

UNA NUEVA APROXIMACIÓN ESTADÍSTICA AL PROBLEMA DE HOMICIDIOS EN EL SALVADOR

A new statistical approach to the problem of homicides in El Salvador

Alberto VALIENTE THORESEN

Universidad de Oslo, Noruega

✉ alberto.valiente.thoresen@gmail.com

Fecha de recepción: 15 de diciembre de 2017

Fecha de aceptación y versión final: 11 de diciembre de 2018

RESUMEN: A pesar de la elevada atención prestada a la tasa de homicidios en El Salvador, no se han fundamentado cuantitativamente las correlaciones empíricas de esta variable con otras variables explicativas. Este estudio contribuye a atender esta carencia, calculando regresiones cuantílicas de la tasa de homicidios por 100.000 habitantes por municipio en El Salvador en el percentil 3, 50 y 98 respectivamente, con datos poblacionales de 2007. Los modelos resultantes demuestran que la tasa de homicidios está significativamente correlacionada: positivamente con la cantidad de pandillas por municipio, en el percentil 3 y 50; positivamente con el índice de urbanidad y negativamente con el acceso a agua potable, en el percentil 50; así como positivamente con la cantidad de pandilleros por 100.000 habitantes, en el percentil 98. Estos resultados pueden usarse para orientar las políticas, pero solo explican parte del problema. Urgen mejores estadísticas para subsanar algunas limitantes en futuros estudios.

Palabras clave: tasa de homicidios; El Salvador; regresión cuantílica; regresión simbólica; programación genética.

ABSTRACT: Despite the high attention paid to the homicide rate in El Salvador, the empirical correlations of this variable with other explanatory variables have not been quantitatively substantiated. This study contributes to address this knowledge gap, calculating quantile regressions of the homicide rate per 100,000 inhabitants for each municipality in El Salvador in the 3rd, 50th and 98th percentile respectively, with population data for 2007. The resulting models show that the homicide rate is significantly correlated with the following variables: positively with the number of gangs per municipality, in the 3th and 50th percentiles; positively with the index of urbanity and negatively with access to drinking water, in the 50th percentile; as well

as positively with the number of gang members per 100,000 inhabitants, in the 98th percentile. These results can be used to guide policy, but they only explain part of the problem. There is an urgent need for better statistics to correct some limitations in future studies.

Key words: homicide rate; El Salvador; quantile regression; symbolic regression; genetic programming.

I. INTRODUCCIÓN¹

El objetivo principal de este estudio es responder qué variables explicativas, de acuerdo con teorías generalmente aceptadas, están correlacionadas positiva o negativamente de una manera significativa con la tasa anual de homicidios en El Salvador. Este estudio no propone nuevas conjeturas o interpretaciones teóricas sobre el problema. Simplemente, se limita a probar teorías generalmente aceptadas, con metodologías novedosas para este campo de estudio. El empleo de dichas metodologías impone ciertos requerimientos sobre la calidad de la información estadística en la que se basan. Este estudio no sacrifica estos requerimientos cuando la información estadística necesaria no está disponible. Inevitablemente, esto conlleva a que el estudio debe posponer discusiones e interpretaciones sobre variables que ciertas teorías señalan, pero sobre las cuales todavía no tenemos suficiente información. Esto no implica que estos factores propuestos sean irrelevantes. Al contrario, porque son factores tan importantes de acuerdo con las teorías, deben atenderse con la mayor rigurosidad metodológica.

Para hacer el modelo explicativo se ha escogido el año 2007, ya que es el año más reciente para el cual hay información poblacional disponible a nivel municipal, porque fue durante ese año cuando se efectuó el VI Censo de Población y el V Censo de Vivienda en El Salvador, a cargo de la Dirección General de Estadística y Censos (DIGESTYC). Esta información ha sido complementada, directamente, con información proporcionada por la Policía Nacional Civil (PNC), a través de una petición de acceso a la información pública, registrada como caso C-106-2016 en la Unidad de Acceso la Información Pública de la PNC.

Las variables escogidas del conjunto de información disponible provienen de las teorías generalmente aceptadas para explicar la tasa de homicidios, ya sea en El Salvador o en estudios efectuados en otros países. Este proceso se explica en el apartado teórico.

Para cumplir el objetivo, se han calculado regresiones cuantílicas, las cuales estiman la mediana condicional, el percentil 3 y 98 condicionales de la tasa anual de homicidios en los diversos municipios de El Salvador. Las regresiones cuantílicas de los percentiles 3 y 98 se emplean como curvas de referencia para la tasa anual de homicidios municipal. Estas tres regresiones proporcionan información importante sobre cómo la tasa de homicidios se relaciona con otras variables contenidas en el VI Censo de Población y el V Censo de Vivienda y los datos proporcionados por la PNC.

Para escoger entre los modelos que estiman la mediana condicional obtenidos, y priorizar variables significativas, se ha empleado un método de estimación de regresión

1. Los autores agradecen los comentarios y las sugerencias de dos evaluadores anónimos de *América Latina Hoy, Revista de Ciencias Sociales*, a la primera versión de este artículo.

simbólica, a través de programación genética. El procedimiento técnico se explica en detalle en el apartado metodológico.

La respuesta efectiva a la pregunta de investigación de este estudio tiene implicaciones para la predicción y el diagnóstico de la tasa de homicidios en los diversos municipios de El Salvador, lo cual, a su vez, puede influir en decisiones sobre políticas. Por ello, el estudio concluye con un apartado en donde la interpretación de los resultados se relaciona con esas posibles implicaciones.

1.1. Relevancia

Durante el período 1995-2014, la tasa anual de homicidios intencionales por 100.000 habitantes² en El Salvador tuvo un promedio de 66,44 (Banco Mundial, IBRD-IDA 2016). Durante este período, dicha tasa no bajó de 37,2. Esto quiere decir que dicha tasa se ha mantenido en niveles endémicos relativamente elevados, los cuales para otros países se considerarían epidémicos, tomando en cuenta que la tasa anual de homicidios intencionales por 100.000 habitantes promedio mundial para el período 1995-2014 fue de 8,45 homicidios, de acuerdo con los datos disponibles para el período (Banco Mundial, IBRD-IDA 2016). Esto implica que el promedio de la tasa anual de homicidios en El Salvador equivalió al 785,8 % del promedio mundial anual para el período.

Aunque no es el único factor para medir los niveles de violencia directa en El Salvador, este indicador ha atraído mucha de la atención de los medios de comunicación, especialistas y autoridades. Se han elaborado algunos estudios e informes que pretenden arrojar luz sobre esta situación (Cruz y González 1997; Cruz, Trigueros Argüello y González 2000; Cruz, González, Romano y Sisti 2000; Acevedo 2008; Carcach 2008; Córdova 2013; Bergmann 2015). Sin embargo, las hipótesis sobre la relación de la tasa de homicidios con otras variables para el país en su totalidad, ya sean de variables sociales, económicas, demográficas, políticas o de otro tipo, no han sido conclusivamente fundamentadas de manera cuantitativa.

Es por esto que muchas de las teorías que se utilizan diariamente para justificar las políticas y el accionar de las autoridades que tratan el problema de homicidio en El Salvador están plagadas de lagunas de conocimiento. De manera que no es posible saber a ciencia cierta si las políticas y las operaciones de las autoridades responden a la verdadera funcionalidad del problema. Este estudio pretende ser un aporte para contribuir a suplir tales lagunas, dar algún fundamento cuantitativo empírico a las teorías más conocidas y sugerir nuevas interpretaciones teóricas del problema, dados los datos disponibles. Son grandes ambiciones que el estudio alcanza en alguna medida, a pesar de los muchos retos, principalmente debido a la falta de existencia y disponibilidad de datos confiables.

2. De aquí en adelante, se emplearán intercambiamente los términos «tasa de homicidios», «tasa de homicidios intencionales», «tasa de homicidios intencionales por 100.000 habitantes» y «tasa de homicidios por 100.000 habitantes».

II. TEORÍAS Y VARIABLES SELECCIONADAS

II.1. *Un posible punto de partida: investigaciones sobre la tasa de homicidios en Brasil*

Es conveniente iniciar este apartado citando una revisión sistemática de estudios sobre crimen violento en contextos similares a los de El Salvador. En este caso podemos usar el trabajo de Murray *et al.* (2013), sobre crimen y violencia en Brasil, con un enfoque en tendencias, prevalencia y factores de riesgo. Si bien sería una falacia de composición suponer que lo que se ha probado en Brasil es directamente aplicable en El Salvador, esta revisión sistemática es un punto de partida muy útil para formular las hipótesis a falsar. Especialmente cuando es posible identificar paralelismos entre las variables y teorías presentadas en esta revisión sistemática y las variables y teorías que generalmente son empleadas en la práctica para explicar la tasa de homicidios en El Salvador.

Murray *et al.* (2013) identifican las siguientes posibles explicaciones de las relativamente altas tasas de homicidio en Brasil³.

I) La teoría presentada por Imbusch *et al.* (2011), quienes consideran «el contexto histórico de América Latina, incluyendo las culturas precolombinas, colonización por los portugueses y españoles, esclavitud, dictaduras militares y transiciones a la democracia en contextos de inestabilidad económica». De acuerdo con estos autores:

El incremento en la tasa de homicidio en Brasil que siguió la democratización de la década de los ochenta se debió a una mayor desigualdad, urbanización desorganizada, disponibilidad de armas de fuego, instituciones sociales débiles, tráfico de drogas, características culturales y una democracia que garantizaba derechos políticos, pero no sociales (Imbusch *et al.* 2011, en Murray *et al.* 2013: 480).

