$

Combinando las ecuaciones anteriores, se llega a que

$$E[\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}] = f_j, \quad (2.10)$$

de donde tomando esperanza $E[\hat{f}_j^{(n)}] = f_j$; es decir, $\hat{f}_j^{(n)}$ es un estimador insesgado de f_j .

Mostremos ahora la no correlación entre los estimadores Chain Ladder de los factores de desarrollo. Para esto supongamos, sin pérdida de generalidad, que $j < k$ y que $\hat{f}_j^{(n)}$ es medible respecto a $\mathcal{G}_k^{(n)}$. Por la ecuación (2.10):

$$E[\hat{f}_j^{(n)} \hat{f}_k^{(n)} | \mathcal{G}_k^{(n)}] = \hat{f}_j^{(n)} E[\hat{f}_k^{(n)} | \mathcal{G}_k^{(n)}] = \hat{f}_j^{(n)} f_k$$

y

$$E[\hat{f}_j^{(n)} f_k] = f_j f_k$$

Por tanto,

$$Cov(\hat{f}_j^{(n)}, \hat{f}_k^{(n)}) = E[\hat{f}_j^{(n)} \hat{f}_k^{(n)}] - E[\hat{f}_j^{(n)}] E[\hat{f}_k^{(n)}] = E[\hat{f}_j^{(n)} f_k] - f_j f_k = 0.$$

□

La varianza del estimador Chain Ladder $\hat{f}_j^{(n)}$ no es tan fácil de establecer. Los problemas surgen debido a que $\hat{f}_j^{(n)}$ es el ratio de dos variables aleatorias dependientes. Para los siguientes calculos será conveniente usar la notación siguiente:

$$\tilde{Y}_{i,j} = Y_{i,j+1} - f_j Y_{i,j}$$

donde asumimos que $Var(Y_j) < \infty$ para $j = 1, 2, \dots, n+1-i$. Puesto que $E[\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}] = f_j$, se tiene que:

$$Var(\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}) = E\left[\left(\frac{\sum_{i=1}^{n-j} \tilde{Y}_{i,j}}{\sum_{i=1}^{n-j} Y_{i,j}}\right)^2 | \mathcal{G}_j^{(n)}\right] = \frac{\sum_{i=1}^{n-j} Var(\tilde{Y}_{i,j} | \mathcal{G}_j^{(n)})}{(\sum_{i=1}^{n-j} Y_{i,j})^2} + \frac{2 \sum \sum_{1 \leq i < k \leq n-j} Cov(\tilde{Y}_{i,j}, \tilde{Y}_{k,j} | \mathcal{G}_j^{(n)})}{(\sum_{i=1}^{n-j} Y_{i,j})^2}$$

Se observa que:

$$Var(\tilde{Y}_{i,j} | \mathcal{G}_j^{(n)}) = Var(\tilde{Y}_{i,j} | \mathcal{F}_{i,j})$$

y

$$Cov(\tilde{Y}_{i,j}, \tilde{Y}_{k,j} | \mathcal{G}_j^{(n)}) = E[\tilde{Y}_{i,j} \tilde{Y}_{k,j} | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}],$$

donde $\mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}$ es la σ -álgebra más pequeña que contiene tanto a $\mathcal{F}_{i,j}$ como a $\mathcal{F}_{k,j}$. Note que si $i < k$, entonces:

$$\begin{aligned} E[\tilde{Y}_{i,j} \tilde{Y}_{k,j} | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}] &= E[E[\tilde{Y}_{i,j} \tilde{Y}_{k,j} | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j+1}] | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}] \\ &= E[\tilde{Y}_{k,j} E[\tilde{Y}_{i,j} | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j+1}] | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}] \\ &= E[\tilde{Y}_{k,j} E[\tilde{Y}_{i,j} | \mathcal{F}_{i,j}] | \mathcal{F}_{i,j} \vee \mathcal{F}_{k,j}] = 0, \end{aligned}$$

donde se ha utilizado la independencia de $\tilde{Y}_{i,j}$ y $\mathcal{F}_{k,j+1}$ y la condición **CL2** del modelo de

Mack. Por lo tanto, se ha demostrado que:

$$\text{Var}(\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}) = \frac{\sum_{i=1}^{n-j} \text{Var}(\tilde{Y}_{i,j} | \mathcal{F}_{i,j})}{(\sum_{i=1}^{n-j} Y_{i,j})^2}, \quad (2.11)$$

Es en este punto nos damos cuenta que las condiciones del modelo de Mack no son suficientes para evaluar estas varianzas condicionales, lo cual se evidencia desde que la condición **CL3** sólo se refiere a los esperados condicionales de los $Y_{i,j}$'s. Por consiguiente, introduciremos aquí una condición adicional a las propuestas por Mack, donde seguiremos usando la notación, de no haber confusión, $\tilde{Y}_j = \tilde{Y}_{i,j}$.

CL4: Para todo $i = 1, 2, \dots, n$ y todo $j = 1, 2, \dots, n+1-i$ se cumple que $\text{Var}(Y_j) < \infty$ y

$$\text{Var}(Y_{j+1} | Y_1, \dots, Y_j) = E[(Y_{j+1} - f_j Y_j)^2 | Y_1, \dots, Y_j] = \sigma_j^2 Y_j. \quad (2.12)$$

La razón de esta condición es de nuevo la conveniencia. Se debe de recordar que las condiciones **CL2** y **CL3** del modelo de Mack tenían un sólo propósito: fueron escogidos con la finalidad de explicar la forma en que los estimadores de Chain Ladder $\hat{f}_j^{(n)}$ son utilizados para predecir el número de reclamos o siniestros futuros. Por tanto, la condición **CL4** fue dado por razones pragmáticas.

Como se ha visto en los cálculos anteriores, las condiciones **CL1**, **CL2** y **CL3** permiten demostrar resultados asintóticos, de insesgamiento y de no correlación de los estimadores Chain Ladder. Asumiendo la condición **CL4**, estamos ahora en condiciones de continuar los cálculos para el desarrollo de la varianza del estimador Chain Ladder. Concretamente se tiene lo siguiente:

Proposición 2.2.3 *Bajo las condiciones **CL1**, **CL2**, **CL3** y **CL4** del modelo de Mack se tiene:*

$$\text{Var}(\hat{f}_j^{(n)}) = \sigma_j^2 E[(\sum_{i=1}^{n-j} Y_{i,j})^{-1}] \quad (2.13)$$

Demostración: Por (2.12):

$$\text{Var}(\tilde{Y}_{i,j} | \mathcal{F}_{i,j}) = \sigma_j^2 Y_{i,j}$$

Luego de (2.11):

$$\text{Var}(\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}) = \frac{\sum_{i=1}^{n-j} \sigma_j^2 Y_{i,j}}{(\sum_{i=1}^{n-j} Y_{i,j})^2} = \frac{\sigma_j^2}{\sum_{i=1}^{n-j} Y_{i,j}} \quad (2.14)$$

La varianza incondicional puede ahora evaluarse mediante la aplicación de la siguiente relación:

$$\begin{aligned} \text{Var}(\hat{f}_j^{(n)}) &= E[\text{Var}(\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)})] + \text{Var}(E[\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)}]) \\ &= E[\text{Var}(\hat{f}_j^{(n)} | \mathcal{G}_j^{(n)})] \\ &= \sigma_j^2 E\left[\frac{1}{\sum_{i=1}^{n-j} Y_{i,j}}\right] \end{aligned}$$

□

donde hemos utilizado (2.10).

Estimadores de las cantidades σ_j^2 pueden volver a ser obtenidos por el método de momentos. Tomando esperanzas en la ecuación (2.12):

$$E[(Y_{j+1} - f_j Y_j)^2] = \sigma_j^2 E[Y_j].$$

Reemplazando los esperados de ambas partes por sus versiones muestrales se obtiene:

$$\frac{1}{n-j} \sum_{i=1}^{n-j} (Y_{i,j+1} - f_j Y_{i,j})^2 = \sigma_j^2 \frac{1}{n-j} \sum_{i=1}^{n-j} Y_{i,j},$$

ecuación que nos provee al reemplazarse f_j por su estimador $\hat{f}_j^{(n)}$, los estimadores:

$$(\hat{\sigma}_j^{(n)})^2 = \frac{\sum_{i=1}^{n-j} (Y_{i,j+1} - \hat{f}_j^{(n)} Y_{i,j})^2}{\sum_{i=1}^{n-j} Y_{i,j}}. \quad (2.15)$$

2.2.3. Predicciones en el Modelo de Mack

Asumamos el modelo de Mack con la condición adicional **CL4**. Dado $i = 1, \dots, n$, nuestro primer objetivo consistirá en predecir la severidad futura $Y_{i,n+2-i}$ para un periodo contiguo en base al triángulo de desarrollo de la figura 2.3. El predictor natural se deriva de las condiciones **CL2** y **CL3** del modelo de Mack al reemplazarse el factor de desarrollo desconocido en el predictor a un periodo en adelante:

$$E[Y_{i,n+2-i} | \mathcal{F}_{i,n+1-i}] = f_{n+1-i} Y_{i,n+1-i}$$

por su estimador Chain Ladder \hat{f}_{n+1-i} . Este procedimiento origina el siguiente predictor:

$$\hat{Y}_{i,n+2-i} = \hat{f}_{n+1-i}^{(n)} Y_{i,n+1-i}. \quad (2.16)$$

Debido al reemplazo del estimador Chain Ladder, el predictor $\hat{Y}_{i,n+2-i}$ pierde la propiedad de minimizar del error cuadrático medio $E[(Y_{i,n+2-i} - M)^2]$ sobre la clase de todas las variables aleatorias M que tienen varianza finita y son medibles respecto a la σ -álgebra $\mathcal{F}_{i,n+1-i}$. En vista de la condición **CL4** el mínimo error cuadrático medio viene dado ahora por:

$$\begin{aligned} E[(Y_{i,n+2-i} - f_{n+1-i} Y_{i,n+1-i})^2] &= E[\text{Var}(Y_{i,n+2-i} | \mathcal{F}_{i,n+1-i})] \\ &= E[\sigma_{n+1-i}^2 Y_{i,n+1-i}] = \sigma_{n+1-i}^2 E[Y_{i,n+1-i}]. \end{aligned} \quad (2.17)$$

La varianza $\sigma_{n+1-i}^2 E[Y_{i,n+1-i}]$ es un mínimo para el error predictivo de la cantidad $\hat{Y}_{i,n+2-i}$. Por otro lado, el error cuadrático medio de la predicción de $\hat{Y}_{i,n+2-i}$ viene dada por:

$$\text{err}_Y^{(n)} = E[(\hat{Y}_{i,n+2-i} - Y_{i,n+2-i})^2].$$

Se observa que:

$$\text{err}_Y^{(n)} = E[(Y_{i,n+2-i} \pm f_{n+1-i} Y_{i,n+1-i} - \hat{Y}_{i,n+2-i})^2]$$

$$\begin{aligned}
 &= E[(Y_{i,n+2-i} - f_{n+1-i}Y_{i,n+1-i})^2] + E[((f_{n+1-i} - \hat{f}_{n+1-i}^{(n)})Y_{i,n+1-i})^2] \\
 &\quad + 2E[(Y_{i,n+2-i} - f_{n+1-i}Y_{i,n+1-i})E[(f_{n+1-i} - \hat{f}_{n+1-i}^{(n)})Y_{i,n+1-i}]] \\
 &= \sigma_{n+1-i}^2 E[Y_{i,n+1-i}] + \text{Var}(\hat{f}_{n+1-i}^{(n)}) E[Y_{i,n+1-i}^2],
 \end{aligned}$$

donde se ha utilizado el hecho que $\hat{f}_{n+1-i}^{(n)}$ es por definición independiente de $Y_{i,n+1-i}$ e $Y_{i,n+2-i}$. Por tanto, de la última proposición obtenemos el siguiente error de predicción.

$$\begin{aligned}
 err_Y^{(n)} &= \sigma_{n+1-i}^2 E[Y_{i,n+1-i}] + \text{Var}(\hat{f}_{n+1-i}^{(n)}) E[Y_{i,n+1-i}^2] \\
 &= \sigma_{n+1-i}^2 (E[Y_{i,n+1-i}] + E[(\sum_{k=1}^{i-1} Y_{k,n+1-i})^{-1}] E[Y_{i,n+1-i}^2]).
 \end{aligned}$$

Una comparación entre la varianza del mejor predictor de $Y_{i,n+2-i}$ dada en la ecuación (2.17) y el error de predicción de $Y_{i,n+2-i}$, muestran que el término adicional

$$\text{Var}(\hat{f}_{n+1-i}^{(n)}) E[Y_{i,n+1-i}^2]$$

aparece debido a la incertidumbre del estimador del parámetro $\hat{f}_{n+1-i}^{(n)}$. Las cantidades $E[Y_{i,n+1-i}]$ y $E[Y_{i,n+1-i}^2]$ pueden, además, ser calculadas mediante el uso de las condiciones **CL2**, **CL3** y **CL4**. Por ejemplo:

$$E[Y_{i,n+1-i}] = f_{n-i} \dots f_1 E[Y_1], \quad (2.18)$$

y tomando $\mathcal{F}_j = \sigma\langle Y_{i,1}, \dots, Y_{i,j} \rangle$, se obtiene:

$$\begin{aligned}
 E[Y_{i,n+1-i}^2] &= E[\text{Var}(Y_{i,n+1-i} | \mathcal{F}_{n-i})] + E[E[(f_{n-i}Y_{i,n-i})^2 | \mathcal{F}_{n-i}]] \\
 &= \sigma_{n-i}^2 E[Y_{i,n-i}] + f_{n-i}^2 E[Y_{i,n-i}^2] = \sigma_{n-i}^2 f_{n-i-1} \dots f_1 E[Y_1] + f_{n-i}^2 E[Y_{i,n-i}^2].
 \end{aligned} \quad (2.19)$$

Para señalar los límites del error de predicción $err_Y^{(n)}$ se necesita reemplazar las siguientes cantidades: σ_{n+1-i}^2 , $E[Y_{i,n+1-i}]$ y $E[Y_{i,n+1-i}^2]$ por estimadores apropiados, como podrían ser los estimadores a través del método de momentos. En virtud de la LFGN, estos estimadores son estimadores fuertemente consistentes de sus contrapartes determinísticos. Por otro lado, como la LFGN es un resultado asintótico, uno necesita el promedio de un número grande de observaciones para conseguir estimaciones seguras. Por construcción de estos estimadores, esto implica que $err_Y^{(n)}$ puede solamente ser estimado en una manera razonable si i y n son grandes.

Ya que las severidades $E[Y_{i,n+1-i}]$ y $E[Y_{i,n+1-i}^2]$ que aparecen en $err_Y^{(n)}$ satisfacen las relaciones recursivas de las ecuaciones (2.18) y (2.19), uno podría estar tentado a reemplazar $E[Y_{i,n+1-i}]$ y $E[Y_{i,n+1-i}^2]$ por estas recursiones y luego estimar los parámetros f_k , σ_k^2 y los momentos $E[Y_0]$ y $E[Y_0^2]$. Este método no evita el problema del tamaño de muestra pequeña, mencionada anteriormente. Finalmente, también es necesario estimar la cantidad $E[(\sum_{k=1}^{i-1} Y_{k,n-i})^{-1}]$ que aparece en $err_Y^{(n)}$.

La condiciones **CL2** y **CL3** del modelo de Mack no solo especifican la forma de predicción de un paso adelante $E[Y_{j+1}|Y_1, \dots, Y_j] = f_j Y_j$ de Y_{j+1} , sino que también define al mejor predictor de Y_{j+k} dado por Y_1, \dots, Y_j para cualquier $k \geq 1$. Usando las propiedades del esperado condicional, el predictor de varios pasos hacia adelante o mejor predictor de Y_{j+k}

es dado por:

$$\begin{aligned} E[Y_{j+k}|Y_1, \dots, Y_j] &= E[E[Y_{j+k}|Y_1, \dots, Y_{j+k-1}]|Y_1, \dots, Y_j] \\ &= f_{j+k-1}E[Y_{j+k-1}|Y_1, \dots, Y_j] = f_{j+k-1} \dots f_j Y_j. \end{aligned} \quad (2.20)$$

El correspondiente error de predicción es luego calculado por un argumento similar:

$$\begin{aligned} E[(Y_{j+k} - f_{j+k-1} \dots f_j Y_j)^2] &= \\ E[(Y_{j+k} - f_{j+k-1} Y_{j+k-1})^2] &+ f_{j+k-1}^2 E[(Y_{j+k-1} - f_{j+k-2} Y_{j+k-2})^2] + \dots + f_{j+k-1}^2 \dots f_{j+1}^2 E[(Y_{j+1} - f_j Y_j)^2]. \end{aligned}$$

Ahora, asumiendo las condiciones **CL2**, **CL3** y **CL4**, se obtiene:

$$\begin{aligned} E[(Y_{j+k} - f_{j+k-1} \dots f_j Y_j)^2] &= \\ &= \sigma_{j+k-1}^2 E[Y_{j+k-1}] + f_{j+k-1}^2 \sigma_{j+k-2}^2 E[Y_{j+k-2}] + \dots + f_{j+k-1}^2 \dots f_{j+1}^2 \sigma_j^2 E[Y_j] \\ &= E[Y_j] \sum_{l=0}^{k-1} f_j \dots f_{j+l-1} \sigma_{j+l}^2 [f_{j+l+1} \dots f_{j+k-1}]^2. \end{aligned} \quad (2.21)$$

Si tomamos la relación de la ecuación (2.20) como punto de partida podemos encontrar una aproximación al mejor predictor de $Y_{i,n+k-i}$, con $k > 1$, dado observaciones del pasado y el último periodo con información $Y_{i,1}, \dots, Y_{i,n+1-i}$. El mejor predictor es:

$$E[Y_{i,n+k-i}|\mathcal{F}_{i,n+1-i}] = f_{n-i+k-1} \dots f_{n+1-i} Y_{i,n+1-i}, \quad k > 1.$$

Un candidato para una aproximación se obtiene al reemplazar el parámetro desconocido f_l por sus estimadores Chain Ladder.

$$\hat{Y}_{i,n+k-i} = \hat{f}_{n-i+k-1}^{(n)} \dots \hat{f}_{n+1-i}^{(n)} Y_{i,n+1-i} \quad (2.22)$$

El error cuadrático medio del predictor $\hat{Y}_{i,n+k-i}$ viene dado por:

$$\begin{aligned} err_Y^{(n)} &= E[(Y_{i,n+k-i} - \hat{Y}_{i,n+k-i})^2] \\ &= E[(Y_{i,n+k-i} - E[Y_{i,n+k-i}|\mathcal{F}_{i,n+1-i}])^2] + E[(E[Y_{i,n+k-i}|\mathcal{F}_{i,n+1-i}] - \hat{Y}_{i,n+k-i})^2], \end{aligned} \quad (2.23)$$

donde

$$E[(Y_{i,n+k-i} - E[Y_{i,n+k-i}|\mathcal{F}_{i,n+1-i}])^2] = E[Y_{i,n+1-i}] \sum_{l=n-i}^{n+k-i-1} (f_{n-i+k-1} \dots f_{l+1})^2 \sigma_l^2 f_{l-1} \dots f_{n-i}. \quad (2.24)$$

Un estimador de este error cuadrático medio viene dado por la siguiente proposición.

