



EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

**REVISIÓN A LAS CIFRAS DE INCIDENCIA
DELICTIVA EN MÉXICO USANDO
MACHINE LEARNING, 2010-2021**

RAÚL ANTONIO TIRADO COSSIO

PROMOCIÓN 2021-2023

ASESOR:

AURORA RAMÍREZ ÁLVAREZ

AGOSTO 2023

Contenido

Resumen	III
Introducción	1
1. Revisión de Literatura	3
2. Datos	7
2.1. SESNSP	7
2.2. Cifras Económicas	7
2.3. ENVIPE	8
3. Hechos Estilizados	9
3.1. La Cifra Negra y la Incidencia Delictiva	9
3.2. El Gasto Público y la Incidencia Delictiva	12
4. Aspectos Econométricos	18
4.1. Neural Networks	19
4.2. Random Forest	22
4.3. XGBoost	26
5. Discusión de los Resultados	29
Conclusiones	33
Anexo	35
Referencias	42

Dedicatoria

En memoria de Beto[†] y Eustaquio[†], que siempre me dieron el cariño que necesitaba. Le dedico este esfuerzo a mi madre y hermana, ya que sin ellas no hubiera sido posible encontrar la motivación para seguir adelante.

Agradecimientos

A mi padre, que gracias a su apoyo incondicional logré terminar mi maestría.

También, gracias a todos los que me apoyaron durante el trayecto. Especialmente, a mi asesora Aurora Ramírez que a pesar de sus compromisos personales accedió a guiarme en mi trabajo de investigación.

También, gracias al Dr. Eneas que como Coordinador de la Maestría siempre nos apoyó.

Gracias al Dr. Eduardo Loría, que siempre ha sido un mentor para mí.

Resumen

Este artículo contribuye a la literatura de criminalidad, e introduce herramientas de *Machine Learning* para generar pronósticos que sirven de insumo en la elaboración de mejores políticas de seguridad. Se encuentra que existe una gran diferencia en niveles y tendencias entre las cifras de incidencia delictiva del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública y las cifras de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE). Esto se cumple para los delitos agregados, y en menor medida para cada tipo de delito. El sesgo es provocado por la cifra negra, es decir, la proporción de crímenes que no se reflejan en las cifras oficiales de incidencia delictiva pues no tienen averiguación previa ni carpeta de investigación (INEGI, 2022c).¹

Como principal contribución, se logró generar una serie consistente, que ayuda a reducir el rezago temporal de la ENVIPE, y que puede ser utilizada en otros contextos. Al respecto, se probaron diferentes especificaciones como Redes Neuronales, *Random Forest* y *XGboost*. Finalmente, se escogió el mejor modelo en función de su desempeño dentro y fuera de la muestra, cuidando que la probabilidad de subestimar los delitos fuera mínima.

El mejor modelo para aproximar la incidencia delictiva en nuestro contexto fue *Random Forest*, ya que tiene una probabilidad muy baja de predecir incorrectamente el cambio de la serie de delitos de la ENVIPE ($P(X)=0.05$ dentro de la muestra y $P(X)=0.18$ fuera de la muestra). Las otras dos especificaciones sirven para dar robustez de los resultados, ya que van en el mismo sentido que la literatura de los determinantes de la criminalidad, en donde las variables económicas y laborales ayudan a explicar el movimiento de los delitos en México.

También, comparando la relación que existe entre las cifras de incidencia delictiva del Secretariado y la ENVIPE con el gasto público, se encontró que hay una correlación baja con la primera, pero alta con la segunda. Las pruebas de causalidad de Granger con las cifras del Secretariado apuntan a que el gasto en impartición de justicia tiene un efecto marginal sobre la incidencia delictiva. El gasto en seguridad nacional parece tener una relación bidireccional con la incidencia delictiva. El problema es que el gasto en orden público, que debería ser la variable más importante, porque contempla en gasto en la policía, no parece causar a la incidencia delictiva. Para el caso de las cifras de la ENVIPE, el gasto en orden sí causa a la incidencia delictiva y viceversa.

En futuras investigaciones se podría replicar esta metodología, para aproximar el comportamiento de los delitos agregados por estado. También, se podría revisar el comportamiento de las cifras por tipo de delito, lo que ayudaría a incrementar la precisión con la que el pronóstico fuera de la muestra refleja el comportamiento real de la criminalidad en México.

¹El cálculo de la cifra negra se realiza con un rezago de 1 año. Entonces, aunque los *policy makers* quisieran tomar en cuenta el nivel de denuncias, tendrían que esperar todo un año para contar con información.

Introducción

En México, durante el mes de agosto de 2022, se cometieron 187,803 delitos del fuero común, de los cuales 3,580 fueron homicidios, lo que da un promedio de cerca de 120 asesinatos diarios. El problema de la inseguridad afecta de manera psicológica y económica a la sociedad, esto lo vuelve un tema de vital importancia para la política pública. De acuerdo con la Encuesta Nacional de Victimización de Empresas (ENVE), la pérdida económica, a consecuencia de la inseguridad, asciende a los 120.2 mil millones de pesos, lo que representa un 0.67% del Producto Interno Bruto (PIB) (INEGI, 2022a).

De igual manera, al tratarse de un problema estructural que ha afectado al país desde hace décadas, el gobierno federal ha tenido la oportunidad de implementar diferentes políticas para combatir la inseguridad. La más notable se llevó a cabo entre 2006 y 2012, con la estrategia denominada “combate frontal al narcotráfico”, lo que llevó a un incremento importante en la mayoría de las cifras de incidencia delictiva (Piñeyro, 2012). Recientemente, el gobierno federal actual adoptó una postura opuesta a la anterior, conocida como “abrazos y no balazos”; sin embargo, esto no ha producido los resultados deseados ya que la criminalidad sigue aumentando. En otras palabras, a pesar de los esfuerzos gubernamentales en la política pública de seguridad, la violencia sigue siendo un problema que aqueja cada vez más a la sociedad mexicana. Por ello, para diseñar políticas públicas que mejoren este panorama, es necesario que los hacedores de política conozcan a fondo las causas del problema. Al respecto, existe poca literatura con un enfoque económico (pues dicho enfoque se relaciona en su mayoría con otros problemas sociales) que aborde las causas de la inseguridad.

El problema que se presenta al momento intentar diseñar políticas públicas de seguridad, es que no todos los delitos son reportados, esto quiere decir que hay una importante proporción de crímenes que no se reflejan en las cifras de incidencia delictiva. Así pues, para poder calcular esta proporción de delitos, conocidos como “Cifra Negra”, es necesario utilizar la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE), pues contempla el nivel de subreporte de los delitos. Específicamente, la cifra negra estima la proporción de delitos que no tienen averiguación previa ni carpeta de investigación (INEGI, 2022c).²

En México, no existe literatura que hable sobre el problema de la subestimación de las cifras de incidencia delictiva. Dicha subestimación es de vital importancia, pues el nivel de denuncias afecta directamente el número de delitos que reporta el secretariado. Esto quiere decir que, si bien existen casos en donde pareciera que hay una disminución de los delitos totales, lo que está ocurriendo es que la cifra negra aumentó (hubo menos denuncias). Lo anterior se denominará en adelante como “el problema de la cifra negra”. Además del problema señalado, se debe considerar que el cálculo de la cifra negra se realiza con un rezago de 1 año. Entonces, aunque los *policy makers* intentaran tomar en cuenta el nivel de denuncias para elaborar mejores estrategias de seguridad, tendrían que esperar este tiempo para contar con la información.

Por ello, en este trabajo de investigación propongo una revisión de las cifras de incidencia delictiva utilizando

²De acuerdo con INEGI (2022c) la cifra negra se refiere a los actos delictivos que no son reportados ante el Ministerio Público o que no son objeto de una averiguación previa y, por tanto, no figuran en ninguna estadística.

Machine Learning, para generar un modelo que prediga el comportamiento de la cifra negra de manera oportuna y precisa. Esto permitirá tener series de datos que muestren la verdadera tendencia de la criminalidad para la toma de decisiones. Además, servirá a quienes diseñen políticas de seguridad pública en evitar tomar decisiones erróneas con respecto a la tendencia de los delitos (creer que la criminalidad va disminuyendo).

Las motivaciones de la investigación son: primero, entender la situación actual de la incidencia delictiva y su evolución, comparando las cifras de las dos principales fuentes de información (El Secretariado y la ENVIPE). Segundo, encontrar el papel que tiene el gasto público en la evolución de la criminalidad. Tercero, utilizando algoritmos de *Machine Learning*, estimar una serie que refleje los delitos de la ENVIPE de forma oportuna.

Lo anterior aportará a la literatura de la criminalidad en México, ya que esta investigación genera un método replicable para producir cifras de incidencia delictiva sin el problema de la cifra negra. También, abro la discusión a la inclusión de otras alternativas de generación de datos a través de *Machine Learning*. Finalmente, con esta investigación aportaré más información al diseño de políticas de seguridad pública.

Dentro de los resultados de la investigación tenemos que existe una gran diferencia entre lo que nos dicen las cifras de incidencia delictiva del Secretariado y las cifras de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE). Esto se cumple para los delitos agregados, y en menor medida para cada tipo de delito. También, utilizando pruebas de causalidad de Granger se muestra la baja correlación entre las cifras de incidencia delictiva del Secretariado, a diferencia de las de la ENVIPE, y el gasto programado en seguridad. Finalmente, usando *Random Forest* se logró generar una serie consistente, que ayuda a reducir el rezago temporal de la ENVIPE, y que puede ser utilizada en otros contextos, ya que tiene una probabilidad muy baja de predecir incorrectamente el cambio de la serie de delitos de la ENVIPE [$P(X)=0.05$ dentro de la muestra y $P(X)=0.18$ fuera de la muestra].

El documento se divide de la siguiente manera, además de esta introducción, se hace una revisión de literatura para explicar la importancia de contar con cifras de incidencia delictiva oportunas. Posteriormente, se hace una presentación de los datos, donde se explican las diferencias de cada fuente de información, sus ventajas y desventajas. Después, se incluye un apartado sobre los hechos estilizados, donde se presenta el comportamiento de las principales variables de interés. Posteriormente, en los aspectos econométricos se presentan los tres métodos de *Machine Learning* que se emplearan para estimar y pronosticar el comportamiento de la incidencia delictiva. Finalmente, se presenta la discusión de los resultados donde evaluó el desempeño de los modelos dentro y fuera de la muestra. Finalmente, presento las conclusiones de la investigación.

1. Revisión de Literatura

En este apartado se presenta lo que se ha hecho con respecto al tema, se dividirá la literatura en tres partes. La primera hablará sobre los costos que conlleva un alto índice de criminalidad. La segunda, sobre los determinantes de la incidencia delictiva y, la tercera, sobre los determinantes del nivel de denuncias o subreporte.

Gran parte de la justificación de este trabajo de investigación recae en el costo económico y social que trae consigo la alta criminalidad. México es considerado uno de los países más inseguros. De acuerdo con el último reporte del índice de seguridad de la OECD (2020) el país ocupa lugar 40 de 41 países. Por lo que es imperativo que las autoridades encargadas de hacer políticas públicas de seguridad cuenten con toda la información posible sobre el costo del crimen. En este sentido, contabilizar la pérdida generada por delitos ayuda a mejorar la evaluación de las políticas de seguridad, tal y como proponen Brand & Price (2000). De acuerdo con los autores, existen diversas maneras de clasificar los costos del crimen y proponen tres formas: anticipación, que se refiere al gasto realizado en medidas para reducir el riesgo de victimización; remediales, que mide el costo como consecuencia del delito, ya sea por pérdida de un bien, tiempo, y entre otros; respuesta, que son los costos que resultan posteriores al crimen como el costo policial de atender una emergencia, el costo de procesamiento de los criminales, el costo de mantenerlos en prisión, y algunos otros que no están relacionados con las pérdidas generadas por el delito.

Para México, existe literatura que estima el costo de la inseguridad siguiendo la metodología de Brand & Price (2000), como en el artículo de Romo (2018) que por medio de un panel con efectos fijos encontró que la pérdida asociada a la inseguridad y delincuencia fue de 5.69% del PIB en 2013. Asimismo, el artículo de Verdugo-Yepes et al. (2015) encuentra resultados similares en el periodo 1993-2012 para los estados mexicanos más afectados por el crimen. Para encontrar el efecto mencionado los autores utilizan un Vector Autoregresivo Estructural (SVAR) y a través de un impulso respuesta encuentran que el efecto del PIB es distinto de cero para todas las especificaciones.

Para el último cuartil de la distribución, de la criminalidad, los estados más afectados pierden cerca del 1.6% de su PIB tras un incremento de 1% en la tasa de incidencia delictiva estatal. También, encuentran que la elasticidad en la media de un incremento en la criminalidad sobre la inversión extranjera directa de los estados será de -0.1% durante los próximos 5 años. Siguiendo la misma metodología, Villoro & Teruel (2004) encuentran que para la Ciudad de México, entre 1997 y 1998, el costo del crimen fue de 3.6% de su PIB y, además, estimaron que el costo total de los homicidios en todo el país asciende al 0.6% del PIB total. Robles et al. (2013) utilizan el consumo de energía como *proxy* del PIB de México, y encuentran con un control sintético que después de la “guerra contra el narco” de 2006 las regiones más afectadas por la violencia presentaron una caída de casi 7% en su consumo mensual de energía, comparadas con aquellas que no fueron afectadas por la política de seguridad.

Para otros países latinoamericanos, Jaitman et al. (2015) encuentran que, en 2015, el costo asociado a la incidencia delictiva es cercano al de Verdugo-Yepes et al. (2015) para México. Por ejemplo, en Chile fue de 1.8 % de su PIB, en Costa Rica de 1.9 % y en Uruguay de 2.3 %. Para Honduras y Paraguay, los resultados se aproximan más a los de Villoro & Teruel (2004), con un costo de 4.6 % para el primero y 3.8 % para el segundo. Finalmente, Leo et al. (2022), por medio de una estimación en dos etapas con instrumentos encuentran una elasticidad del ingreso y la tasa de homicidios. De acuerdo con el artículo anterior, las personas estarían dispuestas a aceptar una reducción del 0.58 % de su ingreso mensual por una reducción del 1 % de la tasa de homicidios.

Entonces, existe un aproximado del costo económico de la criminalidad; no obstante, faltaría plantear, ¿por qué en México las tasas de incidencia delictiva son tan altas? En otras palabras, es necesario definir los determinantes del fenómeno. Esto es importante para la investigación, pues permite acotar el universo de variables explicativas que seleccionaré en el modelo. También, sirve de referencia para dar robustez a los resultados de los modelos de predicción de incidencia delictiva. El artículo seminal de Becker (1968), da una aproximación racional a por qué los individuos delinquen, y ocurre cuando las retribuciones económicas superan a las retribuciones de trabajar de forma legal, contemplando también la probabilidad de ser atrapados y la severidad del castigo.

De acuerdo con Fajnzylber et al. (1998), la literatura de los determinantes de la criminalidad se ha centrado en tres vertientes. La primera tiene que ver con la relación que existe entre los ingresos medios de las regiones. La segunda, con la distribución del ingreso de las comunidades, y la última con el nivel de educación de la población. En la actualidad, la literatura se puede seguir ubicando en alguna de las tres vertientes anteriores, sin embargo, la mayor parte de la investigación reciente se ha centrado en la distribución del ingreso como determinante de la incidencia delictiva.

Uno de los primeros artículos en cuestionarse la relación entre que una persona cometiera un delito y su ingreso fue Fleisher (1966). El artículo encuentra que en, áreas de alta delincuencia, el rol del ingreso juega un papel fundamental, ya que la elasticidad del ingreso-crime es de 2.5 %. Este resultado sugiere que el efecto es más que proporcional, y se puede explicar por un cambio en la estructura social en las áreas con alto crimen, y no en las preferencias por delinquir. Por ejemplo, el autor menciona que: un incremento de 500 dólares en la riqueza familiar reduce en 2.7 la tasa de divorcios de mujeres por millar. También, encuentra efectos de la tasa de desempleo sobre la delincuencia y, si bien no es tan grande como el ingreso, sigue manteniendo una elasticidad de 0.60 %. El autor al final advierte que estos resultados son replicables para áreas con alto nivel de delincuencia. De igual manera, Patterson (1991) encuentra resultados que van en la misma dirección, ya que un incremento en los niveles de pobreza aumenta la tasa de criminalidad.

