Simulación

Trabajo Final

Entrega: 10-Dic-15 de 8:00 a 8:30 AM.

Modelos de Pérdida Agregada en Riesgo de Crédito.

Introducción a los Modelos de Pérdida Agregada

La forma de modelar las pérdidas de una cartera crediticia es muy semejante a la forma en que las aseguradoras modelan las pérdidas que sufren a consecuencia de la realización de los siniestros cubiertos en sus pólizas, las cuales utilizan modelos de pérdidas agregadas; modelos que han sido muy estudiados en lo que se llama la *teoría de riesgo*.

Un *modelo de pérdida agregada*, permite obtener la pérdida total de un conjunto de individuos, para ello existen dos maneras de agregar las pérdidas individuales, a saber: *modelo individual* y *modelo colectivo*.

1. ***Modelo Individual***: En el modelo individual la pérdida agregada se define por la siguiente variable aleatoria:

$$L = \sum_{k=1}^{n} Y_k$$

Donde: "n" es el número total de individuos y "Y_k" es la variable aleatoria del monto de la pérdida incurrida por el k-ésimo individuo (donde el valor de la variable puede ser cero en caso de que el individuo no hubiese registrado pérdida). En este modelo cada una de las "Y_k" pueden tener una distribución diferente.

En el caso de Riesgo de Crédito, dado que la pérdida sólo se registra si el individuo incumple al término del horizonte de tiempo analizado, se tiene que la variable de pérdida individual (Y_k) es una variable aleatoria dicotómica que toma el valor de cero si el crédito no incumple, o toma el valor del monto del crédito en caso de incumplir el acreditado, por lo tanto dicha variable aleatoria queda definida como sigue:

$$Y_k = I_k \; f_k \qquad \text{con:} \qquad I_k \sim \text{Bernoulli}(p_k)$$

Donde:
I_k: Variable indicadora del evento de incumplimiento[1].
p_k: Probabilidad de incumplimiento del k-ésimo acreditado.
f_k: Monto prestado al k-ésimo acreditado.

Para que la variable aleatoria de las pérdidas de la cartera quede bien definida, se debe de considerar una cierta estructura de dependencia entre los incumplimientos de la cartera (i.e. la dependencia entre las variables indicadoras "I_k"), la cual se puede modelar con un modelo en el que se calibre la estructura de dependencia con las correlaciones lineales entre las variables indicadoras (i.e. con $\rho_{i,j} = Corr(I_i, I_j)$; $\forall i \neq j$).

Para obtener la distribución de las pérdidas agregadas, bajo este modelo, se tiene que calcular cada uno de los posibles valores de "L", asociando a cada uno de éstos una probabilidad. Tanto el valor "L", como la probabilidad asociada a ese valor va a estar determinado por la combinación de los posibles valores que pueda tomar cada una de las "n" variables de pérdida individuales ("Y_k").

[1] i.e. variable que toma el valor de uno si el k-ésimo acreditado incumple.

2. ***Modelo Colectivo***: En el modelo colectivo la pérdida agregada se define por la siguiente variable aleatoria:

$$L = \sum_{k=1}^{K} Y_k$$

Donde: "K" es la variable aleatoria del número de individuos que incurrieron en pérdidas y "Y_k" es la variable aleatoria del monto de la pérdida incurrida por el k-ésimo individuo que registró pérdida.

A diferencia del modelo individual, el número de términos en la sumatoria que define a "L" ya no es determinístico y las variables "Y_k" ya no pueden valer cero, puesto que representan los montos de cada una de las pérdidas registradas.

Este es un modelo muy utilizado en el ámbito de los seguros para calcular las indemnizaciones surgidas por la realización de siniestros cubiertos en una cartera de asegurados, por lo que se dice que este modelo sigue un enfoque *actuarial*.