Proposición 2.2.4

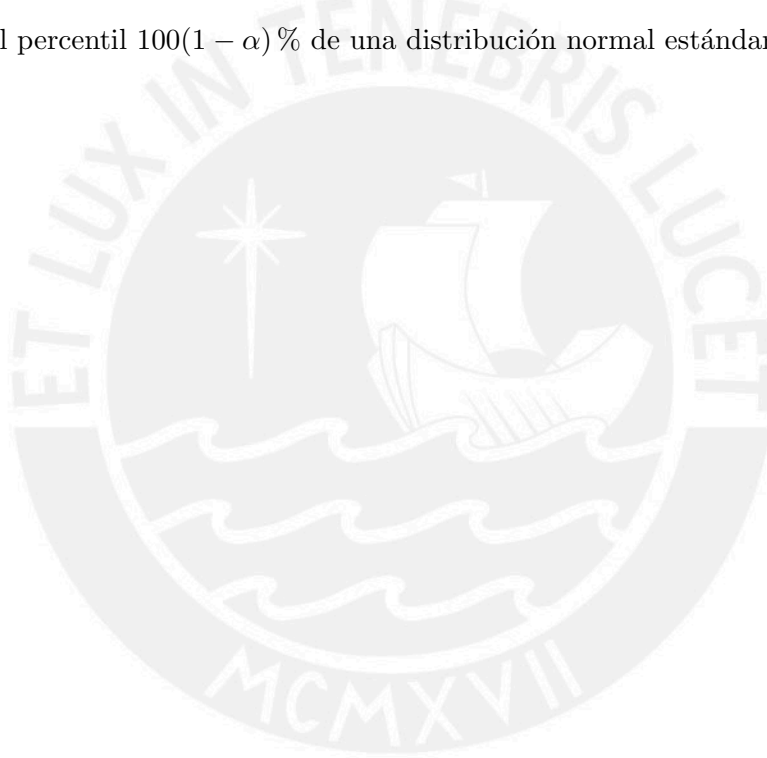
$$e\hat{r}_Y^{(n)} = (\hat{Y}_{i,n+k-i})^2 \sum_{l=n-i}^{n+k-i-1} \frac{\hat{\sigma}_l^2}{\hat{f}_l^2} \left(\frac{1}{\hat{Y}_{i,l+1}} + \frac{1}{\sum_{i=1}^{n-1+i} Y_{i,l+1}} \right) \quad (2.25)$$

Demostración: Véase [Mack \(1997\)](#). □

La estimación en la proposición anterior es muy importante para la construcción de los intervalos de confianza asintóticos de los montos de reservas totales estimados para el periodo de ocurrencia i . Estos vienen dados por:

$$[\hat{Y}_{i,n+k-i} - z_{1-\frac{\alpha}{2}} \sqrt{e\hat{r}_Y^{(n)}}, \hat{Y}_{i,n+k-i} + z_{1-\frac{\alpha}{2}} \sqrt{e\hat{r}_Y^{(n)}}] \quad (2.26)$$

siendo $z_{1-\frac{\alpha}{2}}$ el percentil $100(1 - \alpha) \%$ de una distribución normal estándar.



Capítulo 3

Regresiones Aparentemente no Relacionadas (SUR)

En este capítulo desarrollaremos la teoría de las Regresiones Aparentemente no Relacionadas (**SUR**), la cual fue introducida por Zellner (1962). Esta será utilizada en la extensión multivariada del método Chain Ladder propuesta por Zhang (2010) que expondremos en el siguiente capítulo.

Cuando uno hace referencia a los modelos de ecuaciones simultáneas, inmediatamente uno piensa en aquellos sistemas en los que se especifican variables dependientes en algunas ecuaciones como determinadas por otras ecuaciones del modelo. Si bajo este esquema existe una correlación de los términos de error entre las ecuaciones del sistema, la aplicación del método de mínimos cuadrados ordinarios (MCO) podría resultar en estimadores sesgados y con errores cuadrados medios que pueden ser elevados, especialmente en muestras pequeñas.

Consideremos el siguiente conjunto de M ecuaciones de regresión:

$$\begin{aligned}
 Y_1 &= \mathbf{X}_1\beta_1 + \varepsilon_1 \\
 Y_2 &= \mathbf{X}_2\beta_2 + \varepsilon_2 \\
 &\vdots \\
 Y_m &= \mathbf{X}_m\beta_m + \varepsilon_m \\
 &\vdots \\
 Y_M &= \mathbf{X}_M\beta_M + \varepsilon_M
 \end{aligned} \tag{3.1}$$

donde para cada ecuación Y_m es un vector columna $N \times 1$ de variables endógenas, \mathbf{X}_m es una matriz de orden $N \times k_m$ de variables exógenas (explicativas) no aleatorias, β_m es un vector $k_m \times 1$ de parámetros y ε_m es un vector $N \times 1$ de errores aleatorios con media cero.

Este sistema puede escribirse más compactamente, como:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

ó

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon, \tag{3.2}$$

donde \mathbf{Y} es un vector columna de orden $NM \times 1$, \mathbf{X} es una matriz de orden $NM \times k$, β es un vector columna de orden $k \times 1$ y ε es un vector columna de orden $NM \times 1$, siendo $k = \sum_{m=1}^M k_m$.

Se observa que las variables endógenas Y_m no están como variables predeterminadas en otras ecuaciones del sistema, con lo cual se podría pensar que cada una de las ecuaciones podría ser estimada con el uso del método de mínimos cuadrados ordinarios. Esto sería factible si las ecuaciones fueran completamente independientes en el sentido que la matriz de varianzas-covarianzas de los términos de error en el sistema de ecuaciones fuera diagonal. Lo antes mencionado podría ser incorrecto si se detectara algún tipo de movimiento simultáneo de todas las ecuaciones originado por una supuesta relación contemporánea entre los términos de error. Es decir, las regresiones que no están aparentemente relacionadas, sí lo estarían por medio de correlaciones implícitas entre los términos de error. Esto resulta natural por ejemplo si las variables Y_1, Y_2, \dots, Y_M corresponden a M distintas mediciones hechas sobre los mismos individuos. Aquí los errores en la ecuación (3.1) describen lo no explicado por las variables independientes. Por tanto, si existiera alguna o algunas variables predictoras relevantes no incluidas en el modelo que afecten simultáneamente a las variables dependientes, sería lógico esperar alguna correlación entre estos errores. Es de la idea expresada en esta discusión que proviene el nombre en inglés SUR (de Seemingly Unrelated Regressions) para este sistema de ecuaciones.

Según Araya and Muñoz (1996) existe un aspecto importante por rescatar y este es la posibilidad que existe de cometer un serio error por considerar independientes a varias ecuaciones de regresión, cuando en realidad están asociadas por medio de los términos de error. No existe un procedimiento econométrico que le diga al investigador que tiene que aplicar el método **SUR**, sólo la experiencia y el conocimiento teórico le permitirán identificar la necesidad de utilizarlo.

En el modelo SUR se especifica que la matriz de varianzas-covarianzas del vector de errores ε viene dada por:

$$\mathbf{V} = E[\varepsilon\varepsilon^\top] = \begin{bmatrix} \sigma_{11}I_N & \sigma_{12}I_N & \cdots & \sigma_{1M}I_N \\ \sigma_{21}I_N & \sigma_{22}I_N & \cdots & \sigma_{2M}I_N \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}I_N & \sigma_{M2}I_N & \cdots & \sigma_{MM}I_N \end{bmatrix} = \mathbf{\Sigma} \otimes \mathbf{I}_N \quad (3.3)$$

donde \mathbf{I}_N es la matriz identidad de orden N , \otimes es el producto Kronecker y $\mathbf{\Sigma}$ es la matriz de varianzas-covarianzas:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix} \quad (3.4)$$

Vale recordar que la matriz $\mathbf{\Sigma}$ es simétrica y positiva definida. Asumiremos además que es no singular.

3.1. Estimación de parámetros

Recordemos que si en el modelo (3.2) $\Sigma = \sigma^2 \mathbf{I}_N$, entonces el mejor estimador lineal e insesgado (MELI) para el parámetro β viene dado por el estimador de mínimos cuadrados:

$$\hat{\beta}_0 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Si bien la condición anterior no se cumple en general, es posible extender un resultado similar para un modelo heterocedástico, como (3.2), que especifique que:

$$\text{Var}(\epsilon) = \mathbf{V},$$

donde \mathbf{V} es una matriz simétrica y positiva definida.

En efecto, por el teorema de descomposición singular podemos garantizar la existencia de una matriz ortogonal $\mathbf{P}(\mathbf{P}^\top \mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{I})$ y una matriz diagonal Λ tal que $\mathbf{V} = \mathbf{P}\Lambda\mathbf{P}^\top$. Así multiplicando la ecuación (3.2) por $\Lambda^{-1/2}\mathbf{P}^\top$, tenemos:

$$\Lambda^{-1/2}\mathbf{P}^\top \mathbf{Y} = \Lambda^{-1/2}\mathbf{P}^\top \mathbf{X}\beta + \mathbf{U},$$

donde $\mathbf{U} = \Lambda^{-1/2}\mathbf{P}^\top \epsilon$ puede verificarse que es un término de error de media cero y matriz de covarianza $\sigma^2 \mathbf{I}_N$. Puesto que este modelo satisface los supuestos clásicos uno obtiene entonces el denominado estimador de mínimos cuadrados generalizados:

$$\hat{\beta}_G = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y}.$$

En nuestro caso particular este podría escribirse más explícitamente como:

$$\hat{\beta}_G = (\mathbf{X}^\top (\Sigma^{-1} \otimes \mathbf{I}_N) \mathbf{X})^{-1} \mathbf{X}^\top (\Sigma^{-1} \otimes \mathbf{I}_N) \mathbf{Y}.$$

Un cálculo directo de la matriz de varianzas-covarianza de este estimador muestra que

$$\text{Var}(\hat{\beta}_G) = (\mathbf{X}^\top (\Sigma^{-1} \otimes \mathbf{I}_N) \mathbf{X})^{-1}.$$

Un resultado interesante propuesto por [Milliken and Albohali \(1984\)](#) de cuando el estimador de mínimos cuadrados generalizados colapsa en el MELI de β , viene dado por la siguiente proposición.

Proposición 3.1.1 *Para el modelo lineal (3.2), $\hat{\beta}_G$ es el MELI de β si y solo si :*

$$\mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{I} - \mathbf{W}) = \mathbf{0},$$

donde $\mathbf{W} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$.

Demostración: Con el objetivo de que $\hat{\beta}_G$ sea el MELI de β , \mathbf{X} y \mathbf{V} deben ser tales que $\hat{\beta}_G = \hat{\beta}_0$; es decir,

$$(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y}.$$

Denotemos por \mathbf{U} a la diferencia de estas matrices:

$$\mathbf{U} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y},$$

y expresemos \mathbf{Y} mediante $\mathbf{Y} = \mathbf{WY} + (\mathbf{I} - \mathbf{W})\mathbf{Y}$. Luego, \mathbf{U} se reduce a:

$$\mathbf{U} = -(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{I} - \mathbf{W}) \mathbf{Y}.$$

Por tanto, $\hat{\beta}_G$ es el MELI de β si y solo si $\mathbf{U} = \mathbf{0}$ para todo \mathbf{Y} o, equivalentemente, si y solo si:

$$(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{I} - \mathbf{W}) = \mathbf{0}.$$

Como $(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}$ es no singular, la condición equivale a decir que $\hat{\beta}_G$ es el MELI de β si y solo si:

$$\mathbf{X}^\top \mathbf{V}^{-1} (\mathbf{I} - \mathbf{W}) = \mathbf{0}.$$

□

Es necesario resaltar que todo lo antes mencionado está basado en la asunción que la matriz de varianzas-covarianzas Σ sea conocida, pero por lo general esto no sucede así. Cuando la matriz de varianza-covarianza no es conocida, el correspondiente estimador es el de mínimos cuadrados factible generalizado:

$$\hat{\beta}_{FG} = (\mathbf{X}^\top (\hat{\Sigma} \otimes \mathbf{I}_N)^{-1} \mathbf{X})^{-1} \mathbf{X}^\top (\hat{\Sigma} \otimes \mathbf{I}_N)^{-1} \mathbf{Y}, \quad (3.5)$$

donde $\hat{\Sigma}$ es un estimador de Σ . Existen varias posibilidades para estimar Σ . Estos están basados en el residual $\hat{\varepsilon}$ obtenido de aplicarse el método de mínimos cuadrados ordinarios:

$$\hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta}_0.$$

El más comúnmente utilizado de tales estimadores es:

$$\hat{\Sigma}_0 = \frac{1}{N} (\hat{\varepsilon}^\top \hat{\varepsilon}).$$

Otro estimador de Σ , a tomarse en cuenta, es el propuesto por Zellner and Huang (1962): $\hat{\Sigma}_u = [\hat{\sigma}_{mn}]$, donde:

$$\hat{\sigma}_{nm} = \begin{cases} \hat{\varepsilon}_n^\top \hat{\varepsilon}_n / (N - J_m) & , \forall m = n \\ \frac{\hat{\varepsilon}_n^\top \hat{\varepsilon}_m}{N - J_m - J_n + tr((\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \mathbf{X}_n^\top \mathbf{X}_m (\mathbf{X}_m^\top \mathbf{X}_m)^{-1} \mathbf{X}_m^\top \mathbf{X}_n)} & , \forall n \neq m \end{cases} \quad (3.6)$$

siendo J_m el número de regresores en la m -ésima ecuación y tr la traza de una matriz.

Otro estimador alternativo es $\hat{\Sigma}_g = [\hat{\sigma}_{nm}]$ donde:

$$\hat{\sigma}_{nm} = \frac{\hat{\varepsilon}_n^\top \hat{\varepsilon}_m}{((N - J_n)(N - J_m))^{1/2}} \quad (3.7)$$

Este estimador es insesgado solo si n es igual a m .

CAPÍTULO 3. REGRESIONES APARENTEMENTE NO RELACIONADAS (SUR)

Hasta este punto, se observa que tanto el estimador de Mínimos Cuadrados Generalizados $\hat{\beta}_G$ y su estimador factible $\hat{\beta}_{FG}$ (o FGLS de estimador de mínimos cuadrados factible generalizado) toman en cuenta la correlación entre los términos de error en la estimación. Sin embargo, si esta correlación se ignora, entonces uno obtiene el estimador de Mínimos Cuadrados Ordinarios: $\hat{\beta}_0 = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

Se deben resaltar tres aspectos de los estimadores **SUR** sobre su eficiencia (entendida como varianza mínima). El primer aspecto, es que cuanto más elevada es la *correlación contemporánea* de los términos de error entre ecuaciones, mayor será la ganancia en eficiencia del estimador generalizado. El segundo aspecto se refiere a que si la *correlación contemporánea* es muy baja (los triángulos de la matriz Σ se aproximan a cero), no hay una ganancia importante por aplicar la regresión **SUR** a las ecuaciones en lugar de utilizar los mínimos cuadrados ordinarios en cada ecuación. Finalmente, si cada una de las ecuaciones del sistema tiene las mismas variables exógenas, entonces los estimadores **SUR** son equivalentes a los de mínimos cuadrados ordinarios. En resumen, la varianza será mínima cuando las variables explicativas en las diferentes ecuaciones no están altamente correlacionadas.

Consideramos ahora la eficiencia de los tres estimadores $\hat{\beta}_0$, $\hat{\beta}_G$ y $\hat{\beta}_{FG}$ hasta el momento estudiados. Cuando Σ es conocida, $\hat{\beta}_G$ es el mejor estimador lineal e insesgado. La diferencia entre $\hat{\beta}_G$ y $\hat{\beta}_0$ en términos de eficiencia es:

$$\begin{aligned} \text{Var}(\hat{\beta}_0) - \text{Var}(\hat{\beta}_G) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\Sigma \otimes \mathbf{I}_N) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} - (\mathbf{X}^\top (\Sigma \otimes \mathbf{I}_N)^{-1} \mathbf{X})^{-1} \\ &= \mathbf{P} (\Sigma \otimes \mathbf{I}_N) \mathbf{P}^\top. \end{aligned}$$

donde $\mathbf{P} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - (\mathbf{X}^\top (\Sigma \otimes \mathbf{I}_N)^{-1} \mathbf{X})^{-1} \mathbf{X}^\top (\Sigma \otimes \mathbf{I}_N)^{-1}$ es una matriz que satisface $\mathbf{P} \mathbf{X} = 0$. Dado que $\Sigma \otimes \mathbf{I}_N$ es positiva definida, $\mathbf{P} (\Sigma \otimes \mathbf{I}_N) \mathbf{P}^\top$ es al menos positiva semi definida, por tanto $\hat{\beta}_G$ es al menos tan eficiente como $\hat{\beta}_0$.

Sin embargo, esta conclusión no se sostiene respecto a la comparación entre $\hat{\beta}_{FG}$ y $\hat{\beta}_0$ cuando no se conoce Σ ya que el uso del estimador $\hat{\Sigma}$ podría aumentar la variabilidad del estimador $\hat{\beta}_{FG}$. Desafortunadamente, aunque se reconoce que el estimador FGLS no será universalmente mejor que el estimador MCO, la relación general exacta entre ellos sigue siendo una pregunta abierta. Zellner (1963) consideró un caso especial de un modelo de dos ecuaciones donde los términos de error estaban normalmente distribuidos y los regresores de las diferentes ecuaciones eran ortogonales entre sí; es decir, $\mathbf{X}_1^\top \mathbf{X}_2 = 0$. En este caso especial, Zellner (1963) mostró que, para una muestra con menos de 20 observaciones y valor absoluto de la verdadera correlación contemporánea $\rho_{12} = \sigma_{12} / \sqrt{\sigma_{11} \sigma_{22}}$ menor que 0.3, $\hat{\beta}_{FG}$ es de hecho menos eficiente que $\hat{\beta}_0$. Además, algunos resultados aproximados asintóticos muestran que la eficiencia ganada se reducirá debido a lo siguiente:

1. Una baja correlación entre los términos de error;
2. Una alta correlación entre regresores a través de diferentes ecuaciones;
3. Una alta correlación entre regresores dentro de cada ecuación.

