



Universidad Autónoma del Estado de Hidalgo

---

---

Instituto de Ciencias Básicas e Ingeniería  
Área Académica de Matemáticas y Física

Impacto de la violencia de pareja sobre el  
ingreso laboral de las mujeres mexicanas

T E S I S

que para obtener el título de:

Licenciado en Matemáticas Aplicadas

presenta

**Yei Javier Zepeda Hernández**

Bajo la dirección de la

Dra. Alma Sofía Santillán Hernández

Mineral de la Reforma, Hidalgo.

Septiembre 2024





Universidad Autónoma del Estado de Hidalgo  
Instituto de Ciencias Básicas e Ingeniería  
*School of Engineering and Basic Sciences*

Mineral de la Reforma, Hgo., a 17 de septiembre de 2024

Número de control: ICBI-D/1091/2024  
Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO**  
**DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio le comunico que el Jurado asignado al Egresado de la Licenciatura en Matemáticas Aplicadas **Yei Javier Zepeda Hernández**, quien presenta el trabajo de titulación "**Impacto de la violencia de pareja sobre el ingreso laboral de las mujeres mexicanas**", después de revisar el trabajo en reunión de Sinodales ha decidido autorizar la impresión del mismo, hechas las correcciones que fueron acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente:** Dr. Raúl Temoltzi Ávila

**Secretario:** Dr. Ronald Richard Jiménez Munguía

**Vocal:** Dra. Alma Sofía Santillán Hernández

**Suplente:** Dr. Ricardo Cruz Castillo

Sin otro particular por el momento, reciba un cordial saludo.

Atentamente  
"Amor, Orden y Progreso"

Mtro. Gabriel Vergara Rodríguez  
Director del ICBI



GVR/YCC

Ciudad del Conocimiento, Carretera Pachuca-Tulancingo Km. 4.5 Colonia Carboneras, Mineral de la Reforma, Hidalgo, México. C.P. 42184  
Teléfono: 771 71 720 00 Ext. 40001  
direccion\_icbi@uaeh.edu.mx,  
vergarar@uaeh.edu.mx



uaeh.edu.mx



# Agradecimientos

Quisiera dedicar este espacio para expresar mi más sincero agradecimiento a todos aquellos que han colaborado en la edificación de este trabajo. En primer lugar, a mi familia: mis padres Miriam y Javier, y mis hermanos, Zeltzin y Tonatiuh. Siempre me han brindado su apoyo, desde el día 0 de mi vida, me han acompañado y sé que lo seguirán haciendo hasta el día  $n$ . Gracias a ellos he llegado a ser la persona que soy hoy, son para mí una fuente constante de inspiración y admiración, por su forma de ser, sus valores, su forma de vivir la vida y confrontar las dificultades que se han de presentar, nunca dejaré de aprender de cada uno de ustedes, me motivan a seguir adelante, sin ustedes no habría podido alcanzar este significativo logro.

Asimismo, quiero extender mi agradecimiento a mis catedráticos, quienes han dejado su granito, o incluso, su camión de arena en mi preparación como profesional. Me siento especialmente agradecido con la Dra. Alma Sofía Santillán Hernández, quien no solo se ha convertido en mi mentora, sino que me transmitió su pasión y dedicación por el análisis de datos. Su increíble trayectoria y dedicación han sido fundamentales para mi crecimiento académico, compartiéndome un sinnúmero de conocimientos y guiándome en el tan bonito proceso que transformó un proyecto de una clase optativa en un producto concreto para un artículo, y posteriormente, se desarrolló un trabajo de tesis. Agradezco la confianza que ha depositado en mí, así como cada minuto de su tiempo que dedicó a impartirme valiosas lecciones. Estoy profundamente agradecido por cada conversación y por cada momento de inspiración que me ha brindado.

Quisiera concluir expresando mis agradecimientos a cada uno de los distinguidos sinodales que formaron parte del comité de revisión de este trabajo, el Dr. Raúl Temoltzi Ávila, el Dr. Ronald Richard Jiménez Munguía y el Dr. Ricardo Cruz Castillo. Por hacer un espacio en su valioso tiempo para atender la revisión de este trabajo. Aprecio sinceramente cada uno de sus comentarios, los cuales han contribuido significativamente a mejorar la estructura y claridad de este trabajo. Su experiencia y conocimientos han enriquecido este proceso, y me siento honrado de haber contado con su apoyo en esta etapa.



