



## Modelos de riesgo colectivo en tiempo discreto

Santiago Giraldo Henao

Profesor:  
Norman Diego Giraldo Gomez

Econometría Financiera  
Trabajo 1  
Departamento de Estadística  
Universidad Nacional de Colombia

MARZO 2024

# 1 Introducción

El objetivo del presente trabajo se centra en la aplicación del modelo de riesgo colectivo en tiempo discreto con base en el proceso INAR(1) propuesto por Shi & Wang (2014).

El modelo propuesto sigue el valor de la reserva de una empresa prestadora de servicios (EPS) para cubrir el sueldo de los trabajadores incapacitados de una empresa. Según la ley, las empresas en Colombia pueden pedir a las EPS el reembolso del valor correspondiente al 50% del salario de los trabajadores que se incapaciten entre 3 y 90 días por enfermedad general.

Siguiendo el trabajo de Shi y Wang, el modelo se define por

$$U_n = U_0 + Cn - \sum_{j=1}^{N_n} Y_j, \quad n = 1, 2, \dots \quad (1)$$

donde

- $U_0$  es la reserva inicial.
- $C$  es una constante correspondiente a la prima pagada por la empresa en cada periodo por el servicio de la EPS.
- La serie  $\sum_{j=1}^{N_n} Y_j$  corresponde al valor agregado de los pagos realizados a las empresa, por concepto de salario de trabajadores incapacitados hasta el periodo  $n$ . Siendo  $N_n$  el número de trabajadores incapacitados hasta el periodo  $n$ , y  $Y_j$  el pago individual que genera cada incapacidad.

Siguiendo este modelo, el propósito del trabajo será encontrar valores  $U_0$ , tal que la probabilidad de insolvencia de la reserva se alinee a las políticas de la EPS. Para este fin, se tomará la probabilidad de insolvencia definida en Shi & Wang (2014) como

$$\phi(U_0) = P\left(\bigcup_{n=1}^{\infty} (U_n < 0)\right) \quad (2)$$

En este sentido, el presente trabajo constará de tres partes: El análisis y ajuste de la serie de conteo, para el número de casos de incapacidades, basado en el modelo INAR(1) + Poisson. El análisis y ajuste de la serie de pagos  $Y_j$  basado en el modelo GAR(1) con distribución invariante Gamma. Y, por último, el cálculo de la probabilidad de insolvencia teórica y simulada para un  $U_0$  dado.

Los datos con los que se trabajará corresponden a una serie de pagos realizados entre los años 2000 y 2014 que se asumirán como la totalidad de pagos reembolsados por la EPS  $X$  a una empresa específica  $Y$  a la que presta sus servicios.

## 2 Serie de Conteo

Tal como se explica en la introducción, el modelo sigue de la siguiente forma el agregado de pagos desembolsados hasta el periodo  $n$ :

$$\sum_{j=1}^{N_n} Y_j \quad (3)$$

En esta sección nos ocuparemos de la serie  $N_n$ , que definimos como el número de trabajadores incapacitados hasta el periodo  $n$ . Dada por

$$N_n = \sum_{j=1}^n X_j \quad (4)$$

$X_j$  representa el número de casos en cada periodo  $n$ . Que se comportan de la siguiente manera:

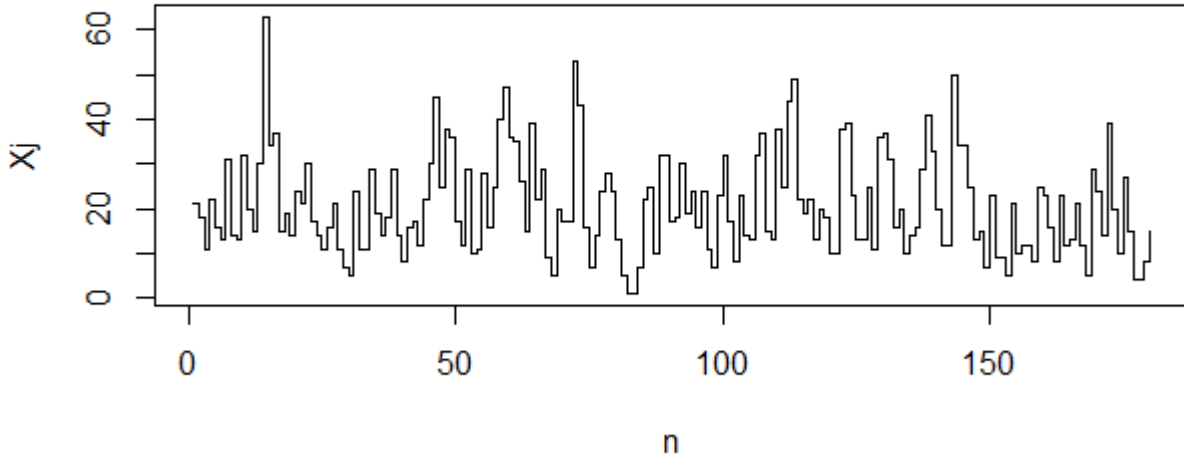


Figure 1: Número de casos por periodo ( $X_j$ )

