

Proyecto de Riesgo de Liquidez



Integrantes:

Leonardo Daniel Rosas Ríos
María Fernanda Hernández Fernández
Martha Lizeth Pastrana Basilio
José Domingo Cortes Sandria

Profesor:

Jorge Luis Reyes García

Tabla de contenido

Introducción	3
Productos Bancarios para el Ahorro	3
Pagares Bancarios.....	3
Certificados de depósito	3
Fondos de inversión	3
Cuentas de Ahorro	4
Descripción de la base de datos y variables utilizadas	4
Base de datos	4
Periodo de análisis	4
Variables utilizadas	4
Modelos de regresión utilizados	5
Regresión Lineal Múltiple.....	5
Regresión Lasso (Least Absolute Shrinkage and Selection Operator)	5
Regresión Ridge	6
Regresión Elastic Net	6
Estandarización de variables.....	6
Validación cruzada	6
Métricas utilizadas	7
Software utilizado.....	7
Resultados.....	8
Tabla comparativa.....	8
Conclusiones	10
Aprendizajes obtenidos	11

Introducción

Las cuentas de captación constituyen uno de los pilares fundamentales del sistema bancario, ya que representan los recursos que las instituciones financieras reciben de personas o empresas y que posteriormente canalizan hacia actividades productivas o de consumo.

El objetivo de este proyecto es desarrollar un modelo macroeconómico de regresión múltiple, para analizar y explicar el efecto que diferentes variables macroeconómicas tienen en las cuentas de captación (Depósito Vista, Depósitos a plazo y Captación tradicional). Con este proyecto se desea identificar las variables que tienen mayor influencia sobre el comportamiento de las cuentas de captación, generando la posibilidad de identificar patrones y predecir cambios en el futuro.

Productos Bancarios para el Ahorro

Los bancos ofrecen una gran variedad de productos destinados a satisfacer la necesidad de las personas de guardar su dinero de forma segura, generando intereses o rendimientos con bajo riesgo, para poder ser usado en un futuro. Los principales productos para el ahorro son: los pagarés, certificados de depósito, fondos de inversión, cuentas de ahorro, cuentas a la vista y seguros de ahorro, etc. Cada uno con diferentes características y beneficios.

Pagares Bancarios

Estos son instrumentos a corto y mediano plazo que ofrecen una tasa fija de interés. Es ideal para ahorradores que buscan una rentabilidad predecible y clara, a cambio de mantener su inversión durante un plazo previamente acordado.

Características

- Tasa fija.
- Plazo fijo (generalmente entre 30 y 365 días).
- Penalización por retiros anticipados.

Certificados de depósito

Son productos de ahorro a plazo con una tasa de interés fija que asegura ganancias por la permanencia del capital invertido durante un periodo previamente acordado.

Características

- Tasa de interés atractiva.
- Plazos variables, superiores a los pagarés.
- Ideal para plazos a largo y mediano plazo.

Fondos de inversión

Es una herramienta que permite a los ahorradores acceder a una cartera de activos administrada por expertos, con diferentes opciones de riesgo y por lo tanto de rendimiento.

Características

- Inversión diversificada.
- Liquidez variables.
- Diferentes opciones de riesgo.

Cuentas de Ahorro

Son una opción de bajo riesgo que permite depositar y retirar dinero con flexibilidad.

Características

- Mayor seguridad de inversión.
- Baja tasa de interés.
- Alta liquidez con acceso inmediato a los fondos.

Descripción de la base de datos y variables utilizadas

Base de datos

Se utilizaron dos fuentes principales de datos en formato Excel:

1. `df_series.xlsx`
Contiene variables macroeconómicas relevantes. Cada fila representa un periodo mensual, con la columna Fecha como clave temporal. Esta base provee indicadores económicos que podrían influir en el comportamiento de los depósitos bancarios.
2. Proyecto 4 Cuentas de Captación 2024.xlsx, contiene la información financiera de captación bancaria para cinco instituciones:
 - TBM
 - Banamex
 - BBVA
 - Santander
 - Banorte

Cada banco está representado en una hoja distinta, y los datos están alineados temporalmente con la base macroeconómica.