# Resumen

En México, el 70.1 % de las mujeres ha sufrido violencia en algún momento de su vida, siendo su pareja el agresor principal. Este tipo de violencia representa un serio problema de salud pública, ya que trae consigo múltiples impactos negativos, no solo para las mujeres que la padecen, sino también para sus hijos y el entorno familiar. A pesar de la gravedad del problema, la relación entre la violencia de pareja y los niveles de ingresos ha sido un tema que, en el caso mexicano, ha recibido poca atención. El presente trabajo tiene como objetivo analizar el impacto de la violencia de pareja sobre el ingreso laboral de las mujeres en México. Para ello, se utilizan datos de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH 2021). Se emplea el método de componentes principales para estimar cinco tipos de violencia: física, emocional, económica, sexual y acoso. Para abordar la endogeneidad entre las variables de ingreso y violencia, se aplica el método de variables instrumentales.

Con excepción del acoso, los resultados muestran una relación negativa entre los ingresos laborales y los otros cuatro tipos de violencia. En particular, la violencia sexual tiene el mayor impacto negativo sobre el ingreso de las mujeres; las víctimas de violencia sexual experimentan, en promedio, una reducción del 17.4 % en su ingreso laboral mensual. En el desarrollo del trabajo, se hace una estimación del costo económico de la violencia, que sugiere que este impacto equivale a una pérdida de más de dos meses de días de trabajo al año, lo que se traduce en una pérdida anual de 16,517 pesos en sus ingresos laborales. En términos agregados, esto representa el 0.60 % del PIB. Estos resultados subrayan la urgencia de orientar las políticas públicas hacia la prevención, atención, sanción y erradicación de todas las formas de violencia contra las mujeres.

# Abstract

In Mexico, 70.1 % of women have suffered violence at some point in their lives, with their intimate partner being the main aggressor. This type of violence represents a serious public health issue, as it brings multiple negative impacts, not only for the women who suffer from it, but also for their children and the family environment. Despite the severity of the problem, the relationship between intimate partner violence and income levels has been a topic that, in the Mexican case, has received little attention. This study aims to analyze the impact of intimate partner violence on women's labor income in Mexico. For this purpose, data from the National Survey on the Dynamics of Household Relationships (ENDIREH 2021) are used. The Principal Component Analysis method is employed to estimate five types of violence: physical, emotional, economic, sexual, and harassment. The Instrumental Variables method is applied to address the endogeneity between income and violence variables.

With the exception of harassment, the results show a negative relationship between labor income and the other four types of violence. In particular, sexual violence has the most significant negative impact on women's income; victims of sexual violence experience, on average, a 17.4 % reduction in their monthly labor income. The study also estimates the economic cost of violence, suggesting that this impact equates to a loss of more than two months worth of working days per year, which translates into an annual loss of 16,517 pesos in their labor income. In aggregate terms, this represents 0.60 % of PIB. These findings underscore the urgency of directing public policies toward the prevention, attention, sanction, and eradication of all forms of violence against women.



# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Revisión de literatura</b>	<b>5</b>
1.1. Teorías económicas . . . . .	5
1.1.1. Negociación dentro del hogar . . . . .	6
1.1.2. Extractiva . . . . .	6
1.1.3. Recurso relativo . . . . .	7
1.2. Teorías sociológicas . . . . .	7
1.2.1. Dependencia marital . . . . .	7
1.2.2. Contragolpe masculino . . . . .	8
1.3. Teoría criminológica . . . . .	8
1.3.1. Reducción a la exposición . . . . .	9
1.4. Evidencia empírica en México . . . . .	10
1.4.1. Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico	10
1.4.2. Public transfers and domestic violence: The roles of private information and spousal control . . . . .	10
1.4.3. Domestic violence and women's earnings in Mexico . . . . .	11
1.5. Endogeneidad de la medida de violencia . . . . .	11
1.5.1. Simultaneidad entre ingreso y violencia . . . . .	12
1.5.2. Características no observables . . . . .	12
1.5.3. Error de medición . . . . .	12
<b>2. Preliminares matemáticos</b>	<b>15</b>
2.1. Análisis de componentes principales . . . . .	15
2.2. Regresión lineal . . . . .	20
2.3. Variables instrumentales . . . . .	27
2.3.1. Validez de los instrumentos . . . . .	28
<b>3. Datos y método</b>	<b>31</b>
3.1. Muestra de estudio . . . . .	31
3.2. Estimación de los índices de violencia . . . . .	32