II) Por su parte, Briceno-León *et al.* (2008) sugieren que, además de la desigualdad, las tasas de homicidio se incrementaron debido a «un mayor desempleo juvenil, la mayor importancia de las aspiraciones consumistas, así como la pérdida de mecanismos tradicionales de control social, especialmente dentro de la familia y las instituciones religiosas» (Briceno-Leon *et al.* 2008, en Murray *et al.* 2013: 480). A esto se suman «los conflictos de la tierra, luchas entre pandillas, violencia policial, y un crecimiento desmesurado de la población» (Reichenheim *et al.* 2011, en Murray *et al.* 2013: 480).

III) En esta misma revisión sistemática realizada por Murray *et al.* (2013) se citan los siete indicadores nacionales de cambios socioeconómicos relacionados con tendencias de homicidio entre 1980 y 2010 (Cerqueira 2010): a) ingreso nacional; b) niveles de desigualdad económica, medidos a través del coeficiente de Gini; c) proporción de hombres jóvenes en la población (entre 15 y 24 años de edad); d) disponibilidad de armas

3. Hay que resaltar, sin embargo, que la tasa promedio anual de homicidios por 100.000 habitantes en Brasil ha sido tan solo de 22,8 en el período 2010-2014, comparada con la tasa de 57,28 en El Salvador para el mismo período (BANCO MUNDIAL, IBRD-IDA 2016). Pero hay que recordar que Brasil representa un área geográfica más grande que El Salvador y que puede haber territorios equivalentes a El Salvador con tasas de homicidio comparables.

de fuego; e) uso de drogas ilegales; f) presencia policial y g) tasas de encarcelamiento (Murray *et al.* 2013: 480).

II.2. Investigaciones sobre la tasa de homicidios en el resto del mundo

Estas variables se repiten en investigaciones e informes realizados sobre el resto del mundo. Un ejemplo es el «Estudio Global sobre Homicidio» cuyos resultados se publicaron por la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC 2013).

De acuerdo con este estudio, el continente americano ha reportado las tasas promedio anuales de homicidio por 100.000 habitantes más elevadas en el período 1955-2012, comparado con el resto de continentes. En América Central y América del Sur, los hombres jóvenes entre 15 y 29 años de edad muestran una tasa de homicidio anual promedio cuatro veces más grande que el promedio global. Los factores catalizadores de los homicidios que se mencionan son la disponibilidad de armas, particularmente las armas de fuego, y el uso de sustancias psicoactivas, como las drogas ilegales y el alcohol. Al mismo tiempo, las tasas de homicidio se relacionan con recientes conflictos sociopolíticos, niveles de desigualdad, pobreza, niveles de urbanización, otros tipos de criminalidad y el funcionamiento de las instituciones (UNODC 2013).

Hay otros ejemplos de la literatura científica reciente, los cuales también han tratado sobre algunas de estas variables, a pesar de que los estudios individuales no incluyan todas estas variables simultáneamente. Por ejemplo, Ouimet (2012) ha estudiado el efecto de la desigualdad económica, la desigualdad del ingreso y el exceso de la mortalidad infantil en la tasa de homicidio en 165 países en 2010. Este estudio, basado en análisis de regresión, ha demostrado que los niveles de renta nacional bruta; la desigualdad, medida por el coeficiente de Gini y la pobreza, medida por mortalidad infantil, son predictores significativos de la tasa de homicidios a nivel mundial. Este es un hallazgo que corroboró un estudio anterior enfocado en las «naciones desarrolladas desde 1975 a 1995» (Jacobs y Richardson 2008). Por su parte, Morris y Graycar (2011) han mostrado cómo los factores sociopolíticos, entre ellos la polarización política, afectan los niveles de tasa de homicidio en zonas urbanas de Jamaica. Hallazgos que han sido corroborados en un estudio posterior (Morris y Maguire 2016). A esto puede sumarse un estudio realizado por David Lester (1991), quien fundamentó la relación entre las tasas de homicidios con la disponibilidad de armas de fuego en países europeos. Con respecto a las variables más complejas sobre control social, sugeridas por Briceno-León *et al.* (2011), Dicristina (2004) ha aportado a la discusión teórica sobre las dificultades inherentes a la realización de estudios empíricos que relacionen algún «desarrollo social» con los niveles de homicidio.

II.3. ¿Qué dicen las investigaciones sobre El Salvador?

Muchas de las variables mencionadas en los estudios sobre la tasa de homicidio a nivel mundial están presentes en los estudios sobre la tasa de homicidios en El Salvador. Sin embargo, no hay investigaciones terminadas que establezcan cuantitativamente

correlaciones poblacionales empíricas entre la tasa anual de homicidios y estas variables para el total de la población salvadoreña.

Por ejemplo, una de las primeras investigaciones de este tipo en El Salvador fue la realizada por Cruz, Trigueros Argüello y González (2000). Si bien este estudio incluye capítulos sobre los «factores asociados a la victimización por crimen violento» y «factores asociados a la violencia homicida y a la reincidencia», los resultados de estos apartados se basan en «mini-encuestas» en el área metropolitana de San Salvador y «entrevistas» con «reclusos del sistema correccional salvadoreño». Si bien este fue un aporte importante, en un período en el cual la disponibilidad de datos estadísticos confiables era todavía más deficiente que en nuestros días, estas encuestas y entrevistas no proporcionan información suficiente y confiable para inferir sobre las variables mencionadas para la totalidad del país. Este estudio incorpora, además, otra variable importante para el caso de El Salvador, la cual no ha sido sugerida por los estudios mencionados sobre otros países: la migración y sus efectos sobre la tasa de homicidio en El Salvador (Cruz, Trigueros Argüello y González 2000: 94).

Recientemente se han publicado otros artículos académicos que prometen el esclarecimiento de «factores clave» o «causas», pero tampoco fundamentan estas hipótesis cuantitativamente de una manera concluyente (Salgado 2012; Sancho 2013). Esta laguna de conocimiento, sobre cómo diversas variables están relacionadas cuantitativamente con la tasa anual de homicidio, implica que el tema se maneja por muchos de una manera editorial, a pesar de no poder cuantificar estas relaciones. Un ejemplo es el editorial de la Universidad Centroamericana «José Simeón Cañas» del 3 de junio de 2015, titulado «Explicaciones de la violencia». En este editorial se repasa una lista de explicaciones comúnmente aceptadas sobre las elevadas tasas de homicidio, entre ellas: la desigualdad económica, la pobreza, las estrategias para combatir el crimen, el narcotráfico, los intereses económicos, el negocio de las armas, la debilidad institucional y la cultura machista. El mismo artículo apunta a la necesidad de «comenzar una reflexión y animar, en otros foros, a profundizar el problema». Esta es una de las pretensiones de este estudio.

II.4. Variables incluidas en el estudio

Con estas consideraciones, es posible presentar una lista de las variables a tomar en cuenta en el estudio. La Tabla I especifica los indicadores de estas variables en relación con las diversas categorías teóricas sugeridas. La Tabla I también especifica qué fuentes estadísticas suplieron la necesidad de información empírica sobre la variable respectiva. Como indican las notas a esta tabla, no todos los datos se encuentran disponibles por municipio. En ciertos casos, en los cuales las variables se consideren importantes y hay argumentos teóricos válidos para hacerlo, se han usado otros datos como aproximaciones a los datos por municipio, cuando estos no están disponibles. Tales aproximaciones no implican retos técnicos, ya que los procedimientos y los cálculos son los mismos que cuando se usan datos exactos, si los hay. Pero debe tenerse cuidado de no olvidar estas aproximaciones a la hora de interpretar los resultados, para no saltar a falsas conclusiones. Ciertas variables relevantes se han obviado, porque no pueden cuantificarse fácilmente, por ejemplo, la polarización política. Tales hipótesis deben tratarse con otros métodos.

TABLA I
 DESCRIPCIÓN DE LAS VARIABLES

VAR.	DESCRIPCIÓN. TODAS LAS VARIABLES SON POR MUNICIPIO PARA 2007	CLASIFICACIÓN	FUENTE
	Incautaciones de narcóticos ilegales por 100.000 habitantes	Criminológica	No disponible
e	Delitos de eficacia (menos homicidios) por 100.000 habitantes	Criminológica	PNC
	Consumo de x-droga por habitante en el municipio	Criminológica	No disponible
p	Habitantes por kilómetro cuadrado	Demográfica	DIGESTYC
c	Tasa de fecundidad	Demográfica	DIGESTYC
i	Índice de mortalidad infantil	Demográfica	DIGESTYC
g	Índice de masculinidad	Demográfica	DIGESTYC
d	Porcentaje de la población en relación de dependencia	Demográfica	DIGESTYC
q	Porcentaje de población urbana	Demográfica	DIGESTYC
t	Extranjeros por 100.000 habitantes	Demográfica	DIGESTYC
j	Población indígena por 100.000 habitantes	Demográfica	DIGESTYC
z	Porcentaje de la población de la tercera edad	Demográfica	DIGESTYC
	Coefficiente de Gini	Socioeconómica	No disponible
l	Habitantes receptores de remesas por 100.000 habitantes*	Socioeconómica	DIGESTYC.
w	Monto promedio de remesas recibido mensualmente por habitante (USD)*	Socioeconómica	DIGESTYC
s	Porcentaje de asistencia escolar	Socioeconómica	DIGESTYC
v	Porcentaje de la población con acceso a agua potable	Socioeconómica	DIGESTYC
u	Porcentaje de la población sin servicios sanitarios	Socioeconómica	DIGESTYC
k	Porcentaje de la población con servicio sanitario conectado a alcantarillado	Socioeconómica	DIGESTYC
f	Porcentaje de viviendas con piso de tierra	Socioeconómica	DIGESTYC
y	Tasa de analfabetismo	Socioeconómica	DIGESTYC
o	Porcentaje de acceso a la electricidad	Socioeconómica	DIGESTYC
	Deportados por 100.000 habitante	Socioeconómica	No disponible
h	Homicidios por 100.000 habitantes (variable dependiente)	Sociopolítica	PNC
	Armas de fuego por 100.000 habitantes	Sociopolítica	No disponible
m	Miembros de pandilla por 100.000 habitantes**	Sociopolítica	PNC
a	Presencia policial por 100.000 habitantes	Sociopolítica	PNC
x	Número de pandillas**	Sociopolítica	PNC

* Los datos están disponibles únicamente por departamento. Por considerarse una variable importante, el promedio por departamento se ha usado para la totalidad de municipios en el departamento.

