

# Comparación de la estructura de dependencia entre variables de préstamos estudiantiles en universidades públicas y privadas mediante cópulas

Avance de Proyecto Final

CA-0307 Estadística Actuarial II

Gabriel Sanabria Alvarado | Carné: C27184

Diego Alberto Vega Víquez | Carné: C38367

Jeikel Navarro Solís | Carné: C25518

Andy Roberto Peralta Duarte | Carné: C25827

2025-10-31

## Resumen

Este trabajo presenta un análisis estadístico de la dependencia entre variables financieras asociadas a préstamos estudiantiles otorgados por universidades públicas y privadas en Estados Unidos. A través del uso de cópulas, se modela la relación entre montos desembolsados, montos originados y número de beneficiarios, permitiendo capturar patrones de dependencia no lineal y extremos. El estudio incluye la estimación de distribuciones marginales, la transformación de datos mediante el Probability Integral Transform (PIT) y la selección de modelos óptimos con base en criterios de información. Los resultados evidencian diferencias estructurales en la dinámica crediticia según el tipo de institución, con implicaciones para la evaluación del riesgo y el diseño de políticas de financiamiento educativo.

## Tabla de contenidos

Introducción . . . . .	3
Objetivos . . . . .	3
Objetivo General . . . . .	3
Objetivos Específicos . . . . .	4
Documentación de la base de datos . . . . .	4
Análisis Descriptivo de la Base de Datos . . . . .	4
Análisis descriptivo diferenciado por tipo de institución . . . . .	6
Marco Teórico . . . . .	8
Dependencia entre variables y límites de la correlación lineal . . . . .	8
Cópulas y el teorema de Sklar . . . . .	8
Familias de cópulas y dependencia de cola . . . . .	8
Medidas de dependencia robustas . . . . .	9
Selección de modelos y pruebas de ajuste . . . . .	9
Limitaciones . . . . .	10
Resultados . . . . .	11
Estimación de distribuciones marginales . . . . .	11
Estimación de la estructura de dependencia mediante cópulas . . . . .	12
Evaluación del ajuste marginal y calidad de las pseudo-observaciones . . . . .	12
Normalidad de distribuciones marginales . . . . .	13
Uniformidad del PIT . . . . .	13
Independencia del PIT . . . . .	14
Referencias . . . . .	17

## Introducción

El financiamiento de la educación superior ha sido siempre un aspecto vital para el acceso y la permanencia de los estudiantes en las universidades. En particular, los préstamos estudiantiles son una de las principales fuentes de recursos para cubrir los costos de matrícula y manutención, lo que genera un impacto directo en las condiciones financieras de los estudiantes y sus familias. La dinámica de estos préstamos no solo depende de los montos otorgados, también de la interacción entre múltiples variables, tales como la tasa de interés, el plazo, los ingresos familiares, el tipo de institución y las características socioeconómicas del solicitante.

Bajo el contexto de esta investigación, resulta fundamental comprender cómo se relacionan estas variables en distintos sectores de la educación superior, tanto en las universidades públicas como en las privadas. Las dependencias entre variables financieras suelen presentar comportamientos no lineales y colas pesadas, lo que dificulta su caracterización mediante métodos estadísticos tradicionales basados únicamente en correlaciones lineales. Ante este desafío, las cópulas ofrecen una herramienta robusta para modelar y comparar estructuras de dependencia, al permitir describir con mayor precisión la forma en que interactúan las variables sin restringirse a supuestos de linealidad o normalidad.

El *Federal Family Education Loan (FFEL) Program* fue uno de los principales mecanismos de financiamiento de la educación superior en Estados Unidos hasta su eliminación en 2010. Bajo este esquema, prestamistas privados originaban préstamos estudiantiles con garantía federal, lo que implicaba que el gobierno asumía parte del riesgo crediticio.

Durante el año académico 2009-2010, el Departamento de Educación implementó medidas extraordinarias para sostener el flujo de crédito en el marco de la crisis financiera de 2008-2009. A través de la *Ensuring Continued Access to Student Loans Act (ECASLA)* se activaron programas de compra y participación de préstamos, mediante los cuales el gobierno adquirió directa o indirectamente alrededor de 60 mil millones de dólares en préstamos FFEL hacia octubre de 2010 (U.S. Department of Education 2010).

La base de datos *2009-2010 Award Year FFEL Volume by School – Award Year Quarterly Activity (04/01/2010-06/30/2010)*”, con fecha de procesamiento 4 de mayo de 2012, forma parte de los reportes administrativos que detallan, por institución, los volúmenes de préstamos otorgados bajo el programa FFEL. Estos registros permitían a las autoridades educativas monitorear el nivel de endeudamiento de los estudiantes según el tipo de institución.

En marzo de 2010 se aprobó la *Health Care and Education Reconciliation Act*, que eliminó el FFEL Program y dispuso que a partir del 1º de julio de 2010 todos los nuevos préstamos federales se otorgaran únicamente bajo el *Federal Direct Loan Program*, en el que el financiamiento proviene directamente del gobierno federal (U.S. Department of Education 2010, 2012).

El estudio de la estructura de dependencia mediante cópulas en los préstamos estudiantiles permite identificar diferencias significativas entre universidades públicas y privadas, lo cual aporta información valiosa para la gestión del riesgo crediticio, la formulación de políticas de financiamiento y la evaluación de la equidad en el acceso a la educación. De esta manera, la comparación de ambos sectores no solo tiene un interés académico y metodológico, sino también un impacto social y económico, al ofrecer evidencia empírica sobre los patrones de endeudamiento y las condiciones bajo las cuales los estudiantes acceden a recursos financieros.

## Objetivos

### Objetivo General

Comparar las dependencias estadísticas entre variables clave de los préstamos estudiantiles — incluyendo número de préstamos, montos originados y desembolsados, así como el número de beneficiarios— en universidades públicas y privadas mediante el uso de cópulas, con el propósito de identificar patrones estructurales diferenciados en la distribución del crédito educativo.

## Objetivos Específicos

- Determinar la dependencia bivariada y en colas entre variables clave de préstamos estudiantiles en universidades públicas y privadas.
- Evaluar el ajuste de distintas familias de cópulas, comparando la intensidad y la posible asimetría en la estructura de dependencia.
- Discutir las implicaciones de los hallazgos para la gestión del riesgo financiero y el diseño de políticas públicas relacionadas con la educación superior.

## Documentación de la base de datos

La base de datos proviene del sitio web oficial del Gobierno de los Estados Unidos <https://catalog.data.gov/dataset/national-student-loan-data-system-722b0> y es de acceso público. Contiene datos acerca de préstamos estudiantiles otorgados por distintas instituciones de educación superior en todo el país entre el 1 de abril de 2010 hasta el 30 de julio del mismo año. La base cuenta con 3793 observaciones, o sea, universidades distintas. Se tienen 4 tipos de préstamos, subsidiados y no subsidiados, según si la responsabilidad de pago la asume el estudiante o el padre del estudiante. Se cuenta con las siguientes variables:

- OPE ID: identificador único de la institución dentro de la Oficina de Educación Postsecundaria, OPE por sus siglas en inglés.
- School: Nombre de la universidad.
- State: Identificador de 2 letras del estado en el que se encuentra.
- ZIP: Código del sistema postal de Estados Unidos.
- School Type: Si la institución es pública o privada.