3.3. Variables para el modelado . . . . .	34
3.4. Estadística descriptiva de la muestra . . . . .	35
3.5. Modelos . . . . .	37
<b>4. Resultados</b>	<b>41</b>
4.1. Omitiendo la endogeneidad de la violencia:	
Modelo de regresión lineal . . . . .	41
4.2. Corrigiendo por la endogeneidad de la violencia:	
Modelo con variables instrumentales . . . . .	43
4.3. Estimación del costo de la violencia . . . . .	48
<b>Conclusiones</b>	<b>51</b>
<b>A. Desarrollo del estimador <math>\hat{\beta}</math></b>	<b>57</b>
<b>B. Pruebas de relevancia y exogeneidad de los instrumentos</b>	<b>69</b>
<b>Bibliografía</b>	<b>72</b>

# Introducción

La violencia contra las mujeres es un grave problema social que aumentó en los años recientes. A nivel nacional, del total de la población de mujeres de 15 años y más el 62.8 % fue víctima de violencia en el año 2011, 66.1 % en 2016 y 70.1 % para 2021 (INEGI, 2022b). En 2021, aproximadamente dos de cada cinco mujeres casadas han sufrido violencia de pareja y 20.7 % la sufrió en el periodo de octubre de 2020 a octubre de 2021. De entre los tipos de violencia que el Instituto Nacional de Estadística y Geografía (INEGI) reporta (psicológica, física, sexual, económica o patrimonial), la violencia psicológica fue el tipo de violencia que más experimentaron las mujeres ya sea por parte de su pareja actual o la última, seguida de la violencia económica (INEGI, 2022a).

No se suelen dimensionar las consecuencias que genera la violencia, sin embargo, esta trae consigo costos monetarios y no monetarios. Dentro de los costos no monetarios se incluyen efectos colaterales que se producen en los miembros de la familia, además del daño psicológico y la pérdida de salud, pues aquellas mujeres que son víctimas de abuso generan problemas tales como trastornos depresivos, problemas crónicos de salud, trastornos gastrointestinales, ginecológicos, entre otros (Plazaola-Castaño y Ruiz-Pérez 2004).

En cuanto a los costos monetarios, se considera el costo por la atención médica que la mujer requiera para atender dichas lesiones y enfermedades que se generen producto de la situación de violencia; también los costos que impliquen las denuncias judiciales que sean necesarias. Otro costo incluye la posible pérdida de ingresos laborales causada por las ausencias laborales y la reducción de la productividad; el elevado porcentaje del decreciente desempeño laboral en las mujeres se relaciona directamente con el incremento del maltrato (Torres-Jacome y Fernand-Desfrancois 2021).

Con respecto a la pérdida de ingresos de las mujeres como consecuencia de sufrir violencia por parte de su pareja, INEGI estimó para el año 2016 que en promedio cada mujer pierde anualmente 29.7 días de trabajado remunerado y 27.8 días de trabajo no remunerado, en términos agregados esta cantidad asciende a 4.4 mil millones de pesos, 1.6 correspondiente al trabajo remunerado y 2.8 en el caso del trabajo no remunerado. En términos relativos esta pérdida representó el 0.03 % del Producto Interno Bruto (PIB) del 2016 (INEGI 2018).

México es un país con ingresos laborales bajos en el que las mujeres reciben en promedio 6% menos ingresos que los hombres y, a pesar de que una gama de factores influyen en la brecha salarial, resulta que al restringir la muestra únicamente a individuos de zonas urbanas esta brecha se explica por las diferencias de género, además aquellas mujeres con mayores salarios potenciales son las que deciden trabajar, (Arceo-Gómez y Campos-Vázquez 2014), por lo que la pérdida de ingresos monetarios a causa de la violencia puede ser un factor que exacerbe las desventajas de las mujeres en términos tanto monetarios como de participación laboral.