En el caso de Riesgo de Crédito, dado que la pérdida sólo se registra si el individuo incumple al término del horizonte de tiempo analizado, se tiene que la variable aleatoria K se define como la sumatoria de las n variables indicadoras de incumplimiento (I_k), y la colección de las variables de pérdida incurrida (Y_k; $k = 1, \ldots, K$) se definirían como un subconjunto de K elementos seleccionados aleatoriamente (con probabilidades equiprobables de seleccionar cada elemento del conjunto) del conjunto que comprende todos los montos de los n créditos de la cartera (i.e. f_k; $k = 1, \ldots, n$). Esto es:

$$K = \sum_{k=1}^{n} I_k \quad \text{con:} \quad I_k \sim \text{Bernoulli}(p_k)$$

$$(Y_1, \ldots, Y_K) = (f_{D_1}, \ldots, f_{D_K}) \quad \text{con:} \quad \{D_1, \ldots, D_K\} \subset \{1, \ldots, n\}$$

Donde:
I_k: Variable indicadora del evento de incumplimiento[2].
p_k: Probabilidad de incumplimiento del k-ésimo acreditado.
D_k: Variable aleatoria que denota al número de acreditado que incumplió.
f_k: Monto prestado al k-ésimo acreditado.

Al igual que en el *modelo individual*, para que la variable aleatoria del número de incumplimientos de la cartera (K) quede bien definida, se debe de considerar una cierta estructura de dependencia entre los incumplimientos de la cartera (i.e. la dependencia entre las variables indicadoras "I_k"), la cual se puede modelar con un modelo en el que se calibre la estructura de dependencia con las correlaciones lineales entre las variables indicadoras (i.e. con $\rho_{i,j} = Corr(I_i, I_j)$; $\forall i \neq j$).

Modelos de Pérdida Agregada en Riesgo de Crédito con un único grupo homogeneo de riesgo

1. **Modelo Individual**: Si se hace el supuesto de que todos los acreditados comparten perfiles similares de riesgo (i.e. todos pertenecen a un mismo grupo homogéneo de riesgo), entonces se tiene que todos los acreditados comparten las mismas probabilidades de incumplimiento (p) y las mismas correlaciones lineales de incumplimientos (ρ), quedando la variable aleatoria de pérdida definida como:

$$L = \sum_{k=1}^{n} I_k \; f_k \quad \text{con:} \quad I_k \sim \text{Bernoulli}(p) \;\; \forall k \quad \text{y} \quad \rho_{i,j} = \rho \quad \forall i \neq j$$

Una forma eficiente de simular la v.a. L es a través de *variables aleatorias latentes* definidas con un *modelo de factor*. Esto es:

[2]i.e. variable que toma el valor de uno si el k-ésimo acreditado incumple.

$$I_i = 1_{X_i \leq u} \quad \text{con:} \quad X_i = \sqrt{\widetilde{\rho}} Z_0 + \sqrt{1-\widetilde{\rho}} Z_i \quad \text{y con:} \quad \{Z_i\}_{i=0}^m \sim iid\ N(0,1) \tag{1}$$

Con la siguiente calibración:

$$u = \Phi^{-1}(p) \;\; \text{y} \;\; \widetilde{\rho} = \{r : \Phi_{(2)}(u, u; r) = p^2 + \rho \cdot p(1-p)\}$$

Donde $\Phi_{(2)}$ representa la *función de distribución* conjunta de un vector aleatorio bivariado con distribución *normal estándar*.

Por lo tanto, para poder simular una realización de la variable L se necesitaría primeramente calibrar los parámetros u y $\widetilde{\rho}$, luego simular las n *variables latentes* para obtener sus correspondientes *variables indicadoras*, para finalmente multiplicarlas por sus respectivos montos y realizar la sumatoria.

2. **Modelo Colectivo**: Si se hace el supuesto de que todos los acreditados comparten perfiles similares de riesgo (i.e. todos pertenecen a un mismo grupo homogéneo de riesgo), entonces se tiene que todos los acreditados comparten las mismas probabilidades de incumplimiento (p) y las mismas correlaciones lineales de incumplimientos (ρ), quedando la variable aleatoria de pérdida definida como:

$$L = \sum_{k=1}^{K} f_{D_k} \quad \text{con:} \quad K = \sum_{k=1}^{n} I_k \quad \text{y con:} \quad I_k \sim \text{Bernoulli}(p) \;\; \forall k \;\; \text{y} \;\; \{D_1, \ldots, D_K\} \subset \{1, \ldots, n\}$$

Donde $\{D_1, \ldots, D_K\}$ es una *muestra sin remplazo* de K elementos del conjunto $\{1, \ldots, n\}$, donde todos los elementos del conjunto comparten la misma probabilidad de ser seleccionados.