Por lo que si los resultados asintóticos todavía se mantienen en muestras finitas, es de esperar que la correlación requerida debe ser aún mayor que 0.3 para ganar eficiencia mediante

el uso de $\hat{\beta}_{FG}$. Así vemos que, en el caso de múltiples triángulos correlacionados, aunque es razonable esperar que las estimaciones más eficientes serán el resultado del desarrollo conjunto de los triángulos teniendo en cuenta la correlación, no es sorpresa que pueda suceder lo inverso debido al uso de un estimador para Σ . Esto es más probable que ocurra cuando se presenten pequeñas correlaciones entre los triángulos.

3.2. El estimador iterativo de Mínimos Cuadrados Factible Generalizado (IFGLS)

Una vez obtenido el estimador $\hat{\beta}_{FG}$ de los coeficientes, este podría ser utilizado para estimar nuevamente la matriz de varianza-covarianza mediante los residuales obtenidos bajo este estimador (en lugar de con los de $\hat{\beta}_0$). A partir de esta nueva matriz, podríamos aplicar nuevamente (3.5) y obtener en una segunda etapa un nuevo estimador SUR factible generalizado. Este a su vez puede ser utilizado nuevamente para estimar una nueva matriz de varianza-covarianza. A partir de dicha matriz, podemos obtener un estimador SUR de tercera etapa; y continuar así de manera iterativa hasta que se produzca una convergencia en el sentido de que la diferencia en norma entre dos estimaciones para etapas consecutivas del vector de coeficientes de regresión sea menor a una tolerancia muy pequeña pre-establecida. El estimador resultante es por tanto conocido como el estimador iterativo de mínimos cuadrados factible generalizado (IFGLS). En adelante denotaremos al estimador IFGLS de l -ésima etapa por $\hat{\beta}_I^{(l)}$. En otras palabras, este estimador viene dado por:

$$\hat{\beta}_I^{(l)} = (\mathbf{X}^\top (\hat{\Sigma}^{(l)} \otimes \mathbf{I}_N)^{-1} \mathbf{X})^{-1} \mathbf{X}^\top (\hat{\Sigma}^{(l)} \otimes \mathbf{I}_N)^{-1} \mathbf{Y},$$

donde $\hat{\Sigma}^{(l)}$ es un estimador consistente de Σ construido con el residual basado en $\hat{\beta}_I^{(l-1)}$. Note que, por construcción, $\hat{\beta}_I^{(0)}$ es el estimador de mínimos cuadrados ordinarios de β y que $\hat{\beta}_I^{(1)}$ es el estimador de mínimos cuadrados factible generalizado.

El procedimiento iterativo, aunque dirigida a mejorar la eficiencia del estimador de mínimos cuadrados factible generalizado (FGLS), no puede evitar una pérdida de eficiencia, similar a la discutida con el estimador FGLS. [Srivastava and Giles \(1987\)](#) mostraron que las iteraciones no siempre pueden ser útiles ya que los estimadores FGLS y IFGLS tienen la misma varianza para muestras grandes, pero para muestras pequeñas, el estimador IFGLS podría dar como resultado una pérdida de eficiencia. Situaciones favorables para el estimador IFGLS resultan cuando se presenta una baja correlación entre los regresores a través de las ecuaciones y una alta correlación entre los términos de error.

Capítulo 4

El modelo General Multivariado Chain Ladder

En este capítulo haremos una presentación de la extensión multivariada propuesta por [Zhang \(2010\)](#) para el modelo Chain Ladder. En ella analizaremos cómo es posible utilizar la relación existente entre múltiples triángulos de desarrollo con el fin de buscar mejores inferencias y predicciones de las reservas.

La motivación de los métodos de reserva multivariados proviene esencialmente de dos fuentes. Una es identificar y entender las conexiones estructurales entre diferentes tipos de data de reclamos y otra es estudiar las correlaciones contemporáneas que pudieran existir entre distintos triángulos de desarrollo.

Un ejemplo que claramente ilustra el problema de no tomar en cuenta las conexiones estructurales es cuando uno dispone de triángulos de desarrollo tanto para los reclamos incurridos como pagados. Como lo indican [Quarg and Mack \(2004\)](#), el tomar por separado un modelo Chain Ladder para cada triángulo conduce a predicciones inconsistentes, razón por la cual estos autores proponen como solución su modelo Munich Chain Ladder. Otro ejemplo en esta línea consiste en la inclusión de data de conteo de reclamos en las proyecciones de los montos de reclamos, véase al respecto [De Alba \(2002\)](#) y [Ntzoufras and Dellaportas \(2002\)](#).

El uso de los métodos de reserva multivariados es, por otro lado, motivado por la presencia de las correlaciones contemporáneas. Es práctica común que la estimación de las severidades de un portafolio este conformado por las inferencias individuales hechas para cada uno de muchos sub-portafolios. Esta correlación es a menudo inducida como consecuencia de la influencia de ciertos factores comunes, tales como la inflación o eventos catastróficos.

Para contabilizar debidamente esta correlación, una distribución multivariada debe ser construida y la estimación conjunta de la correlación de los sub-portafolios debe llevarse a cabo. Desde un punto estadístico, una ventaja importante de la implementación de un método de estimación conjunta es la ganancia de eficiencia en la estimación de los parámetros. El tratamiento actuarial pone un mayor énfasis en la valoración del impacto como resultado del cálculo de la variabilidad de las reservas. Por ejemplo, es posible que sub-portafolios de diferentes líneas de negocios tengan una correlación positiva y como tal, aplicaciones independientes de modelos univariados estocásticos den lugar a una subestimación de la variación de las reservas, representando una imagen de falso optimismo de los riesgos económicos que enfrenta la compañía de seguros. En este contexto, varios papers han hecho extensiones del enfoque univariado Chain Ladder de [Mack \(1993\)](#) para la estimación de las reservas y de su correspondiente variabilidad que tome en cuenta estas correlaciones. Los modelos compren-

didos bajo este marco se conocen como "Modelos Multivariados Chain Ladder".

El modelo General Multivariado Chain Ladder, o simplemente modelo GMCL del acrónimo en inglés General Multivariate Chain Ladder, no solo especifica las correlaciones contemporáneas, sino que también permite conexiones estructurales entre los triángulos de desarrollo. La estructura del modelo GMCL se extiende del modelo Multivariado Chain Ladder (MCL) de una manera natural, sin embargo, esta extensión simple resulta ser ventajosa en la mejora de la adecuación del modelo y en el incremento de la flexibilidad de este. El corazón del modelo GMCL es el empleo de las ya vistas Regresiones Aparentemente no Relacionadas (SURs) para la estimación de los parámetros.

4.1. Formulación del Modelo

A continuación se formulará el modelo GMCL, el cual tiene como objetivo el analizar simultáneamente múltiples triángulos de desarrollo correlacionados. Para esto asumiremos que disponemos de M triángulos de desarrollo del mismo tamaño. Referiremos con $m \in \{1, \dots, M\}$ al m -ésimo triángulo, $i \in \{1, \dots, n\}$ al i -ésimo periodo de origen u ocurrencia y $j \in \{1, \dots, n\}$ al j -ésimo periodo de desarrollo. Denotemos por $\mathbf{Y}_{i,j} = [Y_{i,j}^{(1)} \dots Y_{i,j}^{(M)}]^\top$ al vector columna de severidades (acumuladas) en el periodo de ocurrencia i y desarrollo j , donde (m) hace referencia al m -ésimo triángulo de desarrollo. Cada una de las severidades $Y_{i,j}^{(m)}$ se asumen estrictamente positivas.

Asimismo, denotemos por $\mathcal{D} = \{\mathbf{Y}_{i,j} \mid 1 \leq i \leq n, 1 \leq j \leq n, i+j \leq n+1\}$ al conjunto de todas las severidades (observadas), $\mathcal{D}_{\leq j} = \{\mathbf{Y}_{i,k} \mid 1 \leq i \leq n, k \leq j\}$ al conjunto de todas las severidades hasta el periodo de desarrollo j , y $\mathcal{D}_{i,j} = \{\mathbf{Y}_{i,k} \mid k \leq j\}$ al conjunto de todas las severidades para el periodo de ocurrencia i hasta el periodo de desarrollo j .

Además, denotaremos por $\mathbf{Y}_{\leq,j}^{(m)} = [Y_{1,j}^{(m)} Y_{2,j}^{(m)} \dots Y_{n+1-j,j}^{(m)}]^\top$ al vector columna de las primeras $n+1-j$ severidades para el periodo de desarrollo j del m -ésimo triángulo y por $\mathbf{Y}_{<,j}^{(m)} = [Y_{1,j}^{(m)} Y_{2,j}^{(m)} \dots Y_{n-j,j}^{(m)}]^\top$ al vector columna de las primeras $n-j$ severidades para el periodo de desarrollo j del m -ésimo triángulo. Se usará vec como el operador de vectorización de una matriz de tal manera que si $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_1 \dots \mathbf{a}_n]$ es una matriz $m \times n$, donde cada \mathbf{a}_i es un vector columna de orden m , entonces $vec(\mathbf{A}) = [\mathbf{a}_1^\top \mathbf{a}_1^\top \dots \mathbf{a}_n^\top]^\top$ será un vector columna de tamaño $mn \times 1$.

Como una generalización natural multivariada del modelo Chain Ladder, consideraremos el siguiente modelo de referencia GMCL para el periodo de desarrollo j , el cual hace referencia a la transición entre los periodos de desarrollo j y $j+1$:

$$\mathbf{Y}_{i,j+1} = \mathbf{B}_j \cdot \mathbf{Y}_{i,j} + \epsilon_{i,j}, \quad (4.1)$$

donde $\epsilon_{i,j} = [\epsilon_{i,j}^{(1)} \epsilon_{i,j}^{(2)} \dots \epsilon_{i,j}^{(M)}]^\top$ es un vector columna de orden M de errores aleatorios y

$$\mathbf{B}_j = \begin{bmatrix} \beta_{11} & \dots & \beta_{1M} \\ \vdots & \ddots & \vdots \\ \beta_{M1} & \dots & \beta_{MM} \end{bmatrix}$$

es una matriz de orden $M \times M$ para los factores de desarrollo en el período j . Vale detallar que la matriz $\mathbf{Y}_{i,j+1}$ en (4.1) contiene en la m -ésima fila factores de desarrollo asociados a

CAPÍTULO 4. EL MODELO GENERAL MULTIVARIADO CHAIN LADDER

todos los triángulos de desarrollo:

$$Y_{i,j+1}^{(m)} = \beta_{m1}Y_{i,j}^{(1)} + \beta_{m2}Y_{i,j}^{(2)} + \dots + \beta_{mM}Y_{i,j}^{(M)} + \epsilon_{i,j}^{(m)} \quad (4.2)$$

Cabe también aclarar que, por simplicidad, estamos omitiendo el subíndice j para las entradas de \mathbf{B}_j . Esto desde que ellas serán innecesarias en los cálculos que más adelante veremos.

El modelo (4.1) es usado como modelo de partida ya que no posee un término de intercepto. Más adelante veremos que es posible, sin mayores inconvenientes, el incluir tal término.

Para el modelo (4.1), se tienen los siguientes supuestos:

(GMCL1)

$$E[\epsilon_{i,j}|\mathcal{D}_{i,j}] = \mathbf{0}. \quad (4.3)$$

(GMCL2)

$$Cov(\epsilon_{i,j}|\mathcal{D}_{i,j}) = D(\mathbf{Y}_{i,j})^{1/2}\Sigma_j D(\mathbf{Y}_{i,j})^{1/2}, \quad (4.4)$$

donde $D(\mathbf{Y}_{i,j})^{1/2}$ denota a la matriz diagonal de orden M cuyas entradas están conformadas por las raíces cuadradas de las entradas del vector $\mathbf{Y}_{i,j}$.

(GMCL3) La severidades que provienen de distintos años de origen son independientes entre sí.

(GMCL4) Los errores aleatorios $\epsilon_{i,j}$ están distribuidas simétricamente, siendo

$$\Sigma_j = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \dots & \sigma_{MM} \end{bmatrix}$$

una matriz de orden $M \times M$ simétrica y positiva definida.

Una vez más, por simplicidad, se han omitido los sub-índices j en las entradas de la matriz Σ_j .

Es interesante observar como el supuesto (GMCL2) extiende la asunción (CL4) vista en el modelo Chain Ladder univariado. De manera explícita (4.4) nos dice que para todo $m, q \in \{1, 2, \dots, M\}$:

$$Cov(Y_{i,j+1}^{(m)}, Y_{i,j+1}^{(q)} | D_{i,j}) = Cov(\epsilon_{i,j+1}^{(m)}, \epsilon_{i,j+1}^{(q)} | D_{i,j}) = \sigma_{mp} \sqrt{Y_{i,j}^{(m)} Y_{i,j}^{(q)}}$$

y en particular

$$Var(Y_{i,j+1}^{(m)} | D_{i,j}) = Var(\epsilon_{i,j+1}^{(m)} | D_{i,j}) = \sigma_{mm} Y_{i,j}^{(m)}.$$

Observando detalladamente la estructura del modelo de partida (4.1), se encuentra que ninguna de las variables dependientes (endógenas) se muestran como variables explicativas en las ecuaciones y que las ecuaciones individuales para cada triángulo están vinculados a través de la correlación existente entre los errores de éstas, lo cual sugiere el empleo de las

CAPÍTULO 4. EL MODELO GENERAL MULTIVARIADO CHAIN LADDER

Regresiones Aparentemente no Relacionadas (SUR) para la estimación de los parámetros, punto que ha sido detallado en el capítulo anterior.

Es fácil verificar que el modelo (4.1) puede escribirse, en concordancia a la figura 2.3 y (4.2), como el siguiente sistema de ecuaciones:

$$\begin{aligned} Y_{1,j+1}^{(1)} &= \beta_{11}Y_{1,j}^{(1)} + \beta_{12}Y_{1,j}^{(2)} + \dots + \beta_{1M}Y_{1,j}^{(M)} + \epsilon_{1,j}^{(1)} \\ &\vdots \\ Y_{n-j,j+1}^{(1)} &= \beta_{11}Y_{n-j,j}^{(1)} + \beta_{12}Y_{n-j,j}^{(2)} + \dots + \beta_{1M}Y_{n-j,j}^{(M)} + \epsilon_{n-j,j}^{(1)} \\ Y_{1,j+1}^{(2)} &= \beta_{21}Y_{1,j}^{(1)} + \beta_{22}Y_{1,j}^{(2)} + \dots + \beta_{2M}Y_{1,j}^{(M)} + \epsilon_{1,j}^{(2)} \\ &\vdots \\ Y_{n-j,j+1}^{(2)} &= \beta_{21}Y_{n-j,j}^{(1)} + \beta_{22}Y_{n-j,j}^{(2)} + \dots + \beta_{2M}Y_{n-j,j}^{(M)} + \epsilon_{n-j,j}^{(2)} \\ &\vdots \\ Y_{n-j,j+1}^{(M)} &= \beta_{M1}Y_{n-j,j}^{(1)} + \beta_{M2}Y_{n-j,j}^{(2)} + \dots + \beta_{MM}Y_{n-j,j}^{(M)} + \epsilon_{n-j,j}^{(M)} \end{aligned}$$

Este se expresa matricialmente como:

$$\begin{bmatrix} \mathbf{Y}_{\leq,j+1}^{(1)} \\ \mathbf{Y}_{\leq,j+1}^{(2)} \\ \vdots \\ \mathbf{Y}_{\leq,j+1}^{(M)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} \quad (4.5)$$

donde, para $m = 1, 2, \dots, M$, $\mathbf{Y}_{\leq,j+1}^{(m)}$ es un vector $(n-j) \times 1$ de todas las severidades para el periodo de desarrollo $j+1$ del m -ésimo triángulo de desarrollo, $\mathbf{X}_m = [\mathbf{Y}_{<,j}^{(1)} \mathbf{Y}_{<,j}^{(2)} \dots \mathbf{Y}_{<,j}^{(M)}]$ es una matriz $(n-j) \times M$ de las primeras $(n-j)$ severidades para el periodo de desarrollo j con $\mathbf{X}_1 = \mathbf{X}_2 = \dots = \mathbf{X}_M$, $\beta_m = [\beta_{m1} \dots \beta_{mM}]^\top$ es un vector columna de parámetros o factores de desarrollo en la m -ésima ecuación y $\epsilon_m = [\epsilon_{1,j}^{(m)} \epsilon_{2,j}^{(m)} \dots \epsilon_{n-j,j}^{(m)}]^\top$ es un vector $(n-j) \times 1$ de términos de error para el m -ésimo triángulo de desarrollo.

El modelo (4.5) puede ser escrito más compactamente como:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad (4.6)$$

donde $\mathbf{Y} = \text{vec}([\mathbf{Y}_{\leq,j+1}^{(1)} \dots \mathbf{Y}_{\leq,j+1}^{(M)}])$ es un vector $(n-j)M \times 1$ de variables respuestas; $\mathbf{X} = \text{diag}(\mathbf{X}_1, \dots, \mathbf{X}_M)$ es una matriz diagonal de bloques $(n-j)M \times M^2$, $\beta = \text{vec}([\beta_1 \beta_2 \dots \beta_M])$ es un vector $M^2 \times 1$ de parámetros de desarrollo y $\epsilon = \text{vec}([\epsilon_1 \epsilon_2 \dots \epsilon_M])$ es un vector $(n-j)M \times 1$ de términos de error.