Actualmente la relación entre ingresos y violencia de pareja contra la mujer es un tema que ha sido poco estudiado en México. Un antecedente reciente para el caso mexicano es de Aguirre (2023). Un importante aporte de la autora es que además de emplear la medida convencional de violencia, la cual consiste en variables dicotómicas de que la mujer sufrió algún tipo de violencia, considera medir la violencia mediante un índice para cada tipo de violencia que ella identifica (económica, emocional y física). Al utilizar dichos índices es posible identificar cómo afecta al ingreso laboral la intensidad de violencia que experimenta la mujer.

Para cada tipo de violencia, Aguirre (2023) encuentra evidencia de una relación negativa con los ingresos de las mujeres; tanto para la medida de violencia con variables dicotómicas como para los índices estimados, el efecto más grande se observa para la violencia emocional. No obstante, un aspecto que no considera la autora es la presencia de endogeneidad de la medida de violencia la cual proviene por tres fuentes, la primera es la simultaneidad que existe entre las variables ingreso y violencia, pues resulta que así como el ingreso afecta a la violencia, la violencia también afecta al ingreso; otra fuente de endogeneidad son las características no observables que afectan tanto a la violencia como al ingreso, y la tercera es por el error de medición en las variables.

Este trabajo busca contribuir en el estudio del efecto que tiene la violencia de pareja sobre los ingresos laborales de las mujeres en México, considerando que se va a emplear el método de variables instrumentales para tratar el aspecto de la endogeneidad de la violencia al momento del modelado del problema, para este método se emplean como instrumentos los antecedentes de violencia de la mujer y de su pareja, además del nivel de apoyo que la mujer recibe por parte de terceros, es decir, personas ajenas al hogar.

Para alcanzar el objetivo, se emplea la información recopilada por INEGI en la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) en su edición del año 2021, siendo esta la encuesta representativa a nivel nacional más reciente que permite analizar la relación entre violencia e ingresos. Para el estudio de esta relación se construyen cinco índices de violencia mediante la técnica de componentes principales. Adicionalmente, se hace una estimación del costo de la violencia en términos del número de días laborales perdidos al año a causa de la violencia.

Los resultados obtenidos en este trabajo indican que la violencia de pareja de tipo económica, emocional, física y sexual reducen en promedio los ingresos laborales de las mujeres que se en-

cuentran en una relación de pareja, ya sea casadas o en unión libre. De manera más específica, el mayor impacto se observa en la violencia de tipo sexual, donde se registra una disminución promedio del 17.4% en los ingresos mensuales de las mujeres, lo cual es equivalente a perder anualmente poco más de dos meses de trabajo remunerado. El siguiente efecto más grande se presenta en la violencia física, seguido de la violencia económica y finalmente la violencia emocional.

El primer capítulo de este trabajo presenta la revisión de la literatura acerca de la relación entre la violencia de pareja contra la mujer y su situación económica. Posterior a este, se dedica un capítulo a los preeliminares matemáticos, se aborda la fundamentación matemática del método de componentes principales y de los modelos de inferencia causal que se utilizan para el estudio del problema. Posteriormente se describen los datos y muestra de estudio, se discute además el método de estimación empleado. En el capítulo siguiente se exhiben los resultados obtenidos del efecto que tiene la violencia sobre los ingresos de la mujer, así como la estimación del costo de la violencia. Finalmente se presentan las conclusiones.



# Capítulo 1

## Revisión de literatura

Este trabajo estudia el efecto de la violencia de pareja contra la mujer sobre sus ingresos. Por distintas partes del mundo se han desarrollado teorías que asocian la violencia con los ingresos de la mujer desde una perspectiva económica, sociológica y criminológica, estas teorías se enfocan en analizar cómo es que la violencia contra la mujer se ve afectada a causa de sus ingresos y las conclusiones son mixtas, tal como lo muestra la Tabla 1.1. Por un lado, resulta que la situación económica de la mujer tiene un efecto en la incidencia de violencia que pueda experimentar y por otro, la violencia que sufre la mujer también afecta los ingresos que recibe. Prueba de esto son los modelos teóricos y empíricos que en el presente capítulo se abordan. Este factor de simultaneidad entre violencia e ingresos se debe de considerar en el análisis de la relación, pues influye directamente en el modelado de la problemática. En este capítulo se presentan algunas de las teorías más destacadas de acuerdo al campo de estudio que se toma como enfoque, posteriormente se muestra evidencia empírica de la relación violencia-ingresos en México y finalmente se habla de la endogeneidad de la medida de violencia.