Periodo de análisis

El análisis abarca desde diciembre de 2019 hasta junio de 2024, capturando eventos económicos relevantes, incluyendo la pandemia, recuperación económica y fluctuaciones inflacionarias.

Variables utilizadas

Variables dependientes (a modelar): Estas son las tres formas de captación bancaria, modeladas por separado:

- Depósitos a la vista
- Depósitos a plazo
- Captación tradicional (suma de los anteriores)

Cada tipo de depósito se analiza individualmente para cada banco.

Variables independientes (predictoras macroeconómicas): Estas provienen del archivo df_series.xlsx, incluye variables como:

- "TIIE a 28 días Tasa de interés promedio mensual, en por ciento anual"
- "Tipo de cambio Pesos por dólar E.U.A. Para solventar obligaciones denominadas en moneda extranjera Fecha de determinación (FIX) Cotizaciones promedio"
- "Base Monetaria"
- "Remesas Familiares Total"
- "Índice Nacional de Precios al consumidor Variación mensual"
- "Costo de captación a plazo de pasivos en moneda nacional (CCP) Tasa en por ciento anual"
- "Agregados Monetarios M2 Instrumentos monetarios a plazo en poder de residentes 4/ Captación a plazo En otras instituciones no bancarias Entidades de ahorro y crédito popular 3/"
- "Activos Financieros Internos F1 Instrumentos no monetarios en poder de residentes Fondos de ahorro para la vivienda y el retiro en Banco de México"

La selección de variables se refina mediante técnicas estadísticas dentro de los modelos

Modelos de regresión utilizados

El objetivo fue modelar el comportamiento de los depósitos utilizando diferentes técnicas de regresión para evaluar cuál es más adecuada en cada caso. Se emplearon los siguientes modelos:

Regresión Lineal Múltiple

Modelo base que estima la relación entre una variable dependiente y y múltiples variables independientes x_1, x_2, \dots, x_p :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

- y : variable objetivo (ej. depósitos)
- x_i : variables predictoras (macroeconómicas) • β_i : coeficientes a estimar
- ε : término de error

Regresión Lasso (Least Absolute Shrinkage and Selection Operator)

Este modelo introduce una penalización L1 que elimina coeficientes irrelevantes (los lleva a cero):

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

- λ : parámetro de regularización
- β_j : penalización por el tamaño del coeficiente

Regresión Ridge

Aplica una penalización L2 que reduce la magnitud de todos los coeficientes, útil ante multicolinealidad:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

- La penalización L2 no elimina variables, solo reduce sus efectos.

Regresión Elastic Net

Combinación de Lasso y Ridge. Introduce ambas penalizaciones:

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

- λ_1 : controla el efecto tipo Lasso (selección de variables)
- λ_2 : controla el efecto tipo Ridge (suavización)

Estandarización de variables

Dado que las variables macroeconómicas utilizadas tienen distintas escalas (por ejemplo, tasas de interés vs. montos monetarios), fue necesario estandarizar los datos antes de ajustar los modelos. Esto implica transformar cada variable para que tenga:

- Media = 0
- Desviación estándar = 1

La estandarización es especialmente importante para los modelos Lasso, Ridge y Elastic Net, ya que las penalizaciones que aplican se ven afectadas por la escala de las variables

Validación cruzada

Para asegurar que los modelos tengan buen rendimiento fuera de la muestra (es decir, que no estén sobre ajustados), se empleó la técnica de validación cruzada. En particular, se utilizó:

- Validación cruzada k-fold con $k = 5$: Los datos se dividen en 5 partes iguales. En cada iteración, 4 partes se usan para entrenar el modelo y 1 para validarlo. Esto se repite 5 veces cambiando el conjunto de validación.

Esta técnica permite evaluar el rendimiento promedio del modelo de forma robusta y confiable.

Métricas utilizadas

Para comparar el desempeño de los modelos se utilizó el coeficiente de determinación R^2 :

- R^2 : mide qué proporción de la variabilidad de la variable dependiente es explicada por el modelo.
 - Valor cercano a 1: excelente ajuste.
 - Valor cercano a 0: pobre ajuste.
 - Puede ser negativo si el modelo es peor que una predicción constante.

