
Pronóstico de la tasa de siniestralidad de un seguro de desempleo en Colombia mediante modelos lineales dinámicos.

Forecast of the loss rate of an unemployment insurance in Colombia using dynamic linear models

Erika Valeria Rivera Jimenez.^a
erikarivera@usantotomas.edu.co

Wilmer Dario Pineda Rios.^b
wpinedar@unal.edu.co

Resumen

El desempleo es una de las variables más analizadas y esperadas mensualmente, ya que implica decisiones en política pública que impactan la producción y el modelo de crecimiento del país, de ahí la importancia que las aseguradoras ofrezcan un seguro de desempleo que cubra esta contingencia, por tal motivo, este estudio, propone implementar métodos estadísticos, para generar un pronóstico de la tasa de siniestralidad de un seguro de desempleo en Colombia, describiendo el nivel de exposición que tendrá el producto de desempleo y así contribuir a que la compañía tenga la capacidad de afrontar y mitigar una posible crisis. Dentro de los métodos clásicos de pronóstico, se usó los modelos lineales dinámicos, el error cuadrático medio (RMSE) y el Rhat, utilizados para comparación, selección del mejor modelo y verificación de convergencia de las cadenas; además, el backtesting determinó la eficiencia del modelo en un escenario atípico

Palabras clave: Modelos dinámicos de Poisson, variables económicas, crisis económica.

Abstract

Unemployment is one of the most analyzed and expected monthly variables since it implies deciding on public policy that impact production and the country's growth model, hence the It is important that insurers offer unemployment insurance that covers this contingency, for this reason reason, this study proposes to implement statistical methods, to generate a forecast of the rate of claims of an unemployment insurance in Colombia, describing the level of exposure that will have the product of unemployment and thus contribute to companies having the capacity to face and mitigate a possible crisis. Within the classical forecasting methods, dynamic linear models were used, the root mean square error (RMSE) and the Rhat used for comparison, selection of the best model and string convergence check; In addition, the backtesting determined the efficiency of the model in an atypical scenario.

Keywords: Stress models, economic variables, economic emergency.

1. Introducción

Los riesgos que afectan al mundo han cambiado desde principios del siglo XXI. Según el World Economic Forum “el terrorismo y las pandemias son algunos riesgos de alto impacto mediático y las principales

^aEstudiante

^bDirector

amenazas para la sociedad” [Willige, 2017]. “En 2008, la crisis financiera global hizo tomar conciencia de los riesgos económicos y de la interconexión transfronteriza. Así mismo, desde 2017 los riesgos medioambientales y tecnológicos se han incluido en la lista de los cinco principales riesgos globales” [Buti, 2021]. Además, el Cambridge Global Risk Index revela “que aproximadamente una tercera parte de las pérdidas globales previstas como consecuencia de sucesos traumáticos provienen de catástrofes naturales, y alrededor de una cuarta parte se deben a exposiciones a riesgos en los mercados económicos y financieros” [Deng, 2019].

En el ámbito económico, la amenaza de crisis económica es potencialmente uno de los eventos más significativos y preocupantes en todas partes del mundo, que determinan en un largo plazo el incremento de la ocurrencia de algún evento o índice económico a causa de variables macroeconómicas. De tal manera que en un periodo de tiempo posterior a dicho evento se obtienen afectaciones simultáneas que generan desempleo, que es una de las variables más analizadas y esperadas mensualmente, ya que implica decisiones en política pública. Varios fenómenos afectan la variable de desempleo como la tasa de crecimiento de la población y la estructura económica del país.

Las recientes crisis económicas han incrementado el desempleo en Colombia, sin duda, situación preocupante para la mayoría de personas inclusive para aquellos que cuenten con un contrato de carácter indefinido. Este escenario de desempleo es difícil de sobrellevar, ya que genera un desajuste en el mercado y en las personas que lo padecen, no solo por la ausencia de ingresos que antes percibían, sino también por la incertidumbre de no saber cuánto tiempo durará y que soluciones pueden mitigar este paro. Un ejemplo claro son las medidas para reducir la propagación del Covid-19 en Colombia que posteriormente generó una crisis en la actividad económica, inclusive a nivel mundial. Un gran número de compañías suspendieron sus actividades por prevención, lo cual causó una disminución en los ingresos al punto de reducir su planta de personal o inclusive tuvieron que cerrar de forma definitiva generando un incremento en el desempleo en el país.

Por lo anterior se resalta la importancia que las aseguradoras ofrezcan un seguro de desempleo que cubra al asegurado (trabajador dependiente o independiente) en el pago de las obligaciones crediticias definidas en la póliza, en el caso en que no pueda laborar por un período de tiempo determinado, evitando quedar en mora.

El seguro para trabajador dependiente cubre el riesgo de “pérdida involuntaria del empleo cuando el asegurado quede desempleado involuntariamente durante el término de vigencia de la póliza. La aseguradora pagará la suma indicada en el certificado individual de seguro por cada período de treinta días continuos en el que el asegurado permanezca desempleado, en los siguientes casos; despido sin justa causa, finalización anticipada del contrato y terminación del contrato por mutuo consentimiento, con acuerdo ante el Ministerio de Trabajo” [SCB, 2016]. El amparo de desempleo bajo cualquiera de las causas antes descritas está ligado a los periodos de carencia y reactivación.

- “Para los trabajadores dependientes e independientes se tiene un período de carencia que se define como el período contado a partir del inicio de la vigencia de la póliza y durante el cual el asegurado no tiene derecho al amparo de desempleo, si cuenta con una enfermedad preexistente tendrá un periodo de carencia de doce meses contados a partir del inicio de vigencia de la póliza y solo tendrán cobertura a partir de la primera renovación de la póliza.
- Existe un período mínimo de reactivación en el cual el asegurado ya indemnizado, por un evento y que se ha reincorporado a su actividad económica, debe permanecer en ejercicio de la misma para tener derecho a afectar nuevamente la póliza.” [SCB, 2016]

Adicionalmente, la aseguradora no tiene que responder por ninguna indemnización si:

1. “El hecho que dio lugar a la reclamación, ocurrió como consecuencia de:
 - Guerra (Incluyendo guerra civil).

- Explosión nuclear, radiación nuclear o contaminación nuclear o cualquier riesgo proveniente de materiales nucleares o contaminación tóxica.
 - Huelgas, cierres, motines disturbios y desobediencia civil.
 - Pandemias.
 - Desastres naturales.
2. El asegurado deja de ser residente colombiano.
 3. El asegurado no está al día en las cuotas mensuales del crédito.
 4. El asegurado no acredita haber estado ejerciendo su actividad económica, durante al menos ciento ochenta días previos a la ocurrencia del siniestro.
 5. En la fecha de la solicitud de seguro, o dentro de los treinta días siguientes, existía evidencia clara que podía quedar desempleado.
 6. Licencias no remuneradas y licencia de trabajo.
 7. Renuncia y periodos de prueba.
 8. Si es empleado público y cumple la edad de retiro forzoso, fue destituido o abandonó el cargo.
 9. Fue despedido con justa causa de conformidad con la legislación laboral aplicable, el contrato de trabajo o el reglamento interno de trabajo.
 10. Si queda desempleado por la expiración del plazo del contrato a término fijo o la terminación de la tarea por la cual fue contratado.” [SCB, 2016]

Por otra parte, para los trabajadores independientes se cubre “el riesgo de no poder ejercer la actividad económica (enfermedades graves e incapacidad total o temporal) realizando un desembolso mensual con el fin de que la persona cumpla con sus obligaciones financieras” [SCB, 2016]. Cualquier pago por un siniestro está sujeto a las políticas de la compañía aseguradora y pueden modificarse según el amparo y las cláusulas de la póliza.

El objetivo de este artículo es describir el nivel de exposición que tendrá el producto de desempleo de una compañía aseguradora, realizando unos pronósticos de la tasa de siniestralidad del seguro de desempleo, para evaluar la capacidad de afrontar una crisis macroeconómica utilizando modelos dinámicos de variables económicas, teniendo en cuenta que no todos los ciudadanos colombianos compran un seguro de desempleo y que no necesariamente genera un impacto directo en la siniestralidad de una aseguradora, ya que la compañía, tiene nichos de venta, restricciones de venta y otras estrategias a la hora de minimizar riesgos catastróficos. De igual manera, es de interés:

1. Calcular el impacto de una crisis macroeconómica sobre la cartera de desempleo para detectar si hay una afectación en las actividades de la compañía.
2. Enumerar las características del riesgo en la cartera de desempleo que permita capturar la siniestralidad en escenarios adversos mediante un modelo lineal dinámico.

Para cumplir el objetivo de este trabajo se proponen métodos y modelos Bayesianos, como los MLG que fueron propuestos por Nelder y Wedderburn en 1972 con el objetivo de solucionar los supuestos que no se cumplen en los modelos lineales y en los modelos de regresión lineal múltiple por la naturaleza de la información que se quiere analizar. “Los MLG son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores (Binomiales, Poisson, Gamma, entre otras) y varianzas no constantes” [McCullagh and Nelder, 1983]. Esta clase de modelos facilita la utilización de variables respuesta numéricas o categóricas donde no necesariamente la varianza de los errores tiene un comportamiento constante en el tiempo.