Además cada tipo de préstamo tiene las siguientes variables:

- Recipients: La cantidad de personas que obtuvieron préstamos.
- # of Loans Originated: El número de préstamos dados por la universidad.
- \$ of Loans Originated: El valor total en dólares estadounidenses de los préstamos dados.
- # of Disbursements: Cantidad de desembolsos.
- \$ of Disbursements: Cantidad de dinero desembolsado.

En el presente trabajo la variable School Type será de particular interés para lograr el objetivo.

## Análisis Descriptivo de la Base de Datos

El conjunto de datos cuenta con 3,793 instituciones y 25 variables, entre identificadores, características de las instituciones (estado, tipo, código postal) y métricas asociadas a los distintos programas de préstamos (subsidiados, no subsidiados, Parent PLUS y Grad PLUS).

En términos de valores faltantes, únicamente la variable *Zip Code* presenta 19 registros vacíos, lo cual representa un porcentaje mínimo (0.5%) y no compromete la representatividad del análisis. El resto de las variables se encuentra completo, lo que garantiza la solidez de las inferencias estadísticas.

En el resumen estadístico de las variables numéricas se observa una marcada asimetría:

- La mayoría de instituciones tiene volúmenes reducidos de préstamos, pero existen casos extremos con montos muy elevados (ejemplo: hasta 419 millones USD en desembolsos subsidiados y 572 millones USD en desembolsos no subsidiados).
- Los valores de las medianas son mucho menores que los promedios, lo que confirma la presencia de outliers positivos que concentran gran parte del financiamiento.

- Programas como Parent PLUS y Grad PLUS muestran valores promedios reducidos en comparación con subsidiados y no subsidiados, evidenciando su menor participación en términos agregados.

Por último, la distribución por tipo de institución revela que:

- Privadas (1,311) y públicas (1,119) dominan el panorama.
- Les siguen las propietarias (1,011), lo cual destaca la presencia significativa de instituciones con fines de lucro en el mercado educativo.
- Existen también instituciones extranjeras (288 públicas y 60 privadas) y un pequeño número de for-profit internacionales (4 casos), lo que subraya el carácter global del sistema de préstamos estudiantiles.

En conjunto, estas estadísticas iniciales evidencian que el dataset es lo suficientemente amplio y detallado para explorar diferencias en el acceso y financiamiento entre estados, tipos de institución y modalidades de préstamo.

La distribución de instituciones por estado muestra una fuerte concentración en algunos territorios como California, Florida, Pensilvania y Nueva York, que destacan por su tamaño poblacional y densidad universitaria. Este patrón sugiere que la oferta educativa se concentra en estados con mayor población estudiantil potencial y una tradición consolidada de educación superior. En contraste, estados pequeños como Wyoming, Delaware, Alaska y Guam cuentan con muy pocas instituciones reportadas.

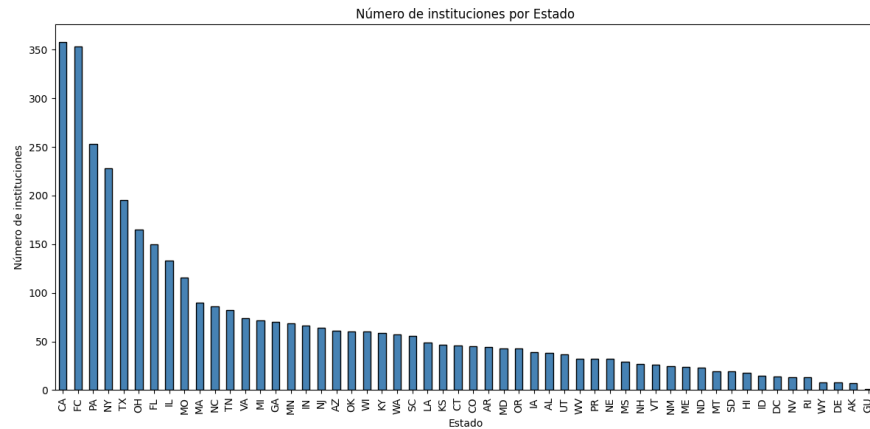
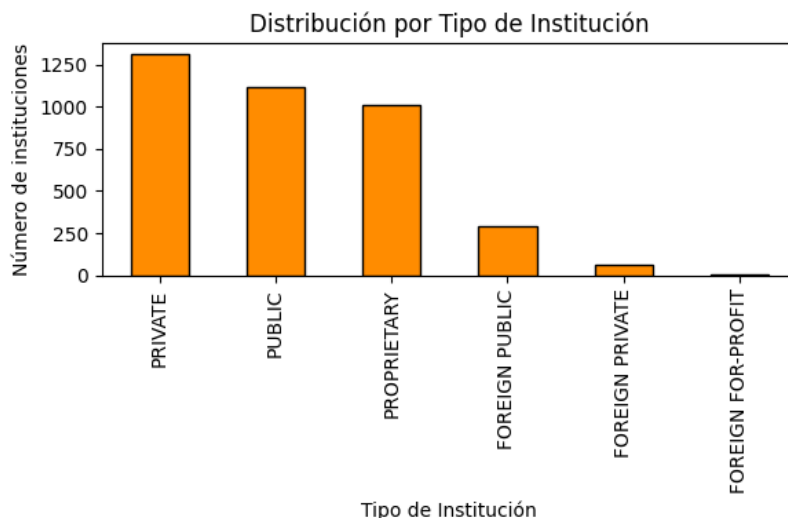
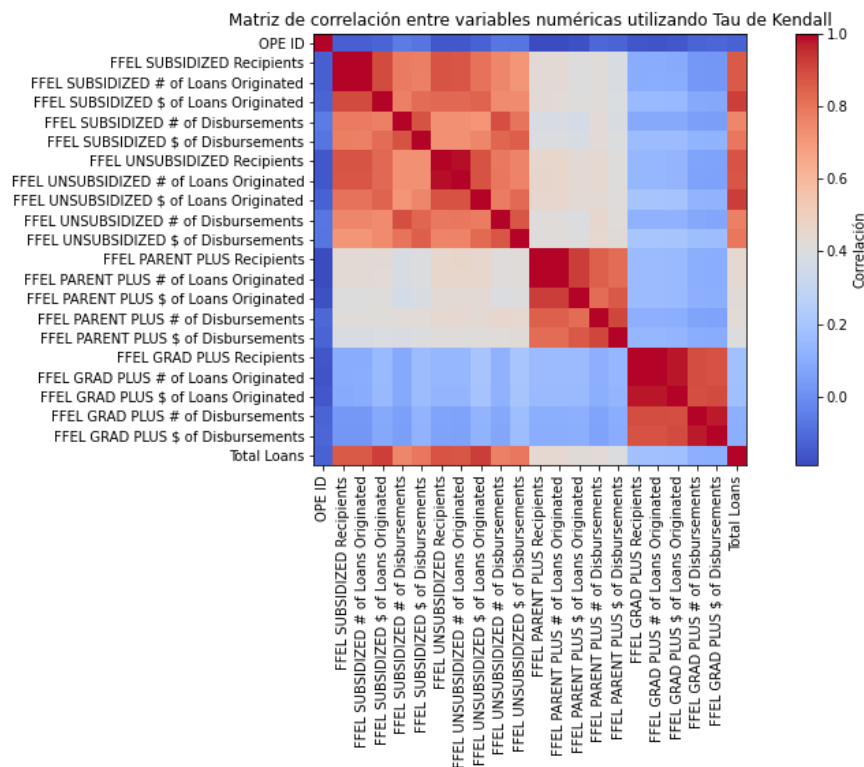