El número de casos en cada periodo sigue un modelo INAR(1) + Poisson que depende de, por un lado, la cantidad de nuevas incapacidades en cada periodo. Definida por una sucesión  $Z_j$  iid  $\sim \text{Poisson}(\lambda)$ . Y, por otro lado, de la cantidad de incapacidades en el periodo anterior que continuarán incapacitados en este periodo, definido como  $\alpha \circ X_{j-1}$ , donde  $\alpha$  es la probabilidad que tiene un trabajador de continuar incapacitado un periodo más.

$$X_j = \alpha \circ X_{j-1} + Z_j \quad (5)$$

Por medio de una regresión lineal podemos estimar los parametros  $\mu$ ,  $\alpha$ . Y comparar la media obtenida de los datos con la media teórica, siguiendo:

$$E(X_n) = \frac{\mu}{(1 - \alpha)} \quad (6)$$

$$E(X_n | X_{n-1}) = \alpha X_{n-1} + \mu \quad (7)$$

Se obtiene:

$\mu$	$\alpha$
13.12403	0.36449

Table 1: Parámetros estimados usando regresión lineal.

media obtenida con los datos	media teórica
20.67222	20.65117

Table 2: Comparación de medias.

Podemos observar que las medias son similares, por lo que concluimos que este es un buen método de estimación.

Usamos la librería de R "gamlss" para encontrar las 5 distribuciones que mejor se ajustan a  $X_n$ . El resultado:

NBII	NBI	GPO	ZANBI	NBF
1353.645	1353.645	1354.814	1355.407	1355.645

Table 3: Distribuciones con mejor ajuste.

Podemos observar que las mejores dsitribuciones según "gamlss" son: Binomial Negativa de segundo grado, Binomial Negativa de primer grado y GPO (Dirtribución Poisson Generalizada). Probaremos el ajuste con Binomial Negativa y con GPO para comparar.

Para el caso de GPO, la media estimada y la media obtenida de los datos son muy similares (20.672219 y 20.67222 respectivamente). Además, gráficamente se ve relativamente un buen ajuste:

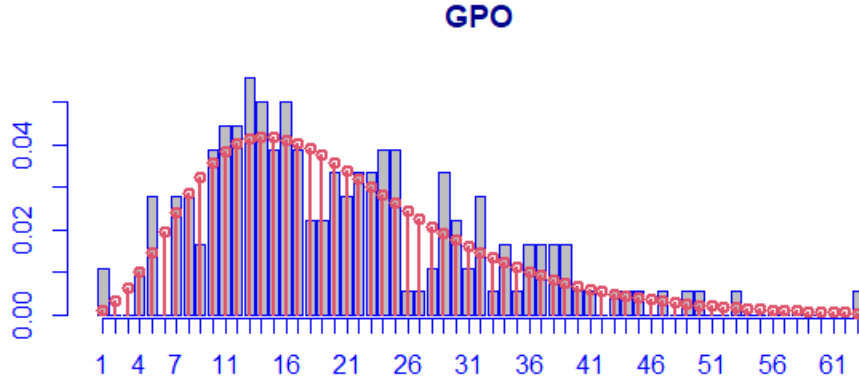


Figure 2: Ajuste GPO.

Para el caso de la Binomial Negativa, podemos intuir que ajusta bien ya que en la prueba de hipótesis el valor-p es de 0.5396. Por lo que no se rechaza la hipótesis nula, indicando que los datos pueden seguir una distribución Binomial Negativa.

Comparamos gráficamente ambos ajustes:

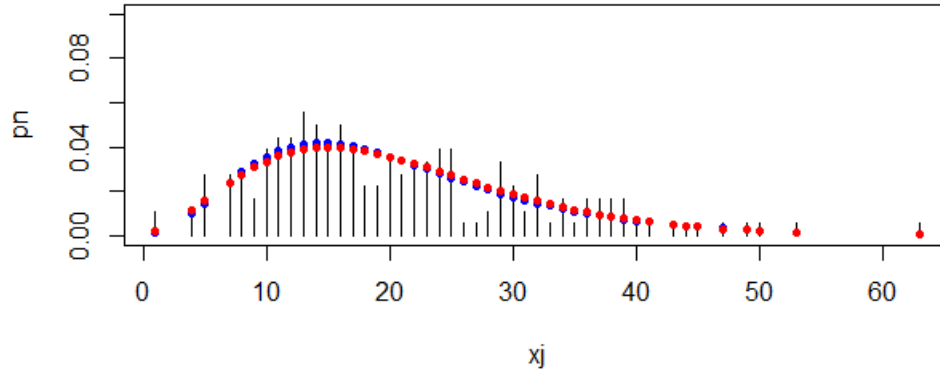


Figure 3: Comparación de ajuste GPO(Azul) y BN(Rojo).

Gráficamente es difícil concluir, ya que ambos parecen ajustar relativamente bien a los datos. En este sentido, podemos simular y comparar versus los valores observados.