Por su parte, Freedman & Owens (2011), realizan un estudio sobre un quasi-experimento de subsidios federales a la vivienda en Estados Unidos. Los autores encuentran que un aumento en la línea de crédito para construir una vivienda, que resulta en un incremento del ingreso disponible, puede reducir la tasa de crimen

local en 2 % en zonas con alta marginación y alta criminalidad. Más recientemente Calnitsky & Gonalons (2021) generaron un diseño experimental para ver si un ingreso garantizado a los individuos tiene efectos sobre las tasas de criminalidad. Los participantes, que eran familias de cuatro personas en el desempleo, recibieron un ingreso mensual garantizado de entre 35, 50 o 75 centavos de cada peso de su salario original. Lo que encontraron es que el ingreso garantizado redujo en 350 delitos por cada cien mil habitantes la tasa de criminalidad y, tomando en cuenta que la tasa de la región donde se aplicó el experimento era de 600 por cada cien mil habitantes, se puede ver una reducción de más del 50 %.

De igual manera, Soares (2004) busca explicar los determinantes de la criminalidad y las diferencias que existen entre países. Para México existe literatura que relaciona el factor del ingreso a la incidencia delictiva, como Ramírez (2014) y Lozano-Cortés et al. (2016) que encuentran que la variable más importante es el desempleo. También, Quiroz et al. (2015) encuentra que los crímenes, como el robo, comparten su ciclo con el de la actividad económica del país. Por último, existe literatura que discute el efecto del crimen sobre el ingreso (Enamorado et al., 2014), lo que podría generar una dinámica de retroalimentación como lo proponen Loría & Salas (2019).

En pocas palabras, la criminalidad genera estragos sociales y económicos, pero existe otro fenómeno que puede amplificar sus efectos negativos. Como se mencionó en la introducción, en México hay un problema de subestimación que puede afectar la estrategia con la que se combate la criminalidad, entonces, es necesario entender que lo provoca los movimientos de la cifra negra y que consecuencias tiene. Existe basta literatura que discute sobre la relevancia de la cifra negra, una de las primeras aproximaciones al concepto fue Biderman & Reiss (1967). Este trabajo advierte sobre la necesidad de tomar en cuenta el universo de delitos que no aparecen en las cifras oficiales. De igual forma, Skogan (1977) menciona que algunas de las consecuencias de la cifra negra son una limitada capacidad de prevenir los delitos, mala locación de los recursos policiales, entre otras consecuencias. Sin embargo, aunque es un dato importante para la política de seguridad pública de cualquier país, no es una tarea sencilla estimar la magnitud de este fenómeno. Penney (2014) menciona las dificultades que existen en la estimación de la cifra negra en Estados Unidos, como es que el nivel de denuncias se ve afectado por el grado de educación. También, que las encuestas de victimización se centran en los hogares, por lo que siguen fallando en capturar los delitos que sufren las personas sin hogar. El autor, además de mencionar otras dificultades, concluye con que es necesario seguir generando información que permita ver el panorama completo y no solo un fragmento.

En resumen, la cifra negra representa una gran área de oportunidad para la política pública de seguridad, es por eso que resulta fundamental entender sus determinantes. Greenberg & Beach (2004) estudian qué es lo que provoca que las personas denuncien los delitos en Estados Unidos y encuentran que son tres factores los más relevantes. El primero es el costo beneficio de realizar la denuncia. El segundo tiene que ver con una explicación emocional, que se deriva de experimentar emociones como el enojo y coraje. Finalmente, el factor social o rol que tiene el contexto en el que se encuentra el denunciante. Los resultados anteriores

replicaron lo encontrado en Greenberg et al. (1979), para el caso de los robos en Estados Unidos. Aunado a esto, Goudriaan & Nieuwebeerta (2007) realizan un experimento en Países Bajos para estudiar cómo afecta el lugar del crimen, si había relación entre la víctima y el victimario, o si el agresor pertenecía a la misma organización que la víctima. Encontraron que las tres variables, y una interacción entre el lugar y pertenencia a la misma organización, reducen dramáticamente la probabilidad de denuncia.

Finalmente, el artículo de Jaitman & Anauati (2019) explora los determinantes de la cifra negra para algunos países de Latinoamérica. En el artículo encuentran que la percepción de corrupción y la baja confianza en la policía se asocian negativamente con el nivel de denuncias. Benavente & Cortés (2006) encuentra para Chile que el ingreso individual, y el nivel de escolaridad incrementan la probabilidad de denuncia, mientras que ser mujer disminuye la probabilidad de denuncia. Para México solo existe el artículo de Carreto & Ramírez-Álvarez (2022), que muestra que la probabilidad de denuncias de los delitos, como robo, extorsiones y fraudes, crece con el nivel de escolaridad. También, la probabilidad de que alguien denuncie incrementa a medida que aumenta el valor de lo perdido a consecuencia del delito. Otro factor importante es que la percepción de corrupción sobre la policía afecta negativamente la probabilidad de denuncia sobre los delitos de robo a casa habitación y fraude.

Según la revisión de literatura, la mayoría de los artículos que modelan la incidencia delictiva, tienen el objetivo de analizar la relación causal de las variables explicativas, y no el de pronosticar la criminalidad (Leo et al., 2022; Verdugo-Yepes et al., 2015). Romo (2018) advierte que en su especificación puede haber relaciones que no sean estables en el tiempo y que, debido a eso, es probable que los pronósticos realizados con el modelo no sean precisos en el largo plazo. Adicionalmente, en ninguna de sus especificaciones se alcanza una bondad de ajuste mayor a .30, lo que deja implícito que las perdiciones pueden diferir en gran medida. Por su parte, existen trabajos que tratan de pronosticar la incidencia delictiva con modelos ARIMA, VAR, SVAR, entre otros, pero advierten que, por la naturaleza de las series, es difícil hacerlo con precisión. Las referencias consultadas comparten la idea de que los pronósticos de incidencia delictiva a través de métodos econométricos tradicionales son fiables para un periodo hacia adelante (Gorr et al., 2003; Shoesmith, 2013). Por último, un artículo que hace lo mismo para China con diversas especificaciones de ARIMA también advierte que el horizonte del pronóstico debe ser corto (P. Chen et al., 2008).

2. Datos

La información sobre los delitos proviene de dos fuentes oficiales, la primera es el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP), de donde se obtienen las cifras de incidencia delictiva. La segunda es el INEGI, que tiene información económica que se utilizará para el ajuste del modelo y que también genera la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE). Todos los datos son de periodicidad mensual, y se obtuvieron para el periodo de 2010M01 a 2021M12. A continuación, se presenta una breve descripción de las bases de datos para cada fuente de información.³

2.1. SESNSP

El SESNSP publica las cifras de incidencia delictiva de manera mensual, que contabilizan los presuntos delitos registrados en averiguaciones previas iniciadas o carpetas de investigación, reportadas por las Procuradurías de Justicia y Fiscalías Generales de las entidades federativas en el caso del fuero común y por la Fiscalía General de la República en el fuero federal.

Los delitos reportados pueden hacer referencia a una o más víctimas, o afectar a más de un bien jurídico. En esta base de datos existe una gran desagregación de la información, ya que contamos con información municipal por delito, sub tipo de delito y hasta el bien agraviado. En 2015 se realizó un cambio de metodología para desagregar algunos delitos, con la intención de contar con una mayor diversidad en la tipificación de los crímenes, véase el cuadro 1A del anexo.

El problema principal con las cifras del SESNSP es que dependen del nivel de denuncias. En otras palabras, todos aquellos delitos que no se denunciaron o que se denunciaron pero no se les inició una carpeta de investigación no aparecen en sus cifras de incidencia delictiva. Por ejemplo, el nivel de denuncias puede variar por factores que poco tienen que ver con la estrategia de seguridad, como el valor de los bienes perdidos, la confianza en las autoridades, u otras cosas. Si analizamos solo los cambios en las cifras del Secretariado, para evaluar o crear políticas de seguridad, estaríamos tomando en cuenta los efectos de las variables que inciden sobre el nivel de denuncias, y no solo los cambios generados por las estrategias de prevención y castigo del delito.

2.2. Cifras Económicas

Para poder pronosticar el valor de las cifras de incidencia delictiva de la ENVIPE es necesario tener información adicional a la que proporciona la encuesta. Las variables económicas que se seleccionaron se pueden dividir en dos bloques, el primer bloque mide la situación económica del país, y el segundo la situación laboral

³En el anexo se incluyen las estadísticas descriptivas de las variables empleadas en la estimación de los modelos, así como pruebas de raíz unitaria.

de las personas. La fundamentación de esto es que a medida que las condiciones económicas y laborales de las personas empeoran, se ven obligados a delinquir. En otras palabras, visto desde la óptica económica los individuos cometen crímenes cuando las recompensas de estos son mayores a los beneficios que obtienen por trabajar de manera legal (Becker, 1968).

Para medir la situación económica del país se utilizan series de alta frecuencia (mensuales) como el Indicador Global de la Actividad Económica (IGAE), un índice que mide el ciclo económico, la inflación anualizada, el indicador coincidente, consumo privado, actividad industrial, inversión fija bruta, confianza de los consumidores, expectativa de los consumidores sobre la situación económica del presente año y el siguiente.

La situación del mercado laboral se mide a través de las tasas de desempleo, informalidad laboral, y condiciones críticas de ocupación. También, se tienen variables ligadas con el desempeño de la mano de obra como las remuneraciones medias, y un índice de productividad manufacturera.

2.3. ENVIPE

El INEGI genera de manera anual la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública con el objetivo de obtener información representativa a nivel nacional y estatal que permita realizar estimaciones de la incidencia delictiva que afectó a los hogares durante el año previo de su levantamiento, así como la cifra negra. También, tiene información sobre la percepción de la seguridad pública, así como el desempeño de las instituciones a cargo de la seguridad pública y justicia. La información que se recaba es sobre las características del delito, el contexto de la victimización y el impacto económico y social del delito.

A través de los microdatos proporcionados por la encuesta podemos hacer estimaciones sobre el número de personas que sufrieron algún delito en el año previo al levantamiento de la encuesta. Con esto podemos calcular el número de delitos ocurridos el año anterior y obtener la “cifra negra” y sus causas. Sin embargo, es importante señalar que no todos los delitos están contemplados en esta encuesta, específicamente podemos rastrear el comportamiento los siguientes delitos:

1. Robo, que incluye; de vehículo, accesorios, refacciones, herramientas, a casa habitación, en la calle o transporte público, banco, cajero automático y otros.
2. Vandalismo.
3. Fraude bancario o al consumidor.
4. Extorsión.
5. Amenazas.
6. Lesiones.
7. Secuestro.
8. Agresión sexual o Violación.
9. Otros delitos distintos a los anteriores.

3. Hechos Estilizados

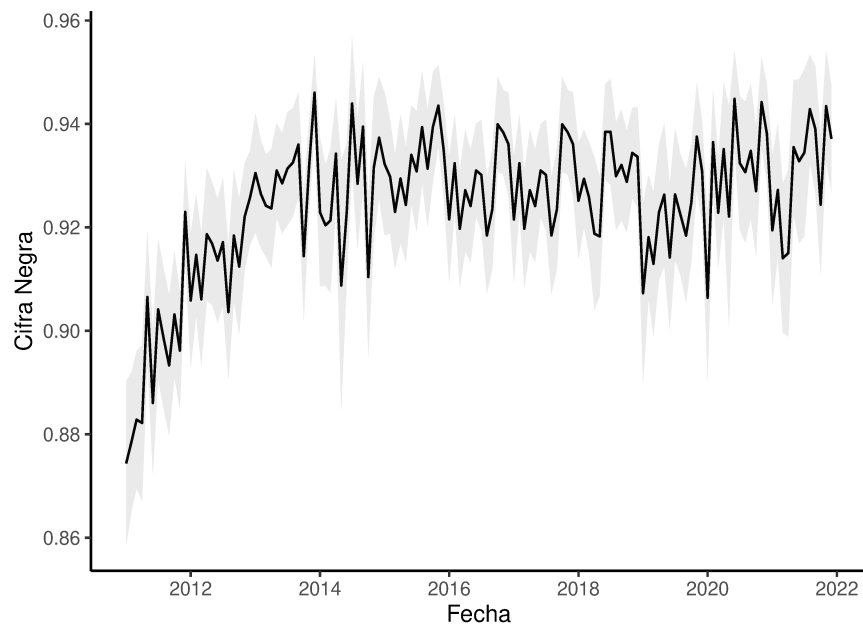
Esta sección se dividirá en dos partes. La primera muestra el comportamiento de la cifra negra y como su movimiento puede explicar las diferencias que existen entre los valores de los delitos que reporta el Secretariado y la ENVIPE. En la segunda parte se presenta la relación que existe entre el gasto público asociado a seguridad (prevención, persecución y castigo) y la incidencia delictiva (Secretariado y ENVIPE) para ver en que sentido va la relación. Por ejemplo, es fundamental saber si incrementos del gasto disminuyen la criminalidad o un aumento de la criminalidad genera una mayor asignación de recursos públicos al gasto en seguridad.

3.1. La Cifra Negra y la Incidencia Delictiva

México tiene un problema de inseguridad, además de la alta incidencia delictiva también tiene una de las cifras negras más altas del continente. De acuerdo con Jaitman & Anauati (2019) la cifra negra para Latino América y el Caribe fue en promedio de 87.9% entre 2004 y 2014, para países más desarrollados como Estados Unidos y el Reino Unido tenemos tasas de apenas 53% y 50.25% respectivamente. Para México, la cifra negra, en 2014 fue de poco más del 92% (INEGI (2022b)).

A continuación, se presenta la evolución de la cifra negra en México calculada de forma mensual, se excluye 2010 debido a que la muestra de la ENVIPE para su primer año de aplicación no permite obtener un valor preciso.

Figura 1: Cifra Negra, 2011-2021

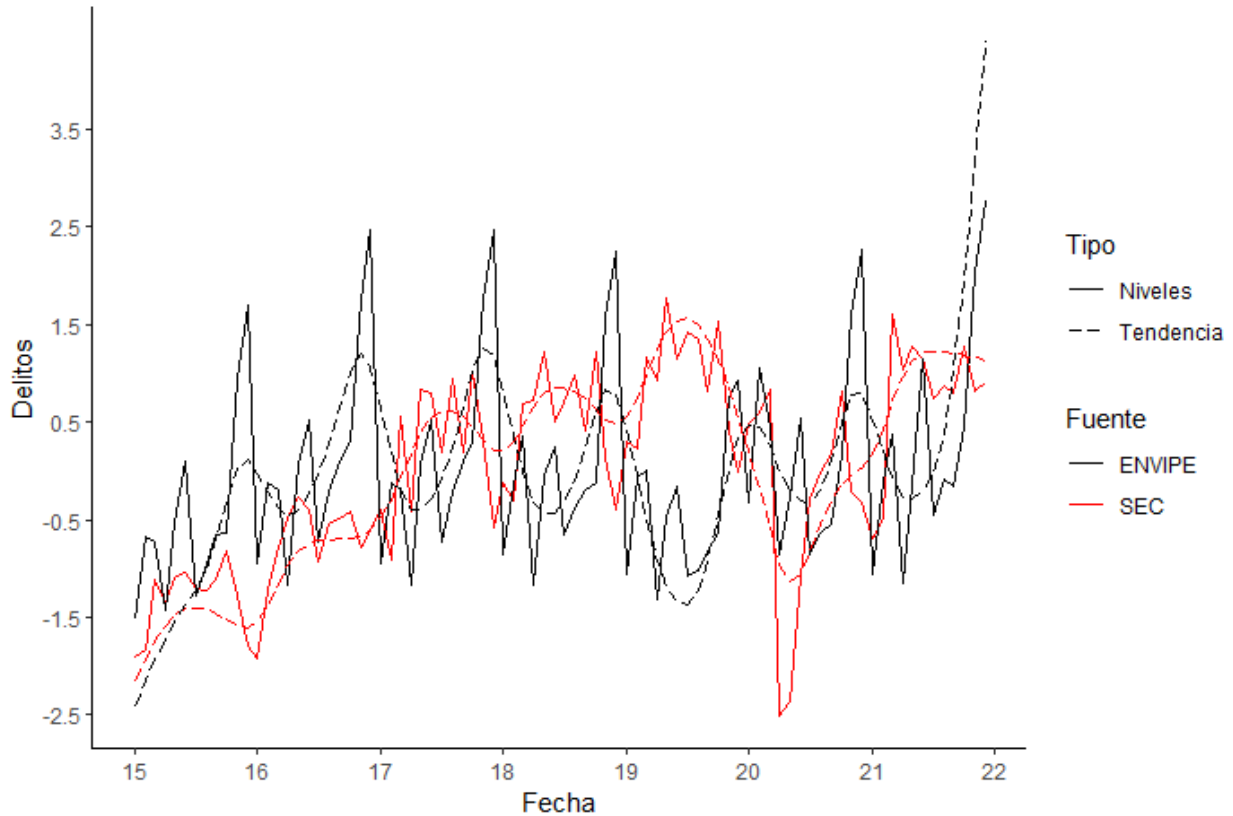


Nota: Elaboración propia con datos de la INEGI (2022b).