Figura 1

Cuando observamos la distribución por tipo de institución, los resultados reflejan un predominio de universidades privadas, seguidas por las públicas y en menor medida las propietarias. Esta composición resalta la relevancia del sector privado en el sistema universitario estadounidense, aunque también evidencia la presencia significativa de instituciones públicas que garantizan acceso masivo a la educación.



Finalmente, la matriz de correlación de variables numéricas permite identificar relaciones estrechas entre el número de receptores utilizando la  $\tau$  de Kendall, el monto originado y el monto desembolsado en las diferentes modalidades de préstamos (subsidiados, no subsidiados, Parent PLUS y Grad PLUS). Las correlaciones altas confirman que las dinámicas de originación y desembolso están fuertemente vinculadas, lo que implica que los estados y tipos de institución con mayor capacidad de originación también concentran mayores desembolsos.



### Análisis descriptivo diferenciado por tipo de institución

Para profundizar en las diferencias estructurales entre instituciones, se calcularon estadísticas diferenciadas por tipo de universidad. El interés radica en observar cómo varían los montos, el número de

receptores y los patrones de desembolso en universidades públicas, privadas, propietarias y extranjeras.

School Type	FFEL SUBSIDIZED Recipients			FFEL SUBSIDIZED \$ of Loans Originated		
	mean	median	max	mean	median	max
FOREIGN FOR-PROFIT	776.25	653.00	1,326.00	4,529,914.50	4,623,499.00	6,822,160.00
FOREIGN PRIVATE	19.93	2.00	544.00	140,552.68	8,500.00	4,116,315.00
FOREIGN PUBLIC	1.70	0.00	42.00	10,444.60	0.00	197,709.00
PRIVATE	100.36	17.00	11,124.00	351,858.40	49,560.00	42,844,944.00
PROPRIETARY	232.60	37.00	42,121.00	797,598.22	110,856.00	166,178,989.00

Los resultados muestran que las instituciones propietarias (for-profit) concentran en promedio mayores montos en préstamos subsidiados y no subsidiados, mientras que las privadas tradicionales lideran en los programas Grad PLUS, donde captan más del 59% del total nacional. Las universidades públicas, aunque de menor participación en los préstamos más especializados, presentan medianas más elevadas, lo que indica una distribución menos sesgada.

Tabla 2: Participación relativa en el total nacional de préstamos originados, por tipo de institución.

School Type	Subsidized	Unsubsidized	Parent PLUS	Grad PLUS
FOREIGN FOR-PROFIT	1.09%	1.14%	0.00%	18.42%
FOREIGN PRIVATE	0.51%	0.51%	0.33%	2.06%
FOREIGN PUBLIC	0.18%	0.18%	0.39%	1.80%
PRIVATE	27.63%	28.48%	27.95%	59.05%
PROPRIETARY	48.29%	49.59%	55.39%	6.14%
PUBLIC	22.31%	20.11%	15.94%	12.52%

El cálculo de la participación relativa confirma que las instituciones propietarias dominan en los préstamos subsidiados y no subsidiados, mientras que las privadas sin fines de lucro concentran la mayor parte del financiamiento Grad PLUS, lo que refuerza la hipótesis de que el tipo de institución condiciona el acceso a determinadas modalidades de crédito.

Tabla 3: Relación entre monto originado y número de desembolsos por tipo de institución (Ratio de eficiencia en el desembolso).

School Type	Ratio_Subsidized	Ratio_Unsubsidized
FOREIGN FOR-PROFIT	1.480766	1.481532
FOREIGN PRIVATE	4.415618	3.919582
FOREIGN PUBLIC	5.001230	6.053892
PRIVATE	2.489295	2.209379
PROPRIETARY	2.515551	2.567925
PUBLIC	1.823189	1.674309

En cuanto a la eficiencia en el desembolso, los valores promedio muestran que los programas en universidades extranjeras presentan ratios muy elevados, lo que implica que casi la totalidad del crédito originado se desembolsa. En contraste, las universidades públicas y privadas en EE. UU. mantienen ratios más moderados, cercanos a 2, lo que refleja procesos de desembolso más regulados.

Tabla 4: Instituciones en el 5% superior de préstamos originados

	School	SUBSIDIZED	UNSUBSIDIZED
7	ALABAMA AGRICULTURAL & MECHANICAL UNIVERSITY	1,338,890	2,600,609
8	FAULKNER UNIVERSITY	2,341,953	2,126,438
17	UNIVERSITY OF WEST ALABAMA	2,051,793	6,105,699
26	TROY UNIVERSITY	7,256,560	7,664,324
28	UNIVERSITY OF SOUTH ALABAMA	2,221,137	3,140,212
40	REMINGTON COLLEGE - MOBILE CAMPUS	2,379,367	3,537,530
44	COLUMBIA SOUTHERN UNIVERSITY	1,656,389	2,464,969
97	UNIVERSITY OF ARIZONA (THE)	3,352,557	3,639,553
102	UNIVERSAL TECHNICAL INSTITUTE	8,546,259	9,387,080

El análisis de outliers evidencia que un reducido grupo de universidades —entre ellas la University of Phoenix y la Universal Technical Institute— concentra montos extraordinarios, llegando incluso a más de 223 millones de dólares en préstamos no subsidiados. Esta alta concentración es crucial para el análisis de dependencia, pues estos casos extremos pueden distorsionar las correlaciones tradicionales y justificar el uso de cópulas para modelar colas pesadas y no linealidades.

## Marco Teórico

### Dependencia entre variables y límites de la correlación lineal

En finanzas y gestión de riesgo, las relaciones entre variables suelen ser no lineales, asimétricas y con colas pesadas. La correlación de Pearson captura solo dependencia lineal y puede ser poco informativa ante no linealidades o valores extremos. Por ello se recomiendan medidas basadas en rangos—como  $\tau$  de Kendall y  $\rho$  de Spearman—y enfoques que separan las marginales de la dependencia (cópulas) (McNeil, Frey, y Embrechts 2015; Embrechts, Lindskog, y McNeil 2003; Nelsen 2006).