Se simula  $X_j = \text{INAR}(1) + \text{Poisson}$  y  $X_j = \text{INAR}(1) + \text{GPO}$  y se comparan con los valores observados:

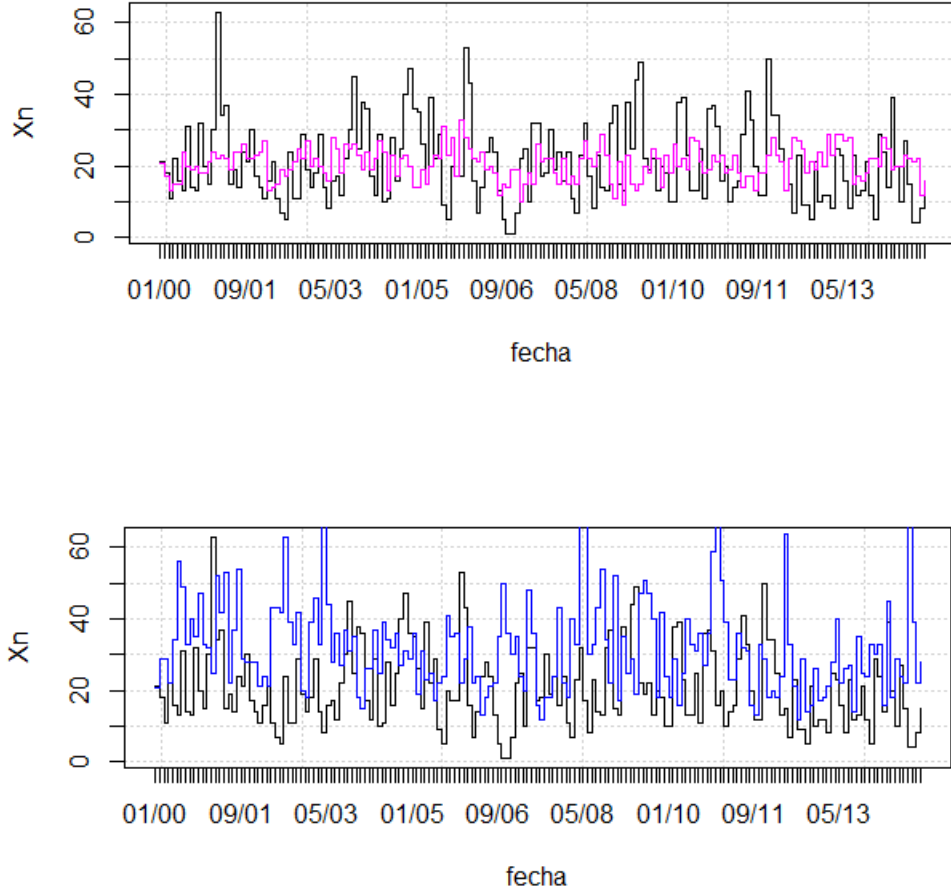


Figure 4: Comparación de Ajuste INAR-BN(Arriba) vs INAR-GPO(Abajo)

Con las gráficas simuladas podemos observar un mejor comportamiento del modelo INAR + BN ya que presenta menor volatilidad y mejor reversion a la media. El modelo INAR + GPO, sobrestima generalmente los valores y llega a valores muy extremos.

### 3 Modelo de Costos

Con el modelo de conteo simulado, pasamos a analizar el modelo de costos dado por

$$S_n = \sum_{j=1}^{X_n} Y_j, \quad n = 1, 2, \dots \quad (8)$$

donde  $S_n$  representa los costos agregados en el período  $n$  y  $Y_j$  los costos individuales de cada trabajador incapacitado.

Los datos se comportan de la siguiente manera:

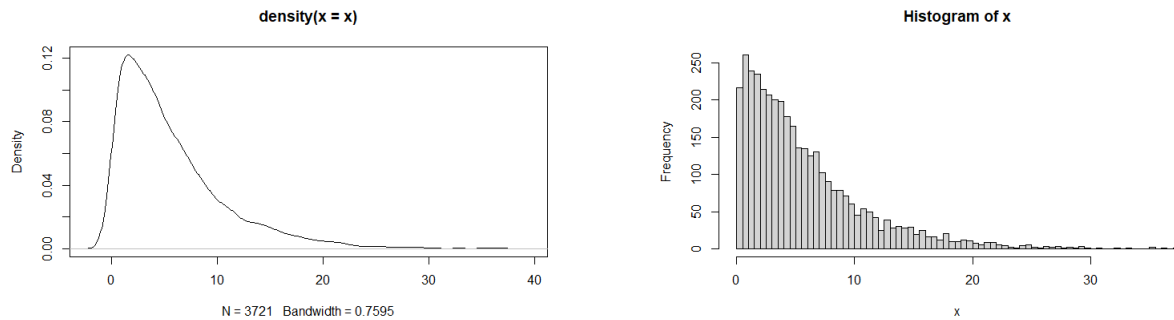


Figure 5: Gráfico de Densidad e Histograma de los datos

Observamos en los datos una distribución asimétrica con una cola muy pesada a la derecha. Estos valores extremos observados corresponden a pagos muy altos que se deben, en este caso, a incapacidades de personal ejecutivo con altos salarios. Para este caso, asumiremos que la EPS se reasegura con las incapacidades médicas del personas ejecutivo de la empresa.