Si denotamos por $\mathbf{W} = \text{vec}([\mathbf{Y}_{<,j}^{(1)} \mathbf{Y}_{<,j}^{(2)} \dots \mathbf{Y}_{<,j}^{(M)}]) = \text{vec}(\mathbf{X}_1)$ al vector $(n-j)M \times 1$ de las primeras $n-j$ severidades observadas para el periodo de desarrollo j , obtendremos de (GMCL2) y (GMCL3) que

$$\text{Cov}(\epsilon_m, \epsilon_q \mid D_{i,j}) = \sigma_{mq} D(\mathbf{Y}_{<,j}^{(m)})^{1/2} \mathbf{I} D(\mathbf{Y}_{<,j}^{(q)})^{1/2}$$

y consecuentemente:

$$Cov(\varepsilon \mid D_{i,j}) = D(\mathbf{W})^{1/2}(\Sigma_j \otimes \mathbf{I})D(\mathbf{W})^{1/2}, \quad (4.7)$$

donde \mathbf{I} es la matriz identidad de orden $(n-j) \times (n-j)$ y D es el operador diagonal definido anteriormente. Multiplicando a ambos lados de (4.6) por $D(\mathbf{W})^{-1/2}$, se obtiene:

$$D(\mathbf{W})^{-1/2}\mathbf{Y} = D(\mathbf{W})^{-1/2}\mathbf{X}\beta + D(\mathbf{W})^{-1/2}\varepsilon,$$

ó

$$\mathbf{Y}^* = \mathbf{X}^*\beta + \varepsilon^*, \quad (4.8)$$

donde

$$\mathbf{Y}^* = D(\mathbf{W})^{-1/2}\mathbf{Y}, \mathbf{X}^* = D(\mathbf{W})^{-1/2}\mathbf{X} \text{ y } \varepsilon^* = D(\mathbf{W})^{-1/2}\varepsilon.$$

Así de (4.7) se tiene que:

$$Cov(\varepsilon^* \mid D_{i,j}) = D(\mathbf{W})^{-1/2}Cov(\varepsilon)D(\mathbf{W})^{-1/2} = \Sigma_j \otimes \mathbf{I}. \quad (4.9)$$

Empleándose luego la ecuación (4.1) con

$$\mathbf{Y}_{i,n-i+2} = \mathbf{B}_{n-i+1} \cdot \mathbf{Y}_{i,n-i+1} + \epsilon_{i,n-i+1},$$

y trabajándose recursivamente, pueden obtenerse los valores finales de las severidades por año de ocurrencia i en función de las últimas severidades observadas $\mathbf{Y}_{i,n-i+1}$:

$$\mathbf{Y}_{i,n} = \left(\prod_{j=n-1}^{n-i+1} \mathbf{B}_j \right) \mathbf{Y}_{i,n-i+1} + \sum_{j=n-i+1}^{n-2} \left(\prod_{l=n-1}^{j+1} \mathbf{B}_l \right) \epsilon_{i,j} + \epsilon_{i,n-1}, \quad (4.10)$$

donde $\prod_{j=n-1}^{n-i+1} \mathbf{B}_j$ representa la productoria de los \mathbf{B}_j desde $j = n-1$ hasta $j = n-i+1$. Usando (4.10) y los supuestos (GMCL1) y (GMCL3), se aprecia que la esperanza de la última severidad para el año de origen i condicional a las severidades observadas \mathcal{D} es:

$$E[\mathbf{Y}_{i,n} \mid \mathcal{D}] = E[\mathbf{Y}_{i,n} \mid \mathcal{D}_{i,n-i+1}] = E[\mathbf{Y}_{i,n} \mid \mathcal{D}_{i,n-i+1}] = \left(\prod_{j=n-1}^{n-i+1} \mathbf{B}_j \right) \mathbf{Y}_{i,n-i+1} \quad (4.11)$$

Finalmente, teniendo la ecuación (4.8) el formato de un modelo SUR, podemos, por lo visto en el capítulo anterior, estimar cada $\beta = \text{vec}(\mathbf{B}_j^T)$ a fin de estimar el esperado de la severidad última mediante:

$$\hat{\mathbf{Y}}_{i,n} = \left(\prod_{j=n-1}^{n-i+1} \hat{\mathbf{B}}_j \right) \mathbf{Y}_{i,n-i+1} \quad (4.12)$$

El lema siguiente resume algunas propiedades del estimador obtenido.

Lema 4.1.1 *Bajo los supuestos (GMCL1)-(GMCL4), se tiene que:*

- (i) $E[\hat{\mathbf{B}}_j | \mathcal{D}_{.,j}] = \mathbf{B}_j$ y $\hat{\mathbf{B}}_j$ es un estimador insesgado de \mathbf{B}_j ;
- (ii) Los estimadores de la matriz de desarrollo para diferentes periodos de desarrollo no están correlacionados; es decir, $E[\hat{\mathbf{B}}_k \hat{\mathbf{B}}_j] = E[\hat{\mathbf{B}}_k] E[\hat{\mathbf{B}}_j] \forall k \neq j$;
- (iii) Condicional a $\mathcal{D}_{.,n-i+1}$, el estimador $\hat{\mathbf{Y}}_{i,n}$ es un estimador insesgado de $E[\mathbf{Y}_{i,n} | \mathcal{D}]$; es decir, $E[\hat{\mathbf{Y}}_{i,n} | \mathcal{D}_{.,n-i+1}] = E[\mathbf{Y}_{i,n} | \mathcal{D}]$;
- (iv) El estimador $\hat{\mathbf{Y}}_{i,n}$ es un estimador insesgado de $E[\mathbf{Y}_{i,n}]$.

Demostración: (i) El hecho que $E[\hat{\mathbf{B}}_j | \mathcal{D}_{.,j}] = \mathbf{B}_j$ se deduce directamente de los resultados dados por [Kakwani \(1967\)](#) bajo el supuesto (GMCL4). Por tanto, $E[\hat{\mathbf{B}}_j] = E[E[\hat{\mathbf{B}}_j | \mathcal{D}_{.,j}]] = E[\mathbf{B}_j] = \mathbf{B}_j$.

(ii) Sin pérdida de generalidad, supongamos que $k < j$, luego

$$E[\hat{\mathbf{B}}_k \hat{\mathbf{B}}_j] = E[E[\hat{\mathbf{B}}_k \hat{\mathbf{B}}_j | \mathcal{D}_{.,j}]] = E[\hat{\mathbf{B}}_k E[\hat{\mathbf{B}}_j | \mathcal{D}_{.,j}]] = E[\hat{\mathbf{B}}_k \mathbf{B}_j] = E[\hat{\mathbf{B}}_k] E[\mathbf{B}_j]$$

donde el último paso utiliza los resultados incondicionales en (i).

(iii) Dado que se observa $\mathcal{D}_{.,n-i+1}$, utilizando los resultados de (i), (ii), [\(4.11\)](#) y [\(4.12\)](#), tenemos que:

$$\begin{aligned} E[\hat{\mathbf{Y}}_{i,n} | \mathcal{D}_{.,n-i+1}] &= E \left[\left(\prod_{j=n-1}^{n-i+1} \hat{\mathbf{B}}_j \right) \mathbf{Y}_{i,n-i+1} | \mathcal{D}_{.,n-i+1} \right] = E \left[\left(\prod_{j=n-1}^{n-i+1} \hat{\mathbf{B}}_j \right) \right] \mathbf{Y}_{i,n-i+1} \\ &= \left[\prod_{j=n-1}^{n-i+1} E(\hat{\mathbf{B}}_j) \right] \mathbf{Y}_{i,n-i+1} = \left(\prod_{j=n-1}^{n-i+1} \mathbf{B}_j \right) \mathbf{Y}_{i,n-i+1} = E[\mathbf{Y}_{i,n} | \mathcal{D}] \end{aligned}$$

(iv) De (iii) y [\(4.11\)](#), tenemos que

$$E[\hat{\mathbf{Y}}_{i,n} | \mathcal{D}_{.,n-i+1}] = E[\mathbf{Y}_{i,n} | \mathcal{D}] = E[\mathbf{Y}_{i,n} | \mathcal{D}_{.,n-i+1}].$$

Tomando esperanza en ambos lados, se tiene que:

$$E[\hat{\mathbf{Y}}_{i,n}] = E[\mathbf{Y}_{i,n}].$$

□

4.2. El modelo GMCL con intercepto

[Barnett and Zelnwirth \(2000\)](#) señalaron que el modelo Chain Ladder resulta ser muchas veces inadecuado, ya que tiende a sobreestimar los valores grandes y subestimar los valores

CAPÍTULO 4. EL MODELO GENERAL MULTIVARIADO CHAIN LADDER

pequeños, y por tanto el gráfico residual con los valores ajustados en el eje de las abscisas muestra una tendencia decreciente. Como Zhang (2010) lo afirma, la causa de este comportamiento es que el modelo en cuestión excluye al término de intersección. Por fortuna, la inclusión de tal término no conlleva mayor dificultad en el SUR y en el modelo GMCL. Esto es lo que precisamente discutiremos en esta sección.

Sea $\mathbf{A}_j = ([\beta_{10} \ \beta_{20} \ \dots \ \beta_{M0}])^\top$ un vector de intercepto $M \times 1$, donde β_{m0} es el intercepto del m -ésimo triángulo. Una vez más omitiremos el indicador del periodo de desarrollo j por simplicidad. El modelo (4.1) puede luego extenderse a:

$$\mathbf{Y}_{i,j+1} = \mathbf{A}_j + \mathbf{B}_j \cdot \mathbf{Y}_{i,j} + \epsilon_{i,j} \quad (4.13)$$

donde los supuestos (GMCL1) - (GMCL3) siguen siendo los mismos. Nos referiremos al modelo (4.13) como el modelo con intercepto, y a \mathbf{B}_j como la matriz de desarrollo. La estimación del modelo (4.13) es sencilla bajo la técnica SUR desarrollada en el capítulo anterior. Tan solo se requiere cambiar la definición de \mathbf{X}_m y β_m en (4.5) para mostrar la inclusión de los interceptos. Para ser más específicos, se definirá $\mathbf{X}_m = [\mathbf{1} \ \mathbf{Y}_{<,j}^{(1)} \ \dots \ \mathbf{Y}_{<,j}^{(M)}]$ como una matriz $n - j \times (M + 1)$, donde $\mathbf{1}$ es el vector columna de 1's, y $\beta_m = [\beta_{m0} \ \beta_{m1} \ \dots \ \beta_{mM}]^\top$ como un vector de parámetros $(M + 1) \times 1$ para estimar el m -ésimo triángulo. Luego se procede del mismo modo que en la sección anterior para obtener los estimadores de los β_m (que engloban ahora tanto a \mathbf{A}_j como a \mathbf{B}_j) y de Σ_j .

Para combinar los coeficientes estimados para cada periodo con la finalidad de obtener la última severidad esperada, se usa el siguiente procedimiento con el fin de mantener la estructura multiplicativa (4.12) de modo que la inclusión de interceptos no tenga efectos en los resultados anteriores. Reescribimos (4.13) como:

$$\begin{bmatrix} 1 \\ \mathbf{Y}_{i,j+1} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{A}_j & \mathbf{B}_j \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \mathbf{Y}_{i,j} \end{bmatrix} + \begin{bmatrix} 0 \\ \epsilon_{i,j} \end{bmatrix} \quad (4.14)$$

donde se ha agregado simplemente una igualdad de constantes en la primera fila. Ahora se denota:

$$\mathbf{Z}_{i,j} = \begin{bmatrix} 1 \\ \mathbf{Y}_{i,j} \end{bmatrix}, \quad \mathbf{E}_j = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{A}_j & \mathbf{B}_j \end{bmatrix}, \quad \mathbf{e}_{i,j} = \begin{bmatrix} 0 \\ \epsilon_{i,j} \end{bmatrix}$$

y (4.14) se convierte en:

$$\mathbf{Z}_{i,j+1} = \mathbf{E}_j \cdot \mathbf{Z}_{i,j} + \mathbf{e}_{i,j}, \quad (4.15)$$

el cual es un modelo con el mismo formato de un modelo sin intercepto. Ahora es fácil ver que (4.10) y (4.11) siguen siendo válidos de substituirse $\mathbf{Y}_{i,j}$, \mathbf{B}_j y $\epsilon_{i,j}$ con $\mathbf{Z}_{i,j}$, \mathbf{E}_j y $\mathbf{e}_{i,j}$, respectivamente, desde que $E[\mathbf{e}_{i,j} \mid \mathcal{D}_{i,j}] = \mathbf{0}$ según (GMCL1). A continuación, se obtendrá el estimador $\hat{\mathbf{Z}}_{i,n} = (\prod_{j=n-1}^{n-i+1} \hat{\mathbf{E}}_j) \cdot \mathbf{Z}_{i,n-i+1}$. Nótese que de extraerse la constante 1 a $\hat{\mathbf{Z}}_{i,n}$ éste resultará ser el estimador $\hat{\mathbf{Y}}_{i,n}$. Nuestro enfoque aumentado no tiene ningún impacto en el lema 4.1.1 ya que se han agregado términos no estocásticos. Específicamente, los puntos (i) y (ii) también se cumplen reemplazando \mathbf{B}_j con \mathbf{E}_j y $\hat{\mathbf{B}}_j$ con $\hat{\mathbf{E}}_j$ desde que $[\hat{\mathbf{A}}_j \ \hat{\mathbf{B}}_j]$ es insesgado. Por tanto los puntos (iii) y (iv) se siguen inmediatamente.

Capítulo 5

Aplicación al Cálculo de Reservas en Compañías de Seguros

5.1. Aspectos generales

En este capítulo mostraremos las aplicaciones de los métodos de cálculos de reservas descritos en el marco teórico: los métodos Chain Ladder determinístico y estocástico; y el método de cálculo de reservas bajo el modelo General Multivariado Chain Ladder. Es necesario señalar que para éste último modelo se necesita más de una data, por lo que trabajaremos en esta aplicación con 3 triángulos de desarrollo de una base de datos proveniente de una compañía de seguros nacional en la rama de salud.

El principal objetivo del capítulo es aplicar y comparar los diferentes métodos de estimación de reservas descritas.

5.1.1. Descripción de las bases de datos

Base de Datos: SALUD

Contar en la actualidad con un seguro de salud es básico y primordial, el seguro de salud es un complemento valioso en la vida de cada persona individual o perteneciente a una familia u otro grupo social, porque suple económicamente parte de los auxilios o servicios accesorios que de otra manera resultarían muy costosos. En el caso de hospitalizaciones, servicios de ambulancia, exámenes médicos y citas con el médico general, el seguro de salud se encarga de suplir una parte del pago para comodidad del cliente que no tiene que preocuparse por cancelar el monto total del servicio.

Se dice que nada es más importante que la salud, y hay mucho de verdad en esto, ya que una pérdida de la salud puede significar una pérdida de ingresos si no son capaces de trabajar, y también puede significar facturas médicas extremadamente caras. El seguro de salud, como se mencionó anteriormente, puede ayudar a evitar algunos de esos gastos y también a ayudar a las personas a llevar vidas más saludables mediante el pago de las medidas preventivas, tales como chequeos regulares y vacunas. Por ello, el tener un seguro de salud debe ser visto como una inversión que sirve para proteger el bienestar personal y familiar, así como un medio para contar con un verdadero respaldo frente a situaciones inesperadas.

Existen varios medios por los cuales una persona puede obtener un seguro de salud, ya sea que ella pague su propio seguro o reciba un seguro de salud como parte de un paquete de empleo. Actualmente, en el Perú existe una variedad de seguros de salud y diversas compañías aseguradoras, las cuales permiten a una persona poder escoger su plan de cobertura.

Para el análisis a realizarse utilizaremos los siniestros reportados en el ramo de salud de una compañía aseguradora local entre los años 2006 y 2012, con una periodicidad anual. Para ello dispondremos de 3 bases de datos, a los que llamaremos:

- Seguro de salud de tipo A, con una cobertura a nivel mundial de atenciones ambulatorias, hospitalarias y de emergencias;
- Seguros de salud de tipo B, con un plan completo a nivel nacional con atenciones ambulatorias, hospitalarias y de emergencias; pero con una protección por persona de hasta un monto limitado;
- Seguros de salud de tipo C, con un plan completo a nivel nacional con atenciones ambulatorias, hospitalarias y de emergencias, pero con una protección por persona de hasta aproximadamente solo la mitad del monto de cobertura de los seguros de tipo B.

Estas 3 bases de datos nos servirán para trabajar con el modelo General Multivariado Chain Ladder, pero también serán utilizadas en la aplicación de los modelos restantes. A continuación se muestra la evolución de los montos en nuevos soles de los siniestros de los 3 tipos de seguros de salud, por año:

Año de ocurrencia	Montos de siniestros reportados	%
2006	29'417,983	13.4 %
2007	35'503,279	16.2 %
2008	28'588,276	13.1 %
2009	31'387,676	14.3 %
2010	26'867,355	12.3 %
2011	30'591,585	14.0 %
2012	36'427,213	16.6 %
Total	218'783,367	100 %

Cuadro 5.1: Distribución de reclamos - Data seguro de salud de tipo A.

Año de ocurrencia	Montos de siniestros reportados	%
2006	5'470,527	9.4 %
2007	6'535,565	11.2 %
2008	7'544,857	12.9 %
2009	7'071,979	12.1 %
2010	8'738,953	15.0 %
2011	10'597,688	18.2 %
2012	12'305,256	21.1 %
Total	58'264,824	100 %

Cuadro 5.2: Distribución de reclamos - Data seguro de salud de tipo B.

Año de ocurrencia	Montos de siniestros reportados	%
2006	4'673,464	8.3 %
2007	6'904,079	12.3 %
2008	8'597,323	15.3 %
2009	7'920,009	14.1 %
2010	7'780,218	13.8 %
2011	8'977,430	15.9 %
2012	11'466,011	20.4 %
Total	56'718,534	100 %

Cuadro 5.3: Distribución de reclamos - Data seguro de salud de tipo C.

Como se puede observar, el seguro de salud de tipo A es el seguro que ha reportado un mayor monto de siniestros en comparación con los demás seguros de salud, seguido por el seguro de tipo B. Sin embargo, los montos de siniestros de los seguros de salud de tipo B y C han mostrado en los últimos dos años un mayor incremento.