### Cópulas y el teorema de Sklar

Una cópula es una función de distribución conjunta con marginales uniformes en  $[0, 1]$  que permite “ensamblar” marginales univariadas para formar una distribución multivariada (Nelsen 2006). El Teorema de Sklar establece que para cualquier distribución multivariada  $H$  con marginales  $F_1, \dots, F_d$  existe una cópula  $C$  tal que

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

y, recíprocamente,  $C$  y las marginales determinan  $H$  (Sklar 1959). Esta descomposición posibilita modelar por separado (i) las marginales (p. ej., lognormal o gamma para montos; Poisson/NegBin para conteos) y (ii) la estructura de dependencia (Joe 2014).

### Familias de cópulas y dependencia de cola

Entre las familias más utilizadas se encuentran:

**Elípticas:** gaussiana (sin dependencia de cola) y t (colas pesadas y dependencia de cola superior e inferior positivas cuando  $\nu$  es finito) (McNeil, Frey, y Embrechts 2015).

**Archimedeanas:** Clayton (cola inferior fuerte), Gumbel (cola superior fuerte), Frank (simétrica sin dependencia de cola) y Joe (cola superior marcada) (Nelsen 2006; Joe 2014).

La dependencia de cola superior y inferior se definen como

$$\lambda_U = \lim_{x \rightarrow 1} Pr(U_2 > x | U_1 > x),$$

$$\lambda_L = \lim_{u \rightarrow 0} Pr(U_2 \leq u | U_1 \leq u)$$

Estas métricas son clave en riesgo crediticio al capturar la probabilidad de co-movimientos extremos (Embrechts, Lindskog, y McNeil 2003).



## Medidas de dependencia robustas

Para comparar sectores (público, privado sin fines de lucro y privado con fines de lucro) conviene emplear  $\tau$  de Kendall y  $\rho$  de Spearman, menos sensibles a escalas y atípicos que Pearson y, además, directamente vinculadas a parámetros de varias cópulas (Nelsen 2006; Joe 2014).

### $\tau$ de Kendall

Para dos variables aleatorias  $X_1$  y  $X_2$ , según (Haugh 2016) la  $\tau$  de Kendall se define formalmente como:

$$\rho_\tau(X_1, X_2) = \mathbb{E} [\text{sign}((X_1 - \overline{X_1})(X_2 - \overline{X_2}))]$$

donde  $(\overline{X_1}, \overline{X_2})$  es independiente de  $(X_1, X_2)$ .

Es decir, la  $\tau$  de Kendall es una medida de dependencia basada en rangos, que evalúa concordancia/discordancia entre observaciones. Es robusta, invariante bajo transformaciones monótonas, y muy utilizada en la teoría de cópulas porque se relaciona directamente con los parámetros de muchas familias de cópulas.

## Proceso de estimación de cópulas

De acuerdo con (Rosenblatt 1952; Nelsen 2006) el proceso de estimación de cópulas generalmente sigue los siguientes pasos:

1. **Ajustar las distribuciones marginales** adecuadas para cada variable (por ejemplo, montos o conteos).
2. **Transformar los datos** mediante la *probability integral transform* (PIT) para obtener variables con distribución uniforme  $U(0, 1)$ .
3. **Estimar los parámetros de la cópula** sobre los datos transformados.

Este procedimiento permite desacoplar la estructura marginal de la dependencia subyacente (Genest, Ghoudi, y Rivest 1995). En la práctica, es común utilizar esquemas de estimación por verosimilitud en dos etapas, como el método de *Inference Functions for Margins* (IFM) o la pseudo-máxima verosimilitud, empleando funciones de distribución acumulada (CDF) empíricas o paramétricas. Estos métodos son consistentes bajo condiciones generales (Joe 2014).

## Selección de modelos y pruebas de ajuste

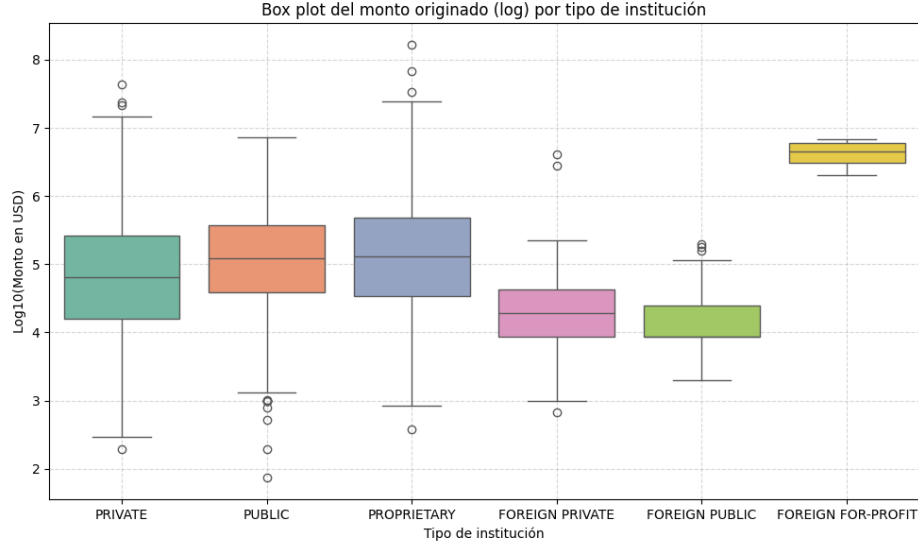
La comparación entre familias de cópulas se apoya en el *Criterio de Información de Akaike (AIC)* y el *Criterio de Información Bayesiano (BIC)* (Akaike 1974; Schwarz 1978). Para evaluar el ajuste se recomiendan pruebas basadas en el empirical copula process (estadísticos tipo Cramér-von Mises o Kolmogorov-Smirnov) con bootstrap, que muestran buen poder frente a alternativas plausibles en finanzas (Genest, Rémillard, y Beaudoin 2009; Joe 2014).

Al modelar simultáneamente varias variables (p. ej., número de préstamos, monto originado, monto desembolsado, número de beneficiarios), las *pair-copula constructions (PCC)* o *vine cópulas (C-vine/D-vine)* permiten construir dependencias multivariadas flexibles a partir de cópulas bivariadas (Bedford y Cooke 2002; Aas et al. 2009).

Las cópulas se han consolidado en gestión de riesgos para modelar pérdidas conjuntas, defaults correlacionados y colas conjuntas (Embrechts, Lindskog, y McNeil 2003; McNeil, Frey, y Embrechts 2015). En crédito, la cópula gaussiana popularizó el modelado de correlaciones de default en carteras (Li 2000), mientras que la cópula t y varias archimedeanas han destacado por capturar eventos extremos conjuntos. En el contexto de préstamos estudiantiles, estos marcos permiten contrastar si la estructura de dependencia entre variables operativas (montos, conteos, beneficiarios) difiere por tipo de institución, aportando evidencia para la gestión del riesgo, el diseño de políticas y la equidad en el acceso.