** Los datos disponibles son tan solo estimaciones de la PNC y para 2015, pero se usan como aproximación a los datos de 2007. Este argumento es válido solo si aceptamos la territorialidad de las pandillas salvadoreñas descritas en la literatura sobre el tema (Portillo 2012).

Fuente: Elaboración propia.

II.5. Hipótesis

La hipótesis nula a falsar, para todas las variables presentadas en la Tabla I, es que las variables explicativas no tienen una correlación significativa con la tasa de homicidios anual municipal.

El presente estudio asume que la gran mayoría de los efectos de las variables independientes en un municipio, si la hay, podrá observarse en la variable dependiente del mismo municipio en cuestión. Este es un supuesto exiguo, pero necesario, por requerimientos operativos y prácticos. Una de las limitaciones que esto implica se menciona brevemente en el apartado de discusión.

III. MÉTODO

Las técnicas estadísticas empleadas en este estudio para estimar modelos explicativos pueden agruparse en dos grandes categorías: regresión cuantílica y regresión simbólica a través de programación genética.

La primera técnica se ha usado para calcular los modelos sobre cuantiles condicionales de la tasa de homicidios por municipios para 2007, dadas las variables explicativas incorporadas en el análisis. La segunda técnica se ha usado para encontrar el modelo de mediana condicional más Pareto-óptimo, es decir, el que consigue alcanzar, simultáneamente, los objetivos de minimizar el error y proporcionar un modelo sencillo y fácil de interpretar.

III.1. Regresión cuantílica

En este estudio se ha optado por hacer uso de regresiones cuantílicas porque, dadas las condiciones del estudio, se considera un análisis más:

Exhaustivo: porque proporciona información sobre diversos puntos de la distribución de frecuencia de la variable dependiente, de manera que es posible hacer inferencias en diferentes niveles, no solo en torno a la media.

Robusto: la regresión cuantílica minimiza el efecto de valores atípicos en el conjunto de datos.

Flexible: es un método válido para hacer inferencias estadísticas sobre la población con base en la regresión estimada, incluso cuando los errores condicionales no están distribuidos normalmente, o cuando se viola el supuesto de homocedasticidad.

III.1.1. Teoría básica de la regresión cuantílica

Koenker y Hallock (2001) proporcionan una concisa explicación de la regresión cuantílica, como un problema de optimización. Mientras el método de mínimos cuadrados se presenta como una solución aproximada a un sistema sobredeterminado, que se obtiene al resolver:

$$\min_{\beta \in R} \sum_{i=1}^n (y_i - \mu)^2$$

donde y es la variable dependiente, i cada observación individual y μ la media condicional estimada por el modelo en cuestión, dadas las variables independientes x con sus respectivos coeficientes β . Es decir que $\mu(x_i, \beta)$ será la media condicional a las variables independientes x y los coeficientes estimados. El método de mínimos cuadrados implica entonces obtener la función que cumpla:

$$\min_{\beta \in RP} \sum_{i=1}^n (y_i - \mu(x_i, \beta))^2$$

El método de la regresión cuantílica es el mismo, con la diferencia de que en lugar de estimar la media condicional $\mu(x_i, \beta)$, esto se hará en torno a ξ , el cual puede expresarse como el escalar $\xi(x_i, \beta)$, y no en torno al cuadrado de esta diferencia, sino la función inclinada de valor absoluto⁴ ρ_τ , en donde τ proporciona el cuantil condicional en torno al cual se desea estimar la función. En el caso de la mediana τ será igual a 0,5.

De manera que el problema de minimización para las regresiones cuantílicas en general, puede formularse de la siguiente manera:

$$\min_{\beta \in RP} \sum_{i=1}^n \rho_\tau(y_i - \xi(x_i, \beta))$$

Para la regresión cuantílica en torno a la mediana condicional, de la siguiente manera:

$$\min_{\beta \in RP} \sum_{i=1}^n \rho_{0.5}(y_i - \xi(x_i, \beta))$$

El percentil 75:

$$\min_{\beta \in RP} \sum_{i=1}^n \rho_{0.75}(y_i - \xi(x_i, \beta))$$

El percentil 95:

$$\min_{\beta \in RP} \sum_{i=1}^n \rho_{0.95}(y_i - \xi(x_i, \beta))$$

Y así sucesivamente.

Este problema puede resolverse a través de métodos de programación lineal. Con esto, es posible estimar una regresión que estime coeficientes de los parámetros que minimicen el valor absoluto inclinado de la suma de los errores.

III.1.2. Las regresiones cuantílicas de este estudio

En el caso de este estudio, se han estimado regresiones cuantílicas para los cuantiles condicionales 3, 50 y 98 de la tasa de homicidios por 100.000 habitantes por municipio para 2007 en El Salvador, dadas las variables independientes incluidas en el análisis. Estos cálculos se han hecho usando el paquete de software estadístico StataSE, versión 14.1

4. También conocida como función de pérdida de *pinball*, por su forma, que asemeja la trayectoria de una pelota en el juego de *pinball*.

(StataCorp 2015). Inicialmente con el comando de «*qreg2*», el cual permite incluir el módulo de la prueba de heterocedasticidad de Machado-Santos Silva «*MSS*». Posteriormente, las regresiones se han calculado con el comando «*bsqreg*», el cual calcula regresiones cuantílicas que computan errores estándar robustos de los coeficientes en condiciones de heterocedasticidad, a través de la técnica de «*bootstrapping*»⁵ (Machado y Santos 2013). En este estudio se ha hecho uso de *bootstrapping* con 10.000 repeticiones.

La elección de los cuantiles 3, 50 y 98 puede parecer arbitraria, pero se debe a dos cuestiones. Por un lado, el estudio usa la mediana, por tratarse de un valor central en el rango de valores posibles que la tasa de homicidios asumió en 2007 en los diversos municipios de El Salvador. Esto se debe a convenciones y a la facilidad de comprender los resultados de la regresión relacionados con la mediana. Sin embargo, hay que recordar que esto no dice toda la historia de la tasa de homicidios en los diversos municipios de El Salvador. Por otro lado, la estimación de regresiones cuantílicas para los percentiles 3 y 98 sigue la lógica de los límites de referencia comúnmente empleados en la bioquímica clínica. De acuerdo con Marshall *et al.* (2014), los límites de referencia se escogen de manera que se crea un intervalo de referencia que comprende el 95% de la muestra de referencia⁶. Esto nos permitirá observar relaciones importantes en los extremos del intervalo de referencia de la tasa de homicidios en El Salvador para el año de 2007, el año más reciente para el que se tienen estadísticas poblacionales.

III.2. Regresión simbólica a través de programación genética

El método de regresión simbólica a través de programación genética que se ha empleado en este estudio es el descrito por Schmidt y Lipson (2009) en el artículo «Destilación de leyes naturales de forma libre, a partir de datos experimentales». Esta teoría desembocó en la elaboración de un programa estadístico de regresión simbólica a través de programación genética llamado Eureqa (Schmidt y Lipson 2014). En este estudio, se ha usado la versión Eureqa Pro-Academic.

5. Esta es una técnica estadística basada en el remuestreo con sustitución, de decenas de miles (o más) de nuevas muestras, para simular distribuciones poblacionales, basadas en una muestra inicial. La técnica fue introducida al público en general por B. EFRON (1979), con su artículo *Bootstrap Methods: Another Look at the Jackknife*. El nombre proviene de la expresión en inglés «*pull oneself by one's own bootstraps*», la cual significa que alguien supera un obstáculo, por ejemplo, salta sobre una cerca o un muro, tirando de las correas de sus propias botas. En sentido figurado, esta expresión significa que alguien alcanza una meta, aparentemente imposible, a través de sus propios medios limitados. En el caso de la estadística, *bootstrapping* puede implicar hacer inferencias poblacionales sobre la base de una pequeña muestra inicial. Por ello, se alcanza algo, antes pensado imposible, con medios muy limitados.

6. En el caso de la bioquímica clínica, los casos que caigan dentro de los límites de referencia, que definen el 95 % de la muestra, serán llamados sanos. Si asumimos que El Salvador, en general, tiene niveles de homicidio endémicos elevados desde una perspectiva internacional, muchos de los municipios no serán sanos internacionalmente. Sin embargo, hay que recordar que este estudio se limita a El Salvador, y, por ello, la muestra de referencia, y con ello el intervalo de sanidad, refiere exclusivamente al contexto salvadoreño.