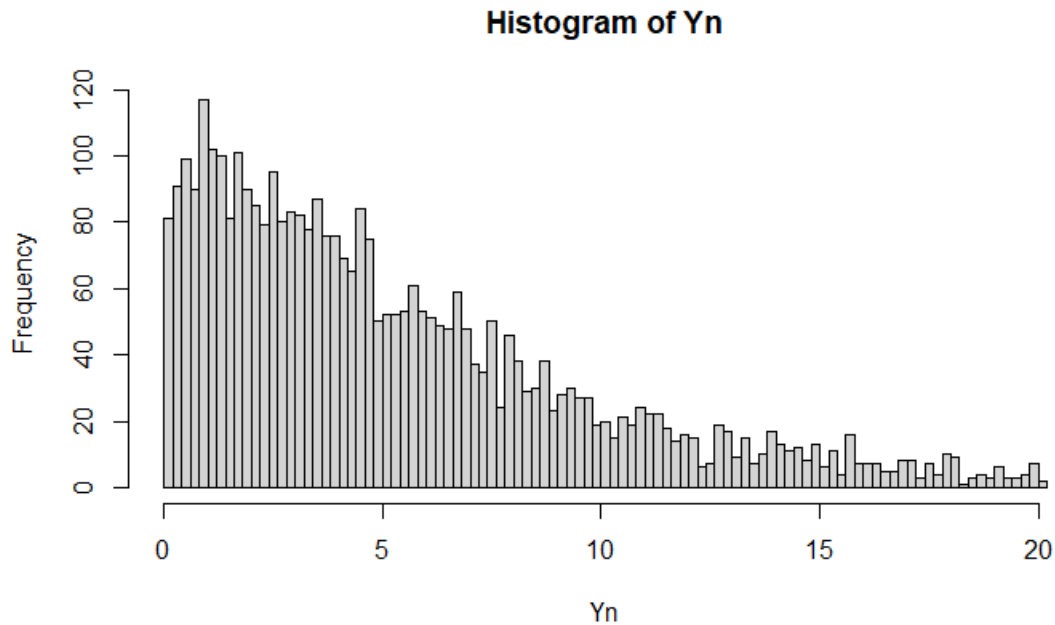


Figure 6: Histograma de los datos.

En total se excluye aproximadamente el 2% de los datos.

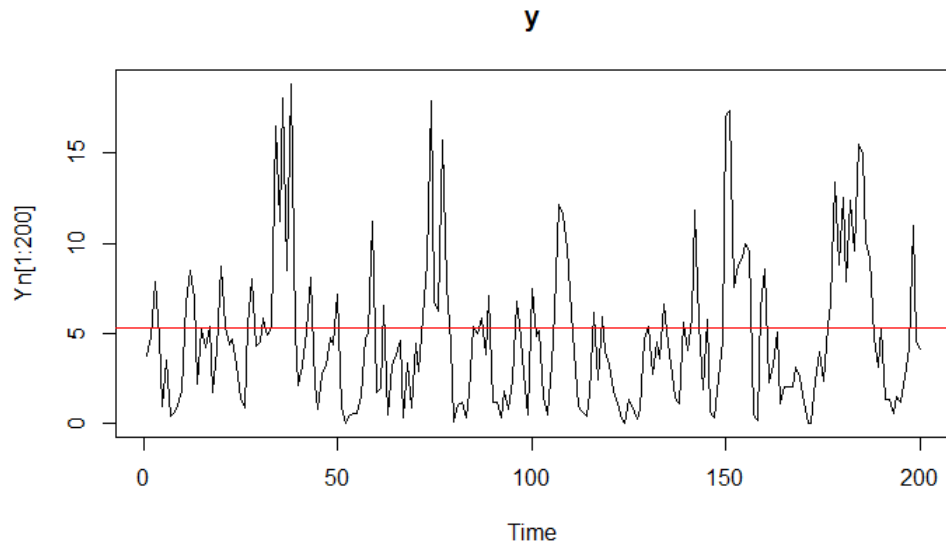


Figure 7:  $Y_n$  primeros 200 períodos

En la Figura 7 se puede observar que el proceso fluctúa al rededor de la media. Además, los datos presentan líneas relativamente suavizadas, lo que indica que puede existir autocorrelación entre los datos. Para obtener mas información sobre esto se grafica la función de autocorrelación.