Es importante destacar el rápido incremento en la cifra negra de 2011 a 2012, para después estancarse en niveles superiores al 90 %, lo que indica que de cada 10 delitos solo se denuncia 1. La serie presenta un comportamiento estacional, ya que de octubre a diciembre vemos un incremento importante de sus valores mientras que de enero a mayo una disminución.

En la siguiente gráfica presento el comportamiento de las series de criminalidad de las dos fuentes principales de información, el SESNSP y la ENVIPE. Sabemos que la serie del Secretariado no reflejan el nivel verdadero de criminalidad debido al subreporte, lo anterior provoca que el nivel de ambas series difiera significativamente, por lo tanto se normalizaron usando el siguiente procedimiento $X_t^n = \frac{x_t - \bar{x}}{\sigma_x}$. La intención es comparar el comportamiento de las series, la del Secretariado que representa una pequeña proporción de los delitos, y la de la ENVIPE que representa el comportamiento real de la incidencia delictiva en México.

Figura 2: Cifra de delitos, 2015-2021



Nota: Elaboración propia con datos del INEGI (2022b) y el SESNSP (2023). La tendencia se obtuvo con el filtro Hodrick-Prescott ($\lambda = 14,400$) (Hodrick & Prescott, 1997).

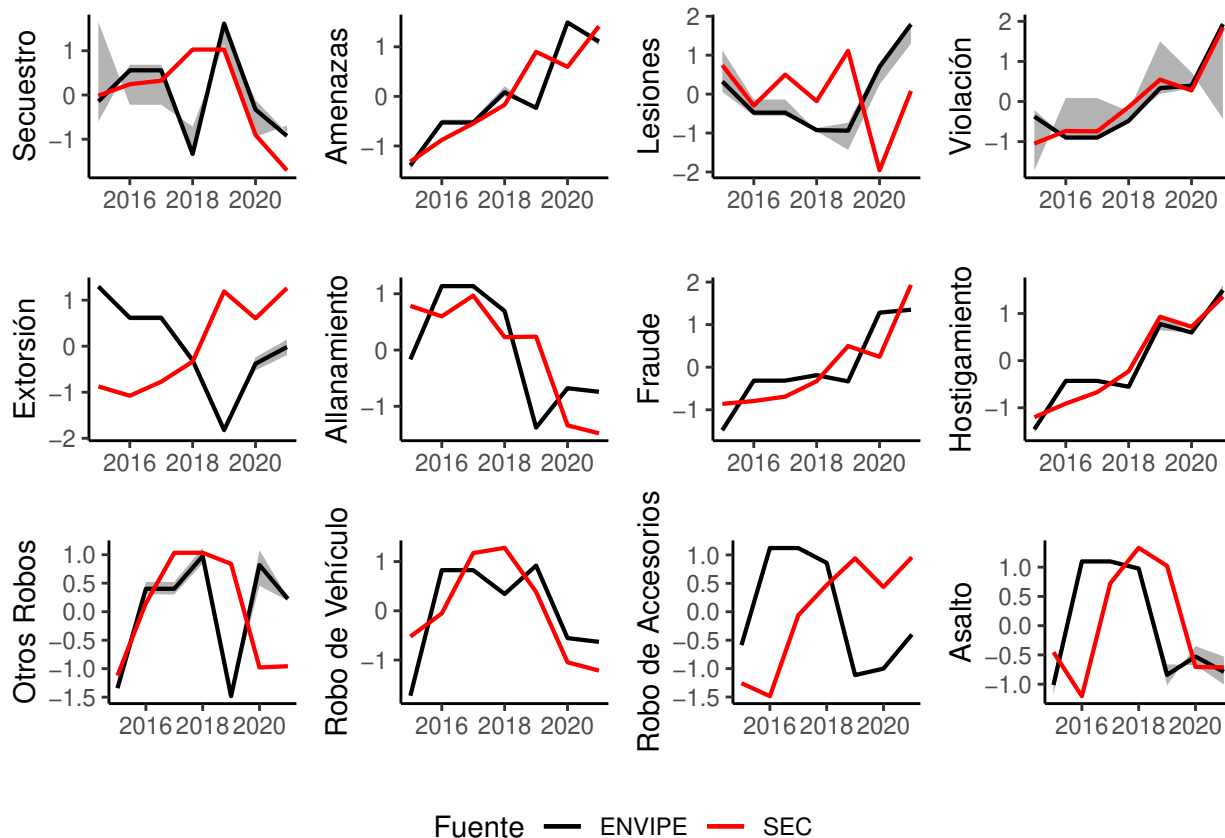
Algo que llama la atención es el comportamiento espejo que tienen las series, ya que solo al inicio del periodo y entre 2020-2021 vemos que sus trayectorias de largo plazo reflejan el mismo comportamiento. Sin embargo,

para los últimos 6 meses de 2015 y hasta el 2020, sus tendencias difieren drásticamente, ya que la ENVIPE muestra en algunos periodos un comportamiento creciente de los delitos, mientras que el Secretariado dice lo contrario. Durante el 2020 volvemos a ver una sincronización de las tendencias, reflejando el mismo comportamiento de incidencia delictiva, pero nuevamente, para el final del periodo las tendencias vuelven a diferir, y lo que llama la atención es el explosivo incremento de la tendencia de los delitos en la ENVIPE. Esto se puede explicar por la sensibilidad del filtro Hodrick-Prescott al inicio y al final de la muestra, por lo que una vez que se incluyan más observaciones hacia adelante debería atenuarse la subida más no modificarse la dirección.

En general podemos concluir que existe un problema de discrepancia en la tendencia de las series, lo que puede generar que la política de seguridad se sesgue si solo toma en cuenta las cifras del Secretariado. Si calculamos un coeficiente de correlación de la diferencia entre las series como $(ENVIPE_t - SEC_t)$ con respecto a la cifra negra tenemos que es $\rho = 0.49$. Lo anterior sugiere que casi el 50% de la diferencia entre las series se explica por el subreporte.

Lo anterior es para el agregado de delitos, pero podemos extender el análisis hacia categorías específicas de la incidencia delictiva como robos, extorsiones, fraudes, entre otros. En la siguiente figura se presentan las series anuales por tipo de delito del Secretariado y la ENVIPE.

Figura 3: Comparación de Tendencias de delitos, 2015-2021



Nota: Elaboración propia con datos del INEGI (2022b) y el SESNSP (2023). En gris se presentan los intervalos de confianza al 95 % para la serie del Secretariado.

Podemos notar que para la gran parte de los delitos la tendencia es similar, sin embargo, existen periodos donde la tasa de crecimiento de la serie es mayor en una u otra fuente. Por ejemplo, en el delito de amenazas tenemos que de 2015 a 2018 el comportamiento de ambas series es creciente pero en 2019 observamos que la cifra del secretariado sigue aumentando mientras que en la ENVIPE disminuyó. Esto se explica principalmente por la variación del nivel de denuncias, que para este delito aumentó 2.5 % ya que la cifra negra pasó de 0.89 a 0.86, lo que provoca que aparentemente suba la incidencia delictiva cuando en la realidad la gente está denunciando más. Siguiendo el mismo ejemplo, vemos que en 2020 la incidencia delictiva parece que cayó pero ahora vemos que la cifra negra pasó de 0.86 a 0.88 lo que representa un incremento de 2.2 %. Este mismo comportamiento se replica en otros delitos, como el de fraude entre 2019 y 2020. La reducción de la cifra negra para el delito de fraude entre 2018 y 2019 fue de 1.2 %, y para el periodo siguiente tuvo un aumento de cerca de 0.47 %.

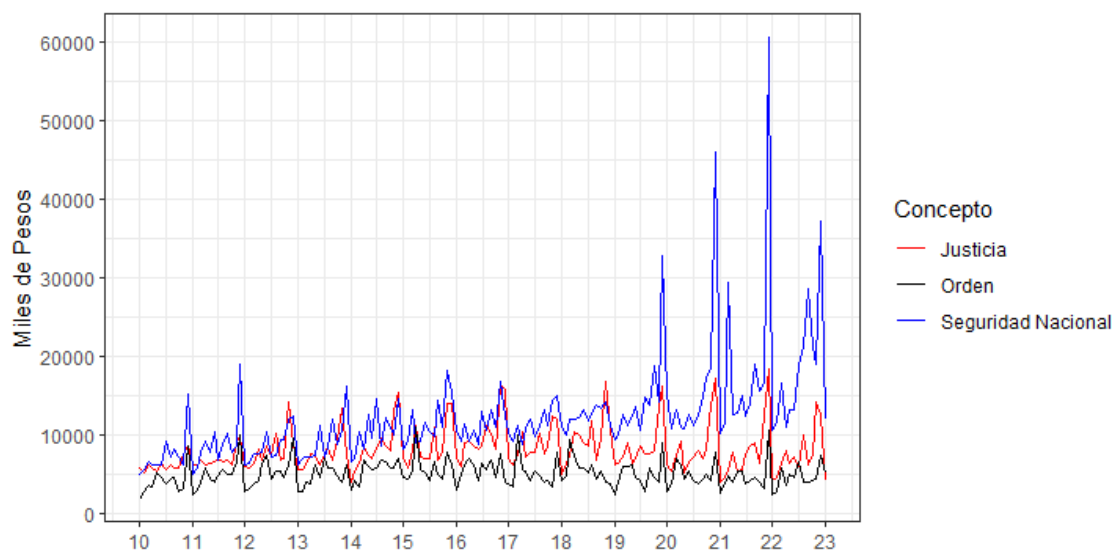
De lo anterior podemos decir que en términos generales el comportamiento de ambas series es distinto entre 2015 y 2021, sin embargo, cuando lo vemos por tipo de delito esta diferencia se concentra solo en algunos periodos y delitos específicos. La explicación es que para algunos delitos la cifra negra es más estable.⁴ Por ejemplo, para los delitos con una menor varianza como amenazas, hospedaje y fraude (0.0009, 0.0003 y 0.0002), podemos notar como el comportamiento de largo plazo es más parecido, pero para delitos como secuestro y lesiones (0.006 y 0.001) vemos como la diferencia es más marcada.

3.2. El Gasto Público y la Incidencia Delictiva

Otro aspecto fundamental es entender el gasto público en seguridad, y ver si éste depende o no de la incidencia delictiva. En la teoría esperaríamos ver que la asignación de recursos depende positivamente de la criminalidad observada. Es importante recordar que tenemos dos fuentes de información sobre la incidencia delictiva, el Secretariado y la ENVIPE. Lo que se puede observar es que de 2015 al primer mes de 2023, el gasto público a precios constantes se ha mantenido estable para los rubros de Justicia y Orden. Para el caso de Seguridad Nacional podemos ver un ligero incremento en su tendencia al final del periodo, que tiene que ver más con la participación de las fuerzas armadas en proyectos de infraestructura y administración que con labores de seguridad. La siguiente gráfica muestra el comportamiento de gasto público en seguridad en estos tres rubros:

⁴Se hizo una prueba de diferencia de medias para verificar que la cifra negra no se comportara como una constante. La variable constante que se seleccionó fue la media muestral de cada respectiva categoría redondeada a dos décimas.

Figura 4: Gasto Público en Seguridad, 2015M01-2023M1

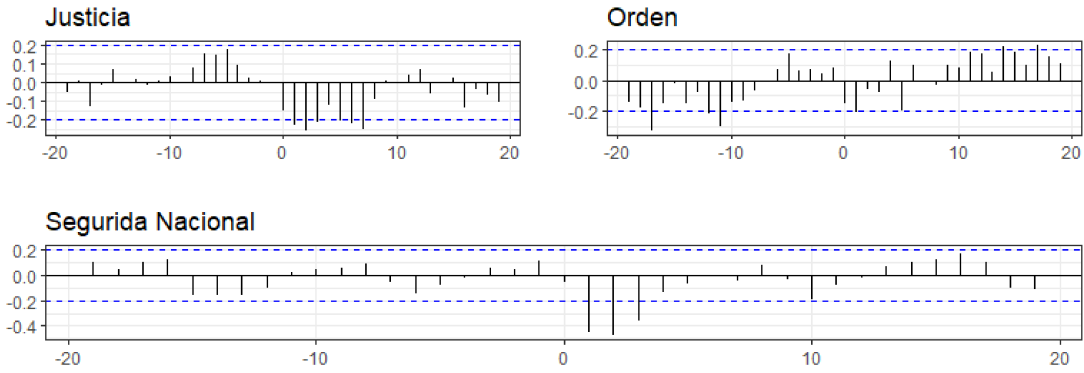


Nota: Elaboración propia con datos de SHCP (2023a).

El gasto en el primer concepto es para la impartición y procuración de justicia, reclusión, readaptación social, y cuidado de los derechos humanos. En el caso de Asuntos de Orden Público y de Seguridad Interior el gasto es para la policía, protección civil, sistema nacional de seguridad pública, y otros asuntos de orden público y seguridad interior. Finalmente, tenemos el gasto destinado a Seguridad Nacional que abarca el presupuesto de la Secretaría de la Defensa, Marina, e inteligencia para la preservación de la seguridad nacional.

Esperaríamos ver que existe una relación positiva y significativa del nivel de incidencia delictiva, independientemente de la fuente, con respecto a la asignación de presupuesto, principalmente con el concepto de Asuntos de Orden Público y Seguridad Interior. Este concepto es el que ayuda a la prevención del delito a través de la policía, y contempla las transferencias realizadas a las entidades federativas y municipios (véase SHCP, 2023b). A continuación, presento las correlaciones cruzadas entre los rubros y el nivel total de delitos reportados por el Secretariado:

Figura 5: Correlación Cruzada, 2015M01-2023M01

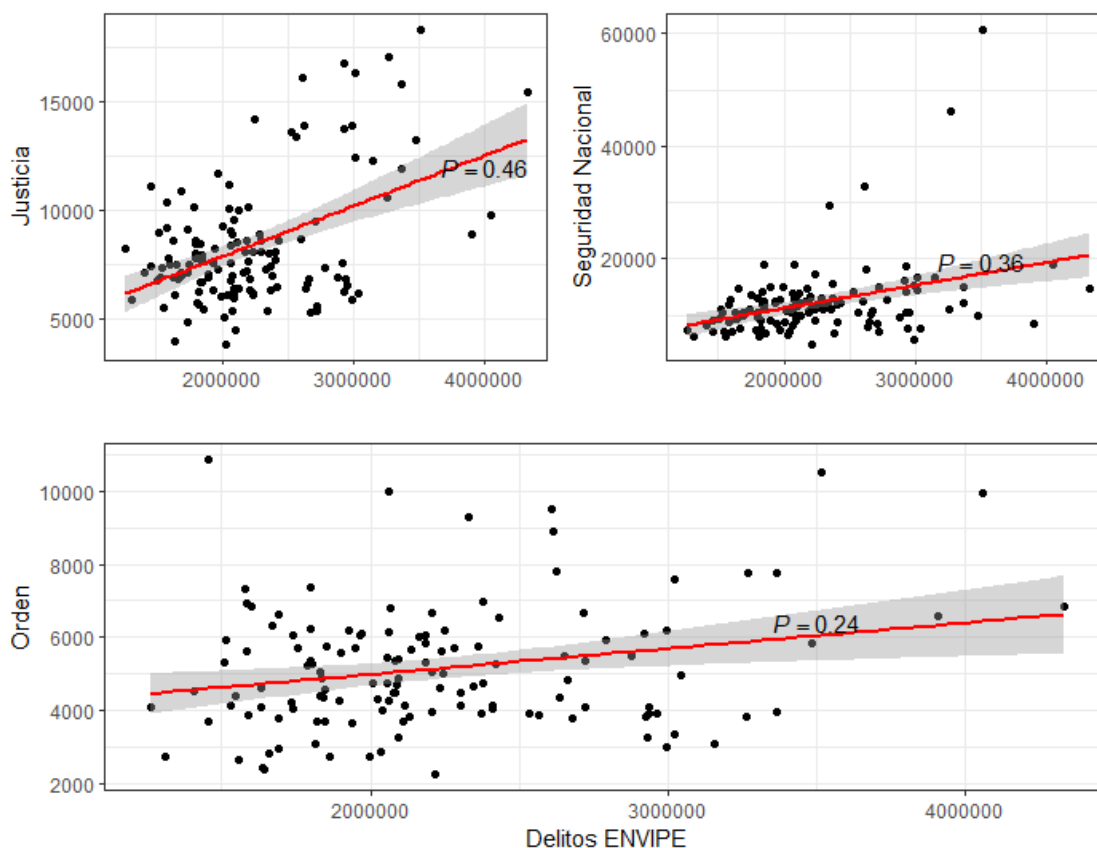


Lags	-5	-4	-3	-2	-1	0	1	2	3	4	5
Justicia	0.174	0.095	0.026	0.012	0.001	-0.154	-0.230	-0.255	-0.210	-0.117	-0.205
Orden	0.170	0.060	0.072	0.044	0.083	-0.156	-0.212	-0.056	-0.077	0.129	-0.196
Seguridad Nacional	-0.074	-0.025	0.061	0.049	0.119	-0.052	-0.453	-0.477	-0.358	-0.131	-0.072

Nota: Elaboración propia con datos de SHCP (2023a) y SESNSP (2023). En el cuadro se presentan los valores de ρ para los rezagos.