Software utilizado

Para este proyecto usamos R y Python, combinando lo mejor de cada lenguaje para aprovechar sus ventajas en distintas etapas del análisis.

Empezamos trabajando en R, ya que es muy práctico para hacer análisis exploratorios, limpiar datos y generar gráficas de forma rápida. También usamos el paquete *glmnet*, que nos permitió ajustar modelos con regularización como Lasso y Elastic Net. Esto nos ayudó a tener una primera idea de cómo se comportaban las variables y qué tan bien se podían explicar los distintos tipos de captación. Además, extraímos las variables macroeconómicas directamente desde la API del Banco de México, lo cual facilitó la obtención de datos actualizados y confiables para el análisis.

Después pasamos a Python, sobre todo porque facilita automatizar procesos y ajustar modelos de forma más ordenada. Usamos la librería *sklearn* para entrenar los modelos de regresión múltiple, Lasso, Ridge y Elastic Net, incluyendo la estandarización de datos y la validación cruzada. Además, Python nos ayudó a comparar los resultados de todos los modelos y generar gráficas para visualizar mejor el desempeño.

La combinación de ambos lenguajes nos permitió validar los resultados, comparar enfoques y trabajar de forma más completa. Nos dio más confianza en las conclusiones porque los hallazgos fueron consistentes en ambos entornos.

Todos los códigos que usamos están disponibles en las carpetas que vienen anexas a este trabajo, organizados por lenguaje y etapa del análisis, por si alguien quiere revisarlos o replicar lo que hicimos.

Resultados

Tabla comparativa

Se aplicaron cuatro modelos de regresión (Múltiple, Lasso, Ridge, Elastic Net) para predecir el comportamiento de tres tipos de depósitos (Vista, Plazo y Total) en cinco bancos. La comparación se realizó utilizando el coeficiente de determinación R^2 para identificar el mejor modelo en cada caso.

Banco	Tipo de Depósito	R^2 Múltiple	R^2 Lasso	R^2 Ridge	R^2 Elastic Net	Mejor Modelo
TBM	Vista	0.9868	0.9867	0.9754	0.9866	Múltiple
	Plazo	0.959	0.9551	0.8462	0.9584	Múltiple
	Total	0.9829	0.9827	0.9618	0.9827	Múltiple
BBVA	Vista	0.964	0.9652	0.9449	0.9651	Lasso
	Plazo	0.8771	0.8735	0.6849	0.8818	Elastic Net
	Total	0.9595	0.9606	0.941	0.9613	Elastic Net
Banamex	Vista	0.8244	0.8264	0.8082	0.8271	Elastic Net
	Plazo	0.7	0.6612	0.6845	0.7034	Elastic Net
	Total	0.7222	0.7222	0.7106	0.7166	Múltiple
Santander	Vista	0.8799	0.8696	0.8548	0.869	Múltiple
	Plazo	0.8742	0.8767	0.7738	0.8724	Lasso
	Total	0.901	0.9084	0.8282	0.8887	Lasso
Banorte	Vista	0.932	0.925	0.922	0.9239	Múltiple
	Plazo	0.955	0.9549	0.6606	0.9566	Elastic Net
	Total	0.941	0.9451	0.8679	0.9445	Lasso