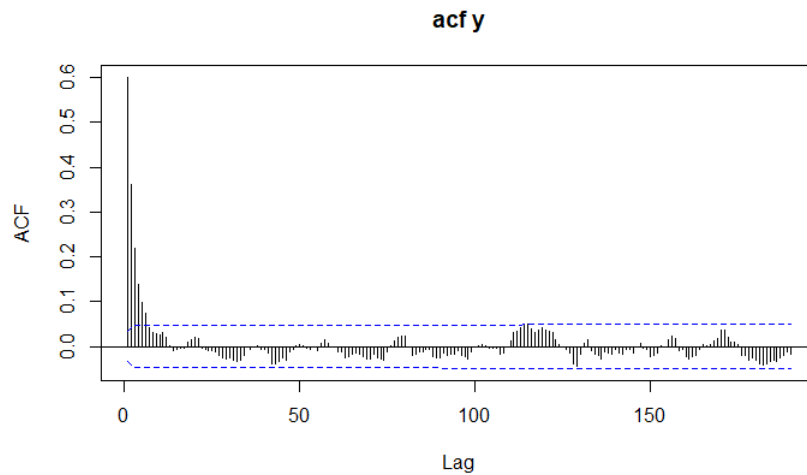


Figure 8: Función de Autocorrelación.

Observamos que efectivamente existe autocorrelación en el proceso, pero esta rápidamente decae a 0. Lo que nos indica que las observaciones en diferentes puntos del tiempo están debilmente correlacionadas y el proceso puede tener estacionariedad débil, consistente con un proceso gaussiano.



Creemos que un modelo GAR(1) con distribución gamma invariante puede ajustar bien a los datos. Para esto estimamos los parámetros  $\alpha$ ,  $\beta$ ,  $\nu$ .

$\alpha$	$\beta$	$\nu$
0.7672598	0.6294715	1.487587

Table 4: Parámetros GAR(1)

Luego estimamos los parámetros de la distribución Gamma invariante

$$shape = \nu \quad (9)$$

$$rate = \alpha(1 - \beta) \quad (10)$$

$shape$	$rate$
1.487587	0.28429

Table 5: Parámetros distribución Gamma invariante

Luego tenemos que

$$E(Y_n) = \frac{\nu}{\alpha(1 - \beta)} \quad (11)$$

$$Var(Y_n) = \frac{\nu}{\alpha^2(1 - \beta)^2} \quad (12)$$

media obtenida con los datos	media teórica
5.2832	5.232609

Table 6: Comparación de medias.

Varianza obtenida con los datos	Varianza teórica
18.76341	18.40578

Table 7: Comparación de varianzas

Observamos que la media y varianza obtenida con los datos son similares. Procedemos entonces a simular con GAR(1) para revisar el ajuste del modelo.

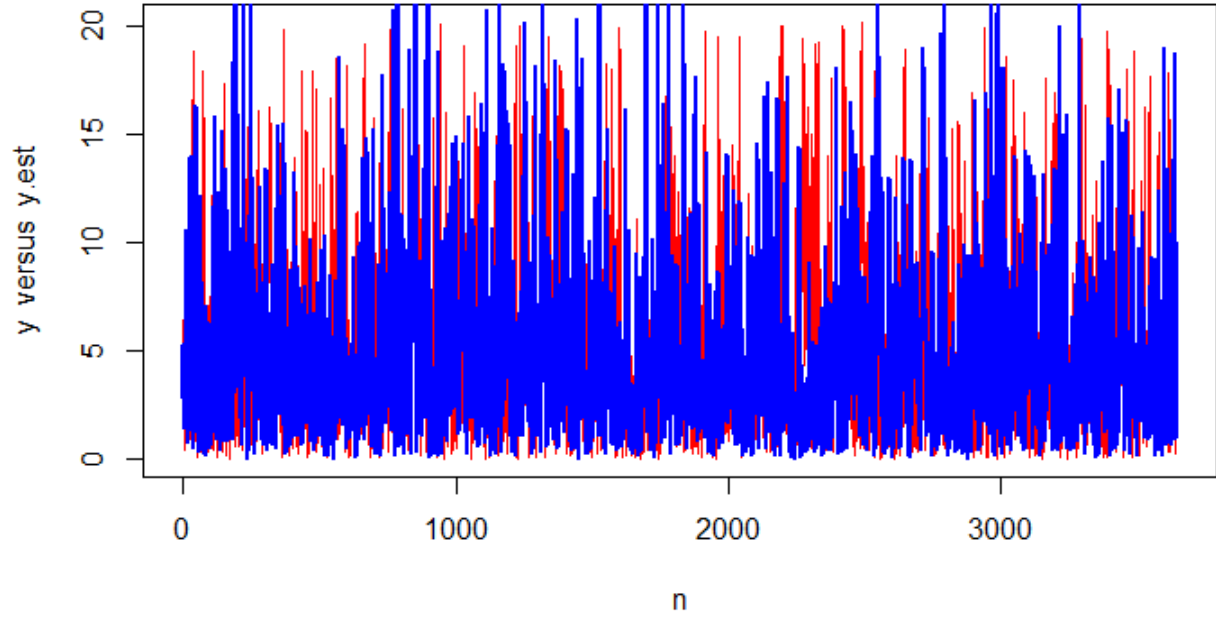


Figure 9: GAR(1) simulado(azul) sobre valores reales(rojo)

De lo que podemos observar en la Figura 10, es posible afirmar que el modelo GAR(1) con distribución invariante gamma, sigue los datos relativamente bien, así como la media y la varianza de la serie. Por lo que concluimos que es un modelo adecuado para continuar con el cálculo de probabilidad de insolvencia. No obstante, antes de pasar a esto, proponemos una alternativa al modelo GAR(1) para los costos.