Una forma eficiente de simular el número de incumplimientos (K) es a través de la suma de variables indicadoras de *variables latentes*, donde éstas últimas variables se modelan con un *modelo de factor* (ver ecuación 1), de tal forma que si se condiciona el factor (Z_0) de las n *variables latentes* (X_i's) entonces se tiene que dichas *variables latentes* son mutuamente independientes, propiedad que se le da el nombre de *independencia condicional*, por lo que la suma de las variables indicadoras de estas *variables latentes* condicionadas en su *factor* se tiene que se distribuye *Binomial* (por ser una sumatoria de v.a. *Bernoullis* independientes), con parámetros dados por el número de acreditados (n) y por la probabilidad condicional al valor del *factor* ($p(z_0)$). Esto es:

$$K|Z_0 = z_0 \sim Binomial(n, p(z_0)) \quad \text{con:} \quad p(z_0) = Pr\{I_i = 1|Z_0 = z_0\} = Pr\{X_i \leq u|Z_0 = z_0\} = \Phi\left(\frac{u - \sqrt{\widetilde{\rho}} z_0}{\sqrt{1-\widetilde{\rho}}}\right)$$

Por lo tanto, para poder simular una realización de la variable L se necesitaría primeramente calibrar los parámetros u y $\widetilde{\rho}$ (exactamente de la misma forma en que se hace bajo el *modelo individual*), luego simular el *factor* para poder simular la v.a. K condicionada en el valor del factor simulado, para con ello simular la *muestra sin remplazo* de K elementos tomados del conjunto $\{f_1, \ldots, f_n\}$, para finalmente sumar dichos valores.

Modelos de Pérdida Agregada en Riesgo de Crédito con varios grupos homogeneos de riesgo

1. **Modelo Individual**: Si se hace el supuesto de que los acreditados se pueden agrupar en m grupos homogeneos de riesgo, entonces se tiene que todos los acreditados pertenecientes a un mismo grupo comparten las mismas probabilidades de incumplimiento y las mismas correlaciones lineales de incumplimientos para cualquier pareja de acreditados tomada dentro del mismo grupo, quedando la variable aleatoria de pérdida definida como:

$$L = \sum_{j=1}^{m} L_j \quad \text{con:} \quad L_j = \sum_{k=1}^{n_j} I_k^{(j)} f_k^{(j)}, \quad I_k^{(j)} \sim \text{Bernoulli}(p_j) \quad \text{y} \quad Corr(I_k^{(j)}, I_i^{(j)}) = \rho_j \quad \forall i \neq j$$

Dónde:
j: Subíndice que denota al número de grupo ($j \in \{1, \ldots, m\}$).

n_j: Número de créditos del grupo j.
$I_k^{(j)}$: Variable indicadora del incumplimiento del acreditado k del grupo j.
$f_k^{(j)}$: Monto del crédito del acreditado k del grupo j.
p_j: Probabilidad de incumplimiento de cualquier acreditado del grupo j.
ρ_j: Correlación lineal de incumplimiento para cualquier pareja de acreditados del grupo j (llamada también *correlación intragrupo*).

Además de las *correlaciones intragrupo* de incumplimiento, exiten correlaciones lineales para parejas de acreditados tomadas de grupos diferentes (llamadas también *correlaciones extragrupo*), que también deben de ser consideradas para modelar todas las posibles correlaciones que pueden existir. Esto es:

$$Corr(I_k^{(j)}, I_r^{(i)}) = \rho_{j,i} \quad \forall i \neq j \text{ y } \forall k \neq r$$

Al igual que en el caso de un único grupo homogeneo de riesgo, se tiene que una forma eficiente de simular la v.a. L es a través de *variables latentes* definidas con un *modelo de factor*, pero ahora utilizando dos *factores* (uno correspondiente a la componente *sistémica* y otro a la componente *sectorial*), utilizando una única correlación ($\widetilde{\rho}$) entre cualquier pareja de *variables latentes* que no pertenezcan al mismo grupo para simplificar el modelo, quedando definidas las variables de la siguiente forma:[3]

$$I_i^{(j)} = 1_{X_i^{(j)} \leq u} \quad \text{con:} \quad X_i^{(j)} = \sqrt{\widetilde{\rho}} \cdot Z_0 + \sqrt{\widetilde{\rho_j} - \widetilde{\rho}} \cdot Z_j + \sqrt{1-\widetilde{\rho_j}} \cdot \epsilon_i \quad \text{y con: } \{Z_j\}_{i=0}^m \text{ y } \{\epsilon_i\}_{i=1}^n \sim iid\ N(0,1) \tag{2}$$