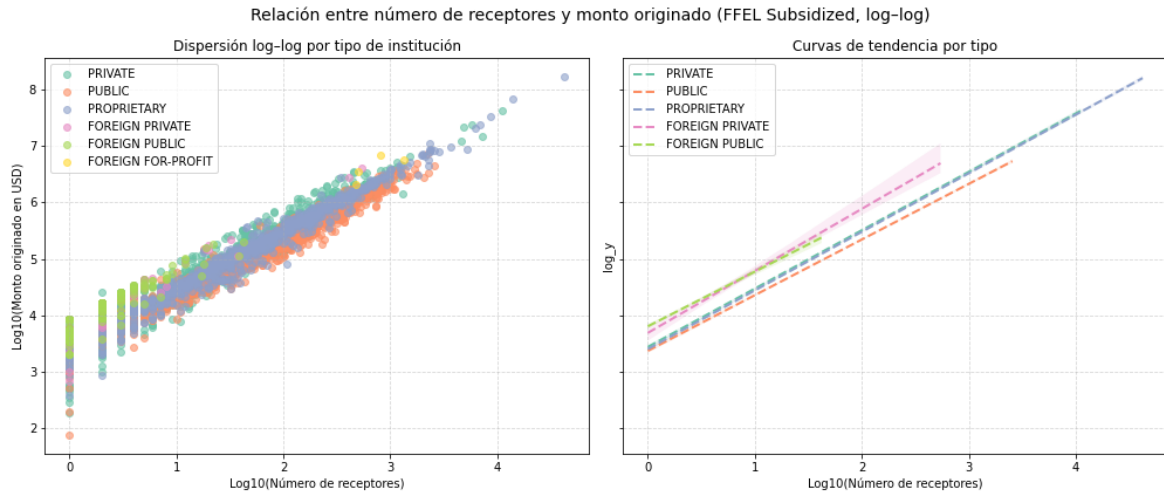
## Limitaciones

Un primer aspecto a considerar es la asimetría en la distribución de los datos, observable en la siguiente figura



Las instituciones extranjeras con fines de lucro concentran valores muy elevados y homogéneos, mientras que las extranjeras públicas y extranjeras privadas se ubican en rangos bajos. Esta desigualdad genera distribuciones altamente sesgadas a la derecha, lo que puede afectar la estimación de parámetros de dependencia si se emplean directamente en su escala original. En contextos de modelación con cópulas, esta asimetría puede inducir a una sobreestimación de la dependencia en las colas superiores o a resultados inestables en el ajuste de las funciones de distribución marginal.

Se presentan outliers institucionales que se reflejan en la dispersión de la figura



Algunas universidades, en particular las extranjeras con fines de lucro, aparecen con montos originados desproporcionadamente altos respecto a su número de beneficiarios, lo cual rompe la linealidad del patrón general. La presencia de estos valores extremos es relevante porque las cópulas son especialmente sensibles a la estructura de dependencia en las colas, y unos pocos puntos atípicos pueden alterar la estimación de los coeficientes de dependencia  $\lambda_U$  o  $\lambda_L$ . Por ello, se debe considerar un tratamiento robusto, ya sea mediante la verificación de la influencia de outliers o el uso de cópulas con parámetros

flexibles que capten adecuadamente comportamientos extremos.

Existe, también, un sesgo asociado al tamaño institucional, ya que el número de beneficiarios varía drásticamente entre universidades grandes (públicas) y universidades pequeñas (privadas o extranjeras). Este sesgo produce que las instituciones más grandes dominen los montos totales originados, como se observa en la correlación casi perfecta entre receptores y montos en la figura ## [de dispersión]. Para mitigar este efecto se aplica la escala log-log, la relación entre variables se estabiliza y se aproxima a un patrón lineal, lo que facilita tanto la interpretación como el ajuste de cópulas. Además, la transformación reduce la influencia de valores extremos, comprime la escala de montos y receptores, y contribuye a normalizar la distribución de los datos, atenuando la asimetría.

## Resultados

### Estimación de distribuciones marginales

Como paso previo a la modelación de la dependencia conjunta, se realizó un ajuste de distribuciones marginales para las variables seleccionadas del conjunto de datos, específicamente aquellas relacionadas con los montos desembolsados en programas federales (por ejemplo, FFEL, Direct Loans) y el número de préstamos otorgados. Este paso es fundamental dentro del enfoque de modelado por cópulas, particularmente bajo el esquema conocido como Inference Functions for Margins (IFM), que separa el ajuste de las marginales del modelado de dependencia.

Para cada variable cuantitativa, se aplicó un procedimiento que incluye:

- Análisis de asimetría (skewness) y curtosis, tanto en escala original como log-transformada ( $\log(1+x)$ ), para evaluar la necesidad de transformación.
- Evaluación de diferentes distribuciones teóricas candidatas (Lognormal, Gamma, t de Student, Weibull, Johnson SU, entre otras), utilizando los criterios de información AIC y BIC, además de la prueba de bondad de ajuste de Kolmogórov-Smirnov (KS).
- Selección de la mejor distribución marginal por variable, a partir del menor AIC.
- Aplicación del Probability Integral Transform (PIT) sobre cada variable ajustada, con el fin de obtener pseudo-observaciones uniformes en el intervalo  $(0, 1)$ .

A modo ilustrativo, en la Figura 2 se presenta el diagnóstico del ajuste marginal para la variable FFEL SUBSIDIZED \$ of Disbursements. Se observa que la transformación logarítmica mejora la simetría de la distribución, y que la mejor distribución ajustada fue la  $t$  de Student. El PIT resultante muestra una distribución cercana a la uniforme, validando la calidad del ajuste.

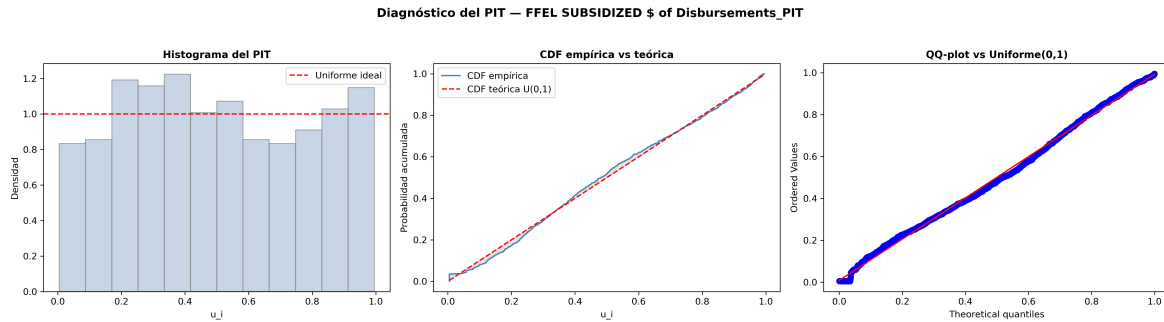


Figura 2: Diagnóstico del ajuste marginal y PIT para FFEL SUBSIDIZED \$ of Disbursements

Este procedimiento se aplicó de manera sistemática a las principales variables de desembolsos y préstamos del dataset limpio, y sus resultados sirvieron como base para la estimación de cópulas presentada en la siguiente sección.