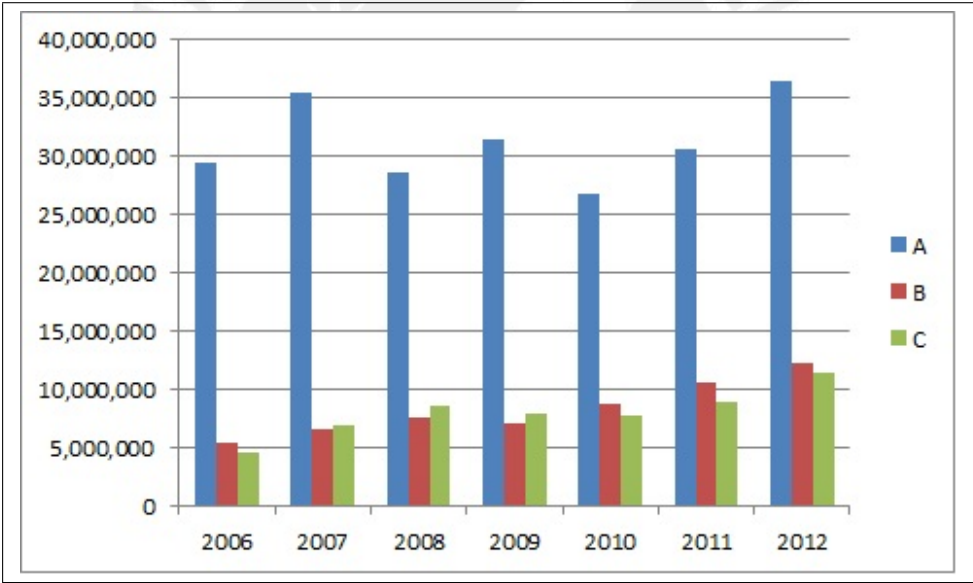


Figura 5.1: Distribución del monto de reclamos por tipo de seguro de salud.

5.2. El Método Chain Ladder determinístico

En esta sección se desarrollará y mostrará la aplicación del método Chain Ladder determinístico a los datos de los 3 seguros de salud. Como se mostró anteriormente, primero se deben construir los triángulos de desarrollo entre los años 2006-2012, últimos que los identificaremos con los números del 1 al 7, respectivamente. En el caso de los seguros de tipo A tenemos:

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	24'373,002	5'010,276	33,491	276	749	100	88
2 (2007)	29'606,879	5'643,946	24,459	1,657	215,950	10,389	
3 (2008)	23'931,287	4'618,818	35,830	1,712	628		
4 (2009)	27'750,549	3'386,216	228,612	22,300			
5 (2010)	23'256,730	3'476,174	134,451				
6 (2011)	25'879,518	4'712,068					
7 (2012)	36'427,213						

Cuadro 5.4: Triángulo de desarrollo (seguro de tipo A).

Cada celda corresponde al monto (en nuevos soles) de todos los reclamos que se presentaron entre un determinado año de ocurrencia y un año de desarrollo.

Luego, se construye el triángulo de severidades acumuladas el cual está dado por:

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	24'373,002	29'383,278	29'416,770	29'417,046	29'417,795	29'417,895	29'417,983
2 (2007)	29'606,879	35'250,825	35'275,284	35'276,941	35'492,890	35'503,279	
3 (2008)	23'931,287	28'550,105	28'585,935	28'587,648	28'588,276		
4 (2009)	27'750,549	31'136,765	31'365,376	31'387,676			
5 (2010)	23'256,730	26'732,904	26'867,355				
6 (2011)	25'879,518	30'591,585					
7 (2012)	36'427,213						

Cuadro 5.5: Triángulo de desarrollo de severidades acumuladas (seguro de tipo A).

El siguiente paso en el método Chain Ladder consiste en, a partir de las severidades acumuladas, obtener los correspondientes factores de desarrollo:

Año de desarrollo	1	2	3	4	5	6
Fact.Des.	1.173435728	1.003024373	1.000208155	1.002329785	1.000161591	1.000002991

Cuadro 5.6: Factores de desarrollo (seguro de tipo A).

A continuación, se completarán las celdas con los montos estimados a partir del triángulo de desarrollo acumulado, aplicando el factor de desarrollo.

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	24'373,002	29'383,278	29'416,770	29'417,046	29'417,795	29'417,895	29'417,983
2 (2007)	29'606,879	35'250,825	35'275,284	35'276,941	35'492,890	35'503,279	35'503,386
3 (2008)	23'931,287	28'550,105	28'585,935	28'587,648	28'588,276	28'592,895	28'592,981
4 (2009)	27'750,549	31'136,765	31'365,376	31'387,676	31'460,803	31'465,887	31'465,981
5 (2010)	23'256,730	26'732,904	26'867,355	26'872,947	26'935,556	26'939,908	26'939,989
6 (2011)	25'879,518	30'591,585	30'684,106	30'690,493	30'761,995	30'766,966	30'767,058
7 (2012)	36'427,213	42'744,993	42'874,270	42'883,194	42'983,103	42'990,049	42'990,177

Cuadro 5.7: Triángulo de desarrollo completo (seguro de tipo A).

Finalmente, las provisiones del seguro de salud de tipo A se calculan de la diferencia entre los montos acumulados estimados y su anterior. Por ejemplo, la reserva para el año de ocurrencia 2007 y el séptimo año de desarrollo (2013) se obtiene de la diferencia entre el monto acumulado de ese periodo y su anterior: $35'503,386 - 35'503,279 = 106$ nuevos soles.

Año de Ocurrencia	Año de Desarrollo							Reservas
	1	2	3	4	5	6	7	
1 (2006)	24'373,002	29'383,278	29'416,770	29'417,046	29'417,795	29'417,895	29'417,983	-
2 (2007)	29'606,879	35'250,825	35'275,284	35'276,941	35'492,890	35'503,279	106	106
3 (2008)	23'931,287	28'550,105	28'585,935	28'587,648	28'588,276	4,620	86	4,705
4 (2009)	27'750,549	31'136,765	31'365,376	31'387,676	73,127	5,084	94	78,304
5 (2010)	23'256,730	26'732,904	26'867,355	5,593	62,608	4,353	81	72,634
6 (2011)	25'879,518	30'591,585	92,520	6,387	71,502	4,971	92	175,473
7 (2012)	36'427,213	6'317,780	129,277	8,925	99,909	6,946	129	6'562,964
Reserva Estim.								6'894,187

Cuadro 5.8: Estimación de reservas (seguro de tipo A).

Por lo tanto, según el método Chain Ladder, se puede decir que al término del año 2012 la provisión final o monto a reservar será de 6'894,187 nuevos soles¹, el cual se obtiene de la suma de las provisiones parciales para cada uno de los periodos de ocurrencia y desarrollo.

A continuación, se desarrollará el mismo método para el resto de tipos de seguros.

¹Más precisamente, este monto corresponde a la reserva estimada necesaria para los años 2013 hasta el año 2018. De este monto total, $6'317,780 + 92,520 + 5,593 + 73,127 + 4,620 + 106 = 6'493,745$ nuevos soles corresponden al año 2013.

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'888,325	577,679	3,794	416	211	55	47
2 (2007)	5'826,166	708,416	775	166	22	20	
3 (2008)	6'474,334	1'061,016	8,202	1,066	238		
4 (2009)	6'395,681	649,734	22,214	4,350			
5 (2010)	7'757,548	969,619	11,786				
6 (2011)	9'326,949	1'270,739					
7 (2012)	12'305,256						

Cuadro 5.9: Triángulo de desarrollo (seguro de tipo B).

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'888,325	5'466,004	5'469,798	5'470,214	5'470,425	5'470,480	5'470,527
2 (2007)	5'826,166	6'534,582	6'535,357	6'535,523	6'535,545	6'535,565	
3 (2008)	6'474,334	7'535,350	7'543,552	7'544,619	7'544,857		
4 (2009)	6'395,681	7'045,415	7'067,629	7'071,979			
5 (2010)	7'757,548	8'727,167	8'738,953				
6 (2011)	9'326,949	10'597,688					
7 (2012)	12'305,256						

Cuadro 5.10: Triángulo de desarrollo de severidades acumuladas (seguro de tipo B).

Año de desarrollo	1	2	3	4	5	6
Fact.Des.	1.128776286	1.00132464	1.000225358	1.000024092	1.000006247	1.000008592

Cuadro 5.11: Factores de desarrollo (seguro de tipo B).

CAPÍTULO 5. APLICACIÓN AL CÁLCULO DE RESERVAS EN COMPAÑÍAS DE SEGUROS 38

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'888,325	5'466,004	5'469,798	5'470,214	5'470,425	5'470,480	5'470,527
2 (2007)	5'826,166	6'534,582	6'535,357	6'535,523	6'535,545	6'535,565	6'535,622
3 (2008)	6'474,334	7'535,350	7'543,552	7'544,619	7'544,857	7'544,904	7'544,968
4 (2009)	6'395,681	7'045,415	7'067,629	7'071,979	7'072,149	7'072,193	7'072,254
5 (2010)	7'757,548	8'727,167	8'738,953	8'740,922	8'741,133	8'741,187	8'741,262
6 (2011)	9'326,949	10'597,688	10'611,726	10'614,118	10'614,374	10'614,440	10'614,531
7 (2012)	12'305,256	13'889,881	13'908,280	13'911,414	13'911,749	13'911,836	13'911,956

Cuadro 5.12: Triángulo de desarrollo completo (seguro de tipo B).

Año de Ocurrencia	Año de Desarrollo							Reservas
	1	2	3	4	5	6	7	
1 (2006)	4'888,325	5'466,004	5'469,798	5'470,214	5'470,425	5'470,480	5'470,527	-
2 (2007)	5'826,166	6'534,582	6'535,357	6'535,523	6'535,545	6'535,565	56	56
3 (2008)	6'474,334	7'535,350	7'543,552	7'544,619	7'544,857	47	65	112
4 (2009)	6'395,681	7'045,415	7'067,629	7'071,979	170	44	61	275
5 (2010)	7'757,548	8'727,167	8'738,953	1,969	211	55	75	2,310
6 (2011)	9'326,949	10'597,688	14,038	2,391	256	66	91	16,843
7 (2012)	12'305,256	1'584,625	18,399	3,134	335	87	120	1'606,700
Reserva Estim.								1'626,296

Cuadro 5.13: Estimación de reservas (seguro de tipo B).

En resumen, según el método Chain Ladder, se puede decir que para el año 2012 la provisión final o monto a reservar para el tipo de seguro B será de 1'626, 296 nuevos soles.

En relación a los seguros de tipo C tenemos:

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'136,332	510,622	14,915	10,445	476	400	275
2 (2007)	6'096,185	807,005	492	279	72	45	
3 (2008)	7'702,008	892,789	2,056	199	270		
4 (2009)	7'200,770	712,275	6,514	450			
5 (2010)	6'980,431	799,667	120				
6 (2011)	8'067,646	909,784					
7 (2012)	11'466,011						

Cuadro 5.14: Triángulo de desarrollo (seguro de tipo C)

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'136,332	4'646,954	4'661,869	4'672,314	4'672,789	4'673,189	4'673,464
2 (2007)	6'096,185	6'903,191	6'903,683	6'903,962	6'904,034	6'904,079	
3 (2008)	7'702,008	8'594,797	8'596,853	8'597,053	8'597,323		
4 (2009)	7'200,770	7'913,045	7'919,559	7'920,009			
5 (2010)	6'980,431	7'780,098	7'780,218				
6 (2011)	8'067,646	8'977,430					
7 (2012)	11'466,011						

Cuadro 5.15: Triángulo de desarrollo de severidades acumuladas (seguro de tipo C).

Año de desarrollo	1	2	3	4	5	6
Fact.Des.	1.11527506	1.0006724	1.000404998	1.000040527	1.000038439	1.000058846

Cuadro 5.16: Factores de desarrollo (seguro de tipo C).

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'136,332	4'646,954	4'661,869	4'672,314	4'672,789	4'673,189	4'673,464
2 (2007)	6'096,185	6'903,191	6'903,683	6'903,962	6'904,034	6'904,079	6'904,485
3 (2008)	7'702,008	8'594,797	8'596,853	8'597,053	8'597,323	8'597,653	8'598,159
4 (2009)	7'200,770	7'913,045	7'919,559	7'920,009	7'920,330	7'920,635	7'921,101
5 (2010)	6'980,431	7'780,098	7'780,218	7'783,369	7'783,685	7'783,984	7'784,442
6 (2011)	8'067,646	8'977,430	8'983,466	8'987,104	8'987,469	8'987,814	8'988,343
7 (2012)	11'466,011	12'787,756	12'796,355	12'801,537	12'802,056	12'802,548	12'803,302

Cuadro 5.17: Triángulo de desarrollo completo (seguro de tipo C).

Año de Ocurrencia	Año de Desarrollo							Reservas
	1	2	3	4	5	6	7	
1 (2006)	4'136,332	4'646,954	4'661,869	4'672,314	4'672,789	4'673,189	4'673,464	-
2 (2007)	6'096,185	6'903,191	6'903,683	6'903,962	6'904,034	6'904,079	406	406
3 (2008)	7'702,008	8'594,797	8'596,853	8'597,053	8'597,323	330	506	836
4 (2009)	7'200,770	7'913,045	7'919,559	7'920,009	321	304	466	1,092
5 (2010)	6'980,431	7'780,098	7'780,218	3,151	315	299	458	4,224
6 (2011)	8'067,646	8'977,430	6,036	3,638	364	345	529	10,913
7 (2012)	11'466,011	1'321,745	8,598	5,182	519	492	753	1'337,290
Reserva Estim.								1'354,762

Cuadro 5.18: Estimación de reservas (seguro de tipo C).

La provisión final o monto a reservar para el año 2012 para el tipo de seguro C será entonces de 1'354,762 nuevos soles.

5.3. El Método Chain Ladder Estocástico

En esta sección mostraremos la aplicación del Modelo Chain Ladder Estocástico para la misma base de datos de Salud, utilizado anteriormente, con la finalidad de obtener los resultados de estimación de las reservas y sobre todo de encontrar los intervalos de confianza de las mismas bajo el modelo estocástico propuesto por Mack (1993). Para esto se utilizará el módulo "MACKCHAINLADDER" del paquete "Chain Ladder" del software libre R.

De manera similar a lo visto en la sección anterior podemos aquí calcular los factores de desarrollo siguientes para los datos del tipo de seguro A; pero adicionalmente se muestran ahora sus respectivas desviaciones estándar:

	2006	2007	2008	2009	2010	2011	2012
Factores Des.	1.173436	1.003024	1.000208	1.002330	1.000162	1.000003	1.000000
Desv. Est.	0.01304542	0.001332711	0.0001686597	0.002090926	0.0001440187	0.00001229168	0.000000

Cuadro 5.19: Factores de desarrollo (seguro de tipo A).

Con ello obtenemos la siguiente proyección para el triángulo de desarrollo:

Año de Ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	24'373,002	29'383,278	29'416,770	29'417,046	29'417,795	29'417,895	29'417,983
2 (2007)	29'606,879	35'250,825	35'275,284	35'276,941	35'492,890	35'503,279	35'503,385
3 (2008)	23'931,287	28'550,105	28'585,935	28'587,648	28'588,276	28'592,896	28'592,981
4 (2009)	27'750,549	31'136,765	31'365,376	31'387,676	31'460,802	31'465,886	31'465,980
5 (2010)	23'256,730	26'732,904	26'867,355	26'872,948	26'935,556	26'939,908	26'939,989
6 (2011)	25'879,518	30'591,585	30'684,105	30'690,493	30'761,995	30'766,966	30'767,058
7 (2012)	36'427,213	42'744,993	42'874,270	42'883,195	42'983,103	42'990,049	42'990,177

Cuadro 5.20: Triángulo de desarrollo completo (seguro de tipo A).

Cabe reiterar que los resultados son los mismos a los obtenidos por el método Chain Ladder determinístico, sin embargo, la gran ventaja de este método es la obtención de los intervalos de confianza, los cuales se muestran a continuación:

Año de Ocurrencia	Reserva Estimada	Mack S.E.	Límite Inferior	Límite Superior
1 (2006)	0	0	0	0
2 (2007)	106	590	-1,050	1,262
3 (2008)	4,705	7,463	-9,922	19,332
4 (2009)	78,304	131,060	-178,574	335,182
5 (2010)	72,634	119,538	-161,660	306,928
6 (2011)	175,473	163,586	-145,156	496,102
7 (2012)	6'562,964	1'113,432	4'380,637	8'745,291
Total	6'894,186			

Cuadro 5.21: Reservas estimadas e intervalos de confianza (seguro de tipo A).

Estas estimaciones de los errores estándar y de las reservas estimadas con sus intervalos al 95 % de confianza se obtienen de la ecuación (2.25) del capítulo 2.

Es importante, por otro lado, obtener algunos gráficos que nos muestren las estimaciones futuras de reservas, como las dadas en la figura 5.2 para el seguro de tipo A. Aquí se aprecia que a medida que los años de desarrollo se incrementan el monto de reserva a estimar es mayor. Adicionalmente se incluyen para el diagnóstico del modelo algunos gráficos de los residuales estandarizados. Ellos deberían de presentar un comportamiento aleatorio sin una tendencia definida a fin de que los supuestos del modelo lineal Chain Ladder Estocástico sean válidos. De lo visto en la figura 5.2, esto efectivamente ocurre.

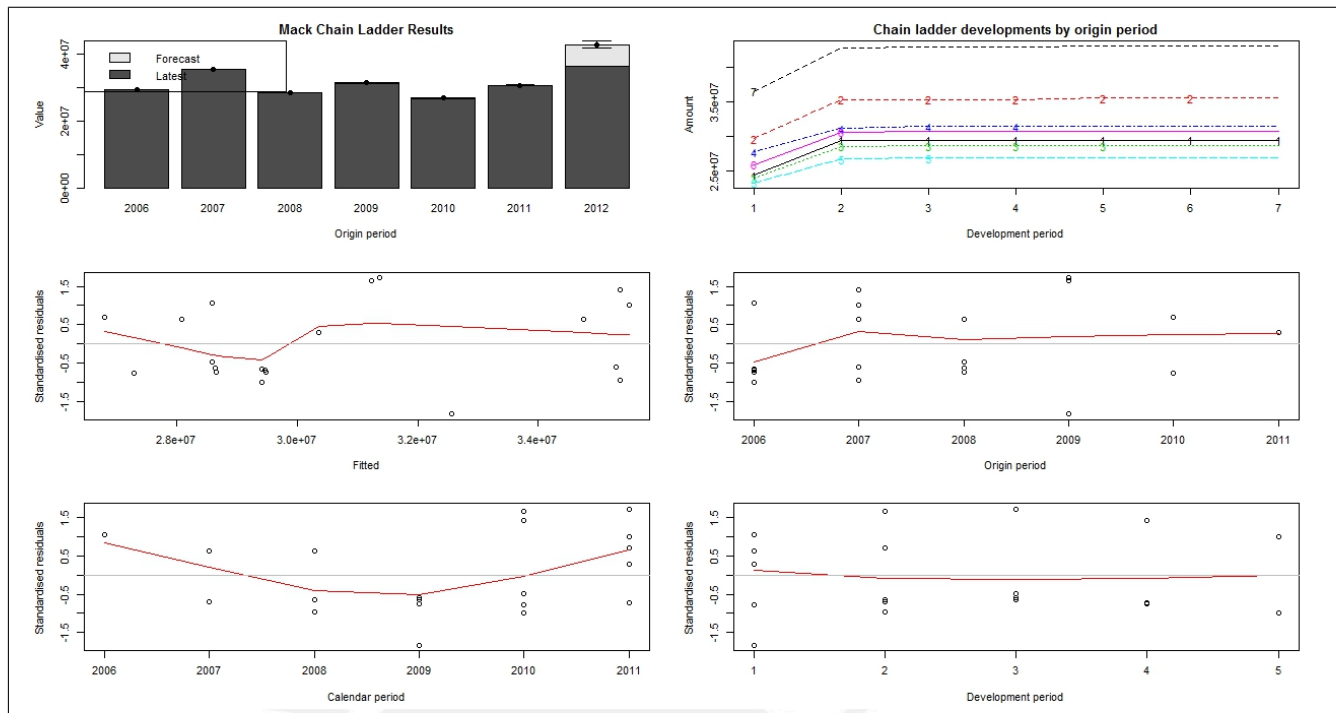


Figura 5.2: Reservas futuras y análisis de residuales para el seguro de tipo A

De igual forma, utilizando el Modelo Chain Ladder Estocástico para el seguro de tipo B, obtenemos:

Año de ocurrencia	Reserva estimada	Mack S.E.	Límite inferior	Límite superior
1 (2006)	0	0	0	0
2 (2007)	56	12	33	79
3 (2008)	112	44	25	199
4 (2009)	275	151	-21	571
5 (2010)	2,310	2,425	-2,443	7,063
6 (2011)	16,843	11,420	-5,539	39,225
7 (2012)	1'606,700	214,025	1'187,211	2'026,189
Total	1'626,296			

Cuadro 5.22: Reservas estimadas e intervalos de confianza (seguro de tipo B).