## 4 Alternativa al modelo GAR(1)

Con la librería "gamlss" de R encontramos la distribución que mejor se ajusta a los datos, lo que da como resultado la distribución Gamma Generalizada(GG).

GG
19265.85

Table 8: Distribución Gamma Generalizada.

Procedemos entonces a ajustar esta distribución a los datos y a comparar los modelos.

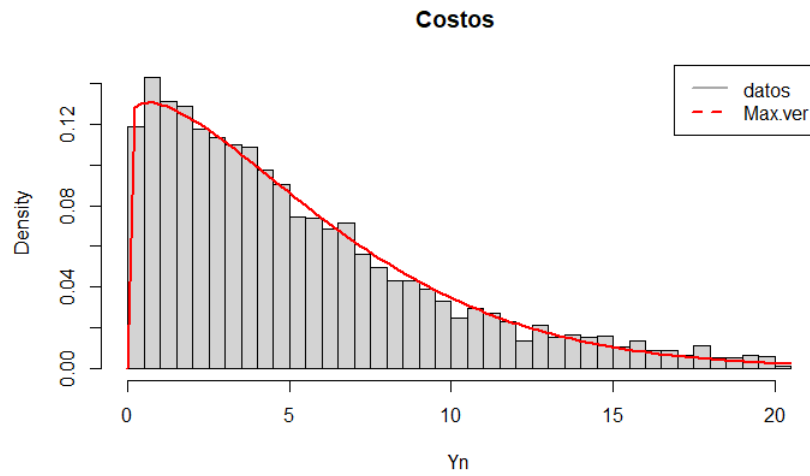


Figure 10: Ajuste GG

Podemos observar que la distribución GG ajusta bien a los datos gráficamente, incluso comparado con el modelo GAR(1) con distribución gamma invariante. Figura 12.

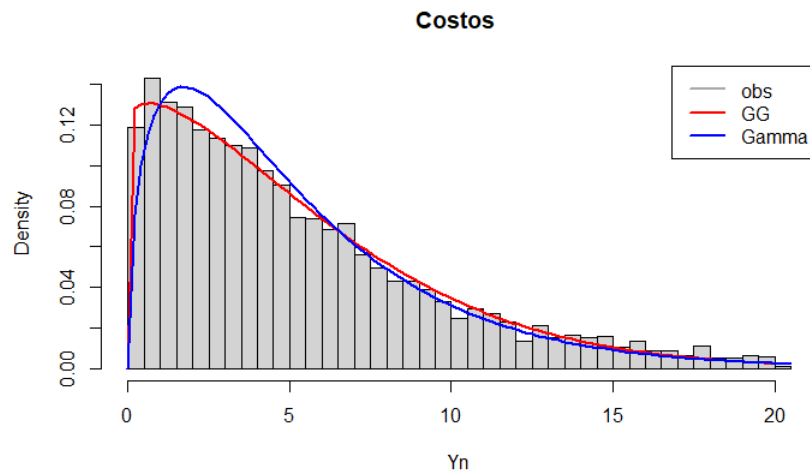


Figure 11: GAR(1) vs GG

Gráficamente es claro que la distribución Gamma invariante sigue mejor los datos que el modelo GAR(1). Para explorar un poco más esto realizamos la comparación de las fda con fda empírica.