En resumen, la intención de Schmidt y Lipson fue la de automatizar el proceso científico de búsqueda de relaciones analíticas de fenómenos, al definir este proceso de búsqueda algorítmicamente como un programa de computadora. Su algoritmo utiliza un principio para la identificación de relaciones que llaman «no-triviales».

De acuerdo con sus experimentos iniciales con el algoritmo, fue posible «descubrir» hipótesis científicas bien establecidas, sin ningún conocimiento *a priori* de estos fenómenos. Estos experimentos fueron utilizados para aprender cómo el algoritmo identifica estas relaciones y esta información fue utilizada para refinar el algoritmo.

Los autores emplean la definición de Koza (1992) de regresión simbólica, como un método de computación evolucionaria, que busca el espacio de expresiones matemáticas que describen analíticamente los datos y simultáneamente minimizan la métrica de error. Se llama regresión simbólica porque se utilizan datos observados en el pasado para encontrar los símbolos matemáticos que mejor describen las relaciones analíticas entre ellos. Pero, a diferencia de los métodos de optimización de la regresión cuantílica o del método aproximado de los mínimos cuadrados, esto se hace a través del ensayo y error, probando millones de expresiones matemáticas una por una, empezando por las expresiones matemáticas más simples e incrementando gradualmente su complejidad.

El método se define como una técnica de computación evolucionaria porque las diferentes expresiones matemáticas compiten en un proceso de selección en el cual solo los modelos más aptos sobreviven. De antemano, el algoritmo define que un modelo más apto es el que minimiza el error⁷, y describe las relaciones analíticas entre las variables incluidas con menor complejidad. Es por ello que se dice que el mejor modelo es el más Pareto-óptimo, en términos de error y simpleza explicativa.

Se ha usado Eureqa para evaluar si hay modelos lineales o no-lineales que puedan describir simbólicamente las relaciones analíticas de los datos de una manera sencilla, sin brindar un error más alto que la regresión cuantílica de la mediana. Para hacer esto, el objetivo de minimización en Eureqa se ha situado en el error absoluto, que es el error que se minimiza en el problema de optimización de la regresión cuantílica de la mediana. Sin embargo, este proceso no es posible de ejecutar con las regresiones cuantílicas de los cuantiles 3 y 98 de la tasa de homicidios, ya que Eureqa no permite minimizar el error absoluto inclinado de estos cuantiles. Es por ello que se incluyen todas las variables en las regresiones cuantílicas en los percentiles 3 y 98, y no es posible comparar estas regresiones con otros modelos. Este procedimiento de comparación solo es válido para la regresión cuantílica estimada en la mediana y puede implicar la aceptación de modelos distintos a la regresión cuantílica sobre la mediana con todas las variables, o la misma regresión cuantílica lineal, pero sin todas las variables.

Los resultados de este ejercicio se presentan después de los resultados de estimación de las regresiones cuantílicas.

IV. RESULTADOS

Las estadísticas descriptivas de las variables consideradas se muestran en la Tabla II.

7. En el programa Eureqa, el tipo de error a minimizar se define por el usuario del programa.

TABLA II
ESTADÍSTICAS DESCRIPTIVAS

VARIABLE	N	MEDIA	ERROR ESTÁNDAR	DESVIACIÓN ESTÁNDAR	VARIANZA	VIF
Homicidios por 100.000 habitantes (variable dependiente)	262	35,831	2,22203	35,96675	1 293,61	
Delitos de eficacia (menos homicidios) por 100.000 habitantes	262	574,1871	21,19518	343,07389	117 699,70	1,828
Habitantes por kilómetro cuadrado	262	456,4147	70,15126	1 135,50	1 289 354,3	1,87
Tasa de fecundidad	262	2,8362	0,03124	0,50572	0,256	2,393
Índice de mortalidad infantil	262	28,6145	0,28861	4,67162	21,824	2,282
Índice de masculinidad	262	91,9895	0,30212	4,89029	23,915	1,834
Porcentaje de la población en relación de dependencia	262	77,9386	0,6022	9,74746	95,013	4,106
Porcentaje de población urbana	262	39,7473	1,51671	24,50003	602,704	4,209
Extranjeros por 100.000 habitantes	262	718,8888	67,44712	1 091,73	1 191 867,8	1,418
Habitantes receptores de remesas por 100.000 habitantes	262	30 835,75	581,48525	9 412,16	8 858 8774,4	3,006
Monto promedio de remesas recibido mensualmente por habitante (USD)	262	38,6701	0,25963	4,20241	17,66	1,468
Porcentaje de asistencia escolar	262	84,6866	0,31067	5,02856	25,286	2,153
Porcentaje de la población con acceso a agua potable	262	69,223	1,10945	17,95805	322,492	1,611
Porcentaje de la población sin servicios sanitarios	262	11,3749	0,69728	11,28646	127,384	2,667
Porcentaje de la población con servicio sanitario conectado a alcantarillado	262	13,1326	1,1988	19,4042	376,523	3,322
Porcentaje de viviendas con piso de tierra	262	27,9953	0,85631	13,86063	192,117	3,913
Población indígena por 100.000 habitantes	262	306,0952	145,72417	2 338,75	5 563 709,9	1,15
Porcentaje de la población de la tercera edad	262	9,9733	0,12579	2,03604	4,145	2,062
Tasa de analfabetismo	262	21,321	0,45272	7,32792	53,698	7,155
Porcentaje de acceso a la electricidad	262	81,9014	0,80526	13,03435	169,894	3,405
Miembros de pandilla por 100.000 habitantes	262	389,9893	30,43399	492,61714	242 671,64	1,814
Presencia policial por 100.000 habitantes	262	137,3512	8,82238	142,80267	20 392,60	1,456
Número de pandillas	262	0,916	0,05648	0,91424	0,836	2,28

Fuente: Elaboración propia.

IV.1. Resultados de las regresiones cuantílicas

Los percentiles de la variable dependiente se calcularon en StataSE 14.1 y se presentan en la Tabla III.

TABLA III
 ESTIMACIÓN DE PERCENTILES DE LA TASA DE HOMICIDIOS EN EL SALVADOR

VARIABLE	NÚMERO DE OBSERVACIONES	PERCENTIL	CENTIL	INTERPRETACIÓN BINOMIAL INTERVALO DE CONFIANZA 95 %	
Homicidios por 100.000 habitantes	262	3	0	0	0
Homicidios por 100.000 habitantes	262	50	27,25	21,74086	32,88504
Homicidios por 100.000 habitantes	262	98	149,3206	113,6056	166,2107

Fuente: Elaboración propia.

Las Tablas IV, V y VI presentan los resultados de las regresiones cuantílicas de la tasa de homicidios con errores estándar calculados con *bootstrap*⁸. Como puede verse en el Diagrama I, calculado con «*grqreg*» en Stata 14.1, los coeficientes de las variables independientes varían dependiendo del cuantil en el que se calcule la regresión cuantílica. La curva de cada subdiagrama representa los valores cambiantes que asumen los coeficientes de las variables independientes en el modelo de regresión cuantílica a lo largo de cada cuantil de la tasa de homicidios en el que se calcule la regresión cuantílica. La línea recta horizontal punteada de cada subdiagrama representa el valor constante del coeficiente de cada variable independiente de una regresión calculada con el método de mínimos cuadrados. Las otras dos líneas horizontales, directamente arriba y debajo del valor del coeficiente del método de mínimos cuadrados, su intervalo de confianza respectivo. El intervalo de confianza de los coeficientes cuantílicos está dado por la zona gris en torno a la curva. Este diagrama muestra cómo las pendientes de las regresiones cuantílicas varían dependiendo del cuantil τ que se use para calcularlas. Esto implica también su error estándar y las inferencias que pueden hacerse con el modelo.

8. Esta técnica se explica en términos generales en el apartado metodológico.

TABLA IV
REGRESIÓN MEDIANA. VARIABLE DEPENDIENTE «H», CON TODAS LAS VARIABLES
EXPLICATIVAS *BOOTSTRAP* (10 000) ERROR ESTÁNDAR. NÚMERO DE OBSERVACIONES 262.
SUMA DE DESVIACIONES BRUTA 3 569,355 (CERCA DE 27,23)
MIN DE LA SUMA DE DESVIACIONES 2 526,327 PSEUDO R2 = 0,2922