Los resultados de la figura 5 coinciden con los reportados en INEGI (2018), ya que no se encuentra evidencia de que exista una relación estrecha entre el gasto público en seguridad y el nivel de la incidencia delictiva del Secretariado. Las gráficas de correlación cruzada revelan que las relaciones entre la incidencia delictiva, medida como el agregado de todos los delitos, y los tres rubros de gasto en seguridad no son significativas de forma contemporánea. Sin embargo, para los casos del gasto en justicia y seguridad nacional vemos como existe una relación negativa y significativa de sus *leading lags* con la criminalidad, pero la magnitud es diminuta. Para el rubro de Asuntos de Orden Público y Seguridad Interior, tenemos que existe una relación negativa y significativa en el rezago 11 y 12, pero también es baja en relación a lo que esperaríamos teóricamente. A diferencia de las cifras del Secretariado, la incidencia delictiva reportada por la ENVIPE sí guarda una relación positiva y significativa de forma contemporánea con el gasto en seguridad. A continuación, se presentan los *scatter plots* por rubro del gasto:

Figura 6: Correlación Gasto y Delitos ENVIPE, 2011M01-2021M12



Nota: Elaboración propia con datos de SHCP (2023a) y INEGI (2022b). Todos los coeficientes de correlación son significativos al 99% de confianza.

El valor de los coeficientes de correlación de los delitos de la ENVIPE con todos los rubros del gasto, no es estadísticamente significativo para todos los rezagos. Sin embargo, para aquellos donde sí lo es tenemos valores negativos, mostrando que el efecto el gasto es negativo sobre los delitos. Como se observó en la gráfica anterior, la relación contemporánea es positiva lo que sugiere que los delitos presentes incrementan el gasto, y posteriormente este incremento genera una reducción paulatina de la criminalidad.

Otra pregunta que es necesario responder es el sentido de la causalidad, porque hasta donde sabemos la dirección puede ir en ambos sentidos. Es decir, es lógico pensar que el gasto en prevención y persecución del delito reduzca la incidencia delictiva, pero también puede ser el caso que a medida que incrementan los delitos se asignen más recursos a la policía. Para esto realicé el test de causalidad de C. W. J. Granger (1969) para ambas fuentes de información:

Variable en Diferencias	Secretariado		ENVIPE	
	Estadístico F	P-Value	Estadístico F	P-Value
12 Lags				
Delitos no causa a Justicia	0.92941	0.5246	2.317120	0.0121
Delitos no causa a Orden	1.65580	0.1014	1.410090	0.1748
Delitos no causa a Seguridad Nacional	2.33277	0.0163	0.339230	0.9797
Justicia no causa a Delitos	2.00732	0.0398	4.825470	0.0121
Orden no causa a Delitos	0.76209	0.6856	1.911827	0.0415
Seguridad Nacional no causa a Delitos	3.62974	0.0004	2.791330	0.0027
6 Lags				
Delitos no causa a Justicia	1.40008	0.2256	3.483660	0.0034
Delitos no causa a Orden	2.09356	0.0637	2.717610	0.0167
Delitos no causa a Seguridad Nacional	1.39620	0.2271	2.830770	0.0132
Justicia no causa a Delitos	2.80952	0.0160	2.992980	0.0094
Orden no causa a Delitos	0.47499	0.8249	2.066850	0.0626
Seguridad Nacional no causa a Delitos	0.74250	0.6171	1.331590	0.2489
4 Lags				
Delitos no causa a Justicia	1.52542	0.2024	3.101980	0.0181
Delitos no causa a Orden	2.12352	0.0852	2.309980	0.0618
Delitos no causa a Seguridad Nacional	1.19855	0.3178	3.611210	0.0082
Justicia no causa a Delitos	3.73648	0.0076	2.902670	0.0247
Orden no causa a Delitos	0.92660	0.4527	2.734330	0.0321
Seguridad Nacional no causa a Delitos	0.37932	0.8228	2.072910	0.0886

Nota: H_0 = La variable X no causa a Y.

Cuadro 1: Test de Causalidad de Granger, 2015M01-2022M12

La columna que se llama “Secretariado” nos dice (12, 6 y 4 rezagos) que la causalidad va del gasto en justicia a los delitos y no en relación inversa con un 95 % de confianza. Con 12 rezagos el gasto en seguridad nacional causa a los delitos al 95 %, pero los delitos también causan al gasto al 90 %, la relación causal se pierde con 6 y 4 rezagos. Otra cosa que llama la atención es que para el caso restante (gasto en orden), solo con 4 y 6 rezagos los delitos causan el gasto en orden y no en sentido contrario al 90 % de confianza. Es importante recordar que en la figura 5 vemos que los rezagos 1 al 10 las correlaciones no son significativas.

Algo que llama la atención es que la relación entre gasto y delitos se vuelve más estrecha para el caso de las cifras de la ENVIPE. La prueba sobre la relación del gasto en justicia y delitos de la ENVIPE señala que la dirección va en ambos sentidos para todos los rezagos. A diferencia del Secretariado, podemos decir que las cifras de delitos de la ENVIPE sí provocan que haya un incremento del presupuesto asignado a la impartición de justicia. También tenemos que el gasto en orden tiene una relación bidireccional al 10 % de significancia para 4 y 6 rezagos. Sin embargo, para el caso de 12 rezagos, se encontró que solo el gasto en orden causa a los delitos al 95 % de confianza. Finalmente, para el gasto en seguridad nacional no tenemos un relación estable en el tiempo con los delitos de la ENVIPE. Por ejemplo, para 12 rezagos tenemos que el gasto sí causa a Delitos al 99 % y no en caso contrario, pero para 6 rezagos ocurre lo opuesto. Finalmente, para 4 rezagos existe una relación en ambos sentidos con el 90 % de confianza.

Todo lo anterior ayuda sugiere que el gasto público en orden no responde a las variaciones de la incidencia delictiva de las cifras del Secretariado pero sí a las de la ENVIPE. Esto es un punto fundamental que afecta la eficacia de las estrategias de seguridad, ya que generalmente se toman como insumo las cifras del Secretariado. Esto fortalece la idea de que las cifras de delitos de la ENVIPE reflejan mejor el comportamiento de la criminalidad en México que las cifras del Secretariado.

Existe literatura oficial por parte del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública que reconoce el problema y busca generar mejores indicadores sobre la inseguridad para que se tomen mejores decisiones (SSPC-SESNSP, 2020). También, en el DOF (2020) la Secretaría de Protección Ciudadana presenta su Modelo Nacional de Policía y Justicia Cívica en donde mencionan que: “Por ello el Modelo Nacional de Policía considera como fundamentales los siguientes indicadores: a) Incidencia Delictiva; b) Percepción sobre seguridad; c) Homicidio doloso; d) Impunidad”. Entonces podemos decir que el presupuesto al interior de la entidad, exógenamente asignado, sí se emplea tomando en cuenta variables de criminalidad. Pero mientras no sea el presupuesto del rubro de orden el que responda a las fluctuaciones del crimen podemos esperar que se generen trade-offs en las estrategias de seguridad debido a la escasez de recursos. Por ejemplo, si durante el periodo t se asignó un presupuesto x para atender una cantidad y de delitos, tenemos que el presupuesto por delito es $\frac{x}{y}$, pero el siguiente año se asigna la misma cantidad porque el presupuesto es $x \perp y$, de acuerdo a lo visto en las pruebas de causalidad de Granger sobre los delitos del Secretariado. Entonces el año siguiente tenemos $j > y$ delitos. Finalmente, aunque dentro de la institución se asigne presupuesto en función de los delitos, la restricción presupuestaría provoca que se opere con $\frac{x}{j} < \frac{x}{y}$. Lo anterior genera un problema estructural en la estrategia de seguridad pública en México que es importante corregir si esperamos combatir al crimen de manera eficiente.

En el siguiente apartado presento otro indicador de delitos que puede servir para mejorar la toma de decisiones con respecto a la estrategia de seguridad. El indicador es el nivel de delitos mensuales que se obtienen a partir de la ENVIPE, pero además se presenta un pronóstico para evitar el problema del rezago anual de las cifras. Con esto los *Policy Makers* de seguridad pueden ampliar sus herramientas con las que toman decisiones.

4. Aspectos Econométricos

En este apartado se presentan los métodos de Machine Learning (Aprendizaje Automático) que se estarán empleando para calcular y pronosticar el número de delitos en México. Se escogieron estos métodos debido a que presentan diversas ventajas sobre los métodos econométricos convencionales, dentro de las cuales destaca la precisión para replicar y pronosticar el comportamiento de variables de interés (Ray, 2019).

Para validar lo anterior se realizaron diversas pruebas de estimación con VAR y ARIMA (4A y 6A), siguiendo la metodología de la bibliografía que ha realizado pronósticos sobre la incidencia delictiva (P. Chen et al., 2008; Gorr et al., 2003; Shoosmith, 2013). Específicamente, se estimó un VAR(1) con las cifras de incidencia delictiva, la primera diferencia del IGAE, Desempleo y la primera diferencia de TCCO, todas las series son $I(0)$ de acuerdo con el cuadro 2A en el anexo. Para el caso del ARIMA se buscó el modelo que minimizara el criterio de AIC y SIC, que resultó ser un ARIMA(1,0,0). Para ambos modelos se dejó fluctuar todos los parámetros hasta 12 rezagos. El pronóstico no captura correctamente el comportamiento de la serie hacia adelante, y en ambos casos la serie converge en el largo plazo a la media. Estos resultados reflejan que, para pronósticos de largo plazo (más de 3 periodos), se pierde precisión, lo que va en línea con artículos que comparan ML contra modelos tradicionales de pronóstico (Ning et al., 2022; Siami-Namini et al., 2018). Si bien, se puede mejorar la selección de rezagos, variables, y otros factores, eso escapa del objetivo de este trabajo de investigación.

Los métodos de Machine Learning se basan en el uso y desarrollo de sistemas informáticos que pueden aprender y adaptarse sin seguir instrucciones explícitas, mediante el uso de algoritmos y modelos estadísticos para analizar y extraer inferencias de patrones en los datos. Existen cuatro maneras de clasificar estos algoritmos de aprendizaje; supervisado, que quiere decir que para cada observación x_i con $i \in [1, \dots, n]$ existe una respuesta asociada y_i ; no supervisado, que observamos un x_i para cada i pero no contamos con un y_i asociado; semi-supervisado, ocurre cuando observamos x_i y y_i para un $m < n$ y posterior a eso no hay información de la variable de respuesta; y aprendizaje reforzado. Cada tipo de algoritmo tiene sus propios objetivos, y para fines de la investigación solo estaremos trabajando con algoritmos de aprendizaje supervisado, ya que son los que nos permiten generar pronósticos.

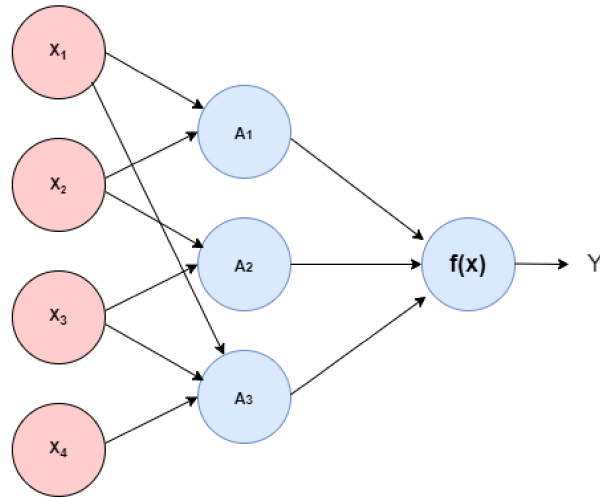
En específico, quiero encontrar un vector de variables que no formen parte de la ENVIPE que expliquen el comportamiento de la serie de delitos proveniente de la ENVIPE. Para ser consideradas deben cumplir con dos criterios: ser de periodicidad mensual, para poder replicar el comportamiento de los delitos con mayor frecuencia; y que estén disponibles de manera oportuna, para reducir el rezago de un año de la publicación de las cifras. En esta investigación se estarán empleando los métodos de Neural Networks (redes neuronales), Random Forest, y XGBoost para ajustar una regresión que nos permita explicar el comportamiento de la cifra negra. La gran parte de la descripción de los algoritmos se obtiene de Tibshirani et al. (2009) y más recientemente de Tibshirani et al. (2021).

4.1. Neural Networks

Las redes neuronales se han popularizado en la última década, con el resurgimiento del deep learning y los problemas de clasificación de imagen, texto y vídeo. Sin embargo, a pesar de que las redes neuronales tienen una gran variedad de aplicaciones, en este trabajo se utilizarán solo para generar un modelo de predicción.

La red neuronal que estimo toma como insumo un vector con 29 variables $X = X_1, X_2, \dots, X_{29}$ y genera una función no lineal $f(x)$ para predecir la respuesta de los delitos reales. Las variables que incluye el vector buscan representar el comportamiento de la economía mexicana, del mercado laboral, de las condiciones sociales, y expectativas de los agentes. Se incluyó el factor de mes para capturar los efectos estacionales de la serie. Lo que diferencia este método de otros, es la construcción de capas que permiten la creación de nuevas variables resultados de combinaciones entre las X_P con $P \in [1, \dots, p]$. A continuación se presenta un ejemplo gráfico de lo que es una red neuronal con $P = 4$.

Figura 7: Red neuronal



Nota: Elaboración propia.

La red anterior tiene la estructura más simple, con una capa de insumo y una capa oculta. La capa de insumo contiene las P variables que alimentarán a la capa oculta. En la capa oculta tenemos neuronas que denominaremos variables de activación que es una transformación no lineal de las variables X_P . Por ejemplo, podemos ver que A_1 resulta de la combinación de $X_1 = PIB$ y $X_2 = Desempleo$, y que A_3 resulta de la interacción de $X_1 = PIB$, $X_3 = Condiciones Críticas$ y $X_4 = Informalidad$ y. En el algoritmo estoy permitiendo que se busque la mejor red neuronal con 1 a 10 capas (c), dejando libres las interacciones. Para definir cual es la mejor red en términos de una regresión tenemos que $f(x) = Delitos ENVIPE$ se define como:

$$f(x) = \beta_0 + \sum_{k=1}^c \beta_c Z_c(x) \quad (1)$$

Tenemos que cada neurona de activación $A_c = Z_c(x) = g(V_{c0} + \sum_{j=1}^p V_{cj}X_j)$, y por lo tanto:

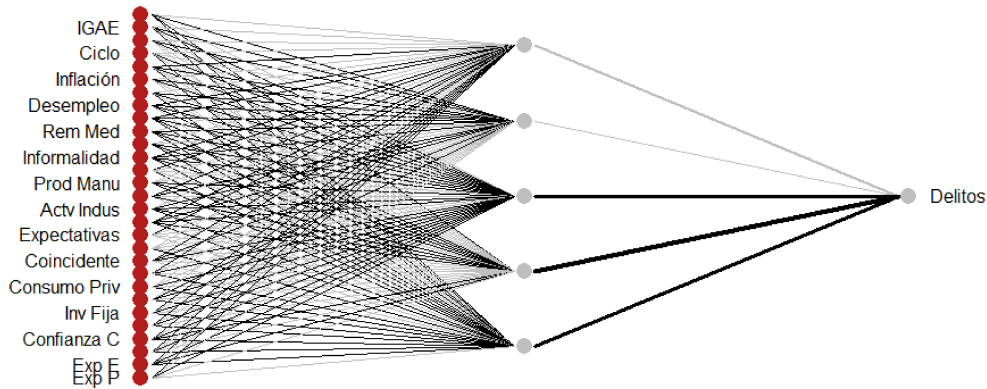
$$f(x) = \beta_0 + \sum_{c=1}^c \beta_c g(V_{c0} + \sum_{j=1}^p V_{cj}X_j) = \beta_0 + \sum_{c=1}^c \beta_c A_c \quad (2)$$

Debido a que se trata de un problema de regresión estaremos utilizando un $g(r)$ denominada *ReLU (rectified linear unit)*, para ver las ventajas de esta función véase Hara et al. (2015), que tiene dos casos $g(r) = 0$ cuando $r < 0$ y $g(r) = r$ en otro caso.

Para encontrar los valores de los β_c , y las interacciones óptimas para cada neurona de activación, el algoritmo recursivo encuentra el punto en el que se minimiza la suma de los errores al cuadrado $\sum_{i=1}^n (y_i - f(x))^2$.