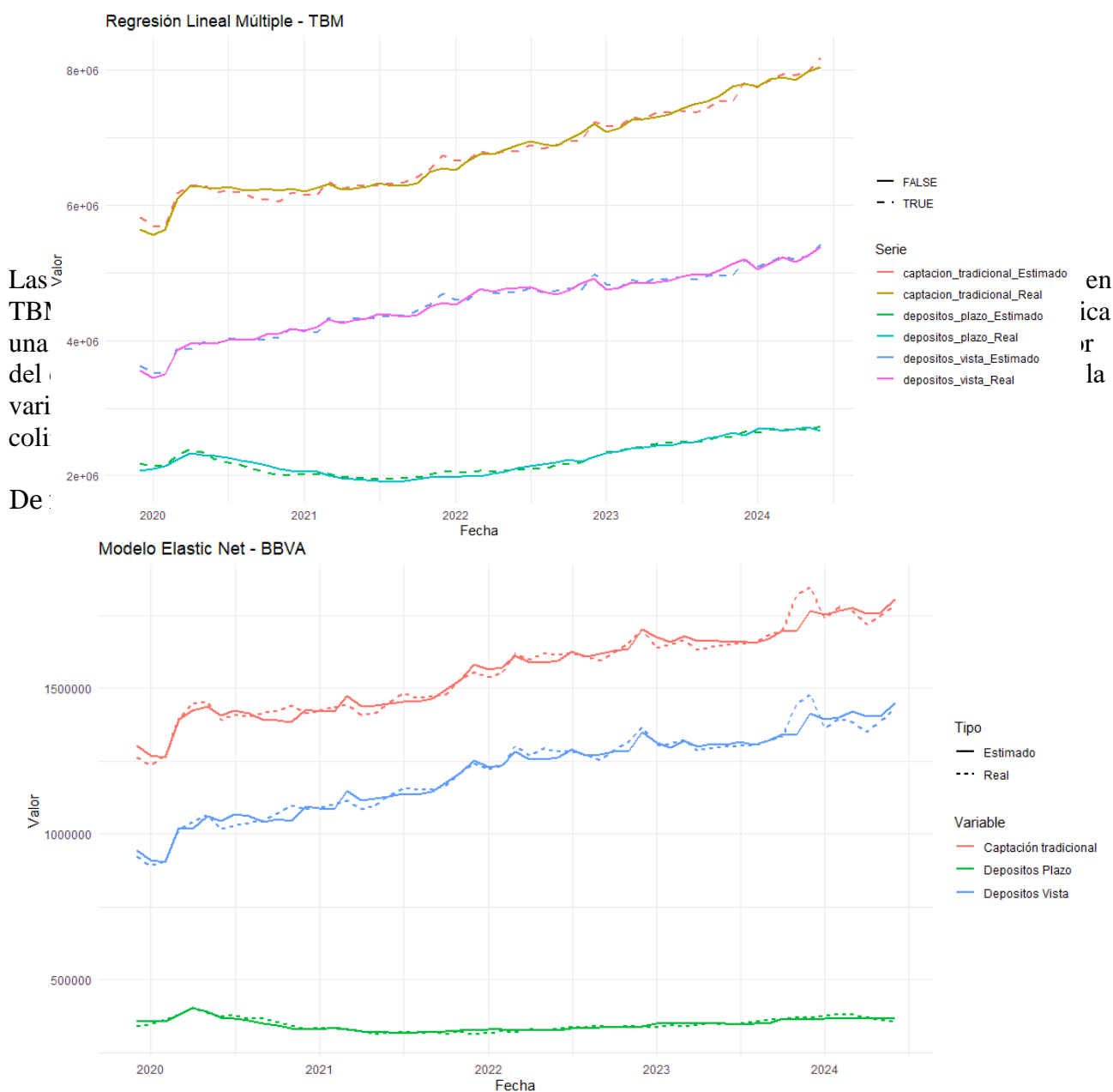
En el análisis de modelos predictivos para explicar la captación bancaria por tipo de depósito y banco, la regresión múltiple se posicionó como la mejor opción en 5 de los 15 casos evaluados. Su desempeño fue particularmente destacado en TBM, así como en el depósito total de Banamex y en los depósitos a la vista de Santander.

Por otro lado, el modelo Elastic Net demostró ser el más robusto en contextos con alta multicolinealidad o con relaciones menos marcadas entre las variables, alcanzando también el mejor desempeño en 5 casos. Este enfoque fue especialmente efectivo para explicar los depósitos en BBVA, Banamex y Banorte.

El modelo Lasso fue el preferido en 4 casos, mostrando una ventaja significativa en bancos como Santander y BBVA, gracias a su capacidad para realizar una selección automática de variables más eficiente y parsimoniosa.

Finalmente, el modelo Ridge no logró destacarse como el mejor en ninguno de los escenarios analizados, lo que sugiere que la penalización L2, utilizada de forma aislada, no fue suficiente para superar el rendimiento de los otros enfoques evaluados.