Con la siguiente calibración:

$$u_j = \Phi^{-1}(p_j) \text{ y } \widetilde{\rho_j} = \{r : \Phi(u_j, u_j; r) = p_j^2 + \rho_j \cdot p_j(1-p_j)\} \ \forall j \in \{1, \ldots, m\}$$

La calibración de $\widetilde{\rho}$ es mucho más compleja por todas las combinaciones posibles que puede haber de grupos, por lo que la calibración se puede realizar por mínimos cuadrados de las diferencias entre las probabilidades de incumplimiento conjunto de las parejas de diferentes grupos que se pueden armar. Esto es:

$$\widetilde{\rho} = \left\{ r : min \left\{ \sum_{i<j} \left[\Phi(u_i, u_j; r) - \left(p_i p_j + \rho_{i,j} \cdot \sqrt{p_i(1-p_i)p_j(1-p_j)} \right) \right]^2 \right\} \right\}$$

Por lo tanto, para poder simular una realización de la variable L se necesitaría primeramente calibrar los parámetros $\{u_j\}_{j=1}^m$, $\{\widetilde{\rho}\}_{j=1}^m$ y ρ, luego simular para cada grupo las n_j *variables latentes* para obtener sus correspondientes *variables indicadoras*, para finalmente multiplicarlas por sus respectivos montos y realizar la sumatoria.

2. **Modelo Colectivo**: Si se hace el supuesto de que los acreditados se pueden agrupar en m grupos homogéneos de riesgo, entonces se tiene que todos los acreditados pertenecientes a un mismo grupo comparten las mismas probabilidades de incumplimiento y las mismas correlaciones lineales de incumplimientos para cualquier pareja de acreditados tomada dentro del mismo grupo, quedando la variable aleatoria de pérdida definida como:

$$L = \sum_{j=1}^m L_j \text{ con: } L_j = \sum_{k=1}^{K_j} f_{D_k}^{(j)} \text{ con: } K_j = \sum_{k=1}^{n_j} I_k^{(j)}$$

$$\text{y con: } \quad I_k^{(j)} \sim \text{Bernoulli}(p_j) \text{ y } \{D_1, \ldots, D_{K_j}\} \subset \{1, \ldots, n_j\} \ ; \ j = 1, \ldots, m$$

Donde $\{D_1, \ldots, D_{K_j}\}$ es una *muestra sin remplazo* de K_j elementos del conjunto $\{1, \ldots, n_j\}$, donde todos los elementos del conjunto comparten la misma probabilidad de ser seleccionados.

Una forma eficiente de simular el número de incumplimientos de cada grupo (K_j) es a través de la suma de variables indicadoras de *variables aleatorias latentes*, donde éstas últimas variables se modelan con un *modelo de factor*

[3]Nótese que bajo este planteamiento existen $m+1$ correlaciones lineales de las *variables latentes*, una correspondiente a la correlación que que se mapea con las *correlaciones extragrupo* y las otras m con las correlaciones que se mapean con las *correlaciones intragrupo*.

utilizando dos factores (ver ecuación 2), de tal forma que si se condicionan los dos factores (Z_0 y Z_j) de las n *variables latentes* (X_i's) entonces se tiene que dichas *variables latentes* son mutuamente independientes, propiedad que se le da el nombre de *independencia condicional*, por lo que la suma de las variables indicadoras de estas *variables latentes* condicionadas en sus dos *factores* se tiene que se distribuye *Binomial* (por ser una sumatoria de v.a. *Bernoullis* independientes), con parámetros dados por el número de acreditados del grupo que se trate (n_j) y por la probabilidad condicional al valor de los *factores* que correspondan. Esto es:

$$K_j|Z_0 = z_0, Z_j = z_j \sim Binomial(n_j, p(z_0, z_j))$$

con: $$p(z_0, z_j) = Pr\{I_i^{(j)} = 1|Z_0 = z_0, Z_j = z_j\} = Pr\{X_i^{(j)} \leq u_j|Z_0 = z_0, Z_j = z_j\} = \Phi\left(\frac{u_j - \sqrt{\rho}\, z_0 - \sqrt{\widetilde{\rho}_j - \rho}\, z_j}{\sqrt{1 - \widetilde{\rho}_j}}\right)$$

