

Modelos de Machine Learning (ML) basados en indicadores macroeconómicos para predecir la pérdida esperada en el sistema crediticio de México

Andrea Rancaño Botaya José Angel Alonso Prieto
Eduardo Daniel Morales Sánchez

Pregunta de investigación

¿Pueden los modelos de Machine Learning, entrenados con indicadores macroeconómicos, anticipar de manera precisa y oportuna la pérdida esperada futura del sistema crediticio mexicano?

Motivación

Al permitir una identificación temprana de señales de deterioro crediticio, estos modelos se convierten en una herramienta clave tanto para las instituciones financieras como para las autoridades regulatorias, al facilitar la toma de decisiones informadas y el diseño de políticas macroprudenciales orientadas a preservar la estabilidad financiera.

Variables

En este análisis se utilizaron diferentes variables para entender cómo evoluciona el riesgo crediticio y qué factores pueden influir en él. Por un lado, consideramos las pérdidas esperadas en distintos tipos de créditos (todos de stage 1); por otro, incluimos variables macroeconómicas que ayudan a explicar los cambios en estos riesgos. A continuación, se presentan las variables dependientes e independientes que forman parte del estudio.

Variables dependientes

Las variables dependientes utilizadas en este análisis corresponden a la pérdida esperada en distintos segmentos del sistema crediticio mexicano, expresadas como porcentaje de los portafolios correspondientes. Separar la variable dependiente según el tipo de portafolio crediticio permite observar la heterogeneidad de riesgo en cada segmento de crédito, lo cual es fundamental tanto para mejorar la capacidad predictiva de los modelos como para hacer recomendaciones puntuales de políticas y estrategias de mitigación que sean más específicas y efectivas. Las variables dependientes son:

1. Pérdida esperada - actividad empresarial o comercial: Representa la pérdida esperada asociada a créditos otorgados para actividades empresariales o comerciales, reflejando el riesgo crediticio en el sector productivo.
2. Pérdida esperada - automotriz: Mide la pérdida esperada en el portafolio de créditos automotrices, es decir, aquellos destinados a la compra de vehículos.
3. Pérdida esperada - entidades financieras: Se refiere a la pérdida esperada en créditos otorgados a otras entidades financieras, capturando el riesgo sistémico entre intermediarios.
4. Pérdida esperada - nómina: Indica la pérdida esperada en créditos de nómina, comúnmente otorgados a trabajadores con ingresos fijos y descontados directamente de su salario.
5. Pérdida esperada - créditos personales: Representa la pérdida esperada en créditos personales, los cuales son generalmente de libre destino y sin una garantía específica.

6. Pérdida esperada - tarjeta de crédito: Mide la pérdida esperada en el portafolio de tarjetas de crédito, segmento caracterizado por alta rotación y riesgo de consumo.

Variables independientes

Por su lado, las variables independientes incluyen diversos indicadores macroeconómicos y financieros que pueden influir en el comportamiento de las pérdidas esperadas en los distintos portafolios crediticios. Estas variables son:

1. IGAE (Indicador Global de Actividad Económica): Índice mensual base 2018=100 aproxima la evolución del PIB mexicano. Se publica en valores originales y ajustados por estacionalidad, para el análisis se han considerado los valores sin ajuste por estacionalidad.
2. INPC: Índice Nacional de Precios al Consumidor (INPC) indicador que mide la variación promedio de los precios de una canasta de bienes y servicios representativa del consumo de los hogares mexicanos a lo largo del tiempo (no se usa la inflación YoY para no tener efectos base).
3. Tasa de referencia (TIIE de fondeo): es la tasa de interés interbancaria de equilibrio utilizada como referencia para el fondeo entre bancos. Su valor actual, justo por debajo de la tasa objetivo, refleja las condiciones de liquidez del sistema financiero.
4. Empleo IMSS: Número total de trabajadores definitivos registrados en el Instituto Mexicano del Seguro Social. Indicador de empleo formal.
5. Confianza empresarial: Resultados de la Encuesta Mensual de Opinión Empresarial (EMOE), mide la percepción de los empresarios sobre la situación actual y futura.
6. S&P 500: Índice bursátil que refleja el comportamiento de 500 empresas grandes en EE.UU. Promedio mensual de precios de cierre diarios.
7. IPC: El Índice de Precios y Cotizaciones (IPC) es el principal indicador de desempeño de la Bolsa Mexicana de Valores. Está compuesto por las acciones de las empresas más grandes y con mayor liquidez de México.
8. Remesas: Ingresos por remesas familiares del exterior, en millones de USD. Publicación mensual del Banco de México.
9. Saldo de la balanza comercial no petrolera: Diferencia mensual entre exportaciones e importaciones de mercancías sin considerar la balanza petrolera.
10. Agregados monetarios M1, M2, M3: M1 es el dinero en efectivo y en cuentas de cheques, es decir, el que puedes gastar al instante. M2 incluye todo lo de M1, más los ahorros y los depósitos a plazo, por lo que es un poco menos líquido, pero todavía fácil de convertir en efectivo. M3 amplía aún más e incorpora lo de M2 más valores emitidos por instituciones no bancarias, por lo que representa dinero menos accesible de inmediato.

Adicionalmente, se incorporan variables obtenidas de Google Trends que reflejan la frecuencia de búsqueda de términos como “deuda” y “empleo”. La inclusión de estas variables responde a la necesidad de captar señales tempranas y complementarias sobre la percepción y las preocupaciones de la población, las cuales pueden anticipar cambios en el comportamiento crediticio antes de que estos se reflejen en los datos tradicionales. Este tipo de información, proveniente de fuentes alternativas, permite enriquecer el modelo con perspectivas que no están presentes en las bases de datos convencionales y que quizás pueden ofrecer una visión más inmediata de cambios en la confianza de los hogares o las condiciones del mercado laboral.