VARIABLE	H	COEF.	STD. ERR.	T	P> t	[95 % CONF. INTERVAL]	
Número de pandillas	X	12,68	3,91	3,24	0,00	4,98	20,38
Porcentaje de población urbana	Q	0,12	0,18	0,68	0,50	-0,23	0,46
Delitos de eficacia (menos homicidios) por 100.000 habitantes	E	0,00	0,01	0,33	0,74	-0,01	0,02
Habitantes por kilómetro cuadrado	P	0,00	0,00	0,06	0,96	-0,01	0,01
Tasa de fecundidad	C	-4,02	5,09	-0,79	0,43	-14,05	6,01
Índice de mortalidad infantil	I	-0,45	0,53	-0,85	0,40	-1,50	0,60
Índice de masculinidad	G	-0,05	0,47	-0,11	0,91	-0,98	0,88
Porcentaje de la población en relación de dependencia	D	-0,58	0,40	-1,42	0,16	-1,37	0,22
Extranjeros por 100.000 habitantes	T	0,00	0,00	-0,64	0,52	0,00	0,00
Habitantes receptores de remesas por 100.000 habitantes	L	0,00	0,00	-0,79	0,43	0,00	0,00
Monto promedio de remesas recibido mensualmente por habitante (USD)	W	-0,98	0,55	-1,78	0,08	-2,06	0,11
Porcentaje de asistencia escolar	S	-0,05	0,55	-0,09	0,93	-1,13	1,03
Porcentaje de la población con acceso a agua potable	V	-0,30	0,12	-2,53	0,01	-0,53	-0,66
Porcentaje de la población sin servicios sanitarios	U	0,00	0,26	-0,02	1,00	-0,52	0,51
Porcentaje de la población con servicio sanitario conectado a alcantarillado	K	0,22	0,20	1,13	0,26	-0,17	0,61
Porcentaje de viviendas con piso de tierra	F	0,07	0,25	0,30	0,77	-0,41	0,56
Población indígena por 100.000 habitantes	J	0,00	0,00	-0,10	0,92	0,00	0,00
Porcentaje de la población de la tercera edad	Z	-1,61	1,55	-1,04	0,30	-4,66	1,44
Tasa de analfabetismo	Y	0,69	0,69	1,01	0,32	-0,66	2,05
Porcentaje de acceso a la electricidad	O	0,13	0,25	0,51	0,61	-0,36	0,61
Miembros de pandilla por 100.000 habitantes	M	0,00	0,01	-0,08	0,94	-0,02	0,02
Presencia policial por 100.000 habitantes	A	0,01	0,02	0,30	0,77	-0,03	0,05
Constante		142,35	68,04	2,09	0,04	8,31	276,39

$$\begin{aligned}
 h = & 12,67667 * x + 0,1187213 * q + 0,0026712 * e + 0,0001898 * p - 4,021062 * c \\
 & - 0,450167 * i - 0,0536578 * g - 0,5756374 * d - 0,0011269 * t - 0,0003058 \\
 & * l - 0,9764386 * w - 0,0516672 * s - 0,2958132 * v - 0,0045232 * u \\
 & + 0,2237101 * k + 0,0733657 * f - 0,0002335 * j - 1,610901 * z + 0,6940198 \\
 & * y + 0,1251835 * o - 0,0006931 * m + 0,0060515 * a + 142,3498
 \end{aligned}$$

Fuente: Elaboración propia.

TABLA V
 REGRESIÓN PERCENTIL 3. VARIABLE DEPENDIENTE «H», CON TODAS LAS VARIABLES
 EXPLICATIVAS *BOOTSTRAP* (10 000) ERROR ESTÁNDAR. NÚMERO DE OBSERVACIONES 262.
 SUMA DE DESVIACIONES BRUTA 281,6319 (CERCA DE 0)
 MIN DE LA SUMA DE DESVIACIONES 246,486. PSEUDO R2 = 0,1248

VARIABLE	H	COEF.	STD. ERR.	T	P> T	[95 % CONF. INTERVAL]
Número de pandillas	X	8,73	3,08	2,84	0,01	2,66 14,79
Porcentaje de población urbana	Q	0,05	0,14	0,39	0,70	-0,22 0,32
Delitos de eficacia (menos homicidios) por 100.000 habitantes	E	0,00	0,01	0,03	0,98	-0,01 0,01
Habitantes por kilómetro cuadrado	P	0,01	0,00	1,87	0,06	0,00 0,01
Tasa de fecundidad	C	3,61	4,03	0,90	0,37	-4,32 11,55
Índice de mortalidad infantil	I	-0,64	0,46	-1,39	0,17	-1,55 0,27
Índice de masculinidad	G	0,12	0,34	0,35	0,72	-0,54 0,78
Porcentaje de la población en relación de dependencia	D	-0,04	0,30	-0,13	0,89	-0,63 0,55
Extranjeros por 100.000 habitantes	T	0,00	0,00	0,48	0,63	0,00 0,00
Habitantes receptores de remesas por 100.000 habitantes	L	0,00	0,00	-0,56	0,57	0,00 0,00
Monto promedio de remesas recibido mensualmente por habitante (USD)	W	-0,18	0,47	-0,38	0,70	-1,11 0,75
Porcentaje de asistencia escolar	S	-0,15	0,33	-0,44	0,66	-0,80 0,51
Porcentaje de la población con acceso a agua potable	V	0,02	0,13	0,14	0,89	-0,23 0,27
Porcentaje de la población sin servicios sanitarios	U	0,16	0,20	0,80	0,43	-0,24 0,56
Porcentaje de la población con servicio sanitario conectado a alcantarillado	K	0,08	0,18	0,43	0,67	-0,27 0,42
Porcentaje de viviendas con piso de tierra	F	0,03	0,19	0,15	0,88	-0,35 0,41
Población indígena por 100.000 habitantes	J	0,00	0,00	0,25	0,80	0,00 0,00
Porcentaje de la población de la tercera edad	Z	0,45	1,10	0,41	0,69	-1,73 2,62
Tasa de analfabetismo	Y	0,07	0,59	0,12	0,90	-1,10 1,24
Porcentaje de acceso a la electricidad	O	0,10	0,21	0,49	0,62	-0,31 0,52
Miembros de pandilla por 100.000 habitantes	M	-0,01	0,01	-1,28	0,20	-0,02 0,00
Presencia policial por 100.000 habitantes	A	0,00	0,01	-0,10	0,92	-0,03 0,02
Constante		-0,15	49,03	0,00	1,00	-96,73 96,43

$$\begin{aligned}
 h = & 8,726373 * x + 0,0527889 * q + 0,0001665 * e + 0,0050513 * p + 3,614258 * c \\
 & - 0,6423433 * i + 0,118632 * g - 0,0400673 * d + 0,0005943 * t - 0,0001338 \\
 & * l - 0,1821194 * w - 0,1473141 * s + 0,0178749 * v + 0,1605296 * u \\
 & + 0,0750166 * k + 0,0286339 * f + 0,0004077 * j + 0,4475492 * z + 0,0718195 \\
 & * y + 0,1030296 * o - 0,0066432 * m - 0,0012509 * a - 0,1524939
 \end{aligned}$$

Fuente: Elaboración propia.

TABLA VI

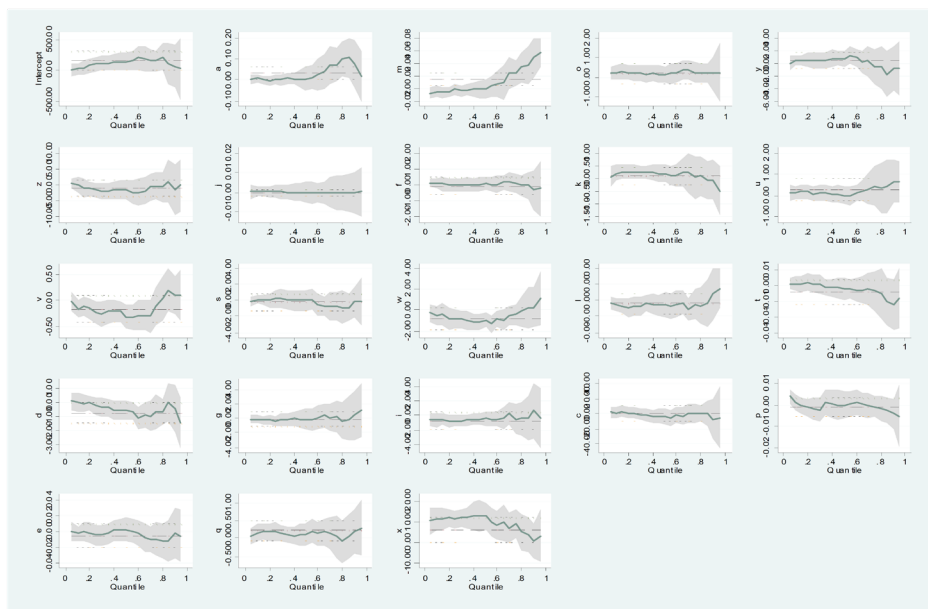
REGRESIÓN PERCENTIL 98. VARIABLE DEPENDIENTE «H», CON TODAS LAS VARIABLES EXPLICATIVAS *BOOTSTRAP* (10 000) ERROR ESTÁNDAR. NÚMERO DE OBSERVACIONES 262.
SUMA DE DESVIACIONES BRUTA 652,6754 (CERCA DE 144)
MIN DE LA SUMA DE DESVIACIONES 325,803 PSEUDO R2 = 0,5008