Se probaron diversas especificaciones con la base de datos de delitos de la ENVIPE (2011-2020)⁵ para encontrar el mejor modelo de redes neuronales, desde *time slice*, hasta *cross validation*. El primer método genera una muestra aleatoria con bootstrap sobre la variable tiempo. Por ejemplo, removemos de la muestra un periodo (mes, trimestre, o año) y vamos evaluando el modelo de forma iterada hasta haber quitado al menos una vez cada periodo. Para el caso de *cross validation* simplemente hacemos una partición de la muestra para entrenamiento y prueba, evaluamos el modelo en la muestra de prueba. Esto se realiza de forma iterativa con bootstrap para encontrar el mejor modelo. Finalmente, se seleccionó la última especificación que generó la siguiente estructura de Red Neuronal:

Figura 8: Estructura de la Red



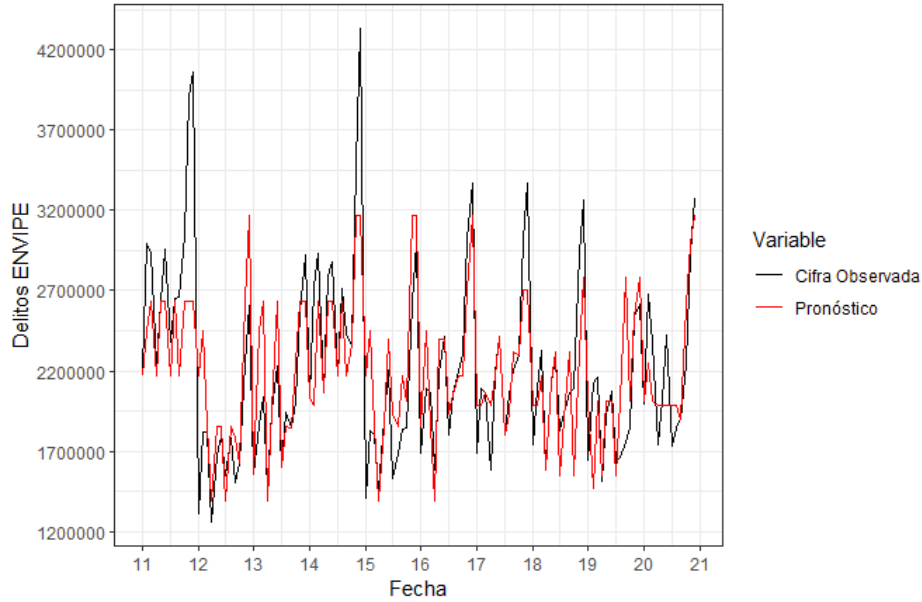
Nota: Elaboración propia. Se eliminó la etiqueta de los nodos insumo correspondientes a factores de mes. Las relaciones positivas se marcan con líneas negras y las negativas con gris.

Lo que podemos rescatar de la estructura anterior es que de los 29 insumos nos quedamos con una segunda

⁵Se quitó 2021 para poder realizar pruebas de robustez del pronóstico fuera de la muestra.

capa de interacciones que contiene solo 5 neuronas para finalmente darnos un valor aproximado de los delitos para cada periodo. Si comparamos lo observado con el pronóstico obtenemos la siguiente figura:

Figura 9: Ajuste de la Red, 2011-2020

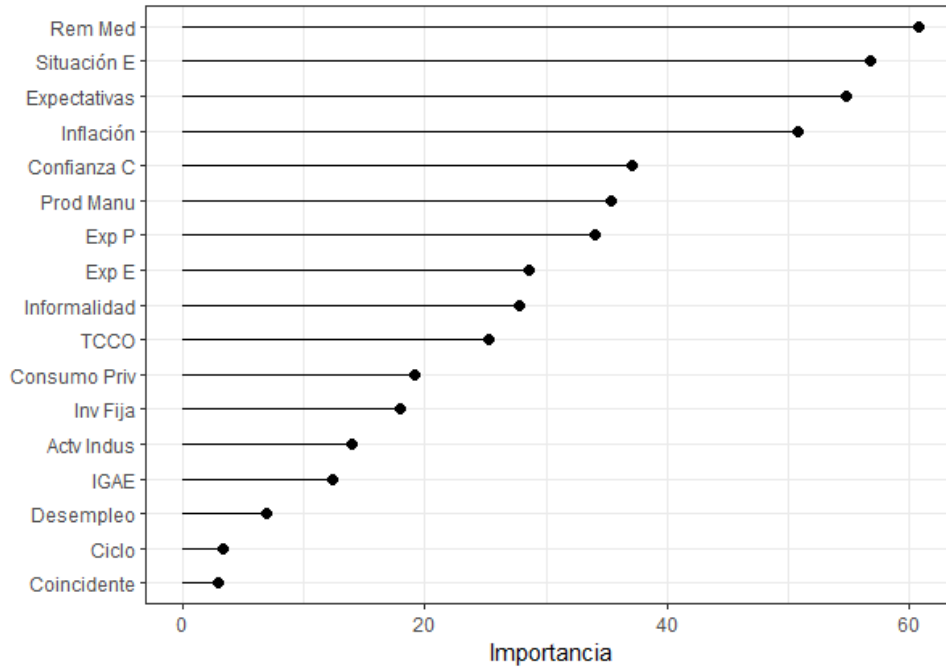


Nota: Elaboración propia con datos de INEGI (2022b) y el SESNSP (2023).

A pesar de que redes neuronales es un poderoso instrumento predictivo, la especificación tiene problemas para capturar la volatilidad de la serie. Esto ocurrió a pesar de que se incluyó un factor de mes para capturar el componente estacional. Pero aún así se puede rescatar la importancia relativa de las variables dentro del modelo. Esto se obtiene a través de medir la sensibilidad del pronóstico a movimientos aleatorios del valor del parámetro asociado al regresor. Intuitivamente, si un predictor es importante entonces asignar otros valores para ese β de forma aleatoria, con todos los demás predictores iguales, debería tener una influencia negativa en la precisión del pronóstico.

Entonces, para calcular los valores de la siguiente figura tomamos la Raíz del Error Cuadrático Medio (RMSE por sus siglas en inglés) con el conjunto de \hat{y}_o original y luego con el conjunto de \hat{y}_p modificado, y los comparamos de forma estandarizada dividiendo por la desviación estándar.

Figura 10: Importancia de variables en la Red



Nota: Elaboración propia con datos de INEGI (2022b) y el SESNSP (2023).

Se dejaron fuera las variables que explican el comportamiento estacional de la criminalidad que son los meses. En otras palabras, la figura anterior solo presenta las variables de interés que permiten hacer análisis económico. Recordando que tenemos tres tipos de variables que miden las de expectativas, la actividad económica y las condiciones laborales, podemos ver que la más relevante es de carácter laboral. Tiene sentido que las remuneraciones medias de la mano de obra sea una variable que prediga la criminalidad, así como las variables de las condiciones económicas, esto, de acuerdo con la literatura sobre los determinantes de la criminalidad (Calnitsky & Gonalons, 2021; Fleisher, 1966; Freedman & Owens, 2011; Patterson, 1991)

4.2. Random Forest

Inicialmente propuesto por Breiman (2001), Random Forest (RF) es otro método que genera ajustes lineales de manera eficiente en presencia de datos con alta volatilidad. Este método está basado en los árboles de decisión, que fueron creados para solucionar los problemas de clasificación. En otras palabras, generar particiones de la información basándonos en ciertos criterios. Para entender como funciona RF tenemos que explicar que es un árbol de decisión.

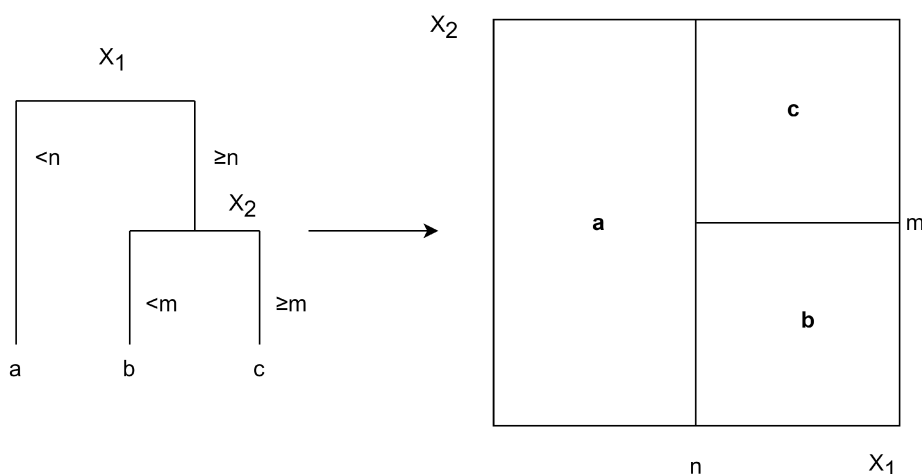
Para definir que es un árbol de decisión en el contexto de la investigación empecemos por considerar un conjunto de vectores con las mismas 29 variables que en la red neuronal anterior $\{(\vec{x}_1, y_1), \dots, (\vec{x}_T, y_T)\}$, entonces $\vec{x}'_t = (x_{t1}, \dots, x_{t29})$ contiene las características del periodo t con $t \in [1, \dots, T]$. y_t corresponde al

valor de los delitos en t de la ENVIPE. Una vez que tenemos los vectores de información, seguimos los siguientes pasos (Tibshirani et al. (2021)):

1. Dividimos el espacio generado por los predictores — esto es, el conjunto de posibles valores para los \mathbf{x}_t con $t \in [1, \dots, T]$ — en J regiones distintas y no sobrepuestas, R_1, \dots, R_j .
2. Para cada observación que cae en la región R_j , hacemos la predicción sobre la media de los valores de la región de la muestra de entrenamiento.

A continuación, se presenta breve ejemplo en \mathbb{R}^2 de como los árboles de decisión generan una partición R_i en el espacio Ω tal que $\bigcup_{i=1}^2 R_i \subseteq \Omega$. Tomemos una $y_t = \text{delitos}_t = [0, \infty]$ para todo t y $\mathbf{x}'_t = (x_{t1}, x_{t2})$ con $x_{t1} = \text{PIB}_t, x_{t2} = \text{Desempleo}_t \in \mathbb{R}$. Supongamos que lo anterior nos genera tres particiones sobre Ω que gráficamente se ven de la siguiente manera:

Figura 11: Árbol de Decisión



Las regiones pueden tener cualquier forma, y se generan de acuerdo a un criterio de minimización de la suma de los errores al cuadrado (RSS por sus siglas en inglés) usando un proceso de división binaria recursiva sobre:

$$\sum_{j=1}^J \sum_{t \in R_j} (\text{delitos}_t - \hat{\text{delitos}}_{R_j})^2 \quad (3)$$

Aplicándolo al ejemplo anterior, tendríamos que seleccionar el predictor PIB o Desempleo junto con un punto de corte s , tal que se minimize el RSS generando las regiones $\{X|X_j < s\}$ y $\{X|X_j \geq s\}$. El proceso recursivo considera a todos los predictores X_j (29 para nuestro caso de estudio) y todos los posibles valores de corte s , generandonos la función objetivo para $j \in [a, b, c] = [1, 2, 3]$ de:

$$\sum_{t:x_t \in R_a(j,s)} (y_t - \hat{y}_{R_a})^2 + \sum_{t:x_t \in R_b(j,s)} (y_t - \hat{y}_{R_b})^2 + \sum_{t:x_t \in R_c(j,s)} (y_t - \hat{y}_{R_c})^2 \quad (4)$$

Es importante entender que es un árbol de decisión, ya que es la base del algoritmo de Random Forest. Como su nombre lo sugiere (Bosque aleatorio) tenemos un número N de árboles de decisión, que se generan a partir de muestras aleatorias tomando $m = \sqrt{J}$ predictores X_j aleatorios con $j \in [1, \dots, J]$.

Similar a un proceso de bootstrap, generamos un predictor promediando el resultado de los N árboles, tal que minimizamos la varianza del estimador:

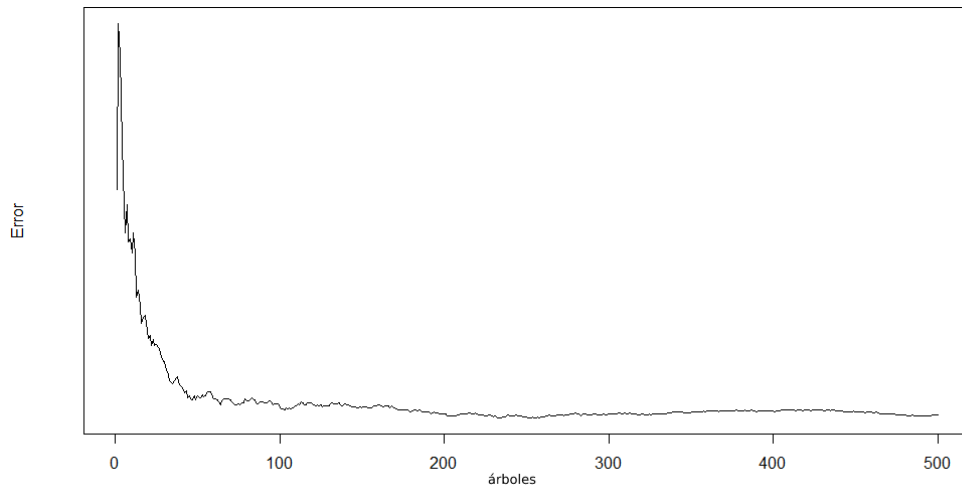
$$\hat{f}_{avg}(x) = \frac{1}{N} \sum_{n=1}^N \hat{f}^n(x) = \text{Delitos} \quad (5)$$

Pero a diferencia de otros métodos basados en árboles de decisión, como Bagging, tenemos que cada árbol es independiente. Esto debido a que para construir el RF, seguí los siguientes pasos:

1. Generé una muestra aleatoria A_n con bootstrap para tener el 40 % como insumo para entrenamiento.
2. Creé un árbol con A_n , y seguí iterando los pasos siguientes hasta encontrar el punto que minimizó la función objetivo RSS.
 - a) Seleccioné todos los predictores posibles (29).
 - b) Encontré las particiones óptimas para cada (X_j, s) .
 - c) Creé ramas dentro de los árboles a partir de los s encontrados.

Lo que produce el procedimiento anterior son árboles $T^b(\mathbf{x})$ no correlacionados, que generan estimadores de los delitos con menor varianza. A continuación se presenta la evolución RMSE a medida que se incrementa el número de árboles en el bosque.

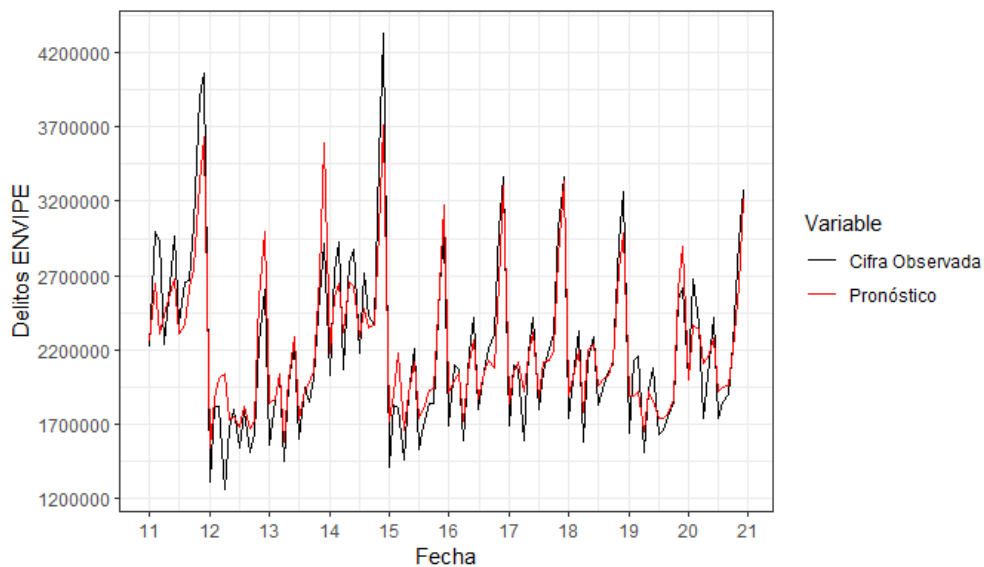
Figura 12: Error vs Árboles



Nota: Elaboración propia. El error de prueba se muestra como una función del número de árboles.

El número óptimo de árboles fue de 500, y por árbol tenemos que el número de nodos es en promedio de 63. Se puede discutir que la mejora en el pronóstico es marginal a partir del árbol 100, ya que el poder computacional que requiere estimar 500 árboles con diferente profundidad es significativamente mayor a 100. Sin embargo, el objetivo de las estimaciones es generar valores lo más cercanos posibles a los observados, por lo que no se toma en cuenta el factor “eficiencia computacional” al momento de especificar un modelo. En la siguiente gráfica se presenta la serie estimada contra las observadas de 2011 a 2020.