11. Google Trends - deuda: Esta variable representa el volumen de búsquedas realizadas en Google relacionadas con el término “deuda” en México, obtenida a partir de Google Trends. Su objetivo es capturar cambios en el interés o preocupación de la población respecto a temas de endeudamiento.
12. Google Trends - empleo: Esta variable corresponde al volumen de búsquedas en Google asociadas con el término “empleo”, también basada en los datos de Google Trends para México. La inclusión de esta variable permite monitorear en tiempo real el interés de la población por cuestiones relacionadas con el empleo.

Datos

Los datos de cartera y riesgo de crédito se tomaron de la Comisión Nacional Bancaria y de Valores (CNBV), Banxico, IMSS e INEGI, y se complementaron con datos de Google Trends (para búsquedas de “deuda” y “empleo”).

Limpieza de datos

Para la elaboración de la base de datos final, primero se procedió con el armado y la integración de las distintas fuentes de datos, asegurando la identificación correcta de variables numéricas, el ordenamiento de las fechas y la estandarización de formatos para garantizar la homologación entre todas las bases. Posteriormente, se trabajó en la homogeneización de los periodos disponibles, de manera que todas las series coincidieran en la misma frecuencia temporal y pudieran ser analizadas de forma conjunta. Finalmente, para preparar las variables para los análisis de series de tiempo y asegurar la estacionariedad, se aplicó la transformación de primeras diferencias utilizando un rezago mensual (lag de un mes), lo cual permitió obtener series listas para el modelado y la comparación de resultados.

Análisis Exploratorio de Datos (EDA)

Descripción general de los datos

Tabla 1: Descripción general de los datos

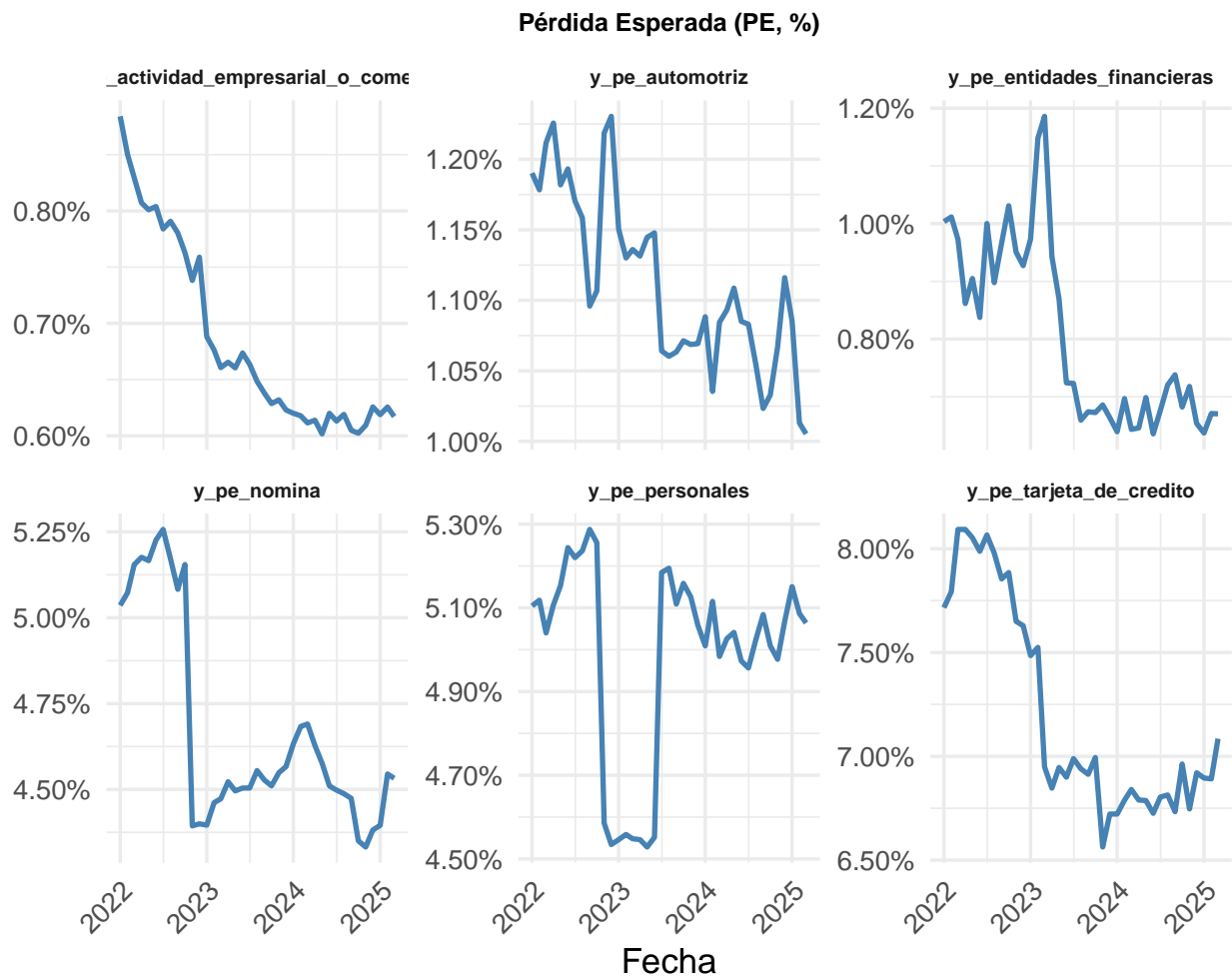
Variable	Descripción	Unidades
llave	Fecha de observación (mensual)	YYYY-MM-DD
y_pe_actividad_empresarial_o_comercial	Pérdida esperada - Actividad empresarial/comercial	Decimal (x100 = %)
y_pe_automotriz	Pérdida esperada - Crédito automotriz	Decimal (x100 = %)
y_pe_entidades_financieras	Pérdida esperada - Entidades financieras	Decimal (x100 = %)
y_pe_nomina	Pérdida esperada - Nómina	Decimal (x100 = %)
y_pe_personales	Pérdida esperada - Crédito personal	Decimal (x100 = %)
y_pe_tarjeta_de_credito	Pérdida esperada - Tarjeta de crédito	Decimal (x100 = %)
usd_mxn	Tipo de cambio USD/MXN	Pesos por USD
tiie_fondeo	Tasa de fondeo bancaria (TIIE a un día)	Decimal (x100 = %)
remesas	Remesas recibidas	Miles de millones USD
m1	Agregado monetario M1	Miles de millones MXN
m2	Agregado monetario M2	Miles de millones MXN
m3	Agregado monetario M3	Miles de millones MXN
balanza	Balanza comercial	Miles de millones USD
inpc	Índice Nacional de Precios al Consumidor	Índice
igoec	Indicador Global de la Actividad Económica (IGOE)	Índice
igae	IGAE - Base 2018 = 100	Índice
google_deuda	Tendencia de búsqueda Google: 'deuda'	Índice
google_empleo	Tendencia de búsqueda Google: 'empleo'	Índice
nacional	Puestos de trabajo registrados IMSS	Millones
s_p_bmv_ipc	Índice bursátil S&P/BMV IPC	Índice
puestos_trabajo	Puestos de trabajo registrados IMSS (copia)	Millones