Este artículo está organizado de la siguiente manera: En la sección 2, se definen los modelos lineales generalizados, sus características principales y se precisa una aplicación: los modelos de regresión en datos de conteo. En la sección 3, se complementa la introducción ofrecida acerca de los modelos lineales dinámicos donde se menciona los antecedentes y el objetivo de implementar esta teoría en el análisis de riesgos. En la sección 4, se dedica un espacio para profundizar en la teoría del riesgo catastrófico y su relación con los seguros comerciales. En la sección 5, se da un resumen que contiene los procedimientos que se implementan para el desarrollo del artículo. En la sección 6, se ejecuta la aplicación de los modelos lineales dinámicos teniendo como insumo la información del seguro de desempleo de una compañía aseguradora colombiana comprendido entre julio de 2016 y enero de 2020. Finalmente, se presentan algunas conclusiones, recomendaciones y mecanismos que puedan mitigar el riesgo a raíz de una crisis financiera mediante el uso de modelos dinámicos con argumentos bayesianos.

2. Modelos lineales generalizados (MLG)

2.1. Familia exponencial

Sea Y una variable aleatoria con una distribución de probabilidad que depende de un parámetro θ . La distribución pertenece a la familia exponencial si puede escribirse de la forma:

$$f(y; \theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

donde a, b, s y t son funciones conocidas. Además, si $a(y) = y$ se dice que la distribución tiene la **forma canónica** y $b(\theta)$ es a veces llamado el **parámetro natural**. Distribuciones como la Poisson, normal y binomial pertenecen a la familia exponencial y se pueden escribir en la forma canónica. [McCullagh and Nelder, 1983]

2.2. Modelos Lineales Generalizados

Un modelo lineal generalizado está definido en “términos de un conjunto de variables aleatorias independientes Y_1, \dots, Y_N cada una con una distribución de la familia exponencial y las siguientes propiedades:

1. La distribución de cada Y_i tiene la forma canónica y depende de un simple parámetro θ_i ” [Cayuela, 2010]. Sin embargo, no todos los θ_i 's tienen que ser el mismo.

$$f(y_i; \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]$$

2. La distribución de todos los Y_i 's es de la misma forma tal que los subíndices b, c y d no son necesarios. Por lo tanto la función de densidad de probabilidad conjunta de Y_1, \dots, Y_N es [Dobson, 2002]

$$\begin{aligned} f(y_1, \dots, y_N; \theta_1, \dots, \theta_N) &= \prod_{i=1}^N \exp [y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ &= \exp \left[\sum_{i=1}^N y_i b(\theta_i) + \sum_{i=1}^N c(\theta_i) + \sum_{i=1}^N d(y_i) \right]. \end{aligned} \quad (1)$$

3. Las componentes de un MLG son:“

- **Componente aleatoria:** Corresponde a la variable de respuesta \mathbf{Y} que sigue una distribución de la familia exponencial uniparamétrica (normal, log-normal, Poisson, gamma,...). Las observaciones $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ son independientes e idénticamente distribuidos y se denota por μ su esperanza matemática. [Agresti.A, 2015]

- **Componente sistemática:** También llamada predictor lineal, se denota por η y corresponde al vector de n componentes, siendo cada una de ellas igual a $\eta_i = \sum_{j=1}^p \beta_j x_{ji} = x'_i \beta$. [Cayuela, 2010]
- **Función de enlace:** También llamada función link, $g(\cdot)$ o función de ligadura, relaciona la esperanza matemática de la variable dependiente con el predictor lineal $\eta_i = g(\mu_i)$ donde $i = 1, \dots, n$. La función de enlace debe ser monótona y diferenciable ” [Cayuela, 2010]. En el caso particular del modelo lineal la función de enlace es la identidad; por otro lado, la función de enlace para una distribución Poisson es la función logarítmica y la función de enlace que mejor se ajusta a una binomial es la logit.

Los MLG permiten aplicar modelos con variables respuesta continuas y categóricas. Algunos ejemplos de su utilidad en la vida cotidiana son:

- Conteo de casos (número de viviendas destruidas, cantidad por especie)
- Conteo de casos expresados como proporciones (porcentaje de robos por localidad, porcentaje de embarazo adolescente, porcentaje de heridos en accidentes)
- Respuestas binarias (éxito o fracaso, mujer u hombre, infectado o no infectado)

2.3. Modelo de regresión para datos de conteo

Un modelo de regresión de Poisson se caracteriza por ser un modelo lineal generalizado (MLG) utilizado para modelar datos de conteo o tasas. Este modelo permite determinar qué variables explicativas X tienen un efecto en la variable respuesta Y (conteo). La variable respuesta sigue la distribución de Poisson, es decir, la probabilidad de que ocurran eventos dentro de un periodo de tiempo determinado, con datos discretos y valores enteros no negativos; además, esta sesgada según los valores del parámetro lambda y la varianza es igual que la media; sin embargo, cuando no se cumple este último supuesto, es decir, hay sobredispersión de los datos, se puede utilizar un modelo de regresión Binomial Negativa. Por otra parte, para transformar la relación no lineal que se puede presentar entre la variable respuesta y las variables explicativas en un modelo de regresión Poisson se utiliza la función de enlace logaritmo, ya que transforma la relación no lineal en forma lineal.

La ecuación general de un modelo de regresión Poisson es:

$$\log(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

donde,

- y es la variable respuesta.
- β_0 y β son coeficientes numéricos, en particular, β_0 es la intersección con el eje y.
- X es la variable predictora o explicativa.

Los coeficientes del modelo se calculan utilizando la función de verosimilitud que se conocen por cumplir las siguientes características: asintóticamente normales, centrados y su matriz de varianzas-covarianzas es la inversa de la matriz de información.

Cuando se quiere modelar tasas la forma matemática general del modelo de regresión de Poisson es:

$$\log\left(\frac{Y}{n}\right) = \beta_0 + \sum \beta_i X_i$$

que se puede reescribir como:

$$\log(Y) - \log(n) = \beta_0 + \sum \beta_i X_i$$

$$\log(Y) = \log(n) + \beta_0 + \sum \beta_i X_i$$

Por lo anterior, los modelos donde la variable respuesta es una tasa incluyen el término $\log(n)$ con un coeficiente igual a 1, conocido como compensación.

3. Modelo lineal dinámico

Un modelo lineal dinámico hace parte de una clase de modelos que se caracterizan por tener parámetros variables en el tiempo, su principal utilidad radica en el modelamiento de series de tiempo y regresiones. El dinamismo de estos modelos permite que en escenarios con cambios externos los parámetros se actualizan gradualmente. Fue introducido por Harrison y Stevens en 1976. Estos modelos están compuestos por un par de ecuaciones, llamadas ecuación de observación y ecuación del sistema (o ecuación de evolución de parámetros). La ecuación de observación describe la relación entre la observación y los regresores, la tendencia, la estacionalidad y otros componentes que toman la forma de una regresión multivariante. Por otro lado, la ecuación del sistema describe la evolución del vector de coeficientes de regresión o parámetros de estado t a través del tiempo.

“Un modelo lineal dinámico se define por una cuádrupla $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$. Para cada t se tendrá una ecuación observacional y una ecuación del sistema respectivamente [Bermúdez, 2012]:

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{F}'_t \theta_t + \nu_t, & \nu_t &\sim N(0, \mathbf{V}_t) \\ \theta_t &= \mathbf{G}_t \theta_{t-1} + \omega_t, & \omega_t &\sim N(0, \mathbf{W}_t), \end{aligned} \quad (2)$$

Suponga para $t = 1, \dots$, que \mathbf{Y}_t es un vector columna de r observaciones. Además:

- \mathbf{F}_t es una matriz de regresión dinámica conocida (de orden $(n \times r)$)
- \mathbf{G}_t es una matriz de estado conocida (de orden $(n \times n)$)
- \mathbf{V}_t es una matriz de varianza observacional conocida (de orden $(r \times r)$)
- \mathbf{W}_t es una matriz de evolución conocida (de orden $(n \times n)$)
- θ_t es un vector de parámetros (de orden $(n \times 1)$)

y en las que las sucesiones de errores ν_t y ω_t son independientes y mutuamente independientes, respectivamente” [Bermúdez, 2012].

3.1. Modelos lineales dinámicos normales univariados y uniparamétricos

Los modelos lineales dinámicos normales univariados y uniparamétricos representados por la cuádrupla $\{\mathbf{F}_t, \lambda, \mathbf{V}_t, \mathbf{W}_t\}$ establecen que para todo t se tiene:“

- Ecuación de observación: $\mathbf{Y}_t = \mathbf{F}_t \mu_t + \nu_t$, con $\nu_t \sim N(0, \mathbf{V}_t)$
- Ecuación del sistema: $\mu_t = \lambda \mu_{t-1} + \omega_t$, con $\omega_t \sim N(0, \mathbf{W}_t)$

- Información inicial: $(\mu_0|D_0) \sim N(m_0, C_0)$

Los errores secuenciales ν_t y ω_t son independientes y mutuamente independientes. Además, son independientes de $(\mu_0|D_0)$. Los valores de las secuencias de varianza \mathbf{V}_t y \mathbf{W}_t pueden ser desconocidas pero la constante λ y los valores relevantes de la secuencia \mathbf{F}_t son conocidos” [West and Harrison, 1997].

3.2. Modelos polinomiales de primer orden

Los modelos polinomiales de primer orden se conocen como los modelos más simples cuando se habla de series temporales. Son una aproximación de una serie de Taylor suavizada en el tiempo de primer orden, también conocida como la tendencia de la serie. Estos modelos de primer orden se definen a partir de la cuádrupla $\{1, 1, \mathbf{V}_t, \mathbf{W}_t\}$, caracterizado por las siguientes ecuaciones: “

- Ecuación de observación: $\mathbf{Y}_t = \mu_t + \nu_t$, con $\nu_t \sim N(0, \mathbf{V}_t)$
- Ecuación del sistema: $\mu_t = \mu_{t-1} + \omega_t$, con $\omega_t \sim N(0, \mathbf{W}_t)$
- Información inicial: $(\mu_0|D_0) \sim N(m_0, C_0)$.

donde los errores secuenciales ν_t y ω_t son mutuamente excluyentes e independientes. Además, son independientes de $(\mu_0|D_0)$ ” [Bermúdez, 2012].