Dados estos resultados tenemos las gráficas de los modelos que más dominaron por banco como es el caso de la regresión lineal múltiple para TBM



En el caso de BBVA, el modelo Elastic Net proporcionó el mejor ajuste, especialmente para los depósitos a plazo y la captación total. Las gráficas reflejan cómo el modelo logra adaptarse a variaciones más complejas en los datos, probablemente debido a la presencia de colinealidad entre las variables predictoras. La combinación de penalizaciones L1 y L2 permitió seleccionar un subconjunto relevante de variables, manteniendo la estabilidad del modelo. Esto destaca la utilidad de Elastic Net en contextos donde las relaciones entre variables son más difusas o no tan lineales.

Conclusiones

El análisis desarrollado permitió identificar patrones en el comportamiento de los depósitos bancarios y evaluar la capacidad predictiva de distintos modelos econométricos en función del banco y el tipo de captación. A lo largo del estudio se observó que la regresión múltiple ofreció un desempeño sólido en varios escenarios, especialmente en TBM y en componentes específicos de Banamex y Santander. No obstante, el modelo Elastic Net mostró un rendimiento comparable, destacando en contextos donde las variables macroeconómicas presentan alta colinealidad, como en los casos de BBVA, Banamex y Banorte. Por su parte, el modelo Lasso resultó particularmente útil para bancos donde la simplicidad del modelo y la automatización en la selección de variables ofrecieron ventajas claras. En contraste, el modelo Ridge no logró posicionarse como el mejor en ningún caso, lo que sugiere que la penalización L2, por sí sola, no fue suficiente para capturar adecuadamente la dinámica de los datos observados.

Desde una perspectiva práctica, los hallazgos indican que Elastic Net es especialmente recomendable en entornos donde existe multicolinealidad entre las variables, al combinar los beneficios de los enfoques Lasso y Ridge. Lasso representa una opción eficiente cuando se busca reducir la complejidad del modelo sin comprometer la capacidad explicativa. En escenarios más estables o con menor ruido estadístico, la regresión múltiple puede seguir siendo una alternativa válida, siempre que se controle adecuadamente el riesgo de sobreajuste.

Si bien los resultados obtenidos son sólidos, es importante reconocer ciertas limitaciones del análisis. El periodo evaluado, de diciembre de 2019 a junio de 2024, aunque abarca eventos macroeconómicos relevantes como la pandemia y la recuperación económica posterior, sigue siendo relativamente acotado y no refleja ciclos financieros de mayor duración. Además, los modelos utilizados parten de la suposición de relaciones lineales entre las variables, lo que podría limitar su capacidad para capturar dinámicas más complejas o estructurales en los datos.

A partir de esta experiencia, se abren distintas líneas para estudios futuros. Una posible extensión sería la incorporación de modelos no lineales o basados en aprendizaje automático, que podrían mejorar la precisión predictiva en entornos económicos más volátiles. También resultaría útil explorar enfoques bayesianos o modelos robustos a distribuciones no normales, así como ampliar el horizonte temporal para incluir periodos de crisis previas o transiciones macroeconómicas estructurales. Finalmente, un análisis segmentado por tipo de institución financiera o perfil de captación podría enriquecer la comprensión del fenómeno y apoyar la toma de decisiones estratégicas a nivel bancario o regulatorio.

Aprendizajes obtenidos

El análisis evidenció que la calidad predictiva de los modelos varía considerablemente según el banco y el tipo de depósito, lo cual refuerza la idea de que no existe un modelo óptimo universal aplicable a todos los casos. En este contexto, los modelos con regularización como Lasso y Elastic Net, mostraron ventajas importantes al mejorar el ajuste, especialmente en presencia de múltiples variables correlacionadas. Asimismo, se identificó que la validación cruzada y la estandarización de las variables fueron pasos fundamentales para asegurar la fiabilidad y estabilidad de los resultados obtenidos. Finalmente, el uso complementario de R y Python permitió aprovechar las fortalezas de ambas plataformas analíticas, facilitando la verificación cruzada de resultados y aumentando la robustez del análisis.