Estadística descriptiva

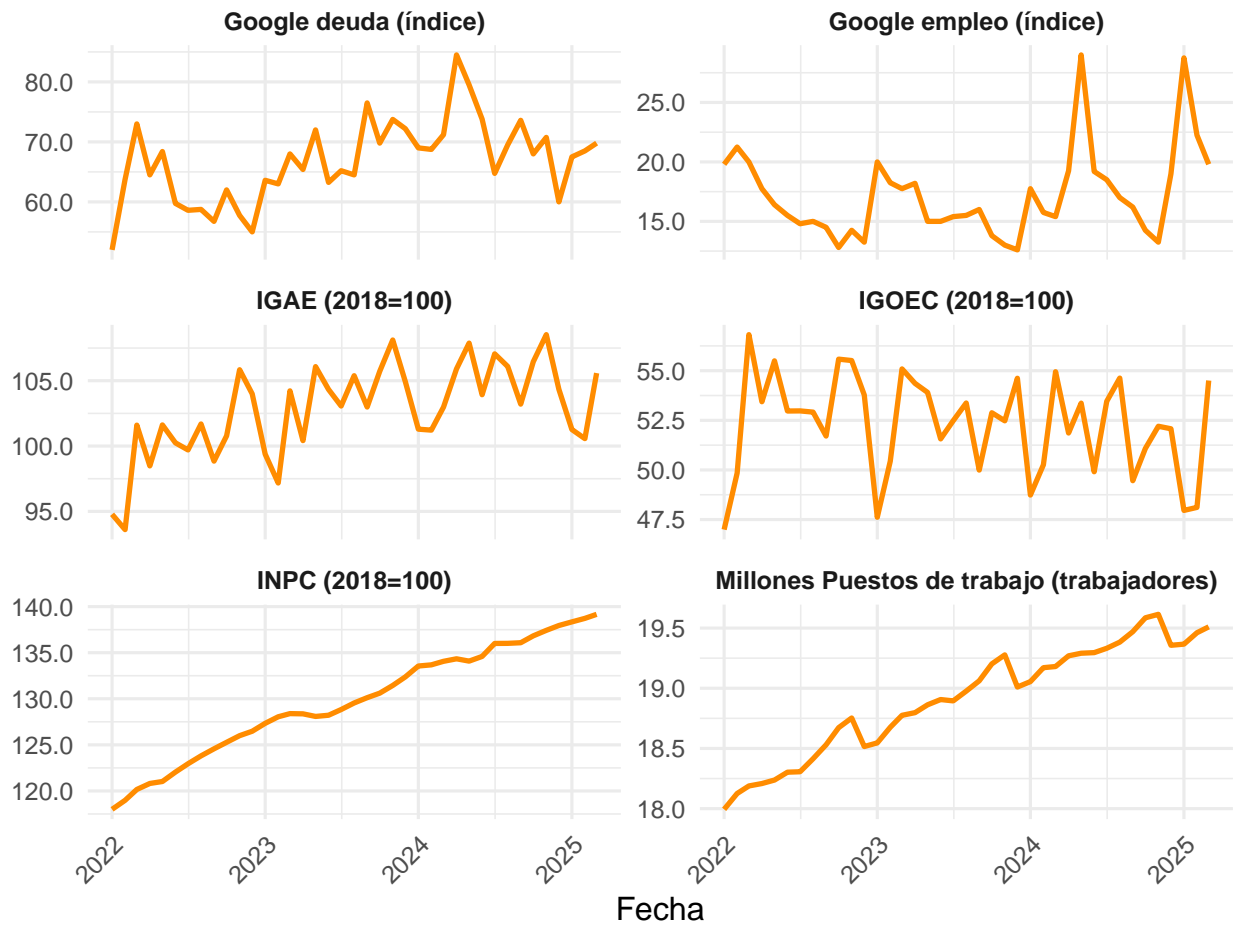
Tabla 2: Resumen estadístico de variables numéricas