VARIABLE	H	COEF.	STD. ERR.	T	P> t	[95 % CONF. INTERVAL]
Número de pandillas	X	7,08	8,59	0,82	0,41	-9,84 24,00
Porcentaje de población urbana	Q	0,14	0,46	0,31	0,75	-0,76 1,05
Delitos de eficacia (menos homicidios) por 100 000 habitantes	E	-0,02	0,02	-0,75	0,45	-0,06 0,03
Habitantes por kilómetro cuadrado	P	-0,01	0,01	-1,12	0,26	-0,02 0,00
Tasa de fecundidad	C	-16,45	18,71	-0,88	0,38	-53,30 20,41
Índice de mortalidad infantil	I	-0,21	1,87	-0,11	0,91	-3,91 3,48
Índice de masculinidad	G	1,00	1,54	0,65	0,52	-2,03 4,02
Porcentaje de la población en relación de dependencia	D	-2,03	1,04	-1,95	0,05	-4,07 0,02
Extranjeros por 100 000 habitantes	T	0,00	0,01	-0,65	0,52	-0,02 0,01
Habitantes receptores de remesas por 100 000 habitantes	L	0,00	0,00	0,24	0,81	0,00 0,00
Monto promedio de remesas recibido mensualmente por habitante (USD)	W	1,54	1,53	1,01	0,31	-1,46 4,55
Porcentaje de asistencia escolar	S	0,32	1,48	0,22	0,83	-2,59 3,23
Porcentaje de la población con acceso a agua potable	V	0,10	0,37	0,27	0,79	-0,62 0,82
Porcentaje de la población sin servicios sanitarios	U	0,07	0,78	0,09	0,93	-1,47 1,61
Porcentaje de la población con servicio sanitario conectado a alcantarillado	K	-0,41	0,64	-0,63	0,53	-1,67 0,86
Porcentaje de viviendas con piso de tierra	F	-0,77	0,88	-0,87	0,38	-2,50 0,97
Población indígena por 100 000 habitantes	J	0,00	0,01	-0,12	0,90	-0,13 0,01
Porcentaje de la población de la tercera edad	Z	0,96	3,29	0,29	0,77	-5,52 7,44
Tasa de analfabetismo	Y	1,43	2,16	0,66	0,51	-2,82 5,68
Porcentaje de acceso a la electricidad	O	-0,43	0,68	-0,63	0,53	-1,77 0,91
Miembros de pandilla por 100 000 habitantes	M	0,05	0,02	3,18	0,00	0,02 0,08
Presencia policial por 100 000 habitantes	A	0,00	0,05	-0,02	0,99	-0,11 0,10
Constante		121,25	249,14	0,49	0,63	-369,54 612,03

$$\begin{aligned}
 h = & 7,079228 * x + 0,1443197 * q - 0,0159534 * e - 0,0064302 * p - 16,4454 * c. \\
 & - 0,2130682 * i + 0,9982828 * g - 2,027734 * d - 0,0037889 * t + 0,0002228 \\
 & * l + 1,544617 * w + 0,3196204 * s + 0,1002174 * v + 0,0676786 * u \\
 & - 0,4064072 * k - 0,7689765 * f - 0,0007441 * j + 0,9599005 * z + 1,429976 \\
 & * y - 0,4272954 * o + 0,0489122 * m - 0,0008602 * a + 121,2476
 \end{aligned}$$

Fuente: Elaboración propia.

DIAGRAMA I
COEFICIENTES DE REGRESIÓN CUANTÍLICA VS. COEFICIENTES DE REGRESIÓN
CON MÍNIMOS CUADRADOS



Nota: Diagrama elaborado automáticamente por Stata 14.1 (comando «*qrreg*»).

IV.2. Resultados de la regresión simbólica a través de programación genética

IV.2.1. Modelo de homicidios

Los genes matemáticos de la búsqueda realizada a través del programa de regresión simbólica por programación genética, Eureka Pro-Academic, se definieron de antemano para que el algoritmo de programación encontrara la regresión simbólica más Pareto-óptima en términos de error absoluto y complejidad. Los genes matemáticos utilizados fueron: constante, variable input, suma, resta, multiplicación, división, función exponencial y potencia. Se pidió al programa ensayar diversas combinaciones de estos genes con todas las variables independientes, para explicar la variable dependiente.

La búsqueda se hizo con 4 núcleos y fue detenida después de 3 horas 43 minutos y 25 segundos. Entonces, el programa había repasado 1.574.753 generaciones, y evaluado $5,7 \times 10^{10}$ fórmulas posibles. El proceso fue detenido en este momento porque las reducciones del error solo se consiguieron a través de complicaciones considerables del modelo, al punto de hacerlo difícil de interpretar, sin muchas posibilidades de hacer inferencias importantes.

El modelo con el menor error, pero también con la mayor complejidad, es el modelo 38 en la lista de modelos encontrados, y tiene la siguiente forma:

$$h = 23,7571948336355 + 13,5932889941802 * x \\ + 0,0016045745400845 * q * u * \exp(0,0626569578500938 \\ * q) + (24,0092794564035 - q - x)/\exp(k) \\ - 7,97414591335625e - 6 * l * v$$

Este modelo tiene un R^2 de 0,42 y un coeficiente de correlación de 0,67. Sin embargo, recibe un *score* de complejidad de 38.

El modelo que parece ofrecer la mayor simplicidad, con un *score* de 7, al mismo tiempo que preserva gran parte del poder explicativo de los otros modelos con un R^2 de 0,29 y un coeficiente de correlación de 0,57 es:

$$h = 17,1 * x + 0,339 * q$$

Considerando que este es un modelo que se ha obtenido minimizando el error absoluto, podrá entenderse fácilmente que es el equivalente de la regresión cuantílica sobre la mediana, calculada en el apartado anterior, con la diferencia de que Eureka Pro-Academic ha destilado las realciones más importantes, descartando variables. Tanto el modelo sobre la mediana con todas las variables presentadas en la Tabla IV como este modelo lineal sugerido por Eureka tienen un coeficiente de determinación de 0,29. Según los resultados del análisis, las variables independientes más importantes son x (cantidad de pandillas diferentes por municipio) y q (índice de urbanidad del municipio).

De acuerdo con los resultados de la regresión cuantílica en la mediana presentados en la Tabla IV, además de x (Beta=12,68, p=0,001) solo la variable v , porcentaje de hogares del municipio con acceso a agua potable (Beta=-0,3, p=0,012), tiene una relación significativa con la tasa de homicidios por 100.000 habitantes por municipio. El porcentaje de acceso a agua potable es la tercera variable independiente más frecuente en los modelos de Eureka.

Desde un punto de vista teórico, dejar o quitar variables, significativas o no, depende de las teorías y los objetivos en cuestión. Para nuestros propósitos, es válido seleccionar el modelo de regresión cuantílica sobre la mediana que relaciona la tasa de homicidios con las variables independientes de cantidad de pandillas por municipio y el índice de urbanidad, dada su simpleza y bajo error relativo⁹. En otras palabras, podemos prescindir de 20 variables, incluyendo la variable de porcentaje de hogares con acceso a agua potable, para alcanzar un modelo con poder explicativo equivalente y mayor simpleza. Pero es importante tomar en cuenta las inferencias que pueden hacerse por ambos modelos.

Dado que no es posible hacer inferencias sobre las variables a través de Eureka, se ha hecho esto en StataSE 14.1, evaluando la regresión cuantílica sobre la mediana, donde la tasa de homicidios por 100.000 habitantes por municipio h es la variable dependiente,

9. Recordemos que Eureka también ha comparado otros modelos que incluyen la variable de porcentaje de acceso a agua potable, y menos variables, pero que brindan un mayor error absoluto.

y las variables independientes son la cantidad de pandillas por municipio x y el índice de urbanidad del municipio q .

StataSE no permite hacer un análisis de regresión con *bootstraps* sin incluir una constante, lo cual es conveniente desde un punto de vista teórico, ya que sabemos que hemos omitido una serie de variables importantes. Por ello, el resultado no será idéntico al del modelo más simple presentado por Eureka, que nos ha proporcionado un modelo Pareto-óptimo desde el punto de vista meramente matemático, pero que no considera la ausencia de variables incógnitas que sí tienen un efecto importante en la práctica, asumido en la constante. Sin embargo, la siguiente opción más compleja en Eureka, que cumple la condición de tener el mismo error, es un modelo de regresión cuantílica sobre la mediana con intercepto. Es decir, $h = n^*x + r^*q - A$, que es justamente lo que StataSE 14.1 nos permite analizar para hacer inferencias. Los resultados de los coeficientes en la regresión en Eureka varían ligeramente a los de StataSE 14.1 debido a aproximaciones y diferencias en los procedimientos de cálculo. Los resultados de regresión cuantílica en StataSE 14.1 con este modelo se presentan en la Tabla VII.

TABLA VII
 REGRESIÓN MEDIANA. VARIABLE DEPENDIENTE «H», VARIABLES EXPLICATIVAS «X» Y «Q»
 BOOTSTRAP (10 000) ERROR ESTÁNDAR. NÚMERO DE OBSERVACIONES 262.
 SUMA DE DESVIACIONES BRUTA 3 569,355 (CERCA DE 34,460823)
 MIN DE LA SUMA DE DESVIACIONES 2 775,756 PSEUDO R2 = 0,2223

VARIABLE	H	COEF.	STD. ERR.	T	P> T	[95 % CONF. INTERVAL]	
Número de pandillas	X	16,62	2,68	6,19	0,00	11,34	21,91
Porcentaje de población urbana	Q	0,37	0,09	4,36	0,00	0,20	0,54
Constante	_cons	-0,93	3,38	-0,28	0,78	-7,59	5,73

$$h = 16,62431 * x + 0,3725954 * q - 0,9310698$$

Fuente: Elaboración propia.

V. DISCUSIÓN

V.1. Límite inferior del rango de referencia

De acuerdo con los resultados presentados, en el percentil 3, solo la variable de cantidad estimada de pandillas por municipio en 2015 tiene una correlación significativa con la tasa de homicidios por 100.000 habitantes en 2007 (Beta=8,73 y p=0,005). Esto puede interpretarse como evidencia de que una mayor conflictividad entre pandillas, según datos de 2015, explica aumentos de la tasa de homicidios en los municipios que muestran tasas de 0 homicidios por 100.000 habitantes. Esto es, siempre y cuando las estimaciones de número de pandillas por municipio de 2015 reflejen de alguna manera estos conflictos en 2007. Quiere decir que no falsificamos la hipótesis nula de que las variables independientes en

el modelo no están correlacionadas significativamente con la variable dependiente de la tasa de homicidios por 100.000 habitantes, con excepción de la cantidad de pandillas por municipio, cuando la regresión cuantílica se calcula sobre el percentil 3, con todas las variables independientes.