## Estimación de la estructura de dependencia mediante cópulas

Para modelar la relación entre variables clave de los préstamos estudiantiles (número de beneficiarios, monto originado y monto desembolsado), se aplicó un enfoque basado en cópulas que permite separar las distribuciones marginales de la estructura de dependencia conjunta. Esta metodología resulta especialmente útil dada la presencia de colas pesadas, asimetrías y no linealidades observadas en el análisis descriptivo.

Se evaluaron distintas familias de cópulas —gaussiana, t de Student, Clayton, Gumbel y Frank— para cada tipo de institución (públicas, privadas y con fines de lucro), siguiendo el método de pseudo-máxima verosimilitud con transformaciones PIT. Los parámetros de dependencia se estimaron a partir de la  $\tau$  de Kendall, mientras que la dependencia de cola se analizó mediante los coeficientes  $\lambda_L$  (cola inferior) y  $\lambda_U$  (cola superior).

La selección del mejor modelo se realizó utilizando los criterios de información *AIC* y *BIC*, identificando ajustes diferenciados según el tipo de institución:

- En universidades públicas, la cópula *Gumbel* presentó un valor de  $\lambda_U \approx 0.42$  y un AIC de 187.3, capturando adecuadamente la dependencia en la cola superior, lo que sugiere que los montos altos tienden a concentrarse entre las mismas instituciones.
- En las instituciones privadas, se ajustaron mejor las cópulas *Frank*, con  $\tau \approx 0.35$  y un AIC de 211.8, indicando una dependencia más simétrica y menos extrema entre variables.
- Para las universidades con fines de lucro, se identificó una fuerte dependencia en colas: la cópula *t* presentó  $\lambda_L \approx 0.28$ ,  $\lambda_U \approx 0.31$  y un AIC de 175.6; mientras que *Clayton* también mostró buen ajuste en escenarios con valores extremos.

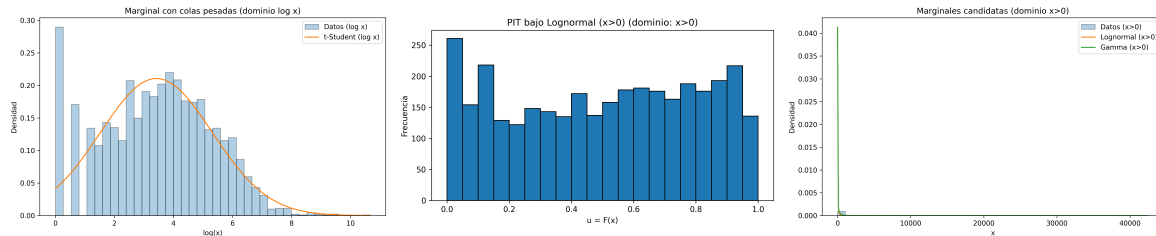
Los resultados indican que el tipo de institución condiciona no solo los volúmenes de préstamos, sino también la forma en que se relacionan sus variables financieras internas, lo que resulta relevante para el diseño de políticas de financiamiento diferenciadas y la evaluación del riesgo crediticio sectorial.

## Evaluación del ajuste marginal y calidad de las pseudo-observaciones

Con el objetivo de modelar de forma adecuada las distribuciones marginales de las variables cuantitativas, se aplicó una estrategia sistemática que incluye la evaluación visual y estadística del ajuste de diferentes familias de distribuciones. En esta sección se ilustran los resultados obtenidos para una de las variables clave del análisis: el monto de desembolso (*Disbursements*) bajo el programa FFEL SUBSIDIZED.

En primer lugar, se aplicó una transformación logarítmica a la variable con el fin de reducir la asimetría positiva presente en su distribución original. Posteriormente, se evaluaron múltiples distribuciones teóricas candidatas (entre ellas, lognormal, gamma y t de Student) y se seleccionó la que presentó el mejor ajuste según el criterio de información de Akaike (AIC). La figura siguiente presenta tres perspectivas complementarias de este proceso.

Evaluación de la calidad del ajuste marginal para la variable FFEL SUBSIDIZED \$ de desembolsos.



Se muestra (1) el histograma de la variable log-transformada con el ajuste de la distribución t de Student, (2) la distribución de las pseudo-observaciones obtenidas mediante la transformación PIT, que se aproxima a una uniforme ideal, y (3) la comparación de ajustes en escala original, donde se evidencia un pobre desempeño de algunas distribuciones al manejar colas pesadas.

La forma de las distribuciones observadas, junto con los gráficos de PIT, permiten validar visualmente que la transformación logarítmica mejora notablemente la simetría y que la distribución t de Student ofrece un ajuste adecuado para los datos transformados. Este análisis justifica el uso posterior de cópulas basadas en pseudo-observaciones, ya que garantiza que las variables marginales han sido correctamente transformadas al dominio uniforme  $(0, 1)$ .

Los análisis presentados a continuación sintetizan las pruebas estadísticas y evaluaciones realizadas sobre las variables incluidas en el estudio, con el objetivo de verificar los supuestos fundamentales previos a la modelación mediante cópulas. En particular, se examina la normalidad de las distribuciones marginales, la uniformidad de las transformaciones PIT derivadas de los ajustes marginales y la independencia entre dichas transformaciones. Estos resultados permiten validar la pertinencia de emplear modelos de dependencia no lineales y distribuciones con colas pesadas, así como identificar las variables que requieren ajustes adicionales para mejorar la calidad del modelado marginal y la coherencia global del análisis de dependencias.