Variable	N	Mean	SD	Q1	Median	Q3	Min	Max	Missing
PE_Comercial	39	0.0068	0.0008	0.0062	0.0065	0.0076	0.0060	0.0088	0
PE_Automotriz	39	0.0111	0.0006	0.0107	0.0110	0.0115	0.0101	0.0123	0
PE_EFinancieras	39	0.0081	0.0016	0.0067	0.0072	0.0095	0.0064	0.0119	0
PE_Nomina	39	0.0467	0.0030	0.0448	0.0453	0.0486	0.0433	0.0526	0
PE_Personales	39	0.0499	0.0024	0.0497	0.0506	0.0514	0.0453	0.0529	0
PE_TC	39	0.0721	0.0051	0.0681	0.0695	0.0768	0.0656	0.0809	0
USD/MXN	39	18.8600	1.3908	17.3422	19.1515	20.1148	16.7918	20.5562	0
TIEF	39	0.0987	0.0179	0.0934	0.1073	0.1124	0.0550	0.1129	0
Remesas	39	5.1586	0.5366	4.8014	5.1894	5.5764	3.9285	6.2069	0
M1	39	7.05e+03	5.78e+02	6.56e+03	6.85e+03	7.48e+03	6.30e+03	8.30e+03	0
M2	39	1.32e+04	1.32e+03	1.21e+04	1.29e+04	1.42e+04	1.12e+04	1.56e+04	0
M3	39	1.57e+04	1.57e+03	1.43e+04	1.53e+04	1.69e+04	1.33e+04	1.84e+04	0
Balanza	39	0.7237	2.1699	-0.0791	0.7251	2.1522	-4.1519	4.4246	0
INPC	39	129.8023	6.0551	125.6365	129.5450	134.4650	118.0020	139.1610	0
IGOE	39	52.2923	2.4729	50.3495	52.8820	54.1340	46.9940	56.8180	0
IGAE	39	102.8016	3.4964	100.6749	103.0628	105.6500	93.6058	108.5259	0
G-Deuda	39	66.8359	6.7719	63.1250	68.0000	70.9750	52.0000	84.5000	0
G-Empleo	39	17.2090	3.7108	14.9000	16.2000	19.1000	12.6000	29.0000	0
IMSS	39	18.9123	0.4611	18.5383	18.9762	19.2934	17.9943	19.6139	0
IPC	39	4.58e+04	5.08e+03	4.27e+04	4.65e+04	4.99e+04	3.46e+04	5.42e+04	0
S&P500	39	3.52e+03	6.56e+02	2.92e+03	3.39e+03	4.01e+03	2.62e+03	4.68e+03	0

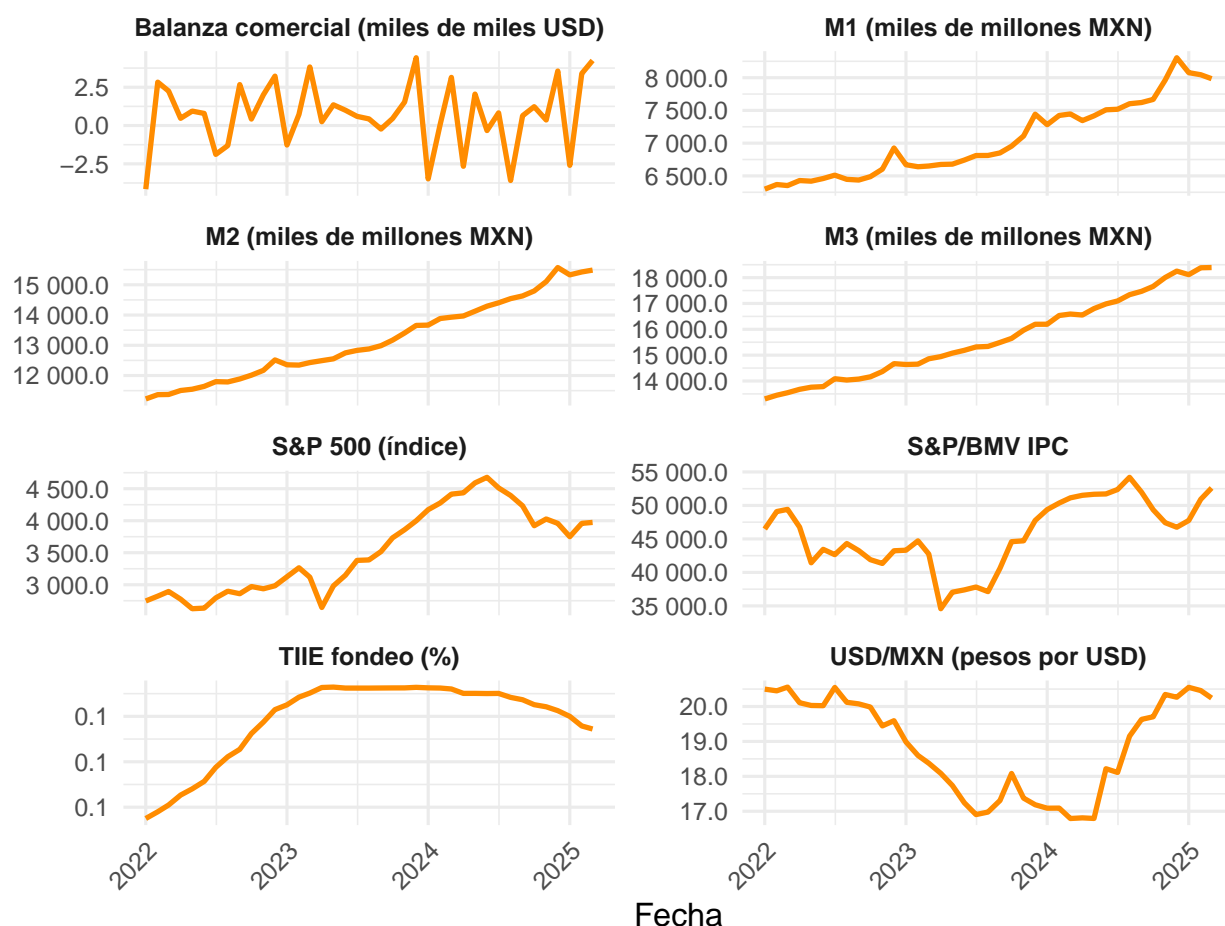
Exploración univariada



Indicadores Macroeconómicos I – Precios, Actividad y Empleo (unidades)