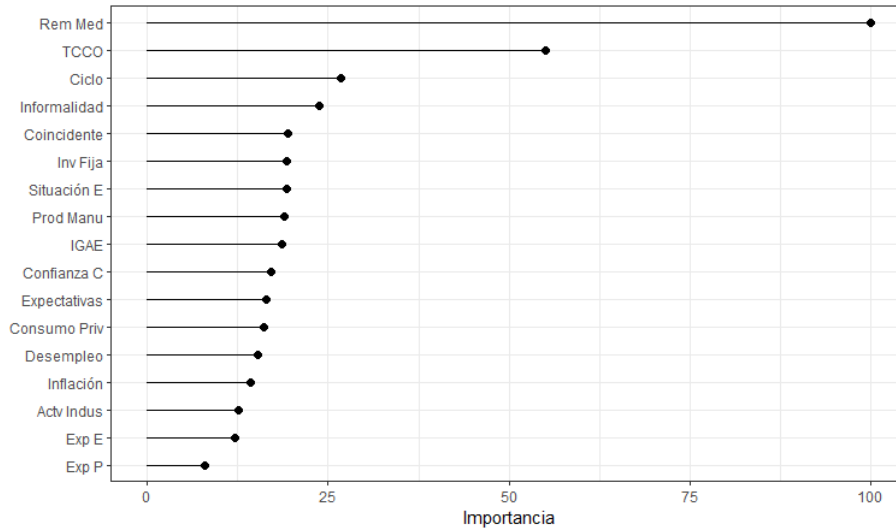
Figura 13: Random Forest, 2011M01-2020M12



Nota: Elaboración propia con datos de INEGI (2022b) y el SESNSP (2023).

Es evidente que el ajuste es mejor al que se tiene con la red neuronal, y otro aspecto importante es que el comportamiento de ambas series es similar. En el contexto de la investigación, nos interesa un modelo que replique los movimientos de la serie original y no exactamente el nivel de la serie. Igual que en el caso anterior, se presenta la figura de la importancia de las variables.

Figura 14: Importancia de variables en el RF



Nota: Elaboración propia.

Algo que le da robustez al resultado de redes neuronales es que también la variable remuneraciones medias fue la más importante, sin tomar en cuenta los meses. En este caso otra variable relevante es la tasa de condiciones críticas de ocupación. En conjunto ambas variables reflejan las condiciones laborales de las personas.

4.3. XGBoost

El algoritmo empleado en este apartado pertenece a la categoría de boosting, que son métodos que mejoran la predicción de los modelos. Boosting puede ser aplicado a diversas herramientas de Machine Learning, pero XGBoost es un algoritmo que se aplica específicamente para los algoritmos de Random Forest (T. Chen & Guestrin, 2016). Lo que genera este procedimiento es que cada árbol se construye de manera secuencial, esto quiere decir que no incluye muestreo por bootstrap ya que cada uno se genera con una versión modificada de la base original.

La idea detrás del procedimiento es que en lugar de ajustar un modelo a una serie de grandes árboles, lo que hace es generar un aprendizaje lento. Esto consiste en usar los residuales de los árboles generados previamente para construir uno nuevo. Los nuevos árboles generados con residuos puede contener pocos nodos para poder mejorar el $f(x)$ lentamente. También, contamos con un parámetro λ que ralentiza la velocidad del proceso, ya que permite que el algoritmo genere una mayor cantidad de árboles con diferentes formas usando los

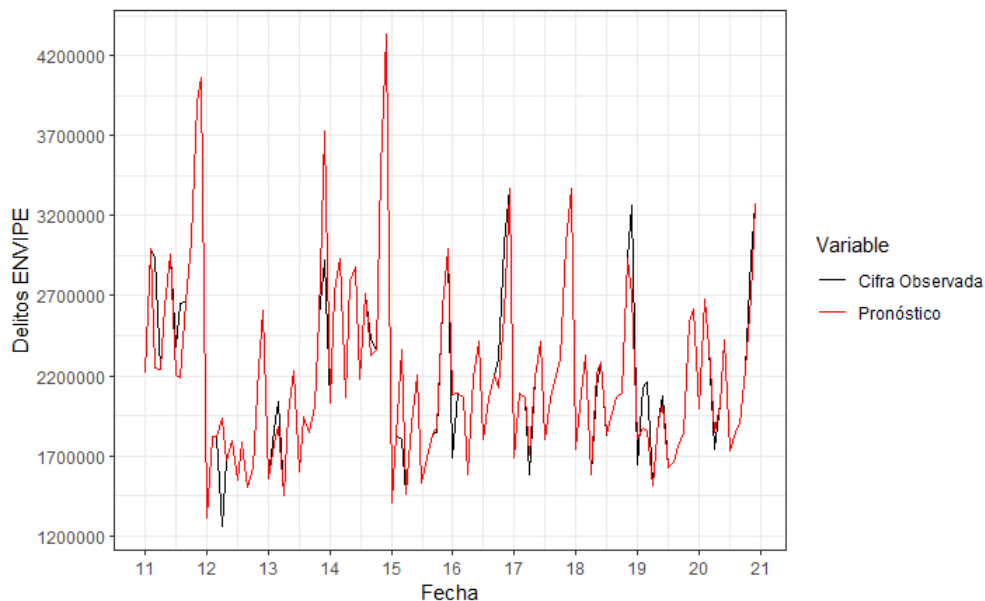
residuos de cada iteración. En este caso no existe independencia entre los árboles, debido a que la forma y tamaño dependen del árbol anterior.

Para generar el modelo de delitos con el procedimiento de boosting seguí los siguientes pasos, de acuerdo con Tibshirani et al. (2021):

1. Fijé $\hat{f}(x) = 0$ y r_i que son residuales de la primera estimación con random forest para todo i en $A(n)$ de entrenamiento (al igual que el caso anterior la proporción es de 60-40).
2. Repetí diez mil veces los siguientes pasos:
 - a) Ajustamos un árbol $\hat{f}^b(x)$ con una profundidad máxima de $d = 100$ particiones ($d + 1$ nodos terminales) en la muestra de entrenamiento.
 - b) Actualizamos $\hat{f}(x)$ añadiendo una versión simplificada del nuevo árbol: $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$.
 - c) Actualizamos los residuales: $r_i^2 \leftarrow r_i^1 - \lambda \hat{f}^b(x)$.
3. El resultado del modelo al que se le aplicó boosting: $\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$.

Algunas consideraciones que hay que tener al usar boosting es que si B es muy grande podemos sobre ajustar el modelo, y para lidiar con ello hay que hacer validación cruzada. El parámetro λ guarda una relación inversa con B , ya que a medida que hacemos más lenta la velocidad de aprendizaje, necesitamos un mayor B para garantizar un buen ajuste. Finalmente, el número de particiones d en cada árbol se refiere a la profundidad de interacción, y controla directamente el orden de interacción de los árboles. Este modelo es el que mejor ajuste dentro de la muestra generó. En la siguiente figura se presenta el pronóstico junto con la serie original.

Figura 15: XGBoost, 2011M01-2020M12

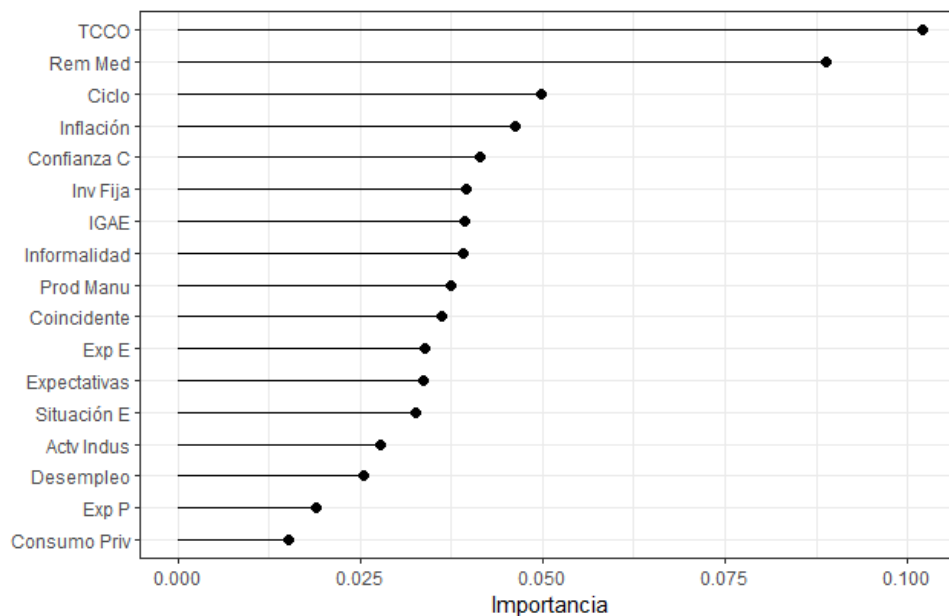


Nota: Elaboración propia con datos de INEGI (2022b).

Para el caso de Random Forest ya teníamos un buen ajuste, pero podemos notar que al aplicarle Boosting logramos mejorar aún más el pronóstico dentro de la muestra. Si obtenemos el coeficiente de correlación entre ambas series en niveles tenemos que $\rho = 0.94$, en tasa de crecimiento ($\frac{x_t - x_{t-1}}{x_{t-1}}$) es de $\rho = 0.90$. Entonces, los valores que se estiman por XGBoost presentan un grado de asociación lineal muy cercano a uno.

En la siguiente gráfica se presenta el cálculo de la importancia de las variables dentro del modelo. Al igual que en los casos anteriores se omiten los factores de mes, ya que solo ayudan a ajustar por el componente estacional de la serie de incidencia delictiva.

Figura 16: Importancia de variables en el XGBoost



Nota: Elaboración propia.

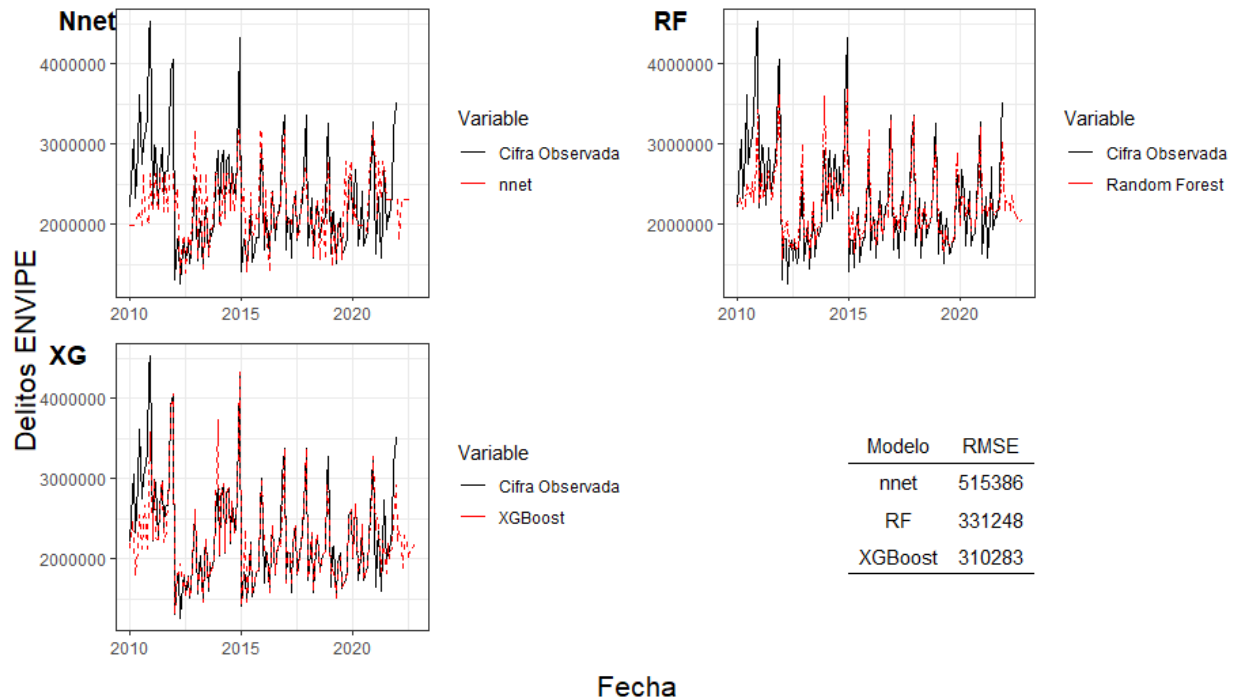
En este modelo las dos variables más relevantes son la Tasa de Condiciones Críticas de Ocupación y las remuneraciones medias. En general, vemos que las variables que describen las condiciones laborales son las más relevantes para explicar el comportamiento de la serie de incidencia delictiva. En este sentido, todas las especificaciones apuntan en la misma dirección, ya que tanto Redes Neuronales como Random Forest coinciden en que las remuneraciones medias son las más importantes. XGBoost y Random Forest contemplan la precariedad de los trabajos, lo que mejoró el ajuste sustancialmente.

Para poder escoger un modelo es necesario compararlos dentro y fuera de la muestra. En el siguiente apartado se presentan los resultados de cada estimación para contrastar su desempeño a la hora de generar un pronóstico.

5. Discusión de los Resultados

En este apartado se presentan los tres modelos estimados para hacer una comparativa de su desempeño a la hora de pronosticar el comportamiento de la incidencia delictiva real en México entre 2010 y 2021. En la siguiente figura presento el RMSE, y las gráficas del valor observado y pronosticado dentro de la muestra (*in-sample*). Esto quiere decir que estamos midiendo qué tan bien se ajustó cada uno de los modelos para valores de incidencia delictiva observados.

Figura 17: Pronóstico Dentro de la Muestra, 2011M01-2020M12



Nota: Elaboración propia con datos de INEGI (2022b).

Como podemos notar, el RMSE del Random Forest no es tan diferente al de XGBoost a pesar de que visualmente pareciera que el último es mejor por mucho. Para dar robustez a la elección del modelo utilicé las dos especificaciones del coeficiente de desigualdad de Theil que permite comparar los tres modelos (Theil, 1958, 1966). En la literatura, existe la convención de denotar U^1 y U^2 a estas dos especificaciones, ambas buscan medir la calidad de los pronósticos generados, pero U^1 nos permite medir las proporciones de sesgo en la media (U^M), varianza (U^S) y covarianza (U^C). El indicador U^1 se presenta de forma similar a la de Pindyck & Rubinfeld (1997):

$$U^1 = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n y_t^2 + \frac{1}{n} \sum_{t=1}^n f_t^2}} \quad (6)$$

Desarrollando el numerador de la expresión anterior tenemos:

$$\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2 = (\bar{y} - \bar{f})^2 + (\sigma_f - \sigma_y)^2 + 2(1 - \rho)\sigma_f\sigma_y \quad (7)$$

$$U^M = \frac{(\bar{y} - \bar{f})^2}{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2}; U^S = \frac{(\sigma_f - \sigma_y)^2}{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2}; U^C = \frac{2(1 - \rho)\sigma_f\sigma_y}{\frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2} \quad (8)$$

Sin embargo existe literatura que argumenta que es mejor usar U^2 para evaluar los pronósticos (Makridakis et al., 1998). En el caso de esta investigación usaremos ambas especificaciones, ya que pueden considerarse indicadores complementarios que si apuntan en el mismo sentido le dan más robustez a los hallazgos (Ahlburg, 1984; Bliemel, 1973; C. Granger & Newbold, 1973). A continuación, se presenta la formula de U^2 :

$$U^2 = \frac{\sqrt{\sum_{t=1}^{n-1} \left(\frac{f_{t+1} - y_{t+1}}{y_t}\right)^2}}{\sqrt{\sum_{t=1}^{n-1} \left(\frac{y_{t+1} - y_{t+}}{y_t}\right)^2}} \quad (9)$$

Podemos notar que para la segunda especificación se normaliza usando el rezago del valor y . Intuitivamente lo que nos dice esta medida es que tanto mejora el pronóstico con respecto a una predicción ingenua. La predicción ingenua es que el valor $y_{t+1} = y_t$, en otras palabras el mejor pronóstico es el último valor observado. Finalmente, el siguiente cuadro resume los valores de todos los indicadores para los tres modelos:

Indicador	NNET	RF	XGBoost
Primero			
U^1	0.0814	0.0481	0.0354
U^M	0.0019	0.0004	0.0016
U^S	0.1556	0.3044	0.0073
U^C	0.8509	0.7036	0.9995
Segundo			
U^2	0.1746	0.1019	0.0857

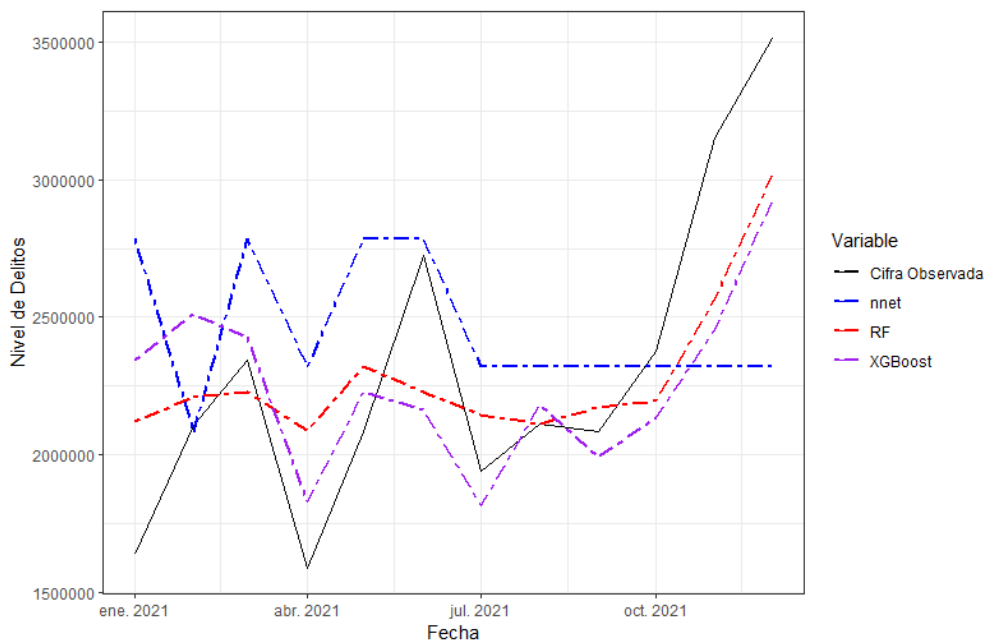
Cuadro 2: Evaluación del Pronóstico Dentro de la Muestra

De acuerdo a lo anterior sabemos que dentro de la muestra el mejor modelo es el XGBoost, ya que tiene menor coeficiente U^1 , U^2 y RMSE. Para el caso de U^M y U^S queremos que el valor sea lo más cercano a cero, mientras que de uno para U^C . Entonces, solo para el caso del sesgo en la media notamos que Random Forest supera a XGBoost, pero esto no cambia que en general el último modelo sigue casi a la perfección el comportamiento de la serie original.