La información inicial es la representación probabilística de la información del pronosticador, creencias sobre el nivel μ_0 en el tiempo $t = 0$. “La media m_0 es una estimación puntual de este nivel, y la varianza C_0 mide la incertidumbre asociada. Cada conjunto de información D_v comprende toda la información disponible en el momento v , incluyendo D_0 , los valores de las varianzas $\{V_t, W_t : t > 0\}$ y los valores de las observaciones Y_v, Y_{v-1}, \dots, Y_1 . Por lo tanto, la única información nueva que llega a ser disponible en cualquier momento t es el valor observado Y_t , de modo que $D_t = \{Y_t, D_{t-1}\}$ [West and Harrison, 1997]. Adicionalmente, μ_t es unidimensional y describe la tendencia del proceso” [Bermúdez, 2012].

3.3. Modelo constante

Un modelo constante se caracteriza por tener una varianza observacional y una varianza del sistema constantes en el tiempo. Es un caso particular de los modelos polinomiales de primer orden y se define a partir de la cuádrupla $\{1, 1, V, W\}$ y definido como: “

- Ecuación de observación: $\mathbf{Y}_t = \mu_t + \nu_t$, con $\nu_t \sim N(0, \mathbf{V})$
- Ecuación del sistema: $\mu_t = \mu_{t-1} + \omega_t$, con $\omega_t \sim N(0, \mathbf{W})$
- Información inicial: $(\mu_0|D_0) \sim N(m_0, C_0)$.” [West and Harrison, 1997]

Una opción que se usa a menudo en la práctica para analizar series de tiempo es la clase de modelos lineales dinámicos, donde la observación depende de un conjunto de componentes y los componentes evolucionan de forma independiente a lo largo del tiempo. Generalmente, asumen observaciones normales, pero hay una generalización a la familia exponencial formalizada por West, Harrison y Migon en 1985, denominados modelos lineales dinámicos generalizados.

3.4. Modelos lineales dinámicos generalizados

Estos modelos son otro caso particular de la clase de modelos dinámicos y una generalización de los modelos lineales generalizados cuando las observaciones son series de tiempo y se permite que los parámetros varíen con el tiempo y las distribuciones muestrales no tienen que ser normales. Estos modelos son cada vez más utilizados en diferentes áreas tales como epidemiología, econometría y marketing.

Quizás un primer intento de analizar el conteo de series de tiempo fue a través del uso de la transformación de datos, como se describe en Stevens en 1976. Esto fue usado para mejorar la aproximación normal para las observaciones transformadas y para que la variación sea independiente de la ubicación. Las desventajas son la dificultad para interpretar los resultados y la insuficiencia de las transformaciones cuando las observaciones son recuentos cercanos a cero. West, Harrison y Migon en 1985 formalizaron la extensión de los modelos dinámicos para permitir observaciones en la familia exponencial, definiendo así los modelos lineales dinámicos generalizados (MLDG). La extensión tiene como base los modelos lineales generalizados propuestos en Nelder y Wedderburn en 1972, que están bien cubiertos en McCullagh y Nelder en 1989.

Es necesario tener en cuenta las distribuciones observacionales no normales cuando los datos probablemente no se puedan modelar asumiendo normalidad, aun cuando se aplicara alguna transformación. En un primer momento, se describe la estructura básica del análisis de la familia exponencial.

3.4.1. Modelo de la familia exponencial- *Distribución observacional en la familia exponencial*

Sea Y_t una serie de tiempo con $(t = 1, 2, \dots)$ donde las variables aleatorias pueden ser discretas o continuas. Los valores que toman estas variables están en el espacio muestral Ω . Si la distribución de la serie de tiempo Y_t está en la *familia exponencial* se obtiene una función de densidad o función de masa de probabilidad de Y_t como se describe a continuación: [West and Harrison, 1997]

“Las cantidades η_t , V_t y las funciones $y_t(Y_t)$, $a(\eta_t)$ y $b(Y_t, v_t)$ son conocidas, por tanto su densidad es

$$p(Y_t|\eta_t, V_t) = \exp \{V_t^{-1}(y_t(Y_t)\eta_t - a(\eta_t))\} b(Y_t, V_t) \quad (Y_t \in \Omega)'' \quad (3)$$

Que cumple las siguientes propiedades [Bermúdez, 2012]:

1. “ η_t es el *parámetro natural* de la distribución, es una cantidad continua.
2. $V_t > 0$ es el *parámetro de escala*: el *parámetro de precisión* de la distribución es definido como $\phi_t = V_t^{-1}$.
3. Como una función del parámetro natural para Y_t fijo, la ecuación (3) puede ser vista como una verosimilitud para η_t , que depende de Y_t a través de la transformación $y_t(Y_t)$.
4. La función $a(\eta_t)$ es doblemente diferenciable en η_t . Se sigue que

$$E(y_t(Y_t)|\eta_t, V_t) = \frac{d}{d\eta_t} a(\eta_t) = \dot{a}(\eta_t) = \mu_t \quad (4)$$

$$V(y_t(Y_t)|\eta_t, V_t) = V_t \ddot{a}(\eta_t) \quad (5)$$

5. Con frecuencia $y_t(\cdot)$ es la función identidad, en cuyo caso

$$p(Y_t|\eta_t, V_t) = \exp \{V_t^{-1}(Y_t\eta_t - a(\eta_t))\} b(Y_t, V_t) \quad (Y_t \in \Omega) \quad (6)$$

para $Y_t \in \Omega$. Adicionalmente,

$$E(Y_t|\eta_t, V_t) = \dot{a}(\eta_t) = \mu_t \quad (7)$$

$$V(Y_t|\eta_t, V_t) = V_t \ddot{a}(\eta_t) \quad (8)$$

”.

3.4.2. Marco de regresión dinámica

En las series temporales frecuentemente se utilizan los modelos **time-varying regression** para describir los *modelos lineales dinámicos generalizados* (DGLM) de la serie Y_t ($t = 1, 2, \dots$) y definirlos así:

$$\text{Modelo observacional } p(Y_t|\eta_t) \text{ y } g(\eta_t) = \lambda_t = \mathbf{F}'_t\theta_t, \quad (9)$$

$$\text{Ecuación de evolución } \theta_t = \mathbf{G}_t\theta_{t-1} + \omega_t \text{ con } \omega_t \sim (0, \mathbf{W}_t) \quad (10)$$

Donde,

- “ θ_t , es un vector de estados n -dimensional en el tiempo t .
- \mathbf{F}_t es un vector conocido de regresión n -dimensional.
- \mathbf{G}_t , es una matriz conocida de evolución $n \times n$.
- ω_t , es un vector n -dimensional de evolución de los errores con media cero y matriz de varianzas conocida \mathbf{W}_t $n \times n$.
- $\lambda_t = \mathbf{F}'_t\theta_t$, una función lineal del vector de parámetros de estado.
- $g(\eta_t)$ es una función monótona continua y conocida de η_t en los reales ” [West and Harrison, 1997].

3.5. Riesgo de desempleo

La sociedad a través de los avances tecnológicos ha logrado evitar los riesgos que podrían generar su extinción. Sin embargo, día a día se generan nuevos riesgos aún más peligrosos que los vistos hasta el momento, posiblemente a causa de estos avances y que exigen herramientas de aseguramiento mejores. De igual manera, la colectividad mundial, genera mucha incertidumbre en los ámbitos económicos, políticos y culturales que pueden ser beneficiosos o riesgosos, exigiendo contratos de seguro y reaseguro que permitan mitigar el riesgo al que se enfrenta el capital social y el bienestar individual.

Por esta razón, los seguros como contratos formales buscar mitigar el daño causado al asegurado en el caso de un evento atípico, así “un seguro es un contrato en virtud del cual una parte (el asegurador) acepta un riesgo de seguro significativo de otra parte (el titular de la póliza) al acordar compensar al titular de la póliza si un evento futuro incierto especificado (el evento asegurado) afecta negativamente al titular de la póliza ” [IFRS, 2016]. Por lo anterior, las aseguradoras toman un papel importante en la sociedad dada la intervención que tiene el capitalismo en muchos aspectos de crecimiento y evolución del hombre, convirtiéndose en un negocio basado en especulaciones que cobra un valor estratégico como instrumento que se utiliza para mitigar las consecuencias económicas y sociales a causa de numerosas contingencias.

“El reaseguro es la operación por medio de la cual una institución de seguro toma a su cargo, parcial o totalmente: a) un riesgo cubierto por otra o, b) el remanente de daños que exceda de la suma asegurada por el asegurador directo” [Minzoni, 2009]. La finalidad del reaseguro es mitigar una posible pérdida por parte del asegurador directo al compartir el riesgo con otra compañía y reducir las pérdidas por parte de la aseguradora y el reasegurador.

“La operación de reaseguro se origina de la necesidad que los mercados imponen a los aseguradores de aceptar riesgos cuyo valor supera el equilibrio técnico de sus carteras, lográndose mediante el reaseguro la segmentación de los riesgos en proporciones tales que la relación entre su magnitud y calidad los convierte en comparables entre sí; en otras palabras, homogeneizar cuantitativamente hablando, la cartera del asegurador”.

Seguro y reaseguro son diferentes, pero es necesario la aseguradora para que exista el reasegurador. El papel que tiene el propietario de la póliza en esta operación es nula, dado que en caso de que exista el siniestro, la única responsable de su indemnización es la compañía con la cual firmo la póliza.