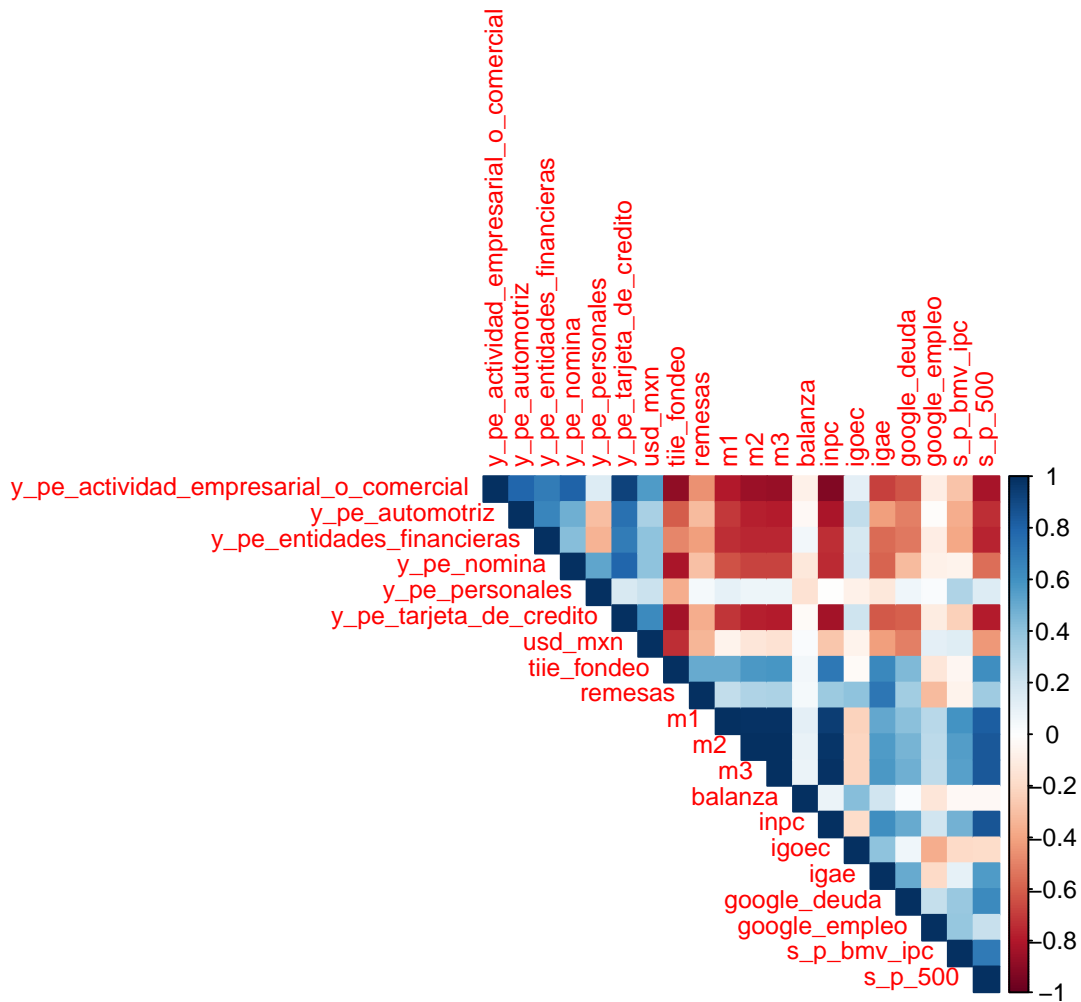
Indicadores Macroeconómicos II – Dinero, Mercados y Tipo de Cambio (Unidades)



Los gráficos univariados revelan las dinámicas temporales que hay que tomar en cuenta antes de modelar las pérdidas crediticias esperadas. Los agregados monetarios (M1, M2, M3), la inflación (INPC) y el empleo (puestos de trabajo) muestran fuertes tendencias al alza, lo que sugiere un crecimiento a largo plazo y la necesidad de transformaciones como la diferenciación. Las tasas de interés (TIIE) y los tipos de cambio (USD/MXN) muestran un comportamiento cíclico alineado con los cambios recientes en la política monetaria, lo que podría tener efectos rezagados en el desempeño crediticio. Los indicadores del ciclo económico, como el IGAE y las variables de Google Trends, muestran fluctuaciones a corto plazo, lo que sugiere su potencial como predictores de alta frecuencia del estrés del prestatario.

En cuanto a las variables de pérdida crediticia, éstas presentan comportamientos diversos según el tipo de producto. Mientras que segmentos como las tarjetas de crédito y la actividad empresarial muestran descensos constantes, posiblemente debido a mejoras en la calidad de la cartera, otros, como la nómina y el personal, revelan cambios estructurales o estancamientos, lo que sugiere posibles cambios de régimen o redefiniciones contables.

Matriz de correlación



La matriz de correlación revela fuertes relaciones entre los indicadores macrofinancieros y las tasas de pérdida esperadas en diferentes productos crediticios. La mayoría de las tasas de pérdida presentan una alta correlación positiva entre sí, lo que sugiere que factores macroeconómicos sistémicos influyen simultáneamente en múltiples segmentos crediticios. Destaca que la variable *Pérdida esperada - Crédito personal* muestra correlaciones bajas o incluso negativas con otras variables de pérdida, lo que indica que podría estar impulsado por dinámicas idiosincrásicas o no observadas. Las pérdidas crediticias también presentan una correlación negativa con los agregados monetarios (M1, M2, M3), el empleo y la inflación, lo que implica que una mayor actividad económica se asocia con un menor riesgo crediticio.

Entre los predictores macroeconómicos, las tasas de interés muestran fuertes correlaciones negativas con todas las variables de pérdida crediticia por lo que tasas más altas coinciden con menores incumplimientos a corto plazo, lo que posiblemente refleja estándares de préstamo más estrictos o efectos de selección. Por el contrario, el tipo de cambio y las tendencias de búsqueda en Google presentan una correlación positiva con las pérdidas crediticias, lo que indica posibles señales de estrés. Estos hallazgos resaltan la importancia del contexto macroeconómico en la modelización del riesgo crediticio.