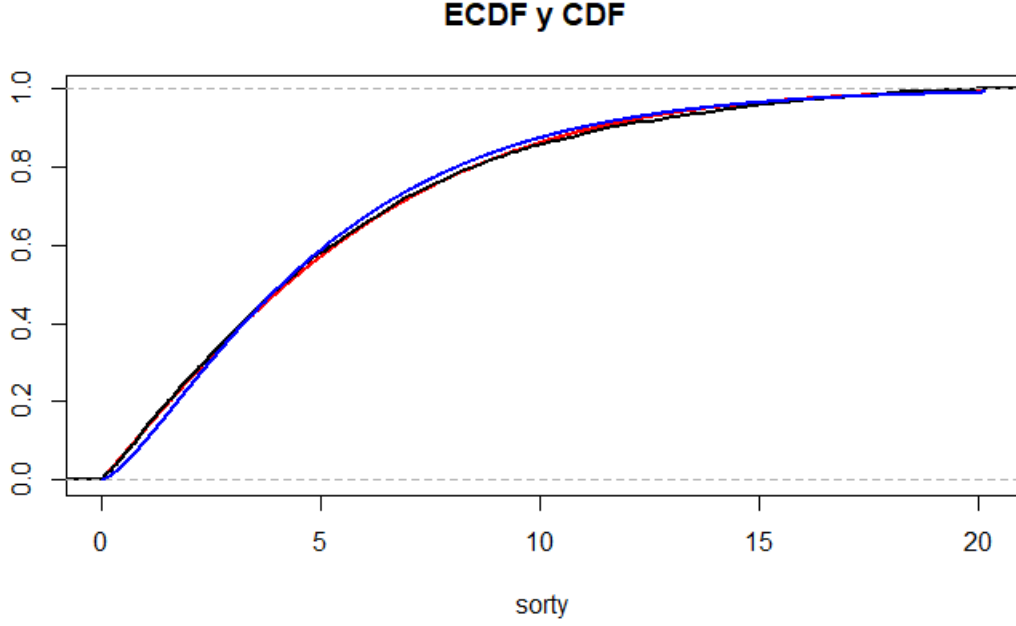


Figure 12: Comparación Fda

A pesar de que las trayectorias son similares, es posible concluir que la distribución GG sigue mejor los datos, por lo que sería valioso realizar el cálculo de las probabilidades de insolvencia modificando el modelo de costos con esta distribución. Dejaremos esta posibilidad para después. Por ahora, continuaremos con el modelo original explicado en la introducción.

## 5 Cálculo de probabilidad de insolvencia

Para el cálculo de la probabilidad de insolvencia, dado por la ecuación 2, asumimos los costos distribuidos GAR(1) con distribución invariante gamma. Mediante la desigualdad de Lundberg se tiene que

$$\phi(U_0) \leq \exp -kU_0 \quad (13)$$

donde el coeficiente de ajuste  $k > 0$  es una constante positiva que es solución de la ecuación

$$\frac{\lambda(1 - \alpha)M_y(k)}{1 - \alpha M_y(k)} = \lambda + Ck \quad (14)$$

Entonces, calculamos la función generadora de momentos de  $Y_n$

$$M_y(t) = \frac{1}{(\frac{1-t}{rate})^\nu}, \quad t < rate \quad (15)$$

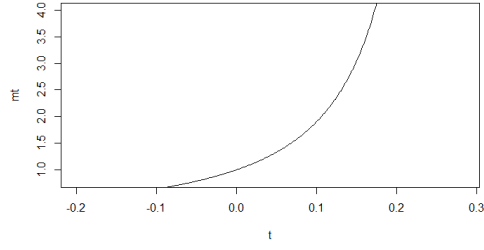


Figure 13: Fgm  $Y_n$

Y la prima  $C$  como

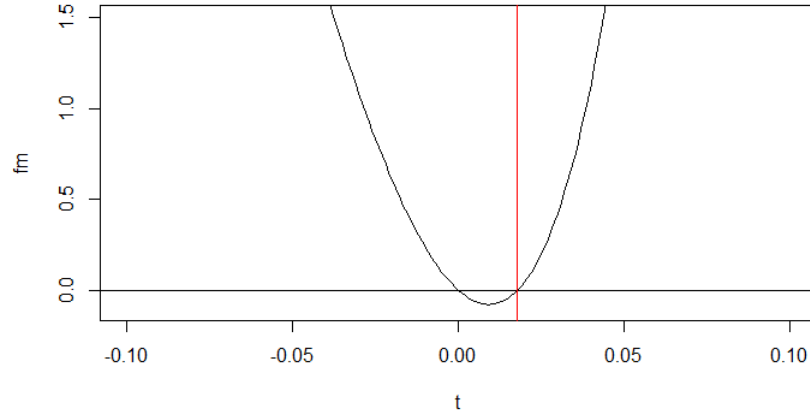
$$C = (1 + \theta) * EX_n * EY_n \quad (16)$$

donde  $\theta = 0.1$

$C$
118.9866

Table 9: Prima  $C$ .

Ahora calculamos  $k$  resolviendo la ecuación 14.



$k$
0.01247723

Table 10: Valor de  $k$ .

Con esto calculado ahora procederemos a calcular las reservas a partir de las probabilidades de insolvencias correspondientes a 0.1%, 1%, 2%, 3%, 4%.

Reservas	Prob. Insolvencias
553.6287	0.001
369.0858	0.01
313.5329	0.02
281.0365	0.03
257.9799	0.04

Table 11: Reservas iniciales por probabilidad de insolvencia.

Podemos observar que a partir de una reserva inicial de 369 (Cifra en Millones de COP), la probabilidad de insolvencia del fondo es baja (1%). Lo cual es muy valioso para la EPS que podría tomar desiciones basada en la política de riesgo de la compañía.

Un último análisis sería simular un número relativamente alto de trayectorias con el fin de calcular la probabilidad de insolvencia. Usaremos el valor de la reserva inicial con probabilidad de 1%.