Al igual que en el seguro directo, también para el reaseguro son válidos los siguientes principios :

- “Debe existir un interés asegurable.
- El contrato de reaseguros es de *máxima buena fe* (utmost good faith) mientras que el contrato de seguro es *uno de buena fe*.
- El contrato es indemnizatorio y la materia del mismo debe existir en el momento de estipularlo” [Minzoni, 2009].

“La función principal del reaseguro es de naturaleza técnica, ya que tiene como objetivo equilibrar la cartera de una institución de seguro al nivelar las exposiciones de la misma. Los aspectos más sobresalientes del reaseguro bajo el perfil técnico, económico y jurídico son” [Minzoni, 2009]:

1. “Económicamente, el reaseguro tiene el efecto de desplazar la incidencia financiera de los siniestros y limita la carga de la empresa aseguradora solamente a la parte de los mismos que ella puede soportar.
2. El reaseguro tiende a mejorar las condiciones técnicas del asegurador, nivelando cuantitativamente la medida de sus exposiciones y permite que alcance un equilibrio entre primas y siniestros, por un lado, o, por otro, contener el pago de excedentes de siniestros dentro de algunos límites también previstos por los contratos.
3. La ayuda técnica se completa con otra muy importante: la de permitir a la empresa aseguradora la suscripción de riesgos de importancia, por lo que aumenta, indirectamente, las posibilidades industriales de la aseguradora.
4. Bajo el punto de vista jurídico, el reaseguro no modifica la relación entre asegurado y aseguradora en el sentido de que el primero es totalmente ajeno a la operación del reaseguro, siendo la segunda la única responsable frente al primero por las obligaciones asumidas por la aseguradora (este principio no se invalida si en algunas ocasiones el asegurado conoce quiénes son los reaseguradores).
5. El reaseguro se considera como un seguro contra pérdidas patrimoniales del asegurador frente a eventuales desviaciones en la siniestralidad de las coberturas otorgadas, lo que originaría un desequilibrio financiero de la aseguradora; se trata, por lo tanto, de un riesgo financiero” [Minzoni, 2009].

El reaseguro se puede entender como una transferencia de riesgos que se puede ver afectada por cinco tipos de riesgo, a saber [Minzoni, 2009]:

1. “**Riesgo de suscripción** (*Underwriting Risk*). Es la incertidumbre relacionada con el costo final del siniestro.
2. **Riesgo de tiempo** (*Timing Risk*). Se refiere a la incertidumbre sobre el tiempo de pago de los siniestros, es decir, el riesgo que se corre cuando los siniestros esperados se liquidan antes de lo establecido, en el convenio, reduciéndose el período por el cual el ingreso por inversiones logra aumentar las aportaciones (primas y adiciones) recibidas (o sea el riesgo de una aceleración en el esquema de pago de los siniestros, consignado en el convenio entre cedente y reasegurador).
3. **Riesgo de activo** (*Asset Risk*). La incertidumbre que los activos empleados e invertidos alcancen los valores esperados debido a tasas de interés insuficientes, en comparación con las hipótesis.

4. **Riesgo de crédito** (*Credit Risk*). Es la inseguridad que se origina si el reasegurado (la cedente o, en su caso, la aseguradora) no logra pagar al reasegurador los recursos pactados y, por lo tanto, este último no puede hacer frente a las obligaciones contraídas.
5. **Riesgo de inversión** (*Investment Risk*). Es aquel que se origina cuando las utilidades de la inversión de los fondos en mano del reasegurador, resultan menores de lo previsto al establecer el convenio, con motivo de cambios en la tasa de interés en el mercado de valores.”

3.5.1. Modelos de análisis de riesgo de desempleo

El trabajo informal, la falta de cobertura frente a un posible riesgo de desempleo y la frecuencia con que se presenta esta situación, influyen económicamente en la estabilidad de un hogar y un país, haciendo la diferencia entre aquellas naciones que cubren todas las necesidades básicas de su población y aquellos que no, en situaciones de crisis financieras la pérdida del empleo puede ser un evento de enormes magnitudes. A nivel económico, la adquisición de un seguro de desempleo constituye una herramienta de administración del riesgo, que a nivel país permite reducir la magnitud de los ciclos económicos, genera una población más equitativa y una mejor asignación de los trabajadores en el mercado laboral. A nivel individual, un seguro de desempleo ayuda a mitigar la pérdida del poder adquisitivo y permite mantener en cierta proporción el sostenimiento personal y familiar, y el pago de sus obligaciones.

En Colombia existen mecanismos que permiten controlar la contingencia del desempleo, algunos de ellos son: Cesantías, seguros privados, indemnizaciones y mecanismos de ahorro, no obstante, los ciclos macroeconómicos de la economía Colombiana han influido fuertemente sobre el mercado laboral, algunos de los eventos más relevantes que han repercutido en el desempleo de Colombia son: Decrecimiento del producto a finales de la década del 90, Reforma laboral Ley 789 de 2002, crisis mundial del 2008, guerras y pandemia.

4.6.1.1 Valoración del costo del seguro de desempleo para el caso colombiano

Se pretende buscar una aproximación del costo de un seguro de desempleo para el caso colombiano, utilizando una lógica contable, como se enuncia a continuación:

“Los cálculos realizados utiliza la siguiente lógica, donde se compara los aportes (A_t) de los beneficios otorgados (B_t) para el periodo t .

Los aportes son:

$$A_t = [AC_t \times w_t \times TS] \times TC$$

,

donde AC es el número de cotizantes, w el salario, TS la tasa de sobrevivencia en el empleo/año y TC la tasa de cotización o prima. Por otro parte los beneficios valen:

$$B_t = [NB_t \times w_t \times TR] \times DB,$$

donde, NB es el número de beneficiarios, w el salario en el empleo previo, TR la tasa de reemplazo y DB la duración del beneficio. Se actualiza estas dos componentes e introduce ciclos para capturar aleatoriedad en los cambios de la tasa de desempleo en el tiempo.” [Bardey, 2009]

Ahora se reescribe la restricción presupuestal planteada antes, teniendo en cuenta que los cálculos de las probabilidades de perder el trabajo (p) no es igual según el nivel de ingresos (w). Por lo anterior, se debe considerar la distribución de densidad de probabilidad conjunta: $f(p, w)$ donde los aportes y los beneficios durante un periodo dado se describen a continuación:

$$A_t = \int_{w_a}^{w_b} (1 - p(\sigma_t)) TC w f(p|w) dw$$

y,

$$B_t = \int_{w_{a'}}^{w_{b'}} p(\sigma_t) \sum_{t+F}^{t+d+F} w \sigma_t \varphi_t(\sigma_t) f(p|w) dw$$

Los límites inferiores y superiores de las integrales en las expresiones anteriores no son necesariamente los mismos. “La función $f(p|w)$ indica la probabilidad de desempleo condicionada al nivel de salario de las personas F representa el periodo de franquicia o carencia (la persona en su primer mes de desempleo no recibe los recursos derivados del seguro de desempleo), mientras d representa la duración de la cobertura. El parámetro σ_t representa la tasa de reemplazo, es decir la fracción del salario que recibe el beneficiario. Para capturar los esfuerzos que los empleados pueden realizar para conservar su trabajo, la probabilidad de perder su empleo depende de generosidad de la tasa de reemplazo. La función $\varphi_t(\sigma_t)$ representa la probabilidad de salir de una situación de desempleo lo que captura la otra fuente de riesgo moral, igualmente llamado “efecto de duración”. De los modelos microeconómicos, esta probabilidad es decreciente en σ_t .

Del análisis se concluyó que la implementación de las ecuaciones de aportes y beneficios durante un periodo de tiempo y la utilización de la función de densidad de probabilidad conjunta, vuelve complejo el cálculo de los parámetros que representan las características del contrato de seguro de desempleo que permiten indicar que es auto financiable. El primer análisis realizado de las variables A_t y B_t son una buena aproximación inicial pero no es suficiente para implementar y garantizar que el seguro de desempleo es sostenible y rentable.” [Bardey, 2009]

4.6.1.2 Percepción del riesgo al desempleo en México. Un estudio cuantitativo

La fuente de información que se utilizó es el Latinobarómetro de 2018 en México. “Se realizó un muestreo probabilístico para obtener los datos, el muestreo se realizó en tres etapas aleatorias y una por cuotas de mil doscientos casos, con margen de error probabilístico ± 2.8 , nivel de confianza de 95 % y con 100 % de representatividad nacional. La población se ubica entre los 18 años en adelante.” [Mejía Reyes, 2021]

La variable dependiente que se utilizó en el estudio fue: ¿Cuán preocupado diría usted que está por quedar sin trabajo o de estar desempleado durante los próximos doce meses? del Latinobarómetro. Las respuestas son 1) Muy preocupado, 2) Preocupado, 3) Poco preocupado, 4) No está preocupado y 5) No tiene trabajo; esta última se omitió del estudio ya que no se ajusta a los objetivos planteados. Se obtuvo respuesta del 98,3 %.

Las variables explicativas que se utilizaron fueron: sexo (Hombre, Mujer), edad (16 a 25 años, 26 a 40 años, 41 a 60 años y 61 años y más), escolaridad (Básica y menos, Secundaria, media técnica y menos, Superior o más), clase social (Casado / Conviviente, Soltero, Separado, divorciado, Viudo) y trabajo que desempeña (Autónomo, Autónomo dueño de negocio, Autónomo agricultor, Autónomo trabajador por cuenta propia, Asalariado, Asalariado alto ejecutivo, Asalariado ejecutivo mando medio y Asalariado).