Pruebas de estacionariedad

Prueba de Raíz Unitaria de Dickey & Fuller

La prueba de Dickey-Fuller Aumentada (DFA) se usa para detectar estadísticamente la presencia de conducta tendencial estocástica en las series temporales de las variables mediante un contraste de hipótesis. En nuestro

caso, se utilizó la prueba DFA para evaluar la presencia de raíz unitaria en las series temporales de nuestra base de datos. Los resultados indican que, en su mayoría, las series analizadas requieren una transformación mediante primeras diferencias para alcanzar la estacionariedad.

Una serie de tiempo debe ser estacionaria para que los modelos econométricos puedan aplicarse correctamente. Si una serie no es estacionaria, sus propiedades estadísticas —como la media, la varianza y la covarianza— cambian con el tiempo, lo cual invalida los supuestos fundamentales de los modelos clásicos. Esto impide usar procedimientos estándar de inferencia estadística y genera relaciones engañosas entre las variables. Como afirma Walter Enders:

“If a time series is nonstationary, it is not possible to model it in terms of a finite number of parameters that summarize its behavior over time. The reason is that the process generating the series is not stable over time. Consequently, any coefficients obtained from fitting a model to the data are not likely to be constant over time” — Walter Enders, *Applied Econometric Time Series*, 4th edition, Chapter 2.3, p. 50 .

Esto significa que los coeficientes estimados pueden no ser válidos en otros periodos, los errores se acumulan y las predicciones pierden validez. Por ello, es indispensable asegurar que las series sean estacionarias antes de modelarlas.

Resultados del Test ADF por Variable Evaluación de Estacionariedad con `tseries::adf.test` ($k = 0$)

variable	adf_stat	adf_pvalue	estacionaria	usable
puestos_trabajo	−3.040	0.166	FALSE	NO
m1	−2.751	0.279	FALSE	NO
inpc	−2.645	0.321	FALSE	NO
y_pe_entidades_financieras	−2.339	0.440	FALSE	NO
m2	−2.290	0.460	FALSE	NO
m3	−2.252	0.475	FALSE	NO
y_pe_personales	−2.129	0.523	FALSE	NO
y_pe_nomina	−1.853	0.631	FALSE	NO
s_p_bmv_ipc	−1.804	0.650	FALSE	NO
y_pe_actividad_empresarial_o_comercial	−1.690	0.695	FALSE	NO
y_pe_tarjeta_de_credito	−1.189	0.891	FALSE	NO
s_p_500	−1.145	0.903	FALSE	NO
usd_mxn	−0.538	0.975	FALSE	NO
tiie_fondeo	−0.342	0.984	FALSE	NO
balanza	−8.164	0.010	TRUE	OK
igoec	−6.230	0.010	TRUE	OK
igae	−4.811	0.010	TRUE	OK
google_deuda	−4.045	0.018	TRUE	OK
remesas	−3.631	0.044	TRUE	OK
google_empleo	−3.614	0.045	TRUE	OK
y_pe_automotriz	−3.614	0.045	TRUE	OK

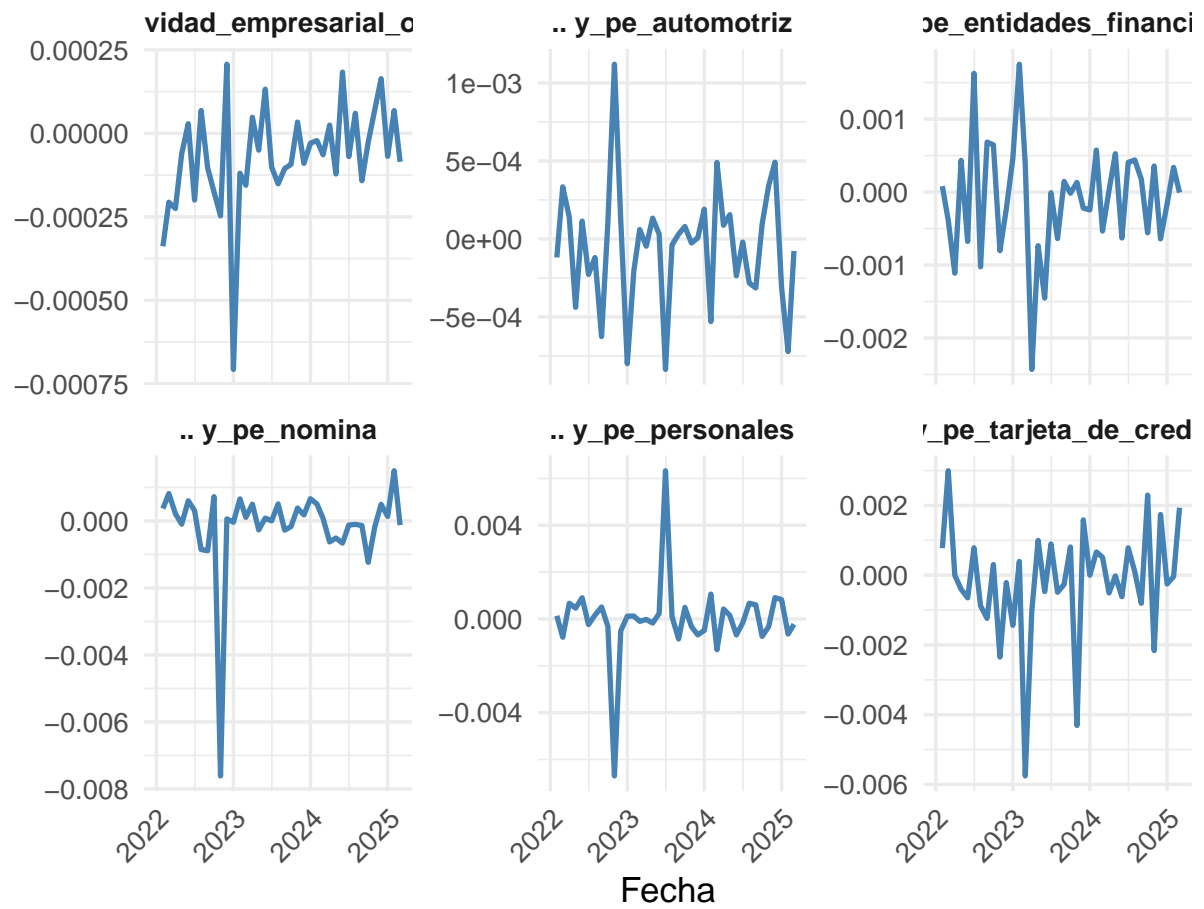
Resultados del Test ADF – Primeras Diferencias Evaluación de Estacionariedad con `tseries::adf.test` ($k = 0$)