V.2. Mediana

El modelo cuantílico con todas las variables en la mediana demuestra que solo la cantidad de pandillas por municipio, según datos de 2015 ($\text{Beta}=12,69$, $p=0,001$) y el porcentaje de hogares del municipio con acceso a agua potable ($\text{Beta}=-0,3$, $p=0,012$) están significativamente correlacionados con la tasa de homicidios.

El análisis posterior con regresión simbólica a través de programación genética muestra que 20 variables, incluida la variable de acceso a agua potable, pueden sustituirse aceptablemente con el índice de urbanidad del municipio, de manera que un modelo que solo incluye la cantidad de pandillas por municipio y el índice de urbanidad no pierde mucho poder explicativo en términos de error absoluto, a pesar de que se han descartado tantas variables. El coeficiente de determinación se mantiene en 0,2922 con 22 variables y con tan solo 2 variables: cantidad de pandillas e índice de urbanidad. Sin embargo, por motivos teóricos y técnicos para la inferencia, se ha optado por un modelo de 2 variables con intercepto, lo cual reduce el coeficiente de determinación a 0,2223. En este modelo, tanto el coeficiente de la cantidad de pandillas por municipio ($\text{Beta}=16,62$ y $p=0,000$) como el del índice de urbanidad del municipio ($\text{Beta}=0,37$ y $p=0,000$) son significativos.

Es decir, que el modelo con todas las variables demuestra, significativamente, que mientras más pandillas por municipio y menor acceso relativo a agua potable en el municipio, mayor será la tasa de homicidios por 100.000 habitantes. Cuando se descartan variables, y el modelo sobre la mediana se calcula solamente con las variables de cantidad de pandillas e índice de urbanidad, podemos falsar la hipótesis nula de que estas variables no están correlacionadas significativamente con la tasa de homicidios por 100.000 habitantes. Por lo tanto, se puede interpretar que el modelo de dos variables, cantidad de pandillas e índice de urbanidad, con intercepto, demuestra que mientras mayor cantidad de pandillas e índice de urbanidad en el municipio, mayor será la tasa de homicidios por 100.000 habitantes.

V.3. Límite superior del rango de referencia

De todos los modelos de regresión calculados, el modelo de la regresión cuantílica múltiple con todas las variables sobre el percentil 98 es el que muestra el mayor coeficiente de determinación, con un valor 0,5008. Esta es una cifra relativamente alta para un fenómeno tan complejo, lo cual indica que, a pesar de que sabemos que nos hacen falta datos sobre variables explicativas importantes de acuerdo con la teoría y que hemos procedido con supuestos incompletos, la información que tenemos sobre las variables independientes hace un buen trabajo para explicar el 50 % de la variación en la variable dependiente de la tasa de homicidios por 100.000 habitantes.

Sin embargo, en este caso, no podemos falsar la hipótesis nula de que las variables independientes en el modelo no están significativamente correlacionadas con la tasa de homicidios por 100.000 habitantes, con excepción de la variable de cantidad de pandilleros por 100.000 habitantes por municipio, de acuerdo con estimaciones de la PNC para 2015 (Beta=0,049, $p=0,002$). Este es un resultado interesante, ya que la variable independiente que tiene correlación significativa con la tasa de homicidios en las regresiones cuantílicas del percentil 3 y 50 fue la cantidad de pandillas por municipio, no la cantidad de pandilleros por 100.000 habitantes.

V.4. La evidencia no corrobora la mayoría de teorías consideradas

De manera que, de acuerdo con la información disponible, este modelo no puede utilizarse como evidencia empírica de la mayoría de postulados que se presentan en el apartado teórico, con excepción de los casos presentados en los apartados anteriores.

Además, el estudio demuestra coeficientes de determinación bajos en el percentil 3 y la mediana. Niveles bajos del coeficiente de determinación son de esperar cuando se modelan fenómenos complejos como este. Pero niveles tan bajos como estos indican que probablemente se han obviado una serie de variables clave, las cuales están correlacionadas con la tasa de homicidios en los cuantiles bajos. Es de suponer que algunas de estas variables excluidas son las que se enumeran en el apartado teórico, por ausencia de datos. También hay que recordar que este estudio presupone que los efectos de las variables independientes se hacen sentir en el municipio en el que estas han sido medidas. Pero, en la práctica, puede ser que factores socioeconómicos en un municipio influyan la tasa de homicidios en otro municipio. De manera que los resultados de las regresiones en este estudio obviarán tales posibilidades, y esto puede también explicar los bajos coeficientes de determinación.

VI. CONCLUSIONES

Este estudio no propone nuevas teorías. Simplemente se limita a probar hipótesis generalmente aceptadas, para demostrar la utilidad de herramientas novedosas al abordar empíricamente el tema de homicidios en El Salvador. Por ello, el estudio evita interpretaciones teóricas y discusiones especulativas, a menos que estas puedan alcanzarse en base a los métodos presentados y la calidad de la información estadística que dichos métodos requieren.

El método de regresión cuantílica es una herramienta muy útil para analizar las relaciones entre diversas variables, al intentar explicar un fenómeno tan complejo como la tasa de homicidios. Este método proporciona flexibilidad en los supuestos, lo cual lo hace más adaptable a este tipo de fenómenos complejos y, al mismo tiempo, es un método que tolera datos atípicos y relaciones cambiantes entre las variables, incluso dentro del mismo modelo de regresión. Este estudio ha probado que los coeficientes de las variables independientes de las regresiones cuantílicas sobre la variable dependiente de la tasa de homicidios por 100.000 habitantes varían dependiendo del percentil de la

variable dependiente en que se calcule la regresión cuantílica (ver el Diagrama I). Quiere decir que un método que presuponga una correlación constante entre las variables, independientemente del cuantil en el que se calcule la regresión, como el método de mínimos cuadrados, será insuficiente para explicar relaciones analíticas entre las variables en este estudio.

El método de regresión simbólica a través de programación genética puede ser muy útil para formular modelos posibles, que describan matemáticamente relaciones analíticas entre las variables independientes y la tasa de homicidios, de acuerdo con diversas teorías.

La cantidad de pandillas por municipio, según estimaciones de la PNC para 2015, es un factor que está significativamente correlacionado de una manera directamente proporcional con la tasa de homicidios por 100.000 habitantes por municipio, tanto en el límite inferior del rango de referencia, si se incluyen suficientes variables de control, y la mediana de la tasa de homicidios por 100.000 habitantes en El Salvador en 2007, en un modelo que además incluya la tasa de urbanidad por municipio. De acuerdo con este estudio, políticas de prevención de homicidios en los municipios con niveles de homicidios por debajo de la mediana, pueden estar enfocadas en reducir la cantidad de pandillas en el municipio y/o la conflictividad entre estas.

En torno a la mediana de la tasa de homicidios por 100.000 habitantes en El Salvador en 2007, el porcentaje de población con acceso a agua potable está inversamente relacionado de una manera significativa con la tasa de homicidios por 100.000 habitantes por municipio, siempre y cuando se incluyan suficientes variables de control. Una política integral de reducción de la tasa de homicidios debe tomar en cuenta este aspecto, que sirve como indicador socioeconómico de los municipios en cuestión, al tratarse del agua, un aspecto tan fundamental para mejorar la calidad de vida.

El índice de urbanidad está significativamente relacionado de una manera directamente proporcional con la tasa de homicidios por 100.000 habitantes por municipio, en el percentil 50 de la tasa de homicidios por 100.000 habitantes por municipio, en un modelo que solo incluye esta variable y la cantidad de pandillas por municipio. Una política integral de reducción de la tasa de homicidios en torno a la mediana debe incluir coordinación con políticas de desarrollo urbano, que permitan el diseño de espacios físicos urbanos menos proclives a la violencia.

En torno al límite superior del rango de referencia de la tasa de homicidios por 100.000 habitantes por municipio en El Salvador en 2007, la cantidad de miembros de pandilla por 100.000 habitantes, según estimaciones de la PNC para 2015, está directamente relacionada de una manera significativa con la tasa de homicidios por 100.000 habitantes en 2007, siempre y cuando se incluyan suficientes variables de control. De acuerdo con este estudio, si el objetivo es reducir los niveles más altos de la tasa de homicidio, parece ser acertado que el enfoque de política se encuentre en las pandillas, tanto con medidas de combate directo a la criminalidad organizada como con medidas de prevención.

Este estudio tiene una serie de limitaciones, algunas de las cuales serán compartidas por cualquier otro estudio de este tipo. Muchas de estas limitaciones han sido resueltas a través de una serie de compromisos metodológicos y teóricos. Sin embargo, los modelos aquí presentados no explican la totalidad del problema y no proporcionan una explicación

de la causalidad de la tasa de homicidios en el sentido clásico de la noción. Esta es una advertencia para evitar que estos resultados se usen equivocadamente para reducir un fenómeno tan complejo a unas pocas variables o como evidencia de las causas de los homicidios o de la violencia en El Salvador.

Se recomiendan estudios futuros que incluyan mejor información estadística, otras variables y técnicas que no se han podido considerar en este estudio, para intentar estimar regresiones sobre la variable dependiente de la tasa de homicidios con un mejor poder explicativo.