Ahora presento los resultados de los modelos *out-of-sample*, que consiste en evaluar el pronóstico con valores que el algoritmo no ha observado. En este caso dejé fuera 2021, que es el último dato disponible de la ENVIPE, para comparar que dice el modelo que hubiera ocurrido y que ocurrió en la realidad. Los resultados fuera de la muestra son fundamentales, ya que lo que se pretende es proveer de información adelantada sobre

inseguridad a los *policy makers*. En la siguiente figura se observa el pronóstico de los modelos y la serie original de la ENVIPE.

Figura 18: Pronóstico Fuera de la Muestra, 2021M01-2021M12



Nota: Elaboración propia con datos de INEGI (2022b).

Podemos notar como el modelo de redes neuronales replica parcialmente la volatilidad de la serie al inicio, sin embargo a partir de julio el mejor pronóstico se vuelve una línea recta. Para el caso de Random Forest, vemos como falla en capturar la variabilidad de la serie observada, siendo una versión más suavizada de la misma. El modelo que parece replicar mejor el comportamiento de la ENVIPE es XGBoost, ya que es muy parecida a la estimada por RF pero con una volatilidad más parecida a la observada. Pero para poder definir objetivamente el mejor modelo es necesario revisar el cuadro 3.

Modelo	Theil					Otros		
	U^1	U^M	U^S	U^C	U^2	MAPE	SMAPE	RMSE
NNET	0.1322	0.0549	0.2523	0.7786	0.2699	0.2021	0.2087	639963
RF	0.0752	0.0039	0.7812	0.3053	0.1481	0.1251	0.1259	351268
XGBoost	0.0876	0.0185	0.4206	0.6501	0.1715	0.1411	0.1399	406580

Cuadro 3: Diagnóstico fuera de la Muestra

Sorpresivamente, de acuerdo con los coeficientes de Theil y las otras medidas de desempeño el modelo que mejor se ajusta *out-of-sample* es el RF. La red neuronal queda descartada como modelo, debido a que no generó una predicción que se acerque a lo observado. Para el caso de *XGBoost*, se puede argumentar que los valores de los indicadores no difieren en gran medida de los de RF, incluso U^S y U^C son mejores.

Lo anterior indica que la serie pronosticada con *XGBoost* tiene marginalmente una mejor correlación con la observada y ajusta mejor su varianza, que la serie pronosticada por RF. Sin embargo, para comparar ambos pronósticos, y asegurarme que RF es el mejor modelo, calculé la probabilidad de predecir correctamente la dirección del cambio sobre la serie de delitos de la ENVIPE dentro y fuera de la muestra. Si la serie observada crece en un periodo t se espera que también el pronóstico lo haga. En este caso la probabilidad de cometer un error a la hora de predecir si la serie crece, o decrece, con RF es de $P(X = Error)_{in_sample} = 0.07$ y $P(X = Error)_{out_sample} = 0.18$ respectivamente. Para *XGBoost*, las probabilidades fueron exactamente las mismas. Entonces, que U^S y U^C no sean mejores en el caso de RF no afecta su capacidad de seguir el movimiento de la serie original.

Los otros indicadores que tomé en cuenta para evaluar dependen del nivel de la serie, entonces en general el pronóstico realizado por RF sigue más de cerca el nivel de la serie observada. Se entiende que los delitos del secretariado están subestimados, por lo que el nivel de la serie resulta irrelevante; sin embargo, lo que no se desea tener es un indicador que subestime el crecimiento de las series. Si bien, se considera que RF y *XGBoost* tienen la misma probabilidad de error, adicionalmente se quiere elegir el modelo que subestime menos veces el crecimiento de la serie observada. En otras palabras, se debe comparar la probabilidad condicionada de que un modelo muestre una caída en los delitos cuando no lo hubo, una vez que se haya equivocado en la dirección del pronóstico. Entonces, $P(X < x|Y = Error)$ siendo $x = 1$, si los delitos crecieron y 0 en otro caso. Sorpresivamente, la probabilidad de ambos modelos es la misma, dentro de la muestra es de $P(X < x|Y = Error)_{in_sample} = 0.05$ y fuera de la muestra es $P(X < x|Y = Error)_{out_sample} = 0.18$.

Por todo lo señalado anteriormente, determiné que el mejor modelo es RF, debido a que fuera de la muestra genera el mejor ajuste, a pesar de que no supera *XGBoost* en U^S y U^C . Entonces, el modelo puede ayudar a pronosticar correctamente el cambio en las series de delitos de la ENVIPE un año hacia adelante el 82% de las veces. Esto se suma a las herramientas disponibles de los *policy makers* para generar políticas en seguridad públicas consistentes con el contexto de inseguridad.

Conclusiones

La incidencia delictiva en México es un fenómeno que aqueja a toda la sociedad y, por lo tanto, debería ser considerada prioridad en la agenda de los tomadores de decisiones. Este artículo contribuye a la literatura de la criminalidad e introduce herramientas de *Machine Learning*, para generar pronósticos que sirven de insumo en la elaboración de mejores políticas de seguridad. Como se vio a lo largo de la investigación, existe poca literatura que trate de corregir el problema de la discrepancia de las cifras de incidencia delictiva con la realidad.

Como principal hallazgo, demostré que existe una gran diferencia entre lo que indican las cifras de incidencia delictiva del Secretariado y las cifras de la ENVIPE. Esto se cumple para los delitos agregados, y en menor medida para cada tipo de delito. El sesgo es provocado por el problema de la cifra negra, que se utilizó como motivación para obtener una serie de delitos que se aproxima mejor a la realidad.

Adoptar herramientas de *Machine Learning*, para modelar fenómenos, amplía el rango de soluciones disponibles. En este caso se logró generar una serie consistente, que ayuda a reducir el rezago temporal de la ENVIPE, y que puede ser utilizada en otros contextos. Para ello, revisé diversas especificaciones como Redes Neuronales, *Random Forest* y *XGboost*. Finalmente, escogí el mejor modelo en función de su desempeño dentro y fuera de la muestra, cuidando que la probabilidad de subestimar los delitos fuera mínima.

A raíz de lo anterior, encontré que el mejor modelo para aproximar la incidencia delictiva en México es *Random Forest*, ya que tiene una probabilidad muy baja de predecir incorrectamente el cambio de la serie de delitos de la ENVIPE. Las otras dos especificaciones sirven para dar robustez de los resultados, pues van en el mismo sentido que la literatura de los determinantes de la criminalidad, en donde variables económicas y laborales ayudan a explicar el movimiento de los delitos en México.

Otro hallazgo importante de este artículo es la baja relación que existe entre las cifras de incidencia delictiva del Secretariado y el gasto programado en seguridad. Esto es preocupante ya que refleja la baja efectividad de las políticas públicas para atender el problema. La causalidad apunta a que el gasto en impartición de justicia sí tiene un efecto marginal sobre la incidencia delictiva. El gasto en seguridad nacional parece tener una relación bidireccional con la incidencia delictiva. El problema es que el gasto en orden público, que debería ser la variable más importante porque contempla en gasto en la policía, no parece causar a la incidencia delictiva. Lo anterior sugiere que aumentos en el gasto en orden no generan disminuciones en la incidencia. Esto se explica principalmente porque las cifras de delitos del Secretariado no capturan la criminalidad real, a diferencia de la ENVIPE que muestra una relación más estrecha con los movimientos del gasto en los tres rubros. Es por eso que se sugiere que los delitos de la ENVIPE son un mejor indicador que los del Secretariado.

Queda abierta la posibilidad de replicar esta metodología para aproximar el comportamiento de los delitos agregados por estado, lo que reforzaría las estrategias de seguridad locales, dotándolas de un mayor número

de indicadores sobre su desempeño. También, otra posibilidad es revisar el comportamiento de las cifras por tipo de delito, no se realizó en este trabajo, debido a que se necesita estimar un modelo por delito y es probable que sus determinantes sean distintos. Sin embargo, tener un modelo por delito ayudaría a incrementar la precisión con la que el pronóstico fuera de la muestra refleja el comportamiento real de la criminalidad en México.

Anexo

1A. Tipos de Delitos

Aborto	Despojo	Incesto	Otros delitos que atentan contra la vida y la integridad corporal
Abuso de confianza	Electorales	Incumplimiento de obligaciones de asistencia familiar	Rapto
Abuso sexual	Evasión de presos	Lesiones	Robo
Acoso sexual	Extorsión	Narcomenudeo	Secuestro
Allanamiento de morada	Falsedad	Otros delitos contra el patrimonio	Tráfico de menores
Amenazas	Falsificación	Otros delitos contra la familia	Trata de personas
Contra el medio ambiente	Feminicidio	Otros delitos contra la sociedad	Violación equiparada
Corrupción de menores	Fraude	Otros delitos del Fuero Común	Violación simple
Daño a la propiedad	Homicidio	Otros delitos que atentan contra la libertad personal	Violencia de género en todas sus modalidades distinta a la violencia familiar
Delitos cometidos por servidores públicos	Hostigamiento sexual	Otros delitos que atentan contra la libertad y la seguridad sexual	Violencia familiar

Note:

Elaboración propia con Información del SESNSP(2022).

2A. Prueba de Raíz Unitaria, 2010M01-2021M12

		Cifra	d(Cifra)	Desempleo	d(Desempleo)	IGAE	d(IGAE)	Inf	d(Inf)	Prod Manu	d(Prod Manu)	Rem Med	d(Rem Med)	TCCO	d(TCCO)	TIL	d(TIL)
ADF																	
Con C	t-Statistic	-3.50	-3.26	-2.27	-1.68	-1.75	-3.62	-2.53	-8.20	-3.38	-11.14	-0.01	-3.93	-0.70	-12.91	-1.81	-12.03
	Prob.	0.01	0.02	0.18	0.44	0.40	0.01	0.11	0.00	0.01	0.00	0.95	0.00	0.84	0.00	0.37	0.00
Con C&T	t-Statistic	-3.43	-3.35	-2.74	-2.29	3.45	-4.27	-2.49	-8.17	-5.53	-11.08	-2.53	-3.94	-2.63	-10.23	-7.20	-11.98
	Prob.	0.05	0.06	0.22	0.44	1.00	0.00	0.33	0.00	0.00	0.00	0.31	0.01	0.27	0.00	0.00	0.00
Sin C&T	t-Statistic	-0.22	-3.32	-0.85	-1.74	0.67	-3.55	-0.70	-8.23	-0.32	-11.18	2.44	-3.01	0.75	-12.84	-0.64	-12.05
	Prob.	0.61	0.00	0.35	0.08	0.86	0.00	0.41	0.00	0.57	0.00	1.00	0.00	0.87	0.00	0.44	0.00
PP																	
Con C	t-Statistic	-7.18	-33.59	-3.05	-17.18	-3.35	-28.86	-2.18	-7.94	-3.05	-19.90	-9.35	-40.16	-0.20	-13.49	-3.47	-38.46
	Prob.	0.00	0.00	0.03	0.00	0.01	0.00	0.22	0.00	0.03	0.00	0.00	0.00	0.93	0.00	0.01	0.00
Con C&T	t-Statistic	-7.17	-46.88	-3.95	-17.30	-4.65	-45.91	-2.15	-7.91	-5.44	-19.92	-32.11	-40.67	-2.48	-14.52	-7.18	-38.27
	Prob.	0.00	0.00	0.01	0.00	0.00	0.00	0.51	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00
Sin C&T	t-Statistic	-0.77	-35.05	-1.01	-17.21	1.43	-17.19	-0.67	-7.98	-0.01	-19.89	1.13	-38.69	1.47	-13.11	-2.58	-27.68
	Prob.	0.38	0.00	0.28	0.00	0.96	0.00	0.42	0.00	0.68	0.00	0.93	0.00	0.96	0.00	0.01	0.00

Ho: \exists raíz unitaria. C se refiere a la constante, y C&T a constante y tendencia. Para la correcta especificación de la prueba ADF se siguió el criterio de Schwartz (14 rezagos máx.) y para la prueba PP se utilizó el criterio de Bartlett Kernel (3 rezagos máx.). Por el procedimiento de Dolado, Jenkinson y Sosvilla-Rivero (Enders, 2014), se concluye que el IGAE, la inflación, las remuneraciones medias (rem med), y TCCO son I(1) en niveles y estacionarios en sus primeras diferencias, mientras que todas las demás son estacionarias. Fuente: Elaboración propia.

3A. Estadísticas Descriptivas, 2010M01-2021M12

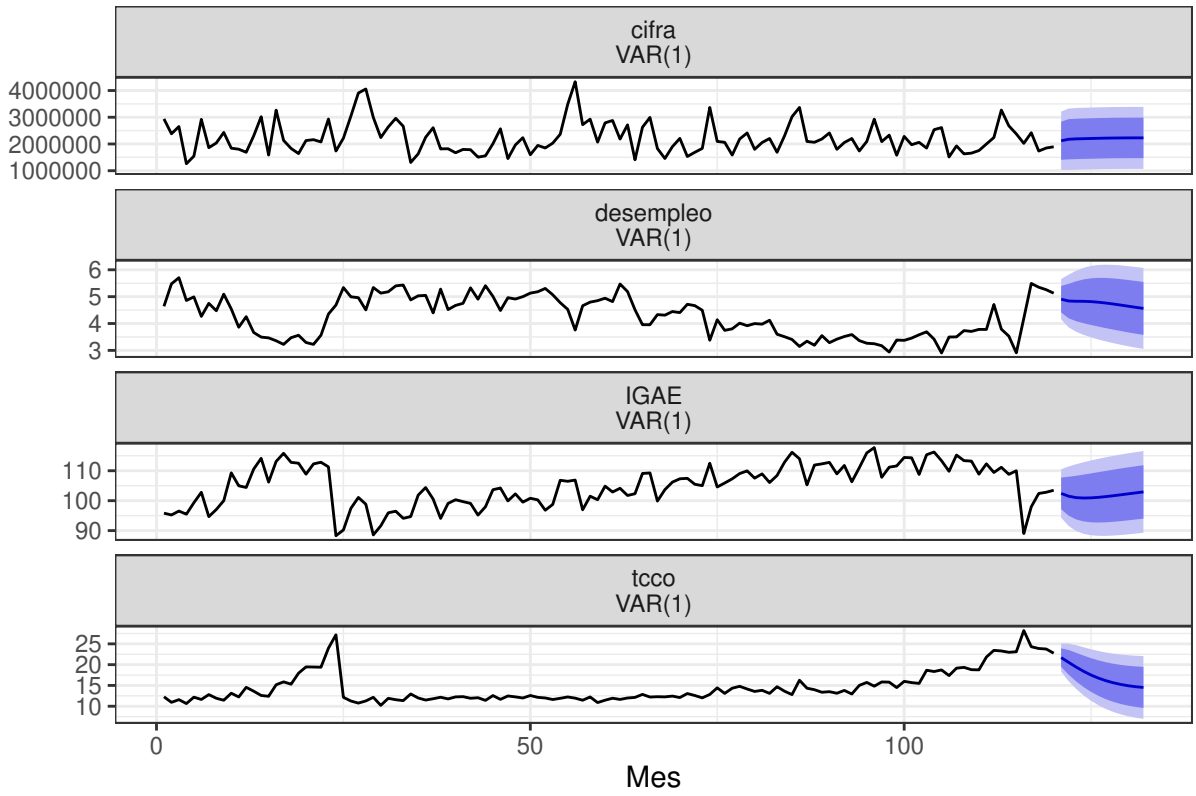
	Var	Obs	Media	Sd	Mediana	Trimmed	MAD	Min	Max	Rango	Asimetría	Kurtosis	Se
Cifra	1	120	2,218,668.63	584,000.67	2,090,878.50	2,160,556.33	482,926.56	1,260,625.00	4,329,149.00	3,068,524.00	1.06	1.15	53,311.72
IGAE	2	120	105.07	6.94	105.70	105.36	8.23	88.29	117.73	29.44	-0.33	-0.70	0.63
Rem Med	3	120	102.66	7.16	101.27	101.37	4.65	92.59	132.04	39.45	1.92	4.08	0.65
Inf	4	120	3.89	1.04	3.74	3.78	0.96	2.13	6.77	4.64	0.86	0.46	0.09
Desempleo	5	120	4.27	0.76	4.34	4.26	1.01	2.91	5.71	2.80	0.00	-1.35	0.07
TCCO	6	120	14.54	3.91	12.84	13.79	1.81	10.21	28.20	17.99	1.57	1.66	0.36
TIL	7	120	57.56	1.87	57.33	57.61	1.51	47.89	60.93	13.04	-1.12	4.88	0.17
Prod Manu	8	120	96.86	3.99	97.65	97.08	3.93	81.10	104.70	23.60	-0.84	1.37	0.36

Fuente: Elaboración propia.