variable	adf_stat	adf_pvalue	estacionaria	usable
diff_y_pe_actividad_empresarial_o_comercial	−8.41468	0.01000	TRUE	OK
diff_y_pe_automotriz	−5.63978	0.01000	TRUE	OK
diff_y_pe_entidades_financieras	−6.41906	0.01000	TRUE	OK
diff_y_pe_nomina	−6.34084	0.01000	TRUE	OK

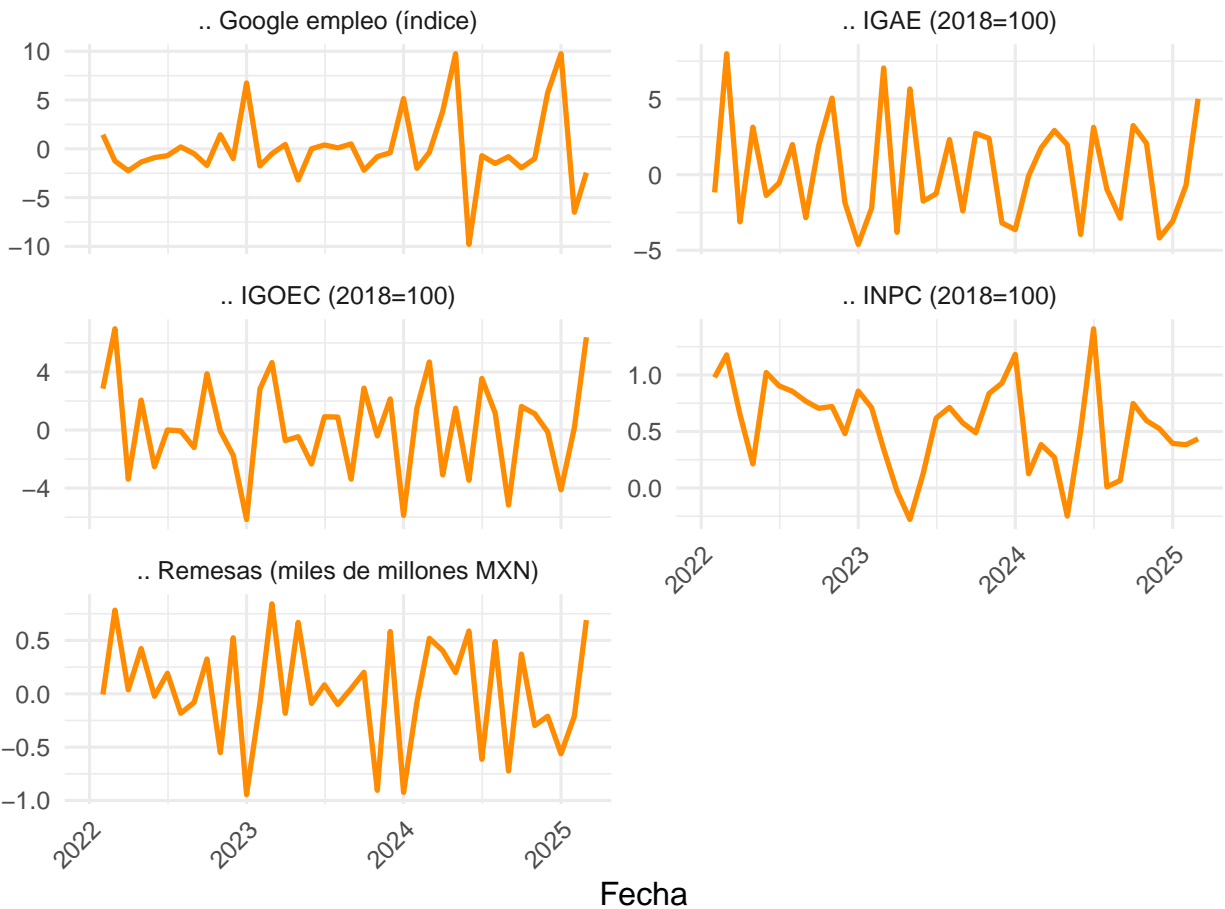
diff_y_pe_personales	-5.56040	0.01000	TRUE	OK
diff_y_pe_tarjeta_de_credito	-7.29581	0.01000	TRUE	OK
diff_usd_mxn	-6.03114	0.01000	TRUE	OK
diff_remesas	-9.05683	0.01000	TRUE	OK
diff_m1	-6.44392	0.01000	TRUE	OK
diff_m2	-6.75203	0.01000	TRUE	OK
diff_m3	-7.65344	0.01000	TRUE	OK
diff_balanza	-11.08885	0.01000	TRUE	OK
diff_inpc	-4.61570	0.01000	TRUE	OK
diff_igoec	-7.27170	0.01000	TRUE	OK
diff_igae	-8.05037	0.01000	TRUE	OK
diff_google_deuda	-8.25949	0.01000	TRUE	OK
diff_google_empleo	-6.76881	0.01000	TRUE	OK
diff_puestos_trabajo	-5.70349	0.01000	TRUE	OK
diff_s_p_bmv_ipc	-5.27137	0.01000	TRUE	OK
diff_s_p_500	-5.04541	0.01000	TRUE	OK
diff_tiie_fondeo	-4.14342	0.01483	TRUE	OK

Tras aplicar la transformación de primeras diferencias a todas las series temporales, se comprobó mediante el Test ADF que las series transformadas cumplen con el criterio de estacionariedad (p-valor < 0.05 en todos los casos). Por tanto, las primeras diferencias de las variables pueden utilizarse de forma adecuada en los modelos de predicción.