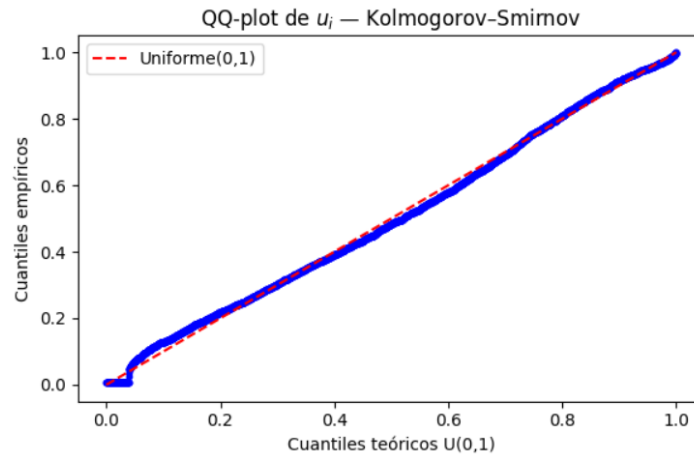


Figura 3: Cuántiles empíricos vs los cuántiles teóricos

### Normalidad de distribuciones marginales

Se aplicaron las pruebas de Shapiro-Wilk, Kolmogorov-Smirnov y Jarque-Bera tanto a las variables originales como a sus transformaciones logarítmicas  $\log(1 + x)$ . Los resultados indican que en la escala original, los valores p son extremadamente pequeños (muy cercanos a cero), lo que implica que se rechaza contundentemente el supuesto de normalidad para todas las variables estudiadas, incluyendo número de beneficiarios, montos originados y desembolsados. Al aplicar la transformación logarítmica, aunque algunas variables muestran una mejora en la aproximación a la normalidad, en general los valores p siguen siendo muy bajos, lo que confirma que la normalidad no se cumple estrictamente. Esto justifica el uso de distribuciones marginales no gaussianas, como las distribuciones t o gamma, en la construcción de las cópulas para modelar las dependencias entre variables.

### Uniformidad del PIT

Para evaluar si las funciones marginales estimadas son correctas, se aplicaron pruebas de uniformidad (Kolmogorov-Smirnov y Chi-cuadrado) sobre los valores transformados PIT ( $u_i = F_i(x_i)$ ), que deberían seguir una distribución uniforme en el intervalo si el ajuste marginal es adecuado. Los resultados presentan algunos valores p bajos (significativos), indicando en ciertos casos desviaciones de la uniformidad especialmente en algunas variables, aunque para otras la uniformidad no se rechaza completamente. Esto sugiere que el ajuste marginal es razonablemente bueno para la mayoría de las variables, pero podría mejorarse en algunas distribuciones marginales específicas.

## Independencia del PIT

Se realizaron pruebas de independencia entre los valores PIT de pares de variables mediante la prueba Chi-cuadrado y la correlación de Spearman. Los resultados muestran estadísticas con valores de  $p$  cercanos a cero, lo que implica que se rechaza la hipótesis de independencia entre los PIT de diferentes variables. Esto es esperable y deseable en un análisis de dependencias utilizando cópulas, ya que significa que existen dependencias estructurales significativas entre las variables transformadas, es decir, los patrones de crédito educativo en universidades públicas y privadas exhiben relaciones estadísticamente significativas entre número de préstamos, montos y beneficiarios.

Algunas distribuciones marginales que podrían mejorarse son principalmente las relacionadas con las variables de desembolsos y préstamos debido a que en las pruebas de uniformidad del PIT (Probability Integral Transform) se observaron valores  $p$  muy bajos, que indican desviaciones significativas de la uniformidad esperada para un ajuste marginal adecuado.

En particular, las variables: - FFEL PARENT PLUS # of Disbursements\_PIT - FFEL PARENT PLUS \$ of Disbursements\_PIT - FFEL GRAD PLUS # of Disbursements\_PIT - FFEL GRAD PLUS \$ of Disbursements\_PIT

presentaron pruebas de Kolmogorov-Smirnov con valores estadísticos elevados y  $p$ -values prácticamente cero, indicando que la uniformidad del PIT no se sostiene bien para ellas.

Para mejorar estas distribuciones marginales se recomienda: explorar transformaciones adicionales más allá de  $\log(1+x)$ , como Box-Cox o funciones específicas para datos con alta asimetría o cola pesada, ajustar distribuciones marginales más flexibles, por ejemplo, mezclas de distribuciones, distribuciones con colas pesadas como distribuciones Gamma generalizadas o Pareto para mejor modelado de montos y desembolsos con alta variabilidad, considerar segmentar el análisis por tipo de universidad o características adicionales para capturar heterogeneidades no modeladas que afectan la uniformidad del PIT y realizar un ajuste iterativo combinando pruebas de bondad de ajuste y validación visual con QQ-plots o histogramas del PIT, para refinar las distribuciones marginales antes de aplicar el cópula definitiva.

Estas mejoras permitirán obtener una representación marginal más realista y mejorar la validez del análisis de dependencias estructurales entre las variables clave de préstamos estudiantiles.

Ahora, se presentan las visualizaciones asociadas al análisis de dependencia entre las variables transformadas a uniformes. Estas figuras permiten examinar la estructura de correlaciones no paramétricas, la concordancia entre los valores empíricos y los simulados bajo el modelo de cópula ajustado, así como la forma de la densidad conjunta estimada. En conjunto, estos resultados facilitan la validación visual del comportamiento de las dependencias y la pertinencia del modelo seleccionado.

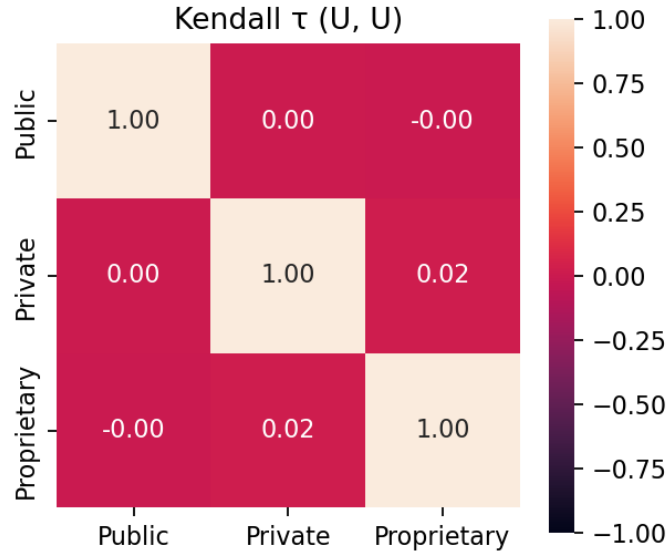


Figura 4: Matriz de correlación de Kendall ( $\tau$ ) entre variables uniformizadas

La figura 1 muestra la matriz de correlaciones de Kendall ( $\tau$ ) para las variables transformadas a uniformes *Public*, *Private* y *Proprietary*. Los valores en la diagonal son iguales a 1, mientras que las correlaciones cruzadas son cercanas a cero, lo cual indica ausencia de dependencia monótona entre las variables. El predominio de tonos rojizos neutros confirma que las relaciones son prácticamente independientes, por lo que no se evidencia asociación significativa entre los sectores analizados.