4A. VAR(1) sobre Cifra Delictiva, 2010M01-2021M12

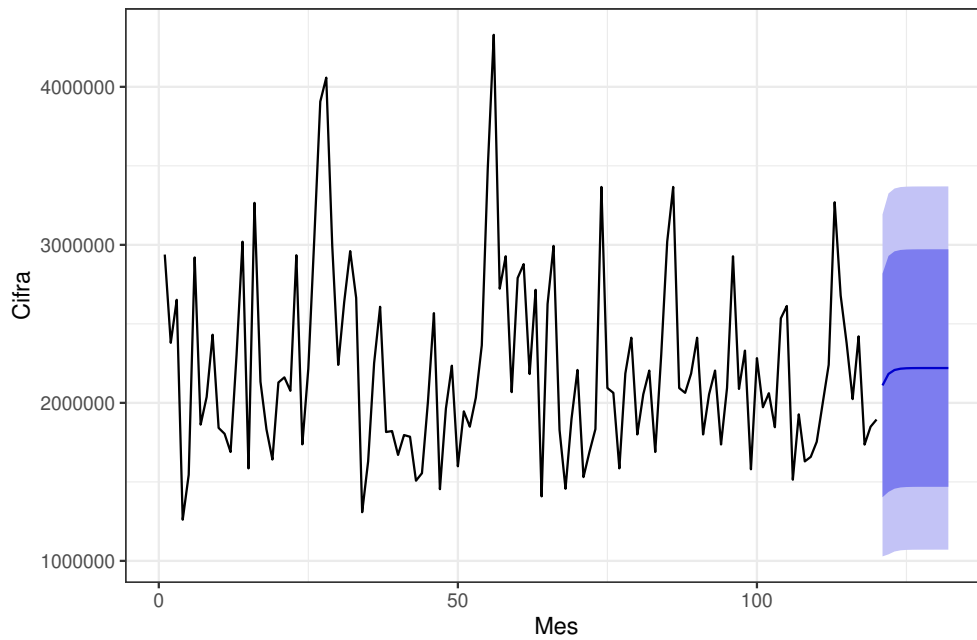
	<i>Variable:</i>
	<i>y</i>
Cifra.l1	0.333*** (0.088)
IGAE.l1	5,878.378 (12,544.790)
Desempleo.l1	85,865.380 (117,905.600)
TCCO.l1	-5,844.917 (14,206.630)
const	574,105.500 (1,767,826.000)
Obs	119
R ²	0.121
Adj R ²	0.091
SE	555,682.300 (df = 114)
Estadístico F	3.936*** (df = 4; 114)
<i>Nota:</i>	*p<0.1; **p<0.05; ***p<0.01, entre paréntesis el SE

5A.Pronóstico Fuera de la Muestra VAR(1), 2021M01-2021M12



Fuente: Elaboración propia.

6A.Pronóstico Fuera de la Muestra ARIMA(1,0,0), 2021M01-2021M12



Fuente: Elaboración propia.

Índice de figuras

1.	Cifra Negra, 2011-2021	9
2.	Cifra de delitos, 2015-2021	10
3.	Comparación de Tendencias de delitos, 2015-2021	11
4.	Gasto Público en Seguridad, 2015M01-2023M1	13
5.	Correlación Cruzada, 2015M01-2023M01	14
6.	Correlación Gasto y Delitos ENVIPE, 2011M01-2021M12	15
7.	Red neuronal	19
8.	Estructura de la Red	20
9.	Ajuste de la Red, 2011-2020	21
10.	Importancia de variables en la Red	22
11.	Árbol de Decisión	23

12. Error vs Árboles	25
13. Random Forest, 2011M01-2020M12	25
14. Importancia de variables en el RF	26
15. XGBoost, 2011M01-2020M12	27
16. Importancia de variables en el XGBoost	28
17. Pronóstico Dentro de la Muestra, 2011M01-2020M12	29
18. Pronóstico Fuera de la Muestra, 2021M01-2021M12	31

Índice de cuadros

1. Test de Causalidad de Granger, 2015M01-2022M12	16
2. Evaluación del Pronóstico Dentro de la Muestra	30
3. Diagnóstico fuera de la Muestra	31

Referencias

- Ahlburg, D. (1984). Forecast evaluation and improvement using theil's decomposition. *Journal of Forecasting*, 3, 345–351.
- Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 169–217. <https://doi.org/http://www.jstor.org/stable/1830482>
- Benavente, J., & Cortés, E. (2006). Delitos y sus denuncias. La cifra negra de la criminalidad en Chile y sus determinantes. *Series Documentos de Trabajo, No. 228*. <https://repositorio.uchile.cl/handle/2250/144557>
- Biderman, A. D., & Reiss, A. J. (1967). On exploring the "dark figure" of crime. <Http://Dx.doi.org/10.1177/00027162673740010374>, 1–15. <https://doi.org/10.1177/000271626737400102>
- Bliemel, F. (1973). Theil's forecasting accuracy coefficient: A clarification. *Journal of Marketing Research*, 10, 444–446.
- Brand, S., & Price, R. (2000). The economic and social costs of crime. *MPRA Paper No. 74968*. https://mpra.ub.uni-muenchen.de/74968/1/MPRA_paper_74968.pdf
- Breiman, L. (2001). *Random forests*. 45, 5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>
- Calnitsky, D., & Gonalons, P. (2021). The impact of an experimental guaranteed income on crime and violence. *Social Problems*, 68, 778–798. <https://doi.org/10.1093/SOCPRO/SPAA001>
- Carreto, C., & Ramírez-Álvarez, A. A. (2022). Why do Mexicans report so little crime? The determinants of crime reporting in Mexico. *Temas de Economía*, 5, 42–68. <https://doi.org/10.16/CSS/JQUERY.DATATABLES.MIN.CSS>
- Chen, P., Yuan, H., & Shu, X. (2008). Forecasting crime using the ARIMA model. *International Conference on Fuzzy Systems and Knowledge Discovery*, 5, 627–630.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939785>
- DOF. (2020). *ANEXO 1 DEL ACUERDO 05/XLVI/2005/XLVI/20*. Secretaría de Seguridad y Protección Ciudadana. https://www.dof.gob.mx/2021/SSPC/SEGURIDADyPC_260121.pdf
- Enamorado, T., López-Calva, L. F., & Rodríguez-Castelán, C. (2014). Crime and growth convergence: Evidence from Mexico. *Economics Letters*, 125, 9–13. <https://doi.org/10.1016/J.ECONLET.2014.07.033>
- Enders, W. (2014). *Applied econometric time series* (4ta edición). John Wiley.
- Fajnzylber, P., Lederman, D., & Loayza, N. (1998). *Determinants of crime rates in Latin America and the world*. The World Bank. <https://doi.org/10.1596/0-8213-4240-1>
- Fleisher, B. (1966). The effect of income on delinquency. *The American Economic Review*, 56, 118–137. <https://www.jstor.org/stable/1821199>
- Freedman, M., & Owens, E. G. (2011). Low-income housing development and crime. *Journal of Urban Economics*, 70, 115–131. <https://doi.org/10.1016/J.JUE.2011.04.001>
- Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal*

- of Forecasting*, 19, 579–594.
- Goudriaan, H., & Nieuwebeerta, P. (2007). Contextual determinants of juveniles' willingness to report crimes: A vignette experiment. *Journal of Experimental Criminology*, 3, 89–111. <https://doi.org/10.1007/S11292-007-9030-4/TABLES/5>
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438. <https://doi.org/10.2307/1912791>
- Granger, C., & Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, 5, 35–47.
- Greenberg, M. S., & Beach, S. R. (2004). Property crime victims' decision to notify the police: Social, cognitive, and affective determinants. *Law and Human Behavior*, 28, 177–186. <https://doi.org/10.1023/B:LAHU.0000022321.19983.EA>
- Greenberg, M. S., Wilson, C. E., Ruback, R. B., & Mills, M. K. (1979). Social and emotional determinants of victim crime reporting. *Social Psychology Quarterly*, 42, 364–372. <https://doi.org/10.2307/3033806>
- Hara, K., Saito, D., & Shouno, H. (2015). Analysis of function of rectified linear unit used in deep learning. *Proceedings of the International Joint Conference on Neural Networks, 2015-September*. <https://doi.org/10.1109/IJCNN.2015.7280578>
- Hodrick, R., & Prescott, E. (1997). Postwar u.s. Business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29, 16. <https://doi.org/10.2307/2953682>
- INEGI. (2022a). *Principales resultados de la ENVE 2022*. Instituto Nacional de Estadística y Geografía. https://www.inegi.org.mx/contenidos/programas/enve/2022/doc/enve_2022_presentacion_ejecutiva.pdf
- INEGI. (2018). *El gasto público frente a los resultados en seguridad pública y justicia*. https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825107260.pdf
- INEGI. (2022b). Encuesta nacional de victimización y percepción sobre seguridad pública (ENVIPE). In 2022. <https://www.inegi.org.mx/programas/envipe/2021/>
- INEGI. (2022c). *Glosario de la encuesta nacional de victimización y percepción sobre seguridad pública 2022*. <https://www.inegi.org.mx/app/glosario/default.html?p=envipe2017>
- Jaitman, L., & Anauati, V. (2019). The dark figure of crime in latin america and the caribbean. *Journal of Economics, Race, and Policy* 2019 3:1, 3, 76–95. <https://doi.org/10.1007/S41996-019-00042-0>
- Jaitman, L., Soares, R., Olavarría-Gambi, M., & Compeán, R. G. (2015). The welfare costs of crime and violence in latin america and the caribbean. In L. Jaitman (Ed.), *The Welfare Costs of Crime and Violence in Latin America and the Caribbean*. Inter-American Development Bank. <https://publications.iadb.org/en/welfare-costs-crime-and-violence-latin-america-and-caribbean>
- Leo, L. R. A., Jurado, J. A., & Ramírez-Álvarez, A. A. (2022). Internal migration and drug violence in mexico. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4127795>

- Loría, E., & Salas, E. (2019). La relación entre robo y desempleo de varones jóvenes en México, 2005-2017. *Revista Mexicana de Economía y Finanzas Nueva Época REMEF*, 14, 433–446. <https://doi.org/10.21919/remef.v14i3.353>
- Lozano-Cortés, René, Maribel, Cabrera-Castellanos, & F, L. (2016). The socioeconomic determinants of crime: The case of Mexico. *Munich Personal RePEc Archive, MPRA Paper No. 68922*.
- Makridakis, S., Wheelwright, S., & Hyndman, R. (1998). *Forecasting: Methods and applications* (3rd ed.). Wiley.
- Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and prophet. *Computers & Geosciences*, 164.
- OECD. (2020). *How's life? 2020: measuring well-being*. OECD. <https://doi.org/10.1787/9870C393-EN>
- Patterson, E. (1991). Poverty, income inequality, and community crime rates. *Criminology*, 29, 755–776. <https://doi.org/10.1111/J.1745-9125.1991.TB01087.X>
- Penney, T. L. (2014). Dark figure of crime (problems of estimation). *The Encyclopedia of Criminology and Criminal Justice*, 1–6. <https://doi.org/10.1002/9781118517383.WBECCJ248>
- Pindyck, R., & Rubinfeld, D. (1997). *Econometric models and economic forecasts* (4th ed.). McGraw-Hill.
- Piñeyro, J. (2012). El ¿saldo? De la guerra de calderón contra el narcotráfico. *El Cotidiano*, 5–14.
- Quiroz, J., Castillo, R., Ocegueda, J., & Varela, R. (2015). Delincuencia y actividad económica en México. *Norteamérica*, 10, 187–209. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-35502015000200187
- Ramírez, L. (2014). Crimen y economía: Análisis de la tasa de homicidio en México a partir de variables económicas (2000, 2005, 2010). *Estudios Sociológicos*, 32, 505–540. <https://www.jstor.org/stable/24368226>
- Ray, S. (2019). A quick review of machine learning algorithms. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 35–39.
- Robles, G., Calderon, G., & Magaloni, B. (2013). Las consecuencias económicas de la violencia del narcotráfico en México. *Documento de Trabajo Del BID*.
- Romo, R. S. (2018). Una estimación del costo de la inseguridad y la delincuencia en México: Análisis comparativo a nivel de las entidades federativas. *Gestión y Política Pública*, 27, 111–147. <https://www.scielo.org.mx/pdf/gpp/v27n1/1405-1079-gpp-27-01-00111.pdf>
- SESNP. (2023). *Incidencia delictiva del fuero común, nueva metodología*. <https://www.gob.mx/sesnp/acciones-y-programas/incidencia-delictiva-del-fuero-comun-nueva-metodologia?state=published>
- SHCP. (2023a). *Estadísticas oportunas de finanzas públicas*. <http://presto.hacienda.gob.mx/EstoporLayout/>
- SHCP. (2023b). *Presupuesto de egresos de la federación*. <https://www.pef.hacienda.gob.mx/>
- Shoemith, G. (2013). Space-time autoregressive models and forecasting national, regional and state crime rates. *International Journal of Forecasting*, 29, 191–201.
- Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A comparison of ARIMA and LSTM in forecasting time series. *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 17, 1394–

1401.

- Skogan, W. G. (1977). Dimensions of the dark figure of unreported crime. *Http://Dx.doi.org/10.1177/001112877702300104*, 23, 41–50. <https://doi.org/10.1177/001112877702300104>
- Soares, R. R. (2004). Development, crime and punishment: Accounting for the international differences in crime rates. *Journal of Development Economics*, 73, 155–184. <https://doi.org/10.1016/j.jdeveco.2002.12.001>
- SSPC-SESNSP. (2020). *Sistema multi-fuente para la estimación de la incidencia delictiva orientada a la inteligencia policial*.
- Theil, H. (1958). Economic forecasts and policy. *Amsterdam: North Holland*.
- Theil, H. (1966). Applied economic forecasting. *Chicago: Rand McNally*.
- Tibshirani, R., Hastie, T., & Jerome, Friedman. (2009). *The elements of statistical learning* (2nd ed.). Springer New York, NY. <https://doi.org/https://doi.org/10.1007/978-0-387-84858-7>
- Tibshirani, R., James, G., Witten, D., & Hastie, T. (2021). *An introduction to statistical learning with applications in r second edition*. https://hastie.su.domains/ISLR2/ISLRv2_website.pdf
- Verdugo-Yepes, C., Pedroni, P., & Hu, X. (2015). Crime and the economy in mexican states: Heterogeneous panel estimates (1993-2012). *IMF Working Paper*. <https://www.imf.org/external/pubs/ft/wp/2015/wp15121.pdf>
- Villoro, R., & Teruel, G. (2004). The social costs of crime in mexico city and suburban areas. *Estudios Económicos*, 19, 3–44. <https://doi.org/https://www.jstor.org/stable/40311480>