Los gráficos, donde se muestran el comportamiento de las severidades según año de desarrollo del seguro de tipo B y el análisis de residuales, se aprecian en la figura 5.2.

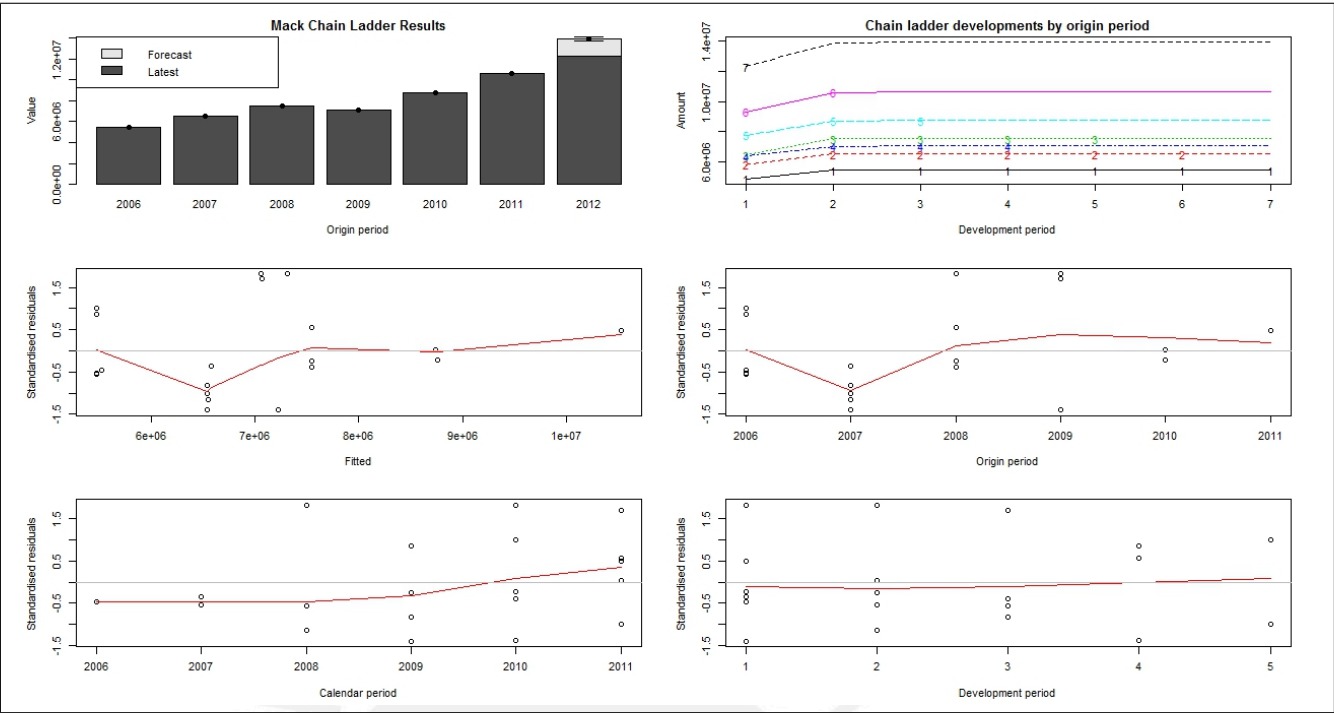


Figura 5.3: Reservas futuras y análisis de residuales para el seguro de tipo B.

Finalmente, los resultados para el seguro de tipo C muestran lo siguiente:

Año de ocurrencia	Reserva estimada	Mack S.E.	Límite inferior	Límite superior
1 (2006)	0	0	0	0
2 (2007)	406	457	-489	1,302
3 (2008)	836	748	-630	2,302
4 (2009)	1,091	792	-461	2,644
5 (2010)	4,224	7,938	-11,334	19,782
6 (2011)	10,913	13,445	-15,439	37,265
7 (2012)	1'337,291	110,591	1,120,533	1'554,049
Total	1'354,762			

Cuadro 5.23: Reservas estimadas e intervalos de confianza (seguro de tipo C).

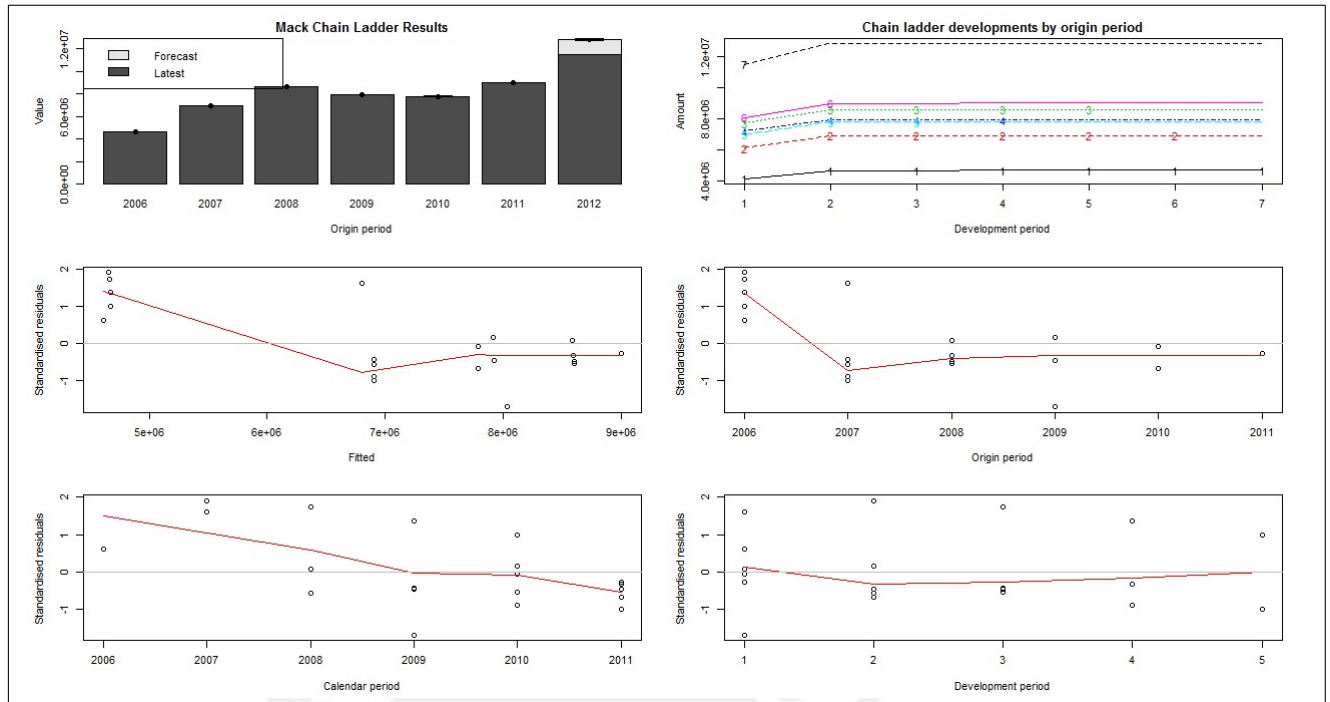


Figura 5.4: Reservas futuras y análisis de residuales para el seguro de tipo C.

Por todo lo expuesto, es posible aseverar que el método estocástico nos provee de una buena herramienta para la toma de decisiones, ya que no sólo nos muestra una estimación única de las reservas, sino que también nos provee de intervalos de confianza para la estimación de estas reservas. Los gráficos residuales permiten además verificar las asunciones establecidas para este tipo de modelos. En nuestra aplicación encontramos un ajuste razonable a estos al modelo Chain Ladder Estocástico.

5.4. El Modelo General Multivariado Chain Ladder

Aplicaremos en esta sección la metodología discutida en el capítulo anterior sobre la base de datos de SALUD. Para esto utilizaremos el módulo “MULTICHAINLADDER” del paquete “Chain Ladder” del software libre R, que se detalla en el apéndice.

Como se mencionó en la parte teórica, una motivación de los métodos multivariados de estimación de reservas es identificar y estudiar las correlaciones contemporáneas que pudieran existir entre distintos triángulos de desarrollo. En nuestro caso particular, deseamos identificar y tomar en cuenta la posible correlación que pudiera existir entre los 3 tipos de seguro de vida considerados y analizar cómo es posible utilizar dicha relación con la finalidad de buscar mejores inferencias y predicciones de las reservas. Queda claro que como nuestros 3 tipos de seguros están referidos a un mismo servicio de salud diferenciados sólo por el monto de cobertura, es esperable que exista una correlación entre sus montos de siniestros. La idea aquí será precisamente poder aprovechar estas correlaciones contemporáneas en la predicción de las reservas futuras.

Dado que contamos con información de sólo 7 años será conveniente, por razones prácticas, dividir nuestros triángulos de desarrollo en 2 partes; la primera compuesta por los años de

desarrollo 1 – 3, en la cuál se utilizará la metodología SUR, y la segunda parte constituida por los años de desarrollo 3 – 7 en donde se utilizará el método tradicional Mack Chain Ladder. Esta práctica, de “robustificar” la estimación en las colas es como lo manifiesta Zhang (2010), no sólo necesaria sino también razonable. Por lo usual la correlación en los términos de error tiende a disminuir para los últimos periodos, lo cual resta eficiencia al método multivariado. Además la data disponible para la aplicación del modelo GMCL es muy exigua en los últimos periodos de desarrollo, imposibilitando la obtención de los respectivos estimadores, debido a la singularidad en las matrices requeridas. En nuestro caso, por citar, solo logramos obtener estos trabajando con la partición descrita.

Es importante señalar, por otro lado, que estamos también considerando interceptos en el modelo y que para no ir en contra del principio de parsimonia, hemos optado por eliminar parámetros que sean no necesarios. Para el proceso de eliminación de parámetros nos basamos en la significancia estadística obtenida por regresión lineal múltiple realizada mediante el método paso a paso (stepwise). Vale, sin embargo, recalcar que se han mantenido, por lógica, los parámetros β_{11} , β_{22} y β_{33} independientemente de los resultados de significancia obtenidos. En resumen, el modelo GMCL en consideración vendrá dado por la ecuación (4.13), en donde para cada periodo de desarrollo j se tendrá la siguiente estructura para los interceptos y factores de desarrollo:

$$[\mathbf{A}_j, \mathbf{B}_j] = \begin{bmatrix} \beta_{10} & \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{20} & \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{30} & \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix}$$

siendo algunos de estos parámetros fijados a 0, según su no significancia estadística. Mas específicamente, en el análisis descrito hemos considerado las matrices siguientes:

$$[\mathbf{A}_1, \mathbf{B}_1] = \begin{bmatrix} \beta_{10} & \beta_{11} & 0 & \beta_{13} \\ \beta_{20} & 0 & \beta_{22} & 0 \\ \beta_{30} & 0 & 0 & \beta_{33} \end{bmatrix}$$

$$[\mathbf{A}_2, \mathbf{B}_2] = \begin{bmatrix} \beta_{10} & \beta_{11} & 0 & 0 \\ \beta_{20} & 0 & \beta_{22} & 0 \\ \beta_{30} & 0 & \beta_{32} & \beta_{33} \end{bmatrix}$$

y las estimaciones de estas mostradas en el cuadro 5.24. Note, como esperabamos, que existe aquí una alta correlación positiva entre los triángulos de desarrollo para los diferentes tipos de seguro de salud.

j	1	2	3	4	5	6
$\hat{\beta}_{10}$	2,366,695	171,604				
$\hat{\beta}_{11}$	1.174	0.997	1.0002	1.002	1.0001	1.000003
$\hat{\beta}_{12}$						
$\hat{\beta}_{13}$	-0.356					
$\hat{\beta}_{20}$	-66,436	1,692				
$\hat{\beta}_{21}$						
$\hat{\beta}_{22}$	1.138	1.001	1.0002	1.00002	1.000006	1.000009
$\hat{\beta}_{23}$						
$\hat{\beta}_{30}$	164,958	36,770				
$\hat{\beta}_{31}$						
$\hat{\beta}_{32}$		-0.0008				
$\hat{\beta}_{33}$	1.090	0.996	1.0004	1.00004	1.00004	1.00006
$\hat{\sigma}_{11}$	38,955	333.21	3.546	407.825	1.346	0.004
$\hat{\sigma}_{22}$	3,385	12.154	0.504	0.002	0.0001	0.00001
$\hat{\sigma}_{33}$	800.35	4.197	6.279	0.012	0.017	0.012
$\hat{\rho}_{12}$	0.863	0.923				
$\hat{\rho}_{13}$	0.726	0.522				
$\hat{\rho}_{23}$	0.517	0.808				

Cuadro 5.24: Parámetros estimados del modelo GMCL.

En base a estos parámetros estimados podemos entonces estimar las siguientes reservas para cada tipo de seguro:

Año de ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	24'373,002	29'383,278	29'416,770	29'417,046	29'417,795	29'417,895	29'417,983
2 (2007)	29'606,879	35'250,825	35'275,284	35'276,941	35'492,890	35'503,279	35'503,385
3 (2008)	23'931,287	28'550,105	28'585,935	28'587,648	28'588,276	28'592,896	28'592,981
4 (2009)	27'750,549	31'136,765	31'365,376	31'387,676	31'460,802	31'465,886	31'465,980
5 (2010)	23'256,730	26'732,904	26'867,355	26'872,948	26'935,556	26'939,908	26'939,989
6 (2011)	25'879,518	30'591,585	30'682,260	30'688,647	30'760,145	30'765,116	30'765,208
7 (2012)	36'427,213	41'049,793	41'112,802	41'121,360	41'217,164	41'223,824	41'223,947

Cuadro 5.25: Triángulo de desarrollo completo (seguro de tipo A).

Año de ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'888,325	5'466,004	5'469,798	5'470,214	5'470,425	5'470,480	5'470,527
2 (2007)	5'826,166	6'534,582	6'535,357	6'535,523	6'535,545	6'535,565	6'535,621
3 (2008)	6'474,334	7'535,350	7'543,552	7'544,619	7'544,857	7'544,904	7'544,969
4 (2009)	6'395,681	7'045,415	7'067,629	7'071,979	7'072,149	7'072,194	7'072,254
5 (2010)	7'757,548	8'727,167	8'738,953	8'740,923	8'741,133	8'741,188	8'741,263
6 (2011)	9'326,949	10'597,688	10'610,638	10'613,030	10'613,285	10'613,352	10'613,443
7 (2012)	12'305,256	13'938,550	13'955,049	13'958,195	13'958,531	13'958,618	13'958,738

Cuadro 5.26: Triángulo de desarrollo completo (seguro de tipo B).

Año de ocurrencia	Año de desarrollo						
	1	2	3	4	5	6	7
1 (2006)	4'136,332	4'646,954	4'661,869	4'672,314	4'672,789	4'673,189	4'673,464
2 (2007)	6'096,185	6'903,191	6'903,683	6'903,962	6'904,034	6'904,079	6'904,485
3 (2008)	7'702,008	8'594,797	8'596,853	8'597,053	8'597,323	8'597,653	8'598,159
4 (2009)	7'200,770	7'913,045	7'919,559	7'920,009	7'920,330	7'920,634	7'921,100
5 (2010)	6'980,431	7'780,098	7'780,218	7'783,369	7'783,684	7'783,984	7'784,442
6 (2011)	8'067,646	8'977,430	8'972,668	8'976,302	8'976,665	8'977,010	8'977,539
7 (2012)	11'466,011	12'668,385	12'647,338	12'652,461	12'652,973	12'653,460	12'654,204

Cuadro 5.27: Triángulo de desarrollo completo (seguro de tipo C).

y obtener de manera resumida las siguientes reservas para cada año de ocurrencia por tipo de seguro y de todo el portafolio:

Año de ocurrencia	Siniestro observado	Siniestro estimado	Reserva estimada	S.E.
1 (2006)	29'417,983	29'417,983	0	0
2 (2007)	35'503,279	35'503,386	106	590
3 (2008)	28'588,276	28'592,981	4,705	7,463
4 (2009)	31'387,676	31'465,981	78,304	131,060
5 (2010)	26'867,355	26'939,989	72,634	119,538
6 (2011)	30'591,585	30'765,208	173,623	170,686
7 (2012)	36'427,213	41'872,696	5'445,483	1'820,862
Total	218'783,367	225'677,553	5'774,856	1'854,140

Cuadro 5.28: Reservas estimadas (seguro de tipo A).

Año de ocurrencia	Siniestro observado	Siniestro estimado	Reserva estimada	S.E.
1 (2006)	5'470,527	5'470,527	0	0
2 (2007)	6'535,565	6'535,621	56	11.8
3 (2008)	7'544,857	7'544,969	112	44.3
4 (2009)	7'071,979	7'072,254	275	151.2
5 (2010)	8'738,953	8'741,263	2,310	2,425.1
6 (2011)	10'597,688	10'613,443	15,755	14,754.8
7 (2012)	12'305,256	14'096,065	1'790,809	280,582.1
Total	58'264,825	60'074,142	1'809,317	281,506.4

Cuadro 5.29: Reservas estimadas (seguro de tipo B).