El análisis multifactorial es el segundo procedimiento que se realizó, ajustando mediante una regresión logística ordinal, el cual tiene por objetivo evidenciar matemáticamente la dependencia de la variable a explicar con un conjunto de variables explicativas indicado por coeficientes en función de la varianza total. Para comprobar la validez de los resultados se utilizó la bondad del ajuste de Pearson de 0.001 y Deviance de 0.000 y el ajuste es favorable conforme la hipótesis nula al 0.000 y el valor del pseudo R. es 0.026.

“Se obtuvo como conclusión que el riesgo laboral en México es dominante entre los encuestados desde 1995 y 2018. No obstante, los cambios porcentuales de encuestados elevan su intranquilidad se elevan en momentos de crisis o recesiones económicas que impactan el país, exceptuando la última de 2018 que se considera desaceleración, fenómeno conceptualmente descrito como menos fuerte.” [Mejía Reyes, 2021].

4. Metodología

El presente artículo se caracteriza por ser una investigación aplicada, descriptiva y explicativa, con una visión de carácter cuantitativo, dado que busca formular un modelo para lograr un objetivo concreto, el cual esta asociado a las predicciones del número de siniestros de un seguro de desempleo sobre la cantidad de vigentes en un mes específico (tasa de siniestralidad), para ser de utilidad en la mitigación de una posible perdida o contingencia de la compañía aseguradora.

La investigación es no experimental, ya que los datos provienen de la fuente real y no es posible la modificación de los mismos, es decir, se observaron situaciones ya existentes, y no fueron provocadas por el investigador.

Se desarrollan los distintos estadísticos descriptivos junto con los gráficos de línea para identificar las características particulares de las variables estudiadas, entre las cuales se resalta la volatilidad y el comportamiento leptocúrtico de los datos. Las variables a utilizar en el presente estudio son:

Variables	Descripción
1. Indicador de seguimiento a la economía (ISE)	Utiliza un conjunto de variables de diferentes sectores económicos, a saber, agropecuario, industria, minería y servicios, con el fin de dar una estimación en un periodo de tiempo corto de la variación de la actividad económica del país. El ISE se caracteriza por “ ser un indicador dinámico, de agregación, coyuntural y como indicador de volumen. Como indicador dinámico, no permite medir en cada momento la magnitud de la producción, sino únicamente su tendencia. Como indicador de agregación, resume sintéticamente una cantidad de información considerable. Como indicador de coyuntura, estima de forma mensual y como indicador de volumen, refleja variaciones en cantidad, pero no de precios” [DANE, 2023].
2. Petroleo real	Es la deflactación del precio del petróleo o variable nominal WTI, es decir, se elimina el efecto que los cambios en los precios (inflación/ deflación) tienen sobre los valores de la serie temporal.
3. Tasa de cambio real	Indica el poder de compra de una moneda frente a la moneda de otro país, que normalmente es el dólar. No obstante, considera los valores de los bienes en el país al que pertenece la moneda, contrario al cambio nominal.
4. Términos de intercambio	Es el coeficiente entre los precios de las exportaciones y las importaciones. El índice de precios de las exportaciones e importaciones permite tener un panorama de la actividad económica de un país, al desagregar en cantidades y montos la variación nominal del comercio exterior, impactando la dinámica de la inflación interna de un país. “La evolución de estos índices facilita el análisis de la competitividad de un país en los mercados internacionales, y la estimación del cambio en las ventas y compras externas ante una variación de los precios” [Banco de la República de Colombia, 2023].
5. Tasa de ocupados	La tasa de ocupados se define como el cociente entre la cantidad de gente ocupada entre 16 a 64 años con cualquier tipo de contrato y el total de la población de 16 a 64 años. Reemplazando a la tasa de desempleo dado que esta si involucra a toda la población que realmente si está laborando con un tipo de contrato
6. Tasa de siniestralidad	Es el cociente entre la cantidad de personas que cobran el seguro de desempleo y el total de personas vigentes o inscritas al seguro ofrecido por la compañía.

La población de estudio son todos los afiliados a un seguro de desempleo que tienen un crédito de

libranza con una entidad financiera y aseguraron el pago de las cuotas de su deuda en caso que tuvieran un siniestro (quedar desempleado) entre julio de 2016 a enero de 2020. La serie de datos fue anonimizada mediante la técnica de agregación [Biblioteca de la CEPAL, 2017], dado que los datos se muestran como totales, esto no permite identificar a los asegurados y son de uso exclusivo para el presente artículo.

Con la población de estudio identificada y el cálculo de la tasa de ocurrencia de siniestros (cantidad de siniestros/ cantidad de vigentes), se investigó y selecciono las variables macroeconómicas que posiblemente se correlacionan con la siniestralidad del seguro de desempleo para el mismo periodo de tiempo mencionado anteriormente. Posteriormente, se elaboró una base de datos unificando la información para realizar un análisis descriptivo de las variables y el estado del seguro de desempleo. Luego, se construyeron modelos que permitieron detectar o explicar la tasa de siniestralidad por medio de las variables macroeconómicas, donde una de las estrategias a utilizar son modelos de regresión para respuestas de conteo y modelos lineales dinámicos generalizados. Para identificar que variables macroeconomicas no eran significativas se observo que su grafica se mantuviera constante en el tiempo y estuviera alrededor del 1.

Finalmente, se hace la selección del mejor modelo según criterios de información, significancia e interpretación adecuada de los coeficientes obtenidos y el poder de predicción de los modelos (backtesting). Definido el modelo y las variables macroeconómicas, se hace una simulación de crisis económica estresando las covariables seleccionadas en el modelo para determinar y cuantificar la afectación y aumento en la tasa de ocurrencia de siniestros.

Con este artículo se espera obtener un modelo que relacione y explique las variables económicas con la siniestralidad del seguro de desempleo ofertado por la compañía y posteriormente pronosticar la tasa de siniestralidad del seguro de desempleo para evitar posibles perdidas o riesgos para la compañía. Se tuvieron en cuenta las variables económicas más conocidas y que fueran significativas para el estudio. Solamente se analizarón el seguro de desempleo porque representa un gran número de afiliados que posiblemente afectarían a la compañía en situaciones adversas.

5. Aplicación

5.1. Análisis descriptivo de los datos

La base contiene el producto asegurado, la cantidad de siniestros ocurridos en el mes, la cantidad de personas aseguradas en ese periodo, la frecuencia de ocurrencia o tasa de siniestralidad (cantidad de siniestros sobre la cantidad de personas aseguradas), los valores para las variables económicas mencionadas en la metodología y la fecha del siniestro desde julio de 2016 a enero de 2020 con una periodicidad mensual. En la tabla 1 se presenta una descripción de la cantidad de siniestros, cantidad de vigentes y la tasa de siniestralidad para el seguro. Se puede observar que los cambios más relevantes son en el tercer y cuarto año de estudio, al tener una disminución en la cantidad de siniestros en el segundo y cuarto trimestre del 2018 al 2019; además, se encontró una disminución en la cantidad de vigentes en estos mismos años.

En la Figura 1 se presenta el total de siniestros ocurridos mensualmente entre el 2016 y 2020 del seguro, se evidencia un periodo de alta siniestralidad en el mes de julio de 2019. Lo anterior, posiblemente este influenciado por factores cómo los presentados en el balance preliminar de las economías en América Latina y el Caribe 2019 del Congreso de la República de Colombia, que reporta “para la mitad del 2019 un déficit a causa del aumento de las importaciones por la demanda interna y la extensión del IVA a las importaciones de bienes de capital, aprobadas por la Ley de Financiamiento de 2018 con el fin de promover la inversión. Para las exportaciones, la disminución de su valor en dólares se debió a los menores precios de los hidrocarburos y a la reducción de la producción y exportación de carbón” [Congreso de la Republica de Colombia , 2019].

De la variable tasa de siniestralidad se observó que la ocurrencia de estos siniestros es aleatoria y no

Fecha	Siniestros	Vigentes	Tasa de siniestralidad
Año 2016			
Trim.2	168	94.681	0,00177438
Trim.3	400	310.838	0,00387357
Trim.4	453	352.842	0,00382703
Año 2017			
Trim.1	503	384.036	0,00392713
Trim.2	567	413.076	0,0041138
Trim.3	632	436.333	0,00434386
Trim.4	675	456.543	0,00442969
Año 2018			
Trim.1	751	468.917	0,00480529
Trim.2	878	480.812	0,00547645
Trim.3	846	493.122	0,00514878
Trim.4	794	496.867	0,00479695
Año 2019			
Trim.1	853	499.114	0,00512629
Trim.2	761	492.244	0,00463833
Trim.3	983	470.884	0,00623063
Trim.4	651	439.295	0,00445343
Año 2020			
Trim.1	425	411.055	0,00309307
Total general	10.340	6.700.659	0,07005868

Tabla 1: Comportamiento seguro desempleo por año y trimestre, producto libranza.

contienen ningún factor que favorezca unas ocurrencias en favor de otras. Además, las ocurrencias de los siniestros están uniformemente distribuidas dentro del intervalo de tiempo utilizado. Por lo tanto, se considera la construcción de un modelo Poisson dinámico que evalúe el comportamiento de la siniestralidad con ciertos rezagos de tiempo.

5.2. Estimación del modelo

Para estimar la tasa de siniestralidad del seguro de desempleo se proponen modelos lineales dinámicos descritos en los capítulos 3 y 4 en los que se incluyen un offset (el logaritmo de la cantidad de asegurados en el tiempo t , es decir, $\log(z[t])$) y como covariables el ISE, el petróleo real, la tasa de cambio real, el término de intercambio y la tasa de ocupados. El modelo se elige bajo los criterios del DIC, el pD y la prueba de Heidelberg & Welch.