VII. BIBLIOGRAFÍA

- ACEVEDO, C. Los costos económicos de la violencia en El Salvador. *América Latina Hoy*, 2008, vol. 50: 71-88.
- ARTIGA GONZÁLEZ, Á. Polarización política: Orígenes, consecuencias y alternativas. En ARTIGA GONZÁLEZ, A. *et al.* (eds.). *La polarización política en El Salvador*. San Salvador: FUNDAUNGO-FLACSO, 2007: 1-22.
- ASAMBLEA LEGISLATIVA DE LA REPÚBLICA DE EL SALVADOR. *Ley de Acceso a la Información Pública. Decreto n.º 534, 8 de abril de 2011*. Disponible en: <http://www.iaip.gob.sv/?q=ley-y-reglamento/ley-de-acceso-la-informaci%C3%B3n-p%C3%ABblica-decreto-no-534>.
- BANCO MUNDIAL, IBRD-IDA. Homicidios intencionales (por cada 100.000 habitantes). *Base de datos de estadísticas de homicidios internacionales de la Oficina de las Naciones Unidas contra la Droga y el Delito*. 2016. Disponible en: <http://datos.bancomundial.org/indicador/VC.IHR.PSRC.P5?end=2014&locations=SV&start=1995>. Fecha de consulta: 26 de Noviembre de 2016.
- BERGMANN, A. Sin razón aparente: conflictos sociales y violencia en la posguerra. En MELÉNDEZ, Ó. y BERGMANN, A. *Violencia en tiempos de paz: Conflictividad y criminalización en El Salvador*. San Salvador: Secretaría de Cultura de la Presidencia, 2015: 221-252.
- BRICENO-LEON, R.; VILLAVECES, A. y CONCHA-EASTMAN, A. Understanding the uneven distribution of the incidence of homicide in Latin America. *International Journal of Epidemiology*, 2008, vol. 37 (4): 751-757.
- CADE, B. S. y NOON, B. R. A Gentle Introduction to Quantile Regression for Ecologists. *Frontiers in Ecology and the Environment*, 2003, vol. 1 (8): 412-420.
- CARCACH, C. A. *El Salvador. Mapa de violencia y su referencia histórica*. San Salvador: Centro de monitoreo y evaluación de la violencia desde la perspectiva ciudadana, 2008.
- CERQUEIRA, D. *Causas e conseqüências do crime no Brasil*. Tesis de doctorado. Rio de Janeiro: Pontificia Universidade Católica do Rio de Janeiro, 2010.
- CÓRDOVA, R. *Atlas de la violencia en El Salvador 2009-12*. San Salvador: Fundación Dr. Guillermo Manuel Ungo (FUNDAUNGO), 2013.
- CRAMÉR, H. *Mathematical methods of statistics*. Princeton: Princeton University Press, 1946.
- CRUZ, J. M. y BELTRÁN, M. A. *Las armas de fuego en El Salvador. Relación e impacto sobre la violencia*. San Salvador: IUDOP, 2000.
- CRUZ, J. M. y GONZÁLEZ, A. Magnitud de la violencia en El Salvador. *Estudios Centroamericanos*, 1997, n.º 588: 953-966.

- CRUZ, J. M.; GONZÁLEZ, L. A.; ROMANO, L. E. y SISTI, E. De la guerra al delito: evolución de la violencia en El Salvador. En LONDOÑO DE LA CUESTA, J. L. y GAVIRIA, A. G. *Asalto al desarrollo: violencia en América Latina*. Washington: BID, 2000.
- CRUZ, J. M.; TRIGUEROS ARGÜELLO, Á. y GONZÁLEZ, F. *El crimen violento en El Salvador. Factores sociales y económicos asociados*. San Salvador: Universidad Centroamericana, «José Simón Cañas», 2000.
- DICRISTINA, B. Durkheim's theory of homicide and the confusion of the empirical literature. *Theoretical Criminology*, 2004, vol. 8 (1): 57-91.
- DIMAGGIO, P.; EVANS, J. y BRYSON, B. Have Americans' Social Attitudes Become More Polarized? *American Journal of Sociology*, 1996, vol. 102 (3): 690-755.
- DURBIN, J. Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables. *Econometrica*, 1970, vol. 38 (3): 410-421.
- DURBIN, J. y WATSON, G. (1951). Testing for serial correlation in least squares regression II. *Biometrika*, 1951, vol. 38 (1-2): 159-178.
- EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 1979, vol. 7 (1): 1-26.
- FOWLER, T. B. The Formality of Reality: Xavier Zubiri's Critique of Hume's Analysis of Causality. *The Xavier Zubiri Review*, 1998, vol. 1: 57-66.
- FUNDAUNGO. Fundación Dr. Guillermo Manuel Ungo. *Factores Asociados a la violencia en El Salvador*. El Salvador: Banco Central de Reserva de El Salvador, 2015. Disponible en: <http://www.bcr.gob.sv/bcrsite/uploaded/content/category/1216398628.pdf>. Fecha de consulta: 28 de Octubre de 2015.
- IMBUSCH, P.; MISSE, M. y CARRIÓN, F. Violence research in Latin America and the Caribbean: A literature review. *International Journal of Conflict and Violence*, 2011, vol. 5 (1): 87-154.
- JACOBS, D. y RICHARDSON, A. M. Economic Inequality and Homicide in the Developed Nations From 1975 to 1995. *Homicide Studies*, 2008, vol. 12 (1): 28-45.
- KALINA, J. Three contributions to robust regression diagnostics. *Jamsi*, 2015, vol. 11 (2): 69-78.
- KOENKER, R. y BASSETT JR., G. Regression Quantiles. *Econometrica*, 1978, vol. 46 (1): 33-50.
- KOENKER, R. y HALLOCK, K. F. Quantile Regression. *Journal of Economic Perspectives*, 2001, vol. 15 (4): 143-156.
- KOZA, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge: MIT Press, 1992.
- KUTNER, M.; NACHTSHEIM, C.; NETER, J. y LI, W. *Applied linear regression models*. Boston: McGraw Hill-Irwin, 2004.
- LESTER, D. Crime as Opportunity. *British Journal of Criminology*, 1991, vol. 31 (2): 186-188.
- MACHADO, J. A. y SANTOS, S. J. *Quantile regression and heteroskedasticity*, 2013. Disponible en: https://jmcss.som.surrey.ac.uk/JM_JSS.pdf. Fecha de consulta: 18 de junio de 2016.
- MARSHALL, W. J.; LAPSLEY, M.; DAY, A. y AYLING, R. *Clinical Biochemistry: Metabolic and Clinical Aspects*. London: Elsevier, 2014.
- MORRIS, P. K. y GRAYCAR, A. Homicide Through a Different Lens. *British Journal of Criminology*, 2011, vol. 51: 823-838.
- MORRIS, P. K. y MAGUIRE, E. R. Political Culture, Neighbourhood Structure and Homicide in Urban Jamaica. *British Journal of Criminology*, 2016, vol. 56: 919-936.

- MURRAY, J.; CERQUERIA, D. R. y KAHN, T. Crime and violence in Brazil: Systematic review of time trends, prevalence rates and risk factors. *Agression and Violent Behaviour*, 2013, vol. 18 (5): 471-483.
- OUIMET, M. A World of Homicides: The Effect of Economic Development, Income Inequality, and Excess Infant Mortality on the Homicide Rate for 165 Countries in 2010. *Homicide Studies*, 2012, vol. 16 (3): 238-258.
- PNUD. Programa de las Naciones Unidas para el Desarrollo. *Armas de fuego y violencia*. San Salvador: PNUD, 2003.
- PORTILLO, N. Estudios sobre pandillas juveniles en El Salvador y Centroamérica: Una revisión de su dimensión participativa. *Apuntes de Psicología*, 2012, vol. 30 (1-3): 397-407.
- REICHENHEIM, M. E.; DE SOUZA, E. R.; MORAES, C. L.; DE MELLO JORGE, M. H.; DA SILVA, C. M. y DE SOUZA MINAYO, M. C. Violence and injuries in Brazil: The effect, progress made, and challenges ahead. *Lancet*, 2011, vol. 377 (9781):1962-1975.
- SALGADO, M. *Factores clave para entender la violencia en El Salvador*. Buenos Aires: CLACSO, 2012.
- SANCHO, E. Causa de la violencia en El Salvador: el apartheid social. *Realidad y Reflexión*, 2013, vol. 38: 101-113.
- SAVIN, N. E. y WHITE, K. J. The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica*, 1977, vol. 45 (8): 1989-1996.
- SCHMIDT, M. y LIPSON, H. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 2009, vol. 324 (5923): 81-85.
- SCHMIDT, M. y LIPSON, H. Genetic programming of implicit regressions. En RIOLO, R. *et al. Genetic Programming. Theory and Practice VII*. New York: Springer US, 2010.
- SCHMIDT, M. y LIPSON, H. Eureqa (Version 0.98 beta) [Software]. Disponible en; www.nutonian.com: <http://www.nutonian.com/research>, 2014.
- STATACORP. *Stata Statistical Software: Release 14*. Texas: StataCorp LP, 2015.
- UNIVERSIDAD CENTROAMERICANA «JOSÉ SIMEÓN CAÑAS». *Explicaciones de la violencia*. Editorial UCA. 3 de Junio de 2015. Disponible en: <http://www.uca.edu.sv/noticias/texto-3683>.
- UNODC. Oficina de las Naciones Unidas contra la Droga y el Delito. *Estudio mundial sobre el homicidio*. Viena: UNODC, 2013.
- YANG, T.-C.; CHEN, V. Y.-J.; SHOFF, C. y MATTHEWS, S. A. Using quantile regression to examine the effects of inequality across the mortality distribution in the U.S. countries. *Social Science and Medicine*, 2012, vol. 74 (12): 1900-1910.