Año de ocurrencia	Siniestro observado	Siniestro estimado	Reserva estimada	S.E.
1 (2006)	4'673,464	4'673,464	0	0
2 (2007)	6'904,079	6'904,485	406	457
3 (2008)	8'597,323	8'598,159	836	748
4 (2009)	7'920,009	7'921,100	1,091	792
5 (2010)	7'780,218	7'784,442	4,224	7,938
6 (2011)	8'977,430	8'977,539	109	13,021
7 (2012)	11'466,011	12'654,204	1'188,193	139,382
Total	56'318,534	57'513,394	1'194,860	141,385

Cuadro 5.30: Reservas estimadas (seguro de tipo C).

Año de ocurrencia	Siniestro observado	Siniestro estimado	Reserva estimada	S.E.
1 (2006)	39'561,974	39'561,974	0	0
2 (2007)	48'942,923	48'943,492	569	746.5
3 (2008)	44'730,456	44'736,110	5,654	7,500
4 (2009)	46'379,664	46'459,335	79,671	131,062
5 (2010)	43'386,526	43'465,694	79,168	119,826
6 (2011)	50'166,703	50'356,189	189,486	181,506
7 (2012)	60'198,480	67'836,889	7'638,409	2'244,139
Total	333'366,726	341'359,683	7'992,957	2'272,295

Cuadro 5.31: Reservas estimadas del portafolio (TIPO A + TIPO B + TIPO C).

Esto nos dice que en el año 2012 debemos prever una reserva de 341'359,683 nuevos soles para cubrir las obligaciones futuras presentes hasta el año 2018 en todo el portafolio.

5.5. Comparación de los modelos multivariados con el Método Chain Ladder univariado

Si bien es difícil mencionar qué modelo es el mejor, ya que todos presentan diferentes insumos importantes para la estimación y análisis, es posible estimar qué modelos podrían brindarnos decisiones futuras más precisas en diferentes situaciones. Para ello compararemos los residuales obtenidos de cada triángulo de desarrollo según distintos modelos. En adición al modelo GMCL con restricciones e interceptos discutido en la sección anterior, al cual referenciaremos como GMCLSALUD1, estaremos interesados en explorar las posibilidades de incluir o no estas restricciones e interceptos. Los residuales obtenidos de cada caso, junto con los del método Chain Ladder univariado, serán luego analizados y comparados, con la intención de poder obtener mayor información o herramientas que nos permitan decidir que método podría ayudarnos a tomar decisiones más precisas y certeras. La construcción de un gráfico simple como el “Boxplot” nos ayudará a poder comparar los residuales obtenidos para cada modelo.

A continuación, se muestran los boxplots resultantes para cada triángulo de desarrollo, donde los modelos a analizar serán los siguientes:

- **MackSALUD**: Se refiere al método Chain Ladder estocástico de Mack;
- **GMCLSALUD1**: Se refiere al modelo GMCL con restricciones y con interceptos;
- **GMCLSALUD2**: Se refiere al modelo GMCL con restricciones y sin interceptos;
- **GMCLSALUD3**: Se refiere al modelo GMCL sin restricciones y con interceptos;
- **GMCLSALUD4**: Se refiere al modelo GMCL sin restricciones y sin interceptos;

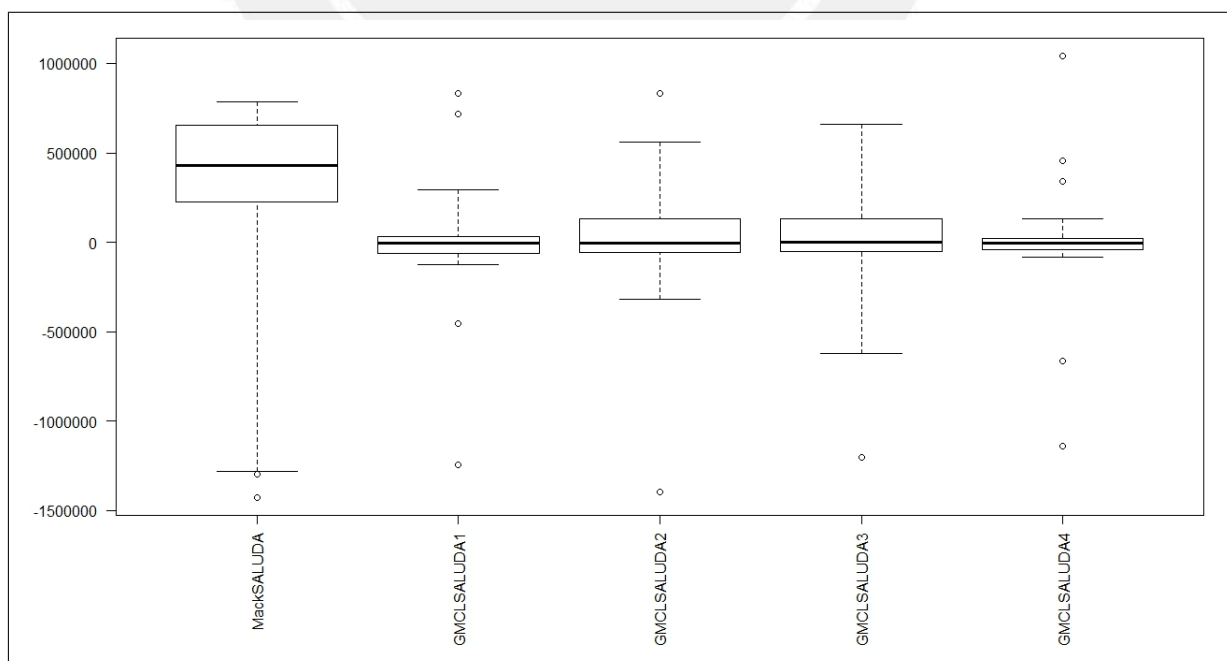


Figura 5.5: Boxplot de residuales para el seguro de tipo A, bajo los distintos modelos en análisis.

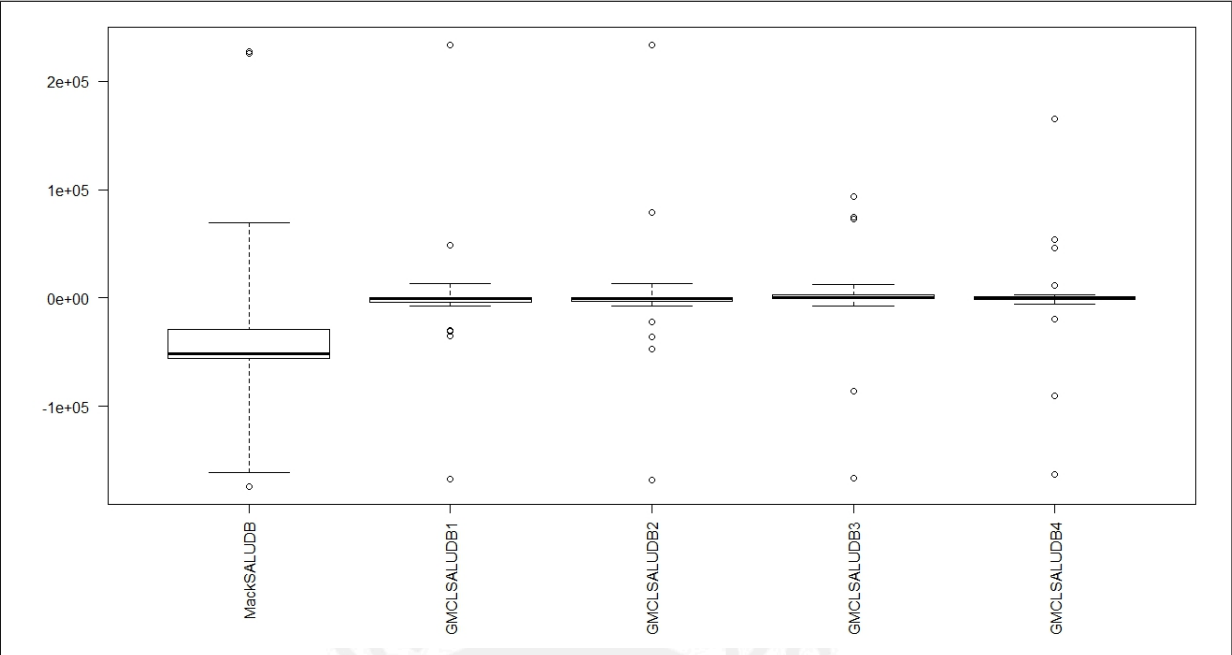


Figura 5.6: Boxplot de residuales para el seguro de tipo B, bajo los distintos modelos en análisis.

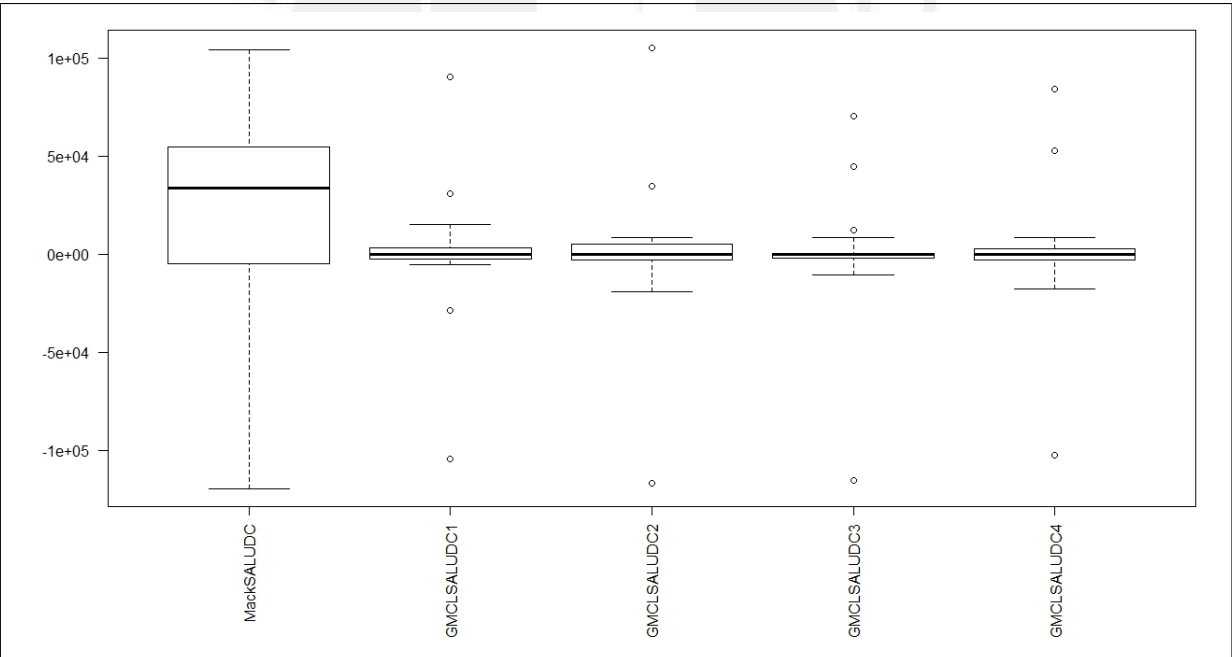


Figura 5.7: Boxplot de residuales para el seguro de tipo C, bajo los distintos modelos en análisis.

Se incluye también en el apéndice B, algunos gráficos de residuales para los distintos modelos. Estos tienen como finalidad verificar los supuestos realizados sobre los modelos de regresión lineal subyacentes a los modelos Chain Ladder utilizados. Por lo encontrado, no detectamos mayores violaciones en estos supuestos.

Como se puede observar, el método Chain Ladder univariado o Estocástico de Mack, pre-

senta uniformemente una distribución asimétrica de residuales, muy alejados en su mediana al valor teórico 0 y con una alta dispersión. Ello es un indicador que dicho modelo no nos brinda una herramienta de elección muy adecuada para el cálculo de las reservas.

De otro lado al observar los boxplots de los modelos GMCL bajo sus diferentes disgregaciones, no solo se observa una mayor simetría y menor dispersión, sino también medianas más cercanas a 0, que es lo ideal. Si bien los residuales son, en la mayoría de casos, pequeños y cercanos a cero, existen evidentemente algunos valores atípicos. Para el tipo de salud A, los residuales atípicos hacia abajo corresponden a la estimación de las reservas del año 2010 que ocurrieron en el año 2009, mientras que los outliers hacia arriba corresponden a estimaciones de reservas para el año 2009 y que ocurrieron en el año 2008. Asimismo, para el caso del seguro de tipo B sucede algo similar, los valores atípicos se observan en los años de ocurrencia del 2008 y 2009, outliers mayores a cero y outliers menores a cero, respectivamente. Finalmente, para el caso del tipo de salud C, los valores atípicos menores a cero se observan en las estimaciones de reservas para el año 2010 y que ocurrieron en el año 2009, mientras que los outliers mayores a cero, se muestran en el año de ocurrencia de 2007.

Por lo expuesto, puede decirse que el modelo GMCL nos ha brindado, en la aplicación, una herramienta en apariencia más certera para la toma de decisiones sobre las reservas futuras. Los datos históricos han mostrado un mejor ajuste a este tipo de modelos, con relación al modelo tradicional, lo cual nos hace vislumbrar que uno podría obtener estimaciones más precisas de las reservas futuras. La diferencia entre los modelos multivariados aquí considerados ha sido, por otro lado, bastante estrecha, destacándose ligeramente los modelos **GMCLSALUD1** y **GMCLSALUD3**, último que presenta menores y no tan extremos valores atípicos.

Capítulo 6

Conclusiones

6.1. Conclusiones

En este trabajo tuvimos como objetivo estudiar los métodos univariados y multivariados Chain Ladder. Partiendo del método Chain Ladder determinístico, el cual es sencillo de implementar y aplicar, mostramos como extender este al modelo Chain Ladder Estocástico propuesto por [Mack \(1993\)](#), el cual a diferencia del anterior nos permite obtener intervalos de confianza para las reservas estimadas.

Puesto que muchas compañías de seguros poseen portafolios constituidos por diferentes sub-portafolios vinculados a un mismo rubro y en otros casos existen de manera natural conexiones estructurales entre los reclamos de estos sub-portafolios, fue de interés en esta tesis estudiar algunos métodos multivariados para el cálculo de las reservas que tomen en cuenta estas relaciones. Concretamente trabajamos con el modelo General Multivariado Chain Ladder (GCML), introducido por [Zhang \(2010\)](#), el cual permite incorporar las características antes mencionadas.

El modelo GCML está basado en la teoría de las Regresiones Aparentemente no Relacionadas (SUR), la cual fue desarrollada en el capítulo 3. Si bien, como se vio en el desarrollo de ese capítulo, el modelo GCML se basa en estimaciones óptimas de los factores de desarrollo, estas deben de sopesarse con la pérdida de eficiencia que uno comete al tener que introducir una estimación de la matriz de varianzas-covarianzas en la formulación de estas estimaciones. Ello hace de que uno no pueda concluir necesariamente que los métodos multivariados funcionen mejor que su contraparte univariada; sin embargo, diversos estudios han mostrado una ganancia en eficiencia cuando los términos de correlación entre los montos acumulados de los siniestros son relativamente grandes.

Otro aspecto relevante de esta investigación, fue la de aplicar los distintos modelos Chain Ladder aquí detallados a una base de datos de seguros del ámbito local. Para ello seleccionamos un portafolio de una compañía de seguros nacional constituido por tres sub-portafolios asociados a tres tipos de seguros de salud, los cuales se diferenciaban básicamente en sus montos de cobertura. Con ello construimos tres triángulos de desarrollo, encontrando, como intuimos, correlaciones relativamente altas entre los montos acumulados de sus reclamos. Mas aún, las estimaciones de reservas obtenidas con el modelo GMCL mostraron, en todas sus variantes, residuales (montos de siniestros observados menos estimados) más cercanos a 0 y con una menor variabilidad que las dadas por el método clásico Chain Ladder Estocástico. Esto nos permite vislumbrar que el modelo GMCL tenderá a proveernos, en nuestra aplica-

ción, de estimaciones de reservas futuras más precisas que las brindadas por el modelo Chain Ladder tradicional.

6.2. Sugerencias para investigaciones futuras

En este trabajo se ha estudiado el modelo General Multivariado Chain Ladder. Este, como su nombre lo indica y [Zhang \(2010\)](#) lo demuestra, incluye como casos particulares a distintos modelos previamente estudiados en la literatura, como los modelos multivariados de [Prohl and Schmidt \(2005\)](#), [Merz and Wuthrich \(2008\)](#) y el modelo estructural bivariado Munich Chain Ladder de [Quarg and Mack \(2004\)](#), entre otros. Sería por tanto de interés comparar la eficiencia de estos modelos con los aquí desarrollados.

En nuestra aplicación sólo explotamos, por las características de nuestro portafolio, las correlaciones contemporáneas existentes entre los triángulos de desarrollo, mas no conexiones de tipo estructural que pudieran existir, por ejemplo, cuando se trabajan con reclamos pagados e incurridos. Sería por tanto interesante encontrar otros ejemplos que involucren ambas características comentadas y ver como estas se ven potenciadas por la aplicación del modelo GMCL.

Nuestro trabajo, de otro lado, estuvo concentrado en la estimación de las reservas y salvo en el caso univariado, no mencionamos detalles de cómo obtener intervalos de confianza para las estimaciones bajo el modelo multivariado. La obtención de los Errores Cuadráticos Medios (MSE) es aún un problema abierto para el modelo GMCL. Por lo que hemos investigado sólo existen algunos desarrollos para casos particulares de este modelo, como los dados por Merz y *Wüthrich* y también por *Pröhl* y Schmidt.

Para finalizar, sería interesante poder realizar un estudio más exhaustivo de los distintos estimadores para las matrices de varianza-covarianza propuestas en el contexto de las regresiones aparentemente no relacionadas. Nosotros en nuestra aplicación sólo utilizamos el estimador $\hat{\Sigma}_g$ descrito en el capítulo 3.

Apéndice A

Programas en R

En este capítulo se presentarán los códigos utilizados para estimar los métodos Chain Ladder estocástico y el modelo General Multivariado Chain Ladder; para tal efecto se utilizó el software R programa.