El modelo inicialmente toma el índice 2 hasta el índice N (cantidad de datos en la base) donde la variable respuesta (tasa de siniestralidad) en el tiempo t , es decir, $y[t]$ se distribuye Poisson con parámetros $\mu[t]$. Luego, se crea un nuevo y que se distribuye Poisson con media $\mu[t]$ que tiene como función de enlace el logaritmo. Tiene un subíndice t que lo convierte en un modelo dinámico generalizado rezagado una vez (11). Los coeficientes deben moverse en el tiempo y su resultado será una función.

$$\log(\mu[t]) = \log(z[t]) + \beta_0[t] + \beta_1[t] * x1_{[t-1]} + \beta_2[t] * x2_{[t-1]} + \dots + \beta_N[t] * xN_{[t-1]} \quad (11)$$

Posteriormente, se calcula los residuales de Pearson que proveen una medida para determinar si la observación pronosticada por el modelo es correcta. El error del modelo es la diferencia entre los conteos observados y los estimados; y las a-prioris van a determinar que la media de cada β_n en el tiempo t va a ser igual a $\beta_{n[t]}$ con $n = 2, \dots, N$ en el tiempo inmediatamente anterior. Todos los β se distribuyen

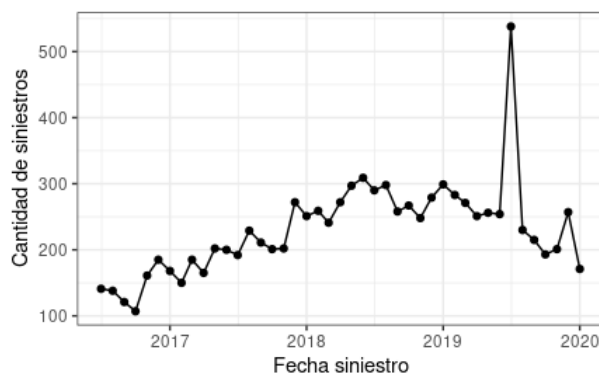


Figura 1: Cantidad de siniestros de 2016 a enero de 2020

con una normal en términos de media y precisión, donde la media será la calculada con las a-priori y las precisiones al ser desconocidas se les impone una distribución a-priori que se distribuyen gamma.

Finalmente, para encontrar los β se utiliza el método de JAGS del software R, el cual permite ajustar modelos que implementan métodos MCMC para generar simulaciones de distribuciones posteriores, especificando únicamente las distribuciones iniciales, la verosimilitud y los datos iniciales.

5.3. Validación de supuestos

Cuando se generan muestras de la distribución posterior mediante MCMC, se busca:

1. “Valores simulados representativos de la distribución posterior, lo cual implica que no deben estar influenciados por el valor inicial (arbitrario) y deben explorar todo el rango de la posterior.
2. Se debe tener suficientes simulaciones de tal manera que las estimaciones sean precisas y estables.

En la práctica se intenta cumplir con los anteriores objetivos. Aunque los métodos MCMC garantizan que cadenas infinitamente largas lograran una representación perfecta, siempre debemos tener un criterio para cortar la cadena y evaluar la calidad de las simulaciones” [Gelman and Hill., 2007].

6.3.1 Representatividad

La finalidad es verificar si el modelo ha alcanzado la distribución objetivo (equilibrio), es decir, la distribución posterior. La convergencia se puede monitorear de diferentes maneras; en el presente estudio se utilizó el diagnóstico de Heidelberg y Welch que calcula un estadístico de prueba para aceptar o rechazar la hipótesis nula: “La cadena que se obtiene es como la posteriori”; si la cadena pasa la primera parte del diagnóstico entonces se genera un nuevo estadístico con la hipótesis alternativa: “El valor al cual converge la cadena es la media de la posteriori”. Utilizar este tipo de pruebas permite obtener la media de una distribución posterior marginal con buena precisión, suponiendo que la cadena de Markov se encuentra en la distribución estacionaria.

Al aplicar el diagnóstico de Heidelberg y Welch en el modelo construido con las variables de estudio y teniendo como parámetros: 3 cadenas de Markov y 100.000 iteraciones totales por cadena, se obtuvo una convergencia total. Se evidenció (Figura 2) que el autocorrelograma cumple con las propiedades (media y varianza) de no variabilidad con respecto al intervalo de tiempo a trabajar, es decir, sin tendencia, sin ciclos y con variaciones aproximadamente homogéneas. Las cadenas convergen y se mantienen alrededor de un solo valor. La densidad de la a-posteriori para cada tiempo se comportan como una normal.

En la Tabla 2 se muestran los resultados del diagnóstico de Heidelberg y Welch donde se observa que todas las variables cumplen las dos condiciones esperadas con una precisión relativa de dos dígitos.

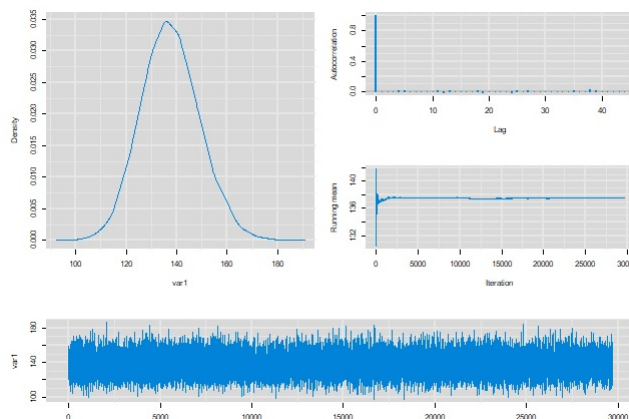


Figura 2: Convergencia de las cadenas

Variable	Primera Parte			Segunda Parte		
	Stationarity test	Start iteration	p-value	Halfwidth test	Mean	Halfwidth
Var 1	passed	1	0,4718	passed	138	0,133
Var 2	passed	11881	0,2623	passed	121	0,16
Var 3	passed	1	0,1505	passed	107	0,118
Var 4	passed	1	0,8071	passed	161	0,143
Var 5	passed	1	0,3315	passed	185	0,153
Var 6	passed	1	0,5004	passed	168	0,143
Var 7	passed	1	0,1987	passed	150	0,14
Var 8	passed	1	0,773	passed	185	0,152
Var 9	passed	1	5579	passed	165	0,147
Var 10	passed	1	0,4912	passed	202	0,162

Tabla 2: Resultado diagnóstico de convergencia Heidelberg y Welch.

6.3.2 Precisión

La finalidad es determinar que la muestra es lo suficientemente grande para realizar buenas estimaciones de la distribución, a partir de la distribución posteriori. Inicialmente, se identifica que la relación entre las variables predictivas y las predichas no es muy diferente. Posteriormente, se revisa que los datos con que se entrena el modelo tienen la capacidad de pronosticar la mayor cantidad de escenarios posibles. Los modelos dinámicos permiten tener en cuenta todos los posibles casos al calibrar el modelo con nuevas variables.

Se utilizan las siguientes medidas para determinar la precisión del modelo:

1. MAPE (Mean Absolute Percent Error) “calcula la diferencia de porcentaje absoluto promedio entre los valores ajustados por el modelo y los valores de datos observados. Se utiliza en el pronóstico para determinar la eficiencia de la predicción” [Agresti.A, 2015].
2. El n.eff (tamaño efectivo de la muestra) “estima el número efectivo de muestras de la posterior que surgen de las cadenas y que en este modelo se obtiene un valor de al menos 100” [McCullagh and Nelder, 1983].
3. El DIC permitió determinar el mejor de los modelos, ya que tuvo el menor criterio de información, correspondiente a 387,7. Este criterio es una estimación del error predictivo esperado.

4. El *Rhat* estima si las cadenas convergen a una distribución estable, un $Rhat < 1.1$ indica convergencia.

Revisados los supuestos, la ecuación matemática del modelo lineal dinámico es:

$$\begin{aligned} \log(\text{cantidad siniestros}_{[t]}) = & \log(\text{cantidad vigentes}_{[t]}) + \beta_{0[t]} \\ & + \beta_{1[t]} \times \text{tasa cambio real}_{[t-1]} \\ & + \beta_{2[t]} \times \text{tasa ocupados}_{[t-1]} \\ & + \beta_{3[t]} \times \text{indice términos de intercambio}_{[t-1]} \end{aligned} \quad (12)$$

El modelo esta compuesto por los β que son los coeficientes a estimar y las variables macroeconomicas significativas para el modelo.

5.4. Pronóstico

El modelo dinámico realizado en este artículo, parte de la información a-priori de la siniestralidad del seguro de desempleo y de las variables macroeconómicas que dentro de los supuestos se asumen constantes en todo el tiempo de análisis y pronóstico, con la cual se elige la distribución del parámetro ; se calcula la varianza y el error en el tiempo t . Finalmente, con esto se calibra la función predictiva con la que se calcula el pronóstico.

Este modelo presenta además una serie de rezagos, que permiten expresar la idea de que el valor de la variable respuesta (tasa de siniestralidad) depende de los valores previos que se hayan observado. La elección implica expresar la variable de interés dependiendo de valores de periodos anteriores. En la tabla 3 se observa la predicción de la tasa de siniestralidad para 3, 4, 5 y 7 meses con 1, 2 y 3 rezagos.

Al calcular el MAPE del pronóstico (Tabla 3) se observa que las predicciones a 3 meses para cualquier rezago alcanzan un error máximo de 12 %, lo cual es bueno. Según [Fretchling, Douglas C, 1996] un cálculo entre el 10 % y el 20 % del MAPE se puede considerar como un pronóstico bueno. Aunque al comparar los pronósticos a corto plazo (3 meses y 4 meses) con los de largo plazo (5 meses y 7 meses) en su mayoría estos no superan el 20 %. Es de resaltar que las predicciones a corto plazo tienen menor distorsión y un mejor pronóstico. Es necesario reentrenar cada 6 meses para obtener mayor efectividad del modelo.