### 1.1. Teorías económicas

La mejora en la situación económica de la mujer está estrechamente relacionada con varios de los Objetivos de Desarrollo Sostenible de las Naciones Unidas, siendo la violencia de pareja una de las problemáticas que abordan estos objetivos (Eggers del Campo y Steinert 2022). No obstante, el vínculo entre el empoderamiento económico de la mujer y el riesgo de violencia de pareja sigue siendo ambiguo de acuerdo a las teorías que se han planteado, se describen a continuación tres de las teorías que destacan al considerar el enfoque económico.

### 1.1.1. Negociación dentro del hogar

Según la teoría de la negociación dentro del hogar, las decisiones económicas y las relaciones de poder dentro de una pareja se caracterizan por procesos de negociación, donde los miembros de la pareja intentan maximizar su propia utilidad o bienestar. Estos procesos de negociación pueden estar influenciados por una variedad de factores, incluidos los recursos económicos disponibles para cada miembro de la pareja, (Becker 1993).

Los ingresos pueden afectar la dinámica de negociación dentro del hogar al cambiar el equilibrio de poder económico entre los miembros de la pareja. Si una mujer tiene ingresos más altos que los percibidos por su pareja, puede aumentar su independencia económica y su capacidad para tomar decisiones autónomas, es decir, podría cambiar la dinámica de poder dentro de la relación al incrementar su poder de negociación y, potencialmente, reducir la violencia recibida dado que la amenaza de abandonar a la pareja se vuelve más creíble, (Baranov et al. 2020).

### 1.1.2. Extractiva

La teoría extractiva predice que la violencia es un medio para extraer dinero de su pareja, por lo tanto, mayores ingresos de la mujer aumentan los recursos que el hombre puede extraer por medio de la violencia (Kibris y Nelson 2022; Rao y Bloch 2002).

Rao y Bloch estudian el caso de tres aldeas del sur de la India y analizan cómo es que a través de la violencia el hombre obtiene mayores recursos. En la India el matrimonio no suele ser una cuestión de elección para las mujeres, sino que surge con base en normas sociales y preferencias parentales. Como parte del acuerdo matrimonial se sigue la práctica del *dote*, que es una cantidad de dinero, bienes o propiedades que la familia de la novia entrega a la familia del novio y si bien la dote ha sido ilegal en la India desde 1961, todavía persiste en muchas partes del país y en diversas comunidades. En Rao y Bloch (2002) se revela que las mujeres cuyas familias pagan dotes más pequeñas sufren un mayor riesgo de violencia marital, mientras que las dotes más grandes reducen la probabilidad de violencia, de hecho, a menudo se paga una dote adicional para evitar que el esposo golpee sistemáticamente a la esposa.

De acuerdo con un estudio que realizaron Kibris y Nelson (2022) utilizando datos del oeste de Turquía de una encuesta levantada en 2019, obtienen que las mujeres con ingresos personales tienen más probabilidades de sufrir abusos por parte de sus maridos que aquellas que no tienen ingresos propios. Además, si bien obtener un ingreso les da a las mujeres cierto poder de negociación, este impacto positivo se ve ampliamente contrarrestado por los incentivos para que sus parejas masculinas utilicen la violencia para extraer recursos.



### 1.1.3. Recurso relativo

La contribución al hogar depende más del aporte económico que del esfuerzo y tiempo dedicado a las actividades dentro del mismo, por lo que la generación de ingresos se ve reflejada en un mayor poder dentro del hogar. Por su parte, las normas sociales de género que la sociedad ha naturalizado influyen en desempoderar a la mujer y, de hecho, se emplean en el hogar cuando previamente no se establecen de manera explícita los acuerdos en la relación de pareja, pues al no discutirlos entonces se convierten en reglas morales a seguir sin cuestionamiento (Covarrubias 2018). La teoría del recurso relativo sugiere que las mujeres con mayores ingresos que su pareja son más susceptibles de ser violentadas, debido a que cuando el ingreso del hombre es menor que el de su pareja siente disminuido su poder y como respuesta para restaurarlo se ve un aumento en la frecuencia y en la gravedad del abuso hacia su esposa (McClosely 1996).