Por lo tanto, para poder simular una realización de la variable L se necesitaría primeramente calibrar los parámetros $\{u_j\}_{j=1}^m$, $\{\widetilde{\rho}\}_{j=1}^m$ y ρ (exactamente de la misma forma en que se hace bajo el *modelo individual*), luego simular el *factor sistémico* y los *factores sectoriales* para poder simular la v.a.'s $\{K_j\}_{j=1}^m$ condicionadas en los valores de los *factores* simulados, para con ello simular la *muestra sin remplazo* de K_j elementos tomados del conjunto $\{f_1, \ldots, f_{n_j}\}$ y sumar dichos valores para obtener la pérdida agregada del grupo (L_j), realizándose lo anterior para cada grupo (i.e. $j = 1, \ldots, m$) para finalmente sumar las pérdidas agregadas de cada uno de los grupos.

Descripción del trabajo:

Se requiere realizar simulaciones de las pérdidas agregadas de una cartera crediticia cuyos acreditados pertenecen a los siguientes grupos homogeneos de riesgo: *industrial*, *construcción*, *comercio* y *servicios*, cuyas *probabilidades de incumplimiento* y *correlaciones intragrupo* se encuentran definidos en el **cuadro 1**, y donde los montos de dichos créditos se encuentran en el archivo **"montos_creditos.xlsx"**. Para lo cual se pide utilizar un *modelo de 2 factores* donde la correlación de cualquier pareja de *variables latentes* que pertenezcan a grupos diferentes sea de **0.5 %** (i.e. no se tiene que calibrar puesto que ya está dada). Utilizando los **dos modelos de pérdida agregada** (*modelo individual* y *modelo colectivo*).

Grupo	**industrial**	**construc.**	**comercio**	**servicios**
p_j	0.70 %	0.90 %	0.65 %	0.60 %
ρ_j	0.09 %	0.04 %	0.05 %	0.07 %

Cuadro 1: Componentes de Riesgo de cada grupo.

Para lo cual se pide lo siguiente:

1. Calcular el tamaño de simulación para estimar la *esperanza* de las pérdidas agregadas de la cartera (L) con una presición de **10.00** y un nivel de confianza del 95 %.
2. Implementar, bajo los **dos modelos de pérdida agregada**, la simulación de L con el tamaño de simulación calculado en (1).
3. Realizar un Análisis Exploratorio de Datos de las L's simuladas, bajo los **dos modelos de pérdida agregada**, realizando lo siguiente:
   a) Histograma y diagrama de caja y brazos (boxplot).
   b) Tabla con media, varianza, coeficiente de sesgo, coeficiente de kurtosis y cuartiles de la distribución empírica simulada.
   c) *VaR* al 95 %
4. Implementar, bajo el *modelo individual*, la técnicas de reducción de la varianza de *variables antitéticas*, calculando el tamaño de la simulación de la misma forma que se hizo en (1), para implementar la simulación que permita estimar la probabilidad de que las pérdidas agregadas de la cartera excedan 2.5 veces su *media teórica*.
5. Realizar una comparación entre el *modelo individual*, el *modelo colectivo* y el *modelo individual* que utiliza las *variables antitéticas*, realizando el comparativo respecto a los siguientes aspectos:
   a) Estimaciones de la característica que se busca estimar en (4).
   b) Tamaños de simulación.
   c) Tiempos de simulación.

Observaciones a considerar:

1. El trabajo a entregar deberá de estructurarse de la siguiente forma:

   a) *Resumen ejecutivo*: Descripción general y resumida del problema y de las metodologías utilizadas (sin entrar a detalles), resumen puntual de los resultados (sin presentar tablas ni gráficos).

   b) *Introducción*: Descripción del problema e introducción del trabajo realizado.

   c) *Metodología*: Descripción detallada de las metodologías utilizadas (supuestos, insumos/parámetros, salidas, pasos a seguir, etc.). Cada metodología deberá de estar claramente referenciada a los códigos que se utilizaron en su implementación los cuales deberán de presentarse en la sección de *Anexos*.

   d) *Resultados*: Presentación detallada de resultados (sustentados con tablas y gráficos).

   e) *Anexos*: Presentación de todos los códigos utilizados en las implementaciones.

2. El trabajo deberá de presentarse impreso.

3. Un aspecto muy relevante a considerarse en la calificación es la **claridad** del documento impreso que se entregue.