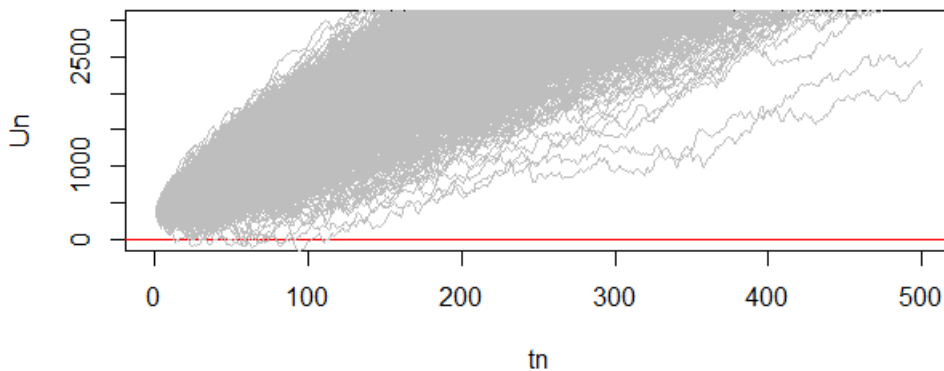


Figure 14: Simulación de 1000 trayectorias

Con 1000 trayectorias, la probabilidad de insolvencia simulada nos da 0.007. Relativamente cercana a la probabilidad calculada para esta reserva inicial de 0.01. Esto podría mejorar al aumentar el número de trayectorias simuladas, por ejemplo, a 10.000.

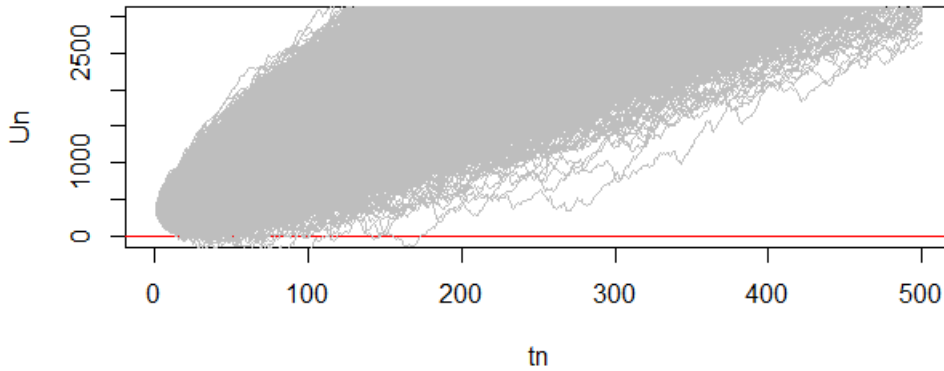


Figure 15: Simulación de 10000 trayectorias

Con la simulación de 10.000 trayectorias la probabilidad de insolvencia nos da 0.0048. Claramente se acerca menos al valor calculado de 0.01. En este punto sería importante revisar el modelo para identificar alternativas que nos permitan obtener mejores resultados. Lo que haremos será ahora usar el modelo Gamma Generalizado (GG) analizado anteriormente para los costos.

## 6 Análisis adicional de probabilidad de insolvencia

Usando los mismos datos del punto anterior y unicamente modificando el modelo GAR(1) por el modelo GG obtenemos los siguientes resultados:

Valor de $k$	Valor de $C$
0.01049007	173.0938

Table 12: Valores actualizados

Reservas	Prob. Insolvencias
658.5044	0.001
439.0029	0.01
372.9264	0.02
334.2741	0.03
306.8499	0.04

Table 13: Reservas iniciales por probabilidad de insolvencia.

Antes de sacar conclusiones apresuradas simulamos con 10000 trayectorias.

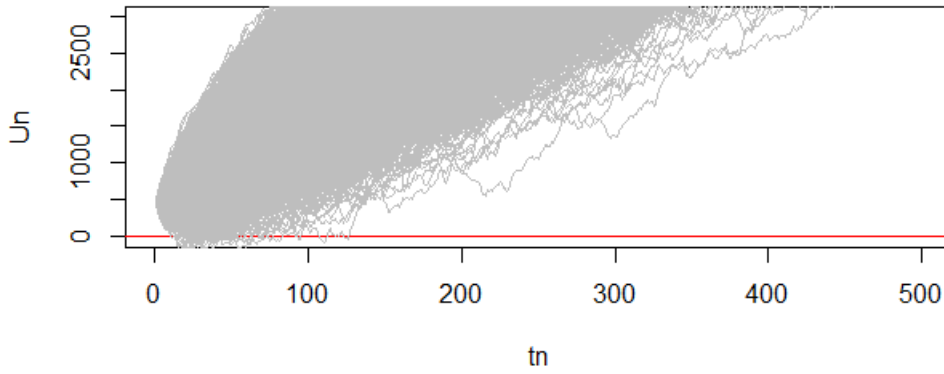


Figure 16: Simulación de 10000 trayectorias

El resultado de la probabilidad de insolvencia mediante la simulación de 10000 trayectorias nos da 0.0042. Lo cual no es muy distinto del resultado obtenido con el modelo original. Por lo que no podemos concluir que esta alternativa sea más apropiada.

## References

Shi, Haifang, & Wang, Dehui. 2014. An Approximation Model of the Collective Risk Model with INAR(1) Claim Process. *Communications in Statistics*, **43:24**.