A.1. Código utilizado para el Método Chain Ladder Estocástico

Para estimar las reservas, los errores cuadráticos medios y sus respectivos intervalos de confianza, se utilizó el módulo “MACKCHAINLADDER” del software libre R.

El código utilizado para la estimación de las reservas basándonos en la data por cada tipo de seguro es:

Tipo de Salud A: SALUDA

```
A<-MackChainLadder(Triangle=SALUDA,est.sigma="Mack")
```

```
A
```

```
A$f
```

```
A$FullTriangle
```

```
A$Models
```

```
A$f.se
```

```
A$F.se
```

```
A$sigma
```

```
A$Mack.ProcessRisk
```

```
A$Mack.ParameterRisk
```

```
A$Mack.S.E
```

```
A$Total.Mack.S.E
```

```
plot(A)
```

```
plot(SALUDA)
```

Tipo de Salud B: SALUDB

```
B<-MackChainLadder(Triangle=SALUDB,est.sigma="Mack")
```

```
B
```

```
B$f
```

```
B$FullTriangle
```

```
B$Models
```

```
B$f.se
```

APÉNDICE A. PROGRAMAS EN R

```
B$F.se
B$sigma
B$Mack.ProcessRisk
B$Mack.ParameterRisk
B$Mack.S.E
B$Total.Mack.S.E
plot(B)
plot(SALUDB)
```

Tipo de Salud C: SALUDC

```
C<-MackChainLadder(Triangle=SALUDC,est.sigma="Mack")
C
C$f
C$FullTriangle
C$Models
C$f.se
C$F.se
C$sigma
C$Mack.ProcessRisk
C$Mack.ParameterRisk
C$Mack.S.E
C$Total.Mack.S.E
plot(C)
plot(SALUDC)
```

A.2. Código utilizado para el Modelo General Multivariado Chain Ladder

El código utilizado para la estimación de las reservas bajo el GMCL se basó en el módulo de "MULTICHAINLADDER" del paquete "Chain Ladder" del software libre R program, el cual se detalla a continuación:

```
SALUD=list(SALUDA,SALUDB,SALUDC)
SALUD
SALUD[1]
SALUD[2]
SALUD[3]
da1<-SALUD[,1:3]
da2<-SALUD[,3:7]
cofr1<-matrix(0,12,7)
pos1<-cbind(c(1:2,4:5,7,9,12),1:7)
cofr2<-matrix(0,12,7)
pos2<-cbind(c(1:2,5,7,9,11,12),1:7)
cofr1[pos1]<-1
```

APÉNDICE A. PROGRAMAS EN R

```

coefr2[pos2] <- - 1
int <- 1:2
restrict.regMat <- c(list(coefr1, coefr2, rep(list(NULL), 4)))
fit1 <- MultiChainLadder(da1, "SUR", int = int, restrict.regMat = restrict.regMat, model
= "GMCL")
fit2 <- MultiChainLadder(da2, "OLS")
fit <- Join2Fits(fit1, fit2)
pred <- predict(fit)
mse <- Mse(fit, pred)
fit.sel <- JoinFitMse(fit, mse)

```

Para la estimación de parámetros, resúmenes estadísticos y gráficos residuales, se utilizan los siguientes códigos:

```

fit.sel@coefficients
fit.sel@residCov
summary(fit.sel)@residCor
summary(fit.sel, portfolio = "1+2+3")@report.summary[[4]]
par(mfrow = c(2, 3))
plot(fit.sel, which.plot = 3:4)
r <- summary(fit.sel)@rstandard
par(mfrow = c(1,3))
for (i in 1:3) {
  hist(r[, i], 20, freq = FALSE, xlim = c(-2, 2),
  xlab = "Standardized Residuals",
  main = paste("Histogram for Triangle", i))
  lines(density(r[, i]))
}

```

Apéndice B

Residuales de los distintos modelos

En este capítulo se mostrarán los gráficos residuales de las distintas variantes del modelo GMCL que se ha trabajado en el capítulo 5, con la finalidad de constatar la no violación de los supuestos del modelo lineal.

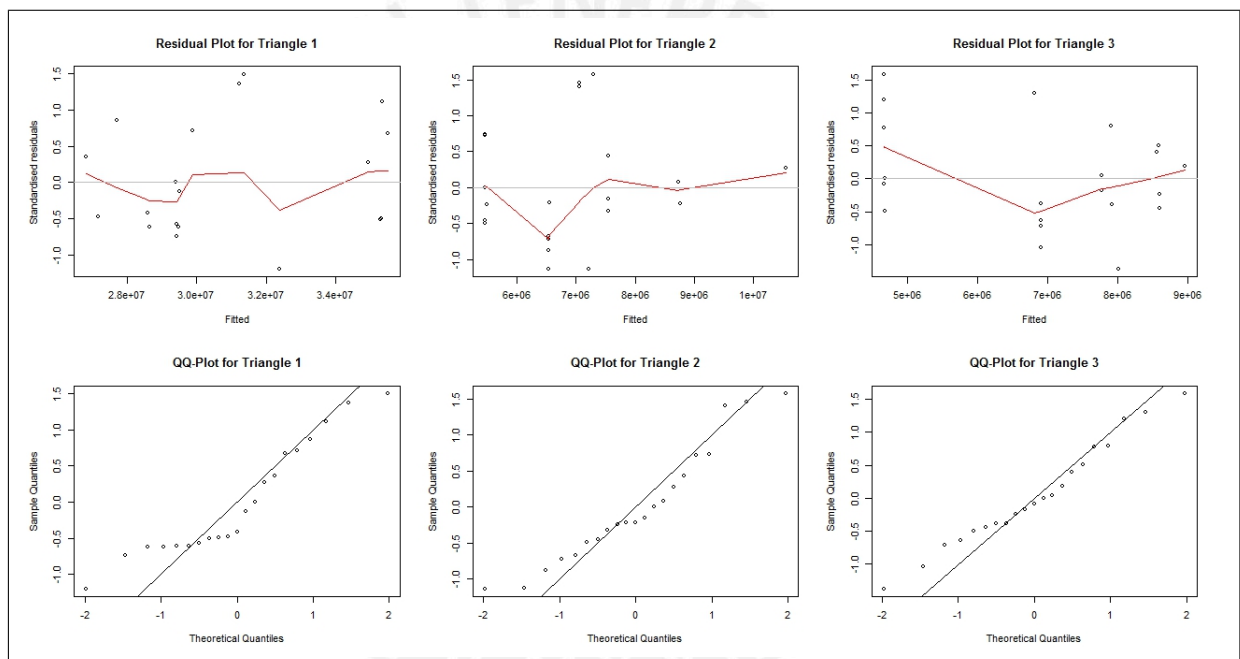


Figura B.1: Gráfico de los residuales del modelo GMCL con restricciones y con intercepto, por tipo de seguro

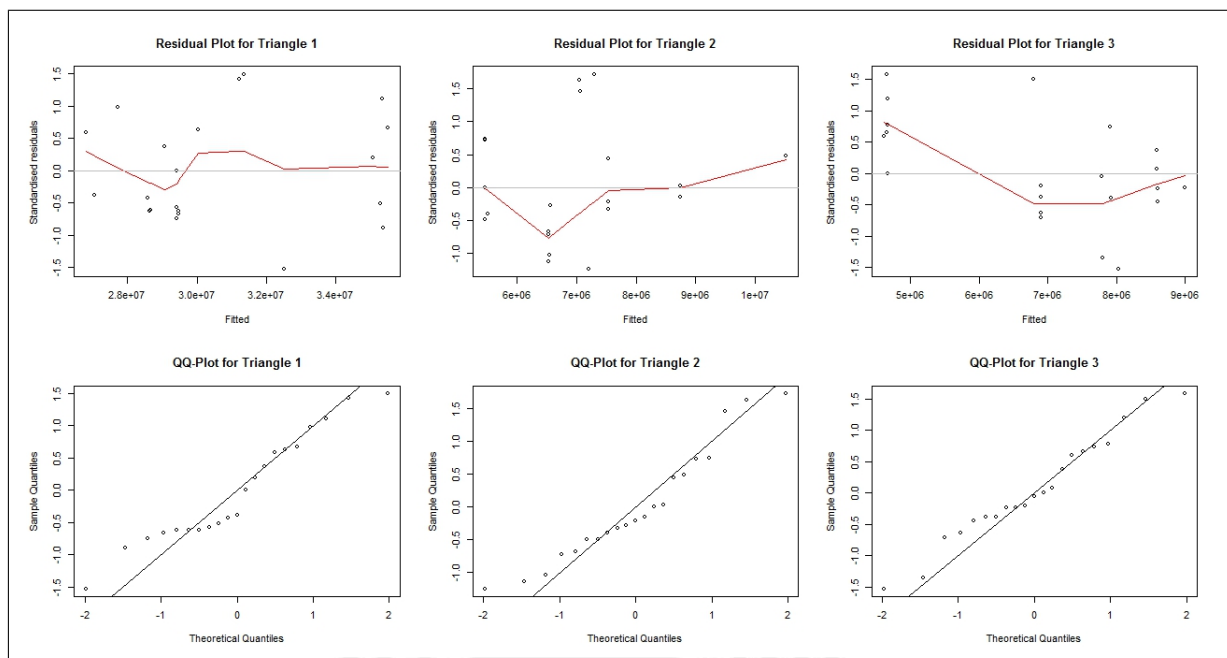


Figura B.2: Gráfico de los residuales del modelo GMCL con restricciones y sin intercepto, por tipo de seguro

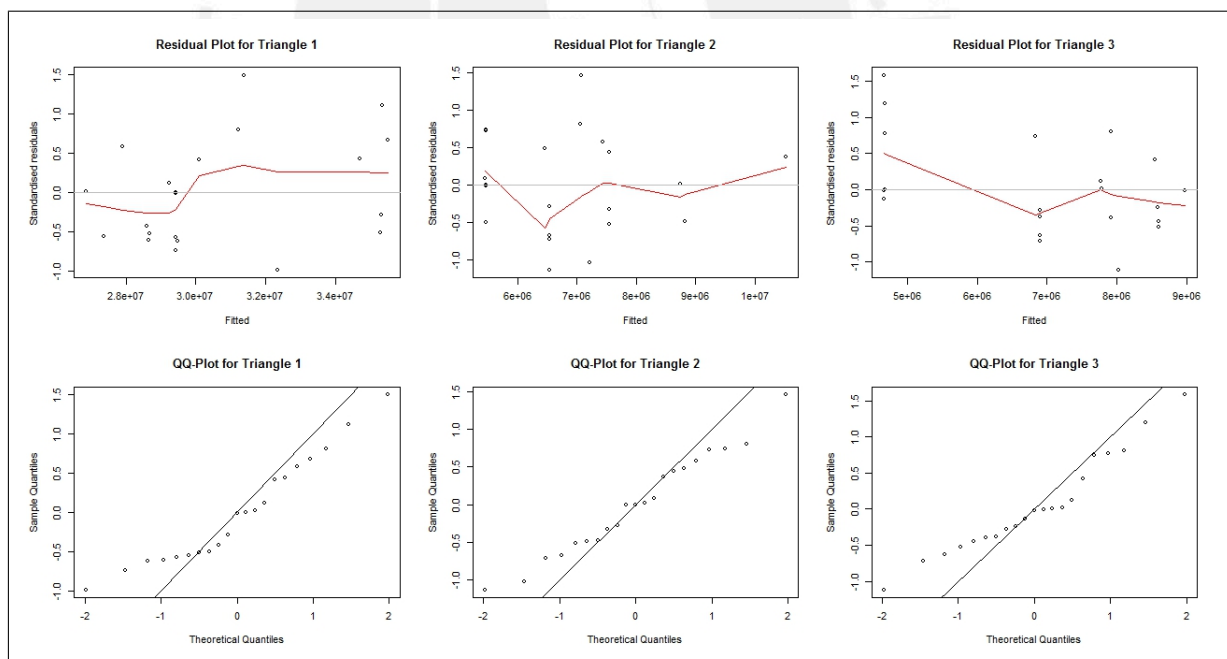


Figura B.3: Gráfico de los residuales del modelo GMCL sin restricciones y con intercepto, por tipo de seguro

APÉNDICE B. RESIDUALES DE LOS DISTINTOS MODELOS

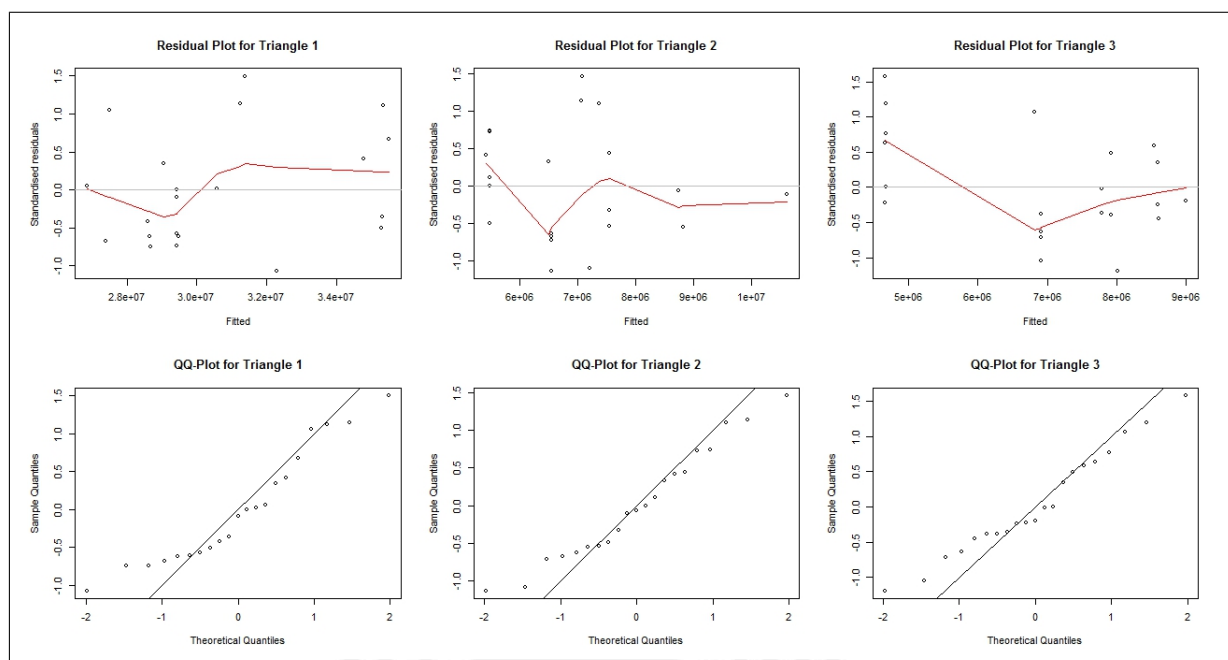


Figura B.4: Gráfico de los residuales del modelo GMCL sin restricciones y sin intercepto, por tipo de seguro

Bibliografía

- Araya, R. and Muñoz, J. (1996). *Regresiones que aparentemente no están relacionadas*, Banco Central de Costa Rica.
- Barnett, G. and Zehnworth, B. (2000). Best estimates for reserves., *Proceedings of the Casualty Actuarial Society* **LXXXVII**: 245–321.
- Beens, F., Bui, L., Collings, S. and Gill, A. (2010). *Stochastic reserving using Bayesian models - can it add value?*, Institute of Actuaries of Australia.
- Boland, P. (2007). *Statistical and Probabilistic Methods in Actuarial Science*, Chapman and Hall.
- De Alba, E. (2002). Bayesian estimation of outstanding claims reserves., *North American Actuarial Journal* **4**: 1–20.
- Jedlicka, P. (2006). *Recent Development in Claims Reserving*, Charles University.
- Kakwani, N. (1967). The unbiasedness of zellner's seemingly unrelated regressions equation estimators., *Journal of the American Statistical Association* **62**: 141–142.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates., *Astin Bulletin* **23**: 213–225.
- Mack, T. (1997). Measuring the variability of chain ladder reserve estimates., *Casualty Actuarial Society* **23**: 101–182.
- Mack, T. and Venter, G. (2000). A comparison of stochastic models that reproduce chain ladder reserve estimates, *Insurance: Mathematics and Economics* **26**: 101–107.
- Merz, M. and Wuthrich, M. (2008). Prediction error of multivariate chain ladder reserving method, *North American Actuarial journal* **12**: 175–197.
- Mikosch, T. H. (2009). *Non-Life Insurance Mathematics: An Introduction with the Poisson Process*, Springer.
- Milliken, A. and Albohali, M. (1984). On necessary and sufficient conditions for ordinary least squares estimators to be best linear unbiased estimators, *Journal of the American Statistical Association* **38**: 298–299.
- Ntzoufras, I. and Dellaportas, P. (2002). Bayesian modelling of outstanding liabilities incorporating claim count uncertainty., *North American Actuarial Journal* **1**: 113–128.

- Ordoñez, E. (2010). Estimación de intervalo de confianza para reservas no reportadas en compañía aseguradoras. Tesis de Maestría en Estadística. PUCP.
- Prohl, C. and Schmidt, K. (2005). *Multivariate chain-ladder*, Dresdner Schriften zur Versicherungsmathematik.
- Quarg, G. and Mack, T. (2004). Munich chain ladder: A reserving method that reduces the gap between ibnr projections based on paid losses and ibnr projections based on incurred losses, *Casualty Actuarial Society* **2**: 266–299.
- Srivastava, V. and Giles, D. (1987). *Seemingly Unrelated Regressions Equations Models*, Marcel Dekker Inc.
- Vegas, J. (1995). Análisis metodológico de los métodos estadísticos en el cálculo de las reservas o provisiones técnicas de prestaciones en los seguros de no vida, *Anales del Instituto de Actuarios Españoles, Tercera Época* **1**: 163–199.
- Verral, J. and England, J. (1996). *Stochastic Claims Reserving in General Insurance*, British Actuarial Journal.
- Wuthrich, M. V. and Merz, M. (2006). *Stochastic Claims Reserving Methods in Non-Life Insurance*, University Tübingen.
- Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests of aggregation bias, *Journal of the American Statistical Association* **57**: 348–368.
- Zellner, A. (1963). Estimators for seemingly unrelated regressions equations: some exact finite sample results., *Journal of the American Statistical Association* **58**: 977–992.
- Zellner, A. and Huang, D. (1962). Further properties of efficient estimators for seemingly unrelated regression equations., *International Economic Review* **3**: 300–313.
- Zhang, Y. (2010). A general multivariate chain ladder model, *Mathematics and Economics* **46**: 588–599.