Utilizar un modelo dinámico bayesiano para calcular la tasa de siniestralidad de un seguro de desempleo es una herramienta muy útil, se obtienen pronósticos reales sin la necesidad de un gran volumen de datos históricos.

5.5. Backtesting

El backtesting es “una metodología utilizada para verificar y diagnosticar la eficiencia y los resultados de un modelo. Esta metodología intenta determinar si las conclusiones del modelo o los cálculos de las variables estimadas son acertadas” [Stevens, R, 2020].

El objetivo con esta metodología es evaluar la eficiencia y los pronostico del modelo. Tomando datos históricos se determina que tan buenas son las predicciones del modelo en periodos anteriores. Esto evita la toma de decisiones sin una certeza de los resultados que genera el modelo, teniendo en cuenta el margen de riesgo o error que el modelo pueda tener.

Para el backtesting se utilizaron datos desde marzo hasta septiembre de 2020 (Figura 3) encontrando un escenario de crisis económica a causa de la Pandemia, lo cual podría dar respuesta a la tasa de siniestralidad en caso de un momento atípico. El comportamiento de los siniestros en el periodo antes mencionado aparece en la Figura 3 donde se observa un comportamiento inusual para el mes de julio y agosto de 2020.

Meses pronosticados/ rezagos	Tasa de siniestralidad real	Rezago 1		Rezago 2		Rezago 3	
		Pronóstico	MAPE	Pronóstico	MAPE	Pronóstico	MAPE
3 meses	201	208	0,12	195	0,10	197	0,10
	257	189		188		187	
	171	157		175		175	
4 meses	193	176	0,16	191	0,11	226	0,15
	201	182		190		220	
	257	172		171		211	
	171	144		163		199	
5 meses	215	284	0,22	234	0,16	250	0,19
	193	245		244		247	
	201	252		231		232	
	257	235		227		226	
	171	206		209		211	
7 meses	538	307	0,34	222	0,20	247	0,18
	230	289		266		198	
	215	317		269		218	
	193	282		207		223	
	201	274		216		180	
	257	264		185		186	
	171	236		178		162	

Tabla 3: Pronóstico tasa de siniestralidad.

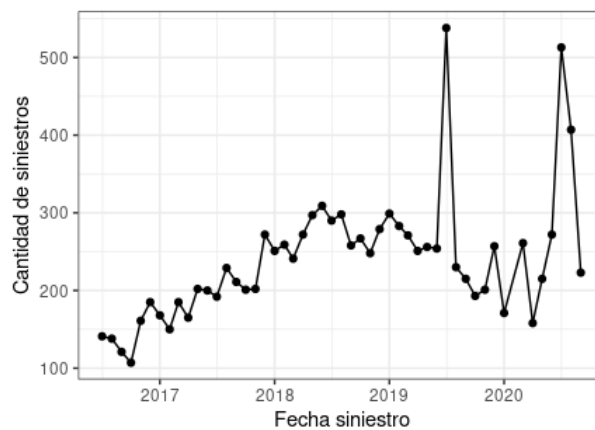


Figura 3: Cantidad de siniestros hasta septiembre de 2020

Al calcular el MAPE del pronóstico en un escenario de crisis (Tabla 4) se observa que las predicciones a 7 meses para cualquier rezago tienen el menor error. Manteniendo la idea de [Fretchling, Douglas C, 1996] se obtienen buenos resultados cuando el modelo tiene dos rezagos y pronostica 7 meses a futuro. A diferencia de los pronósticos del modelo en un escenario económico estable, el backtesting muestra que la tasa de siniestralidad con respecto a las variables económicas (tasa de ocupados, términos de intercambio y tasa de cambio real) no es instantánea ni tampoco se evidencia en un corto plazo. La siniestralidad

Meses pronosticados/ rezagos	Tasa de siniestralidad real	Rezago 1		Rezago 2		Rezago 3	
		Pronóstico	MAPE	Pronóstico	MAPE	Pronóstico	MAPE
3 meses	513	263		283		410	
	407	184	0,4519	267	0,3014	393	0,2511
	223	151		248		339	
4 meses	272	206		417		776	
	513	188	0,4281	383	0,3344	672	0,9536
	407	218		306		624	
5 meses	223	140		290		472	
	215	199		537		177	
	272	191	0,4746	1029	0,533	116	0,4914
7 meses	513	158		826		128	
	407	120		550		145	
	223	88		427		153	
	261	200		179		177	
	158	246	0,3551	191	0,2412	182	0,3227
	215	321		174		190	
	272	323		254		141	
	513	288		260		197	
	407	246		243		208	
	223	183		235		205	

Tabla 4: Pronóstico tasa de siniestralidad-backtesting

del seguro responde a un cambio abrupto de las variables económicas después de 7 meses al obtener una menor distorsión y un mejor ajuste a los datos reales (Figura 4).

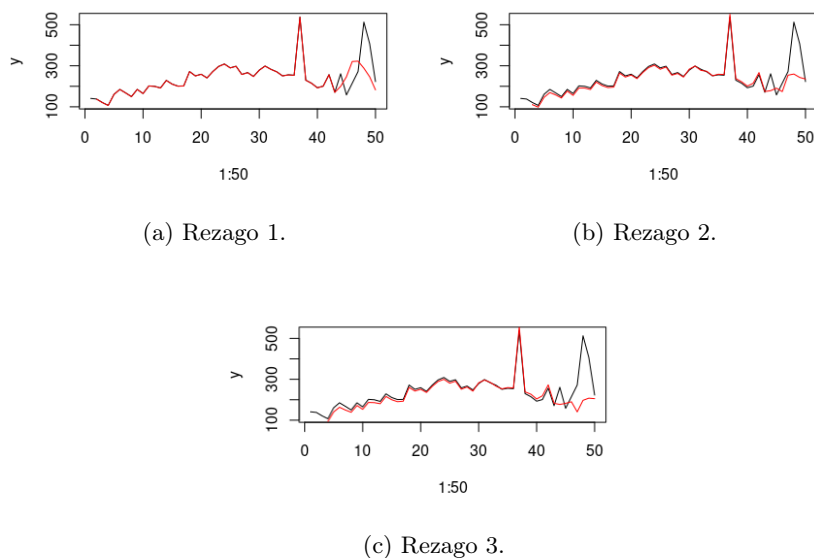


Figura 4: Pronósticos backtesting a 7 meses con diferentes rezagos (En rojo las predicciones y en negro los datos observados)

6. Conclusiones

A manera de conclusión, se podría pensar que en Colombia este tipo de seguros de desempleo son de vital importancia, teniendo presente las diferentes variables que determinan la amenaza de crisis constante y la volatilidad de nuestras políticas económicas, así como la experiencia de los aseguradores frente a los potenciales cambios en los fenómenos de empleabilidad y las consecuencias que se tienen por paros o protestas. Además, es necesario que la población adquiera la cultura de proteger sus bienes o cumplir sus deberes en caso de una crisis a través de los seguros.

Respecto a los pronósticos generados por la implementación del modelo dinámico bayesiano, se obtuvieron las siguientes conclusiones: en un escenario donde la economía es estable, utilizar un modelo dinámico bayesiano para calcular la tasa de siniestralidad de un seguro de desempleo es muy útil. Obteniendo pronósticos reales sin la necesidad de un gran volumen de datos históricos con un MAPE que no supera el 20% y arrojando pronósticos buenos tres meses a futuro. Por el contrario, cuando se realizan pronósticos en un escenario de crisis económica, el modelo muestra que la tasa de siniestralidad con respecto a las variables económicas (tasa de ocupados, términos de intercambio y tasa de cambio real) no es instantánea. La siniestralidad del seguro se ve afectada después de 7 meses. Cuando hay un cambio abrupto de las variables económicas.

Las distribuciones de probabilidad a priori utilizadas en el algoritmo de este artículo se puede modificar fácilmente, permitiendo continuar con esta investigación en otros campos o industrias que no cuenten con un amplio conjunto de datos, generando modelos que se pueden comparar con el actual respecto a su eficiencia.

Utilizar el backtesting para determinar la eficiencia del modelo evita la toma de decisiones al azar o la puesta en marcha de estrategias sin tener una idea clara de los resultados. Además, es una herramienta que se aplica en muchos sectores económicos previniendo malos resultados en las operaciones futuras de la compañía y para comprobar si los datos reales son muy diferentes de los predichos por el modelo.

Se recomienda reentrenar cada 6 meses para la efectividad del modelo y utilizar otras variables económicas que puedan ajustar el modelo y los pronósticos. Esta teoría se puede utilizar para encontrar solución a

problemas del sector bancario e industrial, o en aquellos casos que no cuenten con suficiente información histórica. Sin embargo, se recomienda una maquina con suficiente memoria RAM ya que se necesita un gran número de iteraciones del modelo para la convergencia de las cadenas y obtener pronosticos confiables.