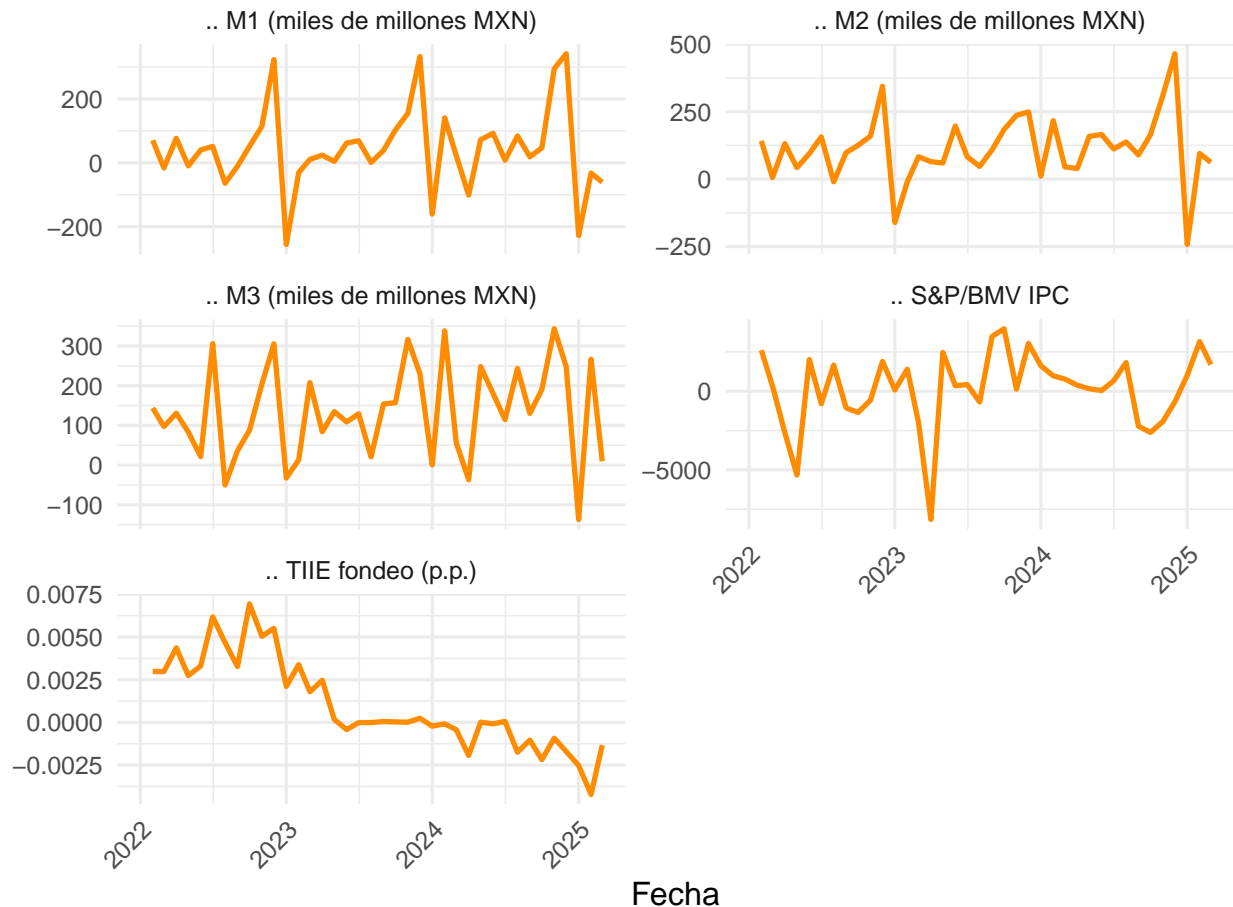
Primeras diferencias: Series con 'y'



Indicadores Macro – Grupo 1 (1 diferencia)



Indicadores Macro – Grupo 2 (1 diferencia)



Modelado (siguientes pasos)

Para el análisis de nuestras variables, planeamos hacer la comparación entre los siguientes modelos:

1. ARIMA

- **Objetivo:** Modelar y predecir la dinámica interna de una sola serie temporal de pérdida esperada, sirviendo como línea base clásica.
- **Ventajas:** Capta patrones autorregresivos, tendencias y ciclos; es ampliamente aceptado y fácil de interpretar.
- **Consideraciones:** La serie debe ser estacionaria; seleccionar correctamente los parámetros (p, d, q); revisar residuos.

2. SARIMA

- **Objetivo:** Incorporar patrones estacionales en series de tiempo, útil si las pérdidas presentan estacionalidad o ciclos repetitivos.
- **Ventajas:** Captura tanto dinámica general como efectos estacionales; permite predecir en presencia de ciclos regulares.
- **Consideraciones:** Identificar y ajustar adecuadamente componentes estacionales (P, D, Q, s); asegurar estacionariedad.

3. XGBoosting

- **Objetivo:** Detectar patrones complejos y no lineales en el riesgo crediticio, integrando variables macroeconómicas, financieras y alternativas (como Google Trends).
- **Ventajas:** Alta precisión, maneja grandes volúmenes de datos, robusto a variables irrelevantes y

- Consideraciones: Evitar fuga de información temporal (data leakage), usar variables rezagadas, validar con divisiones temporales adecuadas (walk-forward).

VAR con Exógenas: Serie Real y Predicción