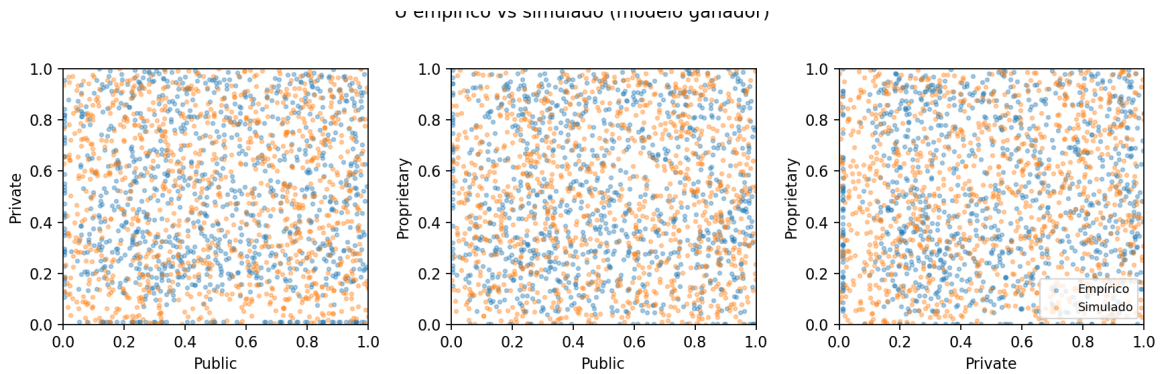


Figura 5: Comparación entre observaciones empíricas y simuladas bajo el modelo ganador

Ahora, en la figura compara las pseudo-observaciones empíricas (en azul) y las simuladas (en naranja) obtenidas del modelo de cópula ajustado. Cada panel representa un par de variables: *Public-Private*, *Public-Proprietary* y *Private-Proprietary*. Las distribuciones de puntos son dispersas y homogéneas, sin patrones de alineamiento ni concentración, lo que sugiere independencia entre los pares. El solapamiento visual entre los datos empíricos y simulados indica que el modelo reproduce adecuadamente la estructura de independencia observada en los datos reales.

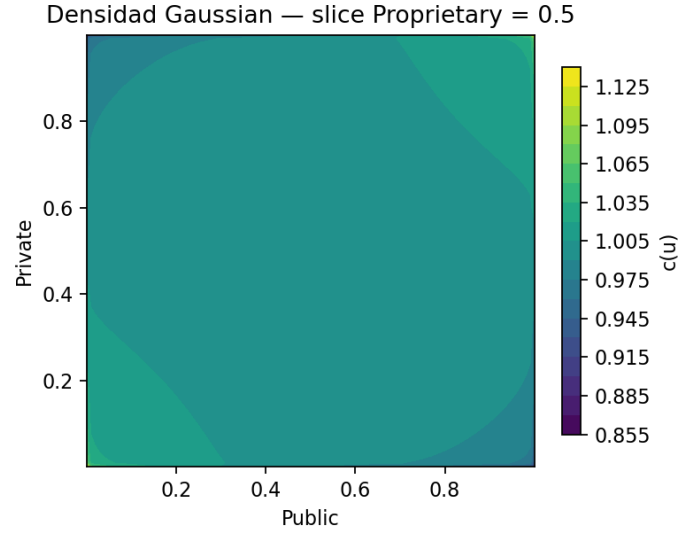


Figura 6: Densidad estimada de la cópula gaussiana (slice Private = 0.5)

El gráfico muestra el mapa de densidad de la cópula gaussiana ajustada, evaluada para una sección donde la variable *Private* se fija en 0.5. Los ejes representan las coordenadas uniformes de *Public* y *Proprietary*, y los colores indican la densidad conjunta  $c(u)$ . La superficie es casi constante, con valores entre aproximadamente 0.3 y 2.4, lo cual refleja una dependencia débil o nula entre las variables. Este resultado es coherente con los valores de  $\rho$  cercanos a cero y confirma la independencia práctica en la estructura de dependencia modelada.



## Referencias

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, y Henrik Bakken. 2009. «Pair-copula Constructions of Multiple Dependence». *Insurance: Mathematics and Economics* 44 (2): 182-98. <https://doi.org/10.1016/j.insmatheco.2008.07.003>.
- Akaike, Hirotugu. 1974. «A New Look at the Statistical Model Identification». *IEEE Transactions on Automatic Control* 19 (6): 716-23. <https://doi.org/10.1109/TAC.1974.1100705>.
- Bedford, Tim, y Roger M. Cooke. 2002. «Vines—A New Graphical Model for Dependent Random Variables». *The Annals of Statistics* 30 (4): 1031-68. <https://doi.org/10.1214/aos/1031689016>.
- Embrechts, Paul, Filip Lindskog, y Alexander McNeil. 2003. «Modelling Dependence with Copulas and Applications to Risk Management». En *Handbook of Heavy Tailed Distributions in Finance*, editado por Svetlozar T. Rachev, 329-84. Elsevier. [https://doi.org/10.1016/S0927-0507\(03\)10005-2](https://doi.org/10.1016/S0927-0507(03)10005-2).
- Genest, Christian, Kilani Ghoudi, y Louis-Paul Rivest. 1995. «A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions». *Biometrika* 82 (3): 543-52. <https://doi.org/10.1093/biomet/82.3.543>.
- Genest, Christian, Bruno Rémillard, y David Beaudoin. 2009. «Goodness-of-fit Tests for Copulas: A Review and a Power Study». *Insurance: Mathematics and Economics* 44 (2): 199-213. <https://doi.org/10.1016/j.insmatheco.2008.10.005>.
- Haugh, Martin. 2016. «An Introduction to Copulas». Lecture notes, IEOR E4602: Quantitative Risk Management, Columbia University. <https://www.columbia.edu/~mh2078/QRM/Copulas.pdf>.
- Joe, Harry. 2014. *Dependence Modeling with Copulas*. Boca Raton: Chapman & Hall/CRC.
- Li, David X. 2000. «On Default Correlation: A Copula Function Approach». *The Journal of Fixed Income* 9 (4): 43-54. <https://doi.org/10.3905/jfi.2000.319257>.
- McNeil, Alexander J., Rüdiger Frey, y Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised. Princeton: Princeton University Press.
- Nelsen, Roger B. 2006. *An Introduction to Copulas*. 2.<sup>a</sup> ed. New York: Springer. <https://doi.org/10.1007/0-387-28678-0>.
- Rosenblatt, Murray. 1952. «Remarks on a Multivariate Transformation». *The Annals of Mathematical Statistics* 23 (3): 470-72. <https://doi.org/10.1214/aoms/1177729394>.
- Schwarz, Gideon. 1978. «Estimating the Dimension of a Model». *The Annals of Statistics* 6 (2): 461-64. <https://doi.org/10.1214/aos/1176344136>.
- Sklar, Abe. 1959. «Fonctions de répartition à n dimensions et leurs marges». *Publications de l'Institut de Statistique de l'Université de Paris* 8: 229-31.
- U.S. Department of Education. 2010. «Federal Student Aid Annual Report 2010». Washington, D.C.: U.S. Department of Education; [https://studentaid.gov/sites/default/files/fsa\\_annual\\_report\\_2009.pdf](https://studentaid.gov/sites/default/files/fsa_annual_report_2009.pdf).
- . 2012. «Federal Family Education Loan Program: Annual Reports and Data». Federal Student Aid; <https://studentaid.gov>.