Referencias

- [Agresti.A, 2015] Agresti.A (2015). *Foundations of linear and Generalized Linear Models*, volume 1. John Wiley & Sons, Inc.
- [Banco de la República de Colombia, 2023] Banco de la República de Colombia, C. (2023). Índice de términos de intercambio de bienes. Technical report, Banco de la República de Colombia.
- [Barbeito and Villalón, 2003] Barbeito, J. and Villalón, J. (2003). *Introducción al cálculo estocástico aplicado a la modelización económico-financiero-actuarial*. Netbiblo.
- [Bardey, 2009] Bardey, D, e. (2009). Seguros de desempleo: Revisión de literatura y propuesta para colombia. *Universidad del Rosario*, pages 28–31.
- [Bermúdez, 2012] Bermúdez, D. (2012). Modelamiento bayesiano de caudal y precipitación aplicando modelos dinámicos y procesos de poisson no homogéneos. *UNAL-Universidad Nacional de Colombia*.
- [Biblioteca de la CEPAL, 2017] Biblioteca de la CEPAL, C. (2017). Gestión de datos de investigación-anonimización de los datos: Técnicas de anonimización - agregación. Technical report, Comisión Económica para América Latina y el Caribe.
- [Buti, 2021] Buti, M. (2021). What the EU has learned from overcoming the financial crisis and covid-19. Technical report, World Economic Forum.
- [Cayuela, 2010] Cayuela, L. (2010). Modelos lineales generalizados (GLM). *Centro Andaluz de Medio Ambiente, ECOLDO- Universidad de Granada*.
- [Congreso de la Republica de Colombia , 2019] Congreso de la Republica de Colombia (2019). Balance preliminar de las economías de américa latina y el caribe 2019. *Comisión Económica para América Latina y el Caribe (CEPAL)*.
- [Contreras, 2012] Contreras, N. (2012). Análisis de votos electorales usando modelos de regresión para datos de conteo. Master's thesis, Pontificia Universidad Católica de Perú.
- [Cárdenas, 2009] Cárdenas, V. (2009). Financiamiento de riesgos catastróficos naturales. *Inter-American Development Bank*.
- [DANE, 2023] DANE (2023). Indicador de seguimiento a la economía. Technical report, Departamento Administrativo Nacional de Estadística.
- [Deng, 2019] Deng, K. (2019). The next decade in risk: Financial and economic risks. *Centre for Risk Studies-University of Cambridge*.
- [Dobson, 2002] Dobson, A. J. (2002). *An introduction to generalized linear models-2nd ed.*, volume 1. Chapman & Hall/CRC.
- [Fretchling, Douglas C, 1996] Fretchling, Douglas C (1996). Practical tourism forecasting. *Oxford*.
- [Garcia Quinteiro, 2014] Garcia Quinteiro, E. (2014). Aplicación de modelos de regresión bivariados a los resultados de los partidos de la liga española de fútbol. *USC-Universidad de Coruña*, pages 25–31.
- [Gelman and Hill., 2007] Gelman, A. and Hill., J. (2007). *Data Analysis Using Regression and Multi-level/Hierarchical*, volume Analytical methods for social research. New York: Cambridge University Press.
- [Harrison and Stevens, 1976] Harrison, P. and Stevens, C. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society*, 38:205–247.
- [IFRS, 2016] IFRS (2016). IFRS 4 Insurance Contracts. Technical report, Fundación NIIF 2021.

- [McCullagh and Nelder, 1983] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman & Hall.
- [Mejía Reyes, 2021] Mejía Reyes, C. (2021). Percepción del riesgo al desempleo en México. un estudio cuantitativo. *Universidad del Zulia*, pages 15–20.
- [Migon, 1984] Migon, H. (1984). *An approach to Non-linear Bayesian Forecasting Problems with Application*. PhD thesis, University of Warwick.
- [Migon and Harrison, 1985] Migon, H. and Harrison, P. (1985). An application of nonlinear bayesian forecasting to television. *Bayesian Statistics*.
- [Minzoni, 2009] Minzoni, A. (2009). *Reaseguro*. Universidad Nacional Autónoma de México-Facultad de Ciencias.
- [Nelder and Wedderburn, 1972] Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *JSTOR*, 135(3).
- [Pérez Fructuoso, 2005] Pérez Fructuoso, M. J. (2005). Modelos estocásticos de estimación de pérdidas en la protección financiera a través de la titulación del riesgo de catástrofes naturales. a juste mediante técnicas de aprendizaje automático. *Fundación MAPFRE Estudios*.
- [Rachev et al., 2005] Rachev, S., Bagasheva, B., and Fabozzi, F. (2005). *Bayesian methods in finance*, volume 153. John Wiley & Sons.
- [RAE, 2004] RAE (2004). *Riesgo catastrófico*. Real Academia Española.
- [SCB, 2016] SCB (2016). *Póliza de Seguro de desempleo*. Seguros Comerciales Bolívar (SCB).
- [Schnieper, 1993] Schnieper, R. (1993). The insurance of catastrophe risks. *Actuarial Science: Cat Risks*.
- [Stevens, R, 2020] Stevens, R (2020). Backtesting:definición y ejemplos. *Rankia*.
- [West and Harrison, 1997] West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dinamic models*. Springer- Verlag.
- [West et al., 1987] West, M., Pole, A., and Harrison, J. (1987). *BATS: Bayesian Analysis of Time Series*. The Professional Statistician.
- [West et al., 1985] West, P., Migon, H., and Harrison, J. (1985). *Dynamic Generalized Linear Models and Bayesian Forecasting*, volume 80. Journal of the American Statistical Association.
- [Willige, 2017] Willige, A. (2017). You live in a world full of risks. These are the biggest. Technical report, World Economic Forum.

7. Anexos

7.1. Código JAGS

El siguiente código es solo para el pronóstico de un rezago a 2 meses. Sin embargo, con una modificación se aplica el mismo código para el cálculo de los rezagos faltantes.

```
#####
                          Modelo lineal din mico
#####

LIB<-variables_final %% filter (PROTECCION=="LIBRANZA")

N<-43
y<-LIB$CANT_SINIESTROS
x1<-LIB$TASA_CRECIMIENTO_ANUAL_ISE
x2<-LIB$TASA_CRECIMIENTO_MENSUAL_ISE
x3<-LIB$VARIACION_ANUAL_TASA_CRECIMIENTO_ANUAL_ISE
x4<-LIB$VARIACION_MENSUAL_TASA_CRECIMIENTO_ANUAL_ISE
x5<-LIB$VARIACION_ANUAL_TASA_CRECIMIENTO_MENSUAL_ISE
x6<-LIB$VARIACION_MENSUAL_TASA_CRECIMIENTO_MENSUAL_ISE
x7<-LIB$PETROLEO_REAL
x8<-LIB$VARIACION_ANUAL_PETROLEO_REAL
x9<-LIB$VARIACION_MENSUAL_PETROLEO_REAL
x10<-LIB$TASA_CAMBIO_REAL_CE
x11<-LIB$VARIACION_ANUAL_TASA_CAMBIO_REAL
x12<-LIB$VARIACION_MENSUAL_TASA_CAMBIO_REAL
x13<-LIB$TASA_OCUPADOS
x14<-LIB$VARIACION_ANUAL_TASA_OCUPADOS
x15<-LIB$VARIACION_MENSUAL_TASA_OCUPADOS
x16<-LIB$INDICE_TERMINOS_INTERCAMBIO
x17<-LIB$VARIACION_ANUAL_TERMINOS_INTERCAMBIO
x18<-LIB$VARIACION_MENSUAL_TERMINOS_INTERCAMBIO
z<-LIB$CANT_VIGENTES

##### pronostico a 2 meses con un rezago #####

y
y1=y[1:40]
N1=40

modelo1=function()
{
  for(t in 2:N1)
  {
    y1[t]~dnorm(mu[t],10000)
    mu[t]~dpois(mu0[t])
    mu0[t]=z[t]*tasa[t]
    log(tasa[t])=b10[t]*x10[t-1]+b13[t]*x13[t-1]+b16[t]*x16[t-1]
  }
  #Prioris
  for(t in 2:N1){
```

```

m.b10[t] <- b10[t-1]
m.b13[t] <- b13[t-1]
m.b16[t] <- b16[t-1]
# los beta se distribuyen con una normal en terminos de media y precision
b10[t] ~ dnorm(m.b10[t],tau.b10)
b13[t] ~ dnorm(m.b13[t],tau.b13)
b16[t] ~ dnorm(m.b16[t],tau.b16)
}

tau.b10~ dgamma(1,0.001)
sigma.b10 <- 1/sqrt(tau.b10)
b010 ~ dnorm(0,0.1)
b10[1] <- b010

tau.b13~ dgamma(1,0.001)
sigma.b13 <- 1/sqrt(tau.b13)
b013 ~ dnorm(0,0.1)
b13[1] <- b013

tau.b16~ dgamma(1,0.001)
sigma.b16 <- 1/sqrt(tau.b16)
b016 ~ dnorm(0,0.1)
b16[1] <- b016

log(mu[1])=n0
n0~dnorm(0,0.001)
ypred[1]=mu[1]
}

data=list(y1=c(y1,rep(NA,3)),N1=N1+3,x10=x10,x13=x13,x16=x16,z=z)
param=c("b10","b13","b16","mu")

resultados1=jags(data=data,parameters.to.save = param,model.file = modelo1,
n.iter=360000,n.thin = 10,n.burnin = 10000,n.chains = 4)

print(resultados1)

# REPRESENTATIVIDAD- Convergencia
library(mcmcplots)
resultados1$BUGSoutput$sims.list$b10
resultados_heidle_mu<-as.mcmc(resultados1$BUGSoutput$sims.list$b10)
heidel.diag(resultados_heidle_mu,pvalue = 0.05)

#pronostico
A=matrix(resultados1$BUGSoutput$summary[1:(43*3),1],ncol=3,byrow=F)
A=A[-1,]
B=cbind(x10[-43],x13[-43],x16[-43])
C1=A*B
y_predichos=exp(apply(C1,1,sum))*z[-1]
mean(abs((y_predichos[40:42]-y[41:43])/y[41:43]))

```

```
y_predichos [40:42]  
  
plot (1:43 ,y , type="l " , col="black ")  
lines (2:43 ,y_predichos , col="red ")
```