Dado que de manera cultural se ha establecido al hombre como el proveedor del hogar, cuando este se ve con la incapacidad de proveer más recursos surgen las discusiones y la agresividad de su parte hacia su pareja (Abramsky et al. 2019). La violencia resultará ser una herramienta de intimidación utilizada para mantener el poder en el hogar, incluso las normas sociales de género la justifican, pues dan legitimidad y autorización al hombre de disciplinar a su pareja si es que esta no cumple con sus roles de género (Contreras, 2008).

## 1.2. Teorías sociológicas

La problemática de la influencia que tiene la situación económica de las mujeres sobre el nivel de violencia del que puedan ser víctimas también se ha estudiado con un enfoque sociológico y al igual que se vio con el enfoque económico, las conclusiones de las teorías son mixtas, una describe al empoderamiento femenino como un mecanismo de protección ante la violencia y la otra señala que el empoderamiento es de hecho un factor potencial que aumenta el nivel de violencia que pueden experimentar las mujeres.

### 1.2.1. Dependencia marital

Esta teoría argumenta que la dependencia económica de una pareja puede crear una situación en la que una parte tiene control sobre la otra, lo que aumenta el riesgo de violencia contra la mujer y condiciona su capacidad para poder salir de la relación (Johnson 2008). Aquellas mujeres que tienen una mayor dependencia económica de sus parejas pueden estar en mayor riesgo de sufrir violencia por parte de ellas debido a que la falta de recursos propios dificulta la capacidad de la mujer para salir de una relación abusiva. Mientras que las mujeres con menores niveles de dependencia marital, es decir, aquellas que tienen ingresos y recursos propios, tienen menos probabilidades de sufrir violencia por parte de sus parejas gracias a que tienen más opciones y

medios para buscar ayuda y escapar de situaciones de abuso (Vyas y Watts 2009).

### **1.2.2. Contragolpe masculino**

El contragolpe masculino usado en sociología argumenta que las mujeres son más propensas a experimentar violencia cuando tienen mayores recursos disponibles. Tanto el empoderamiento de las mujeres como los cambios en las normas de género tradicionales pueden desafiar las estructuras de poder impuestas por la sociedad, lo que genera resistencia y reacciones violentas en aquellos hombres que se sienten amenazados por estos cambios, siendo la violencia una forma en que estos hombres intentan mantener su control sobre las mujeres y reafirman su autoridad en un contexto donde las mujeres están ganando más autonomía y poder (Kimmel 2000).

Después de la Primera Guerra Mundial, el Tratado de Versalles en 1919 otorgó a Francia y al Reino Unido mandatos para administrar y controlar partes del territorio alemán colonial en África, incluido el territorio que ahora es Camerún. Este período de mandato colonial duró hasta que Camerún francés se independizó en 1960 y posteriormente Camerún británico se unió a él en 1961 para formar la República Federal de Camerún, que luego se convirtió en la República Unida de Camerún. Durante ese mandato de 1919 hasta 1961, cada país impuso su propio régimen, en particular, a las mujeres en los territorios británicos se les brindó un sistema educativo universal y obtuvieron oportunidades para ganar salarios en efectivo equitativamente, es decir, sin importar el género. Por su parte, Francia se centró en invertir en el sector de infraestructura dominado por el empleo masculino. Aprovechando que las mujeres que formaron parte de este fenómeno naturalmente causado por la historia convivían en un entorno cultural y geográficamente homogéneo, Guarnieri y Rainer (2018) lo tomaron como objeto de estudio para analizar la relación entre el empleo femenino y la violencia doméstica, encontraron que las mujeres pertenecientes a los territorios de la colonia británica tienen una mayor probabilidad de participar en el mercado laboral que sus contrapartes de la colonia francesa, sin embargo, las mujeres en los antiguos territorios británicos tienen un 30% más de probabilidades de ser víctimas de violencia doméstica que aquellas en los antiguos territorios franceses, pues de acuerdo al contragolpe masculino, los hombres perciben las mejoras económicas de las mujeres como amenaza y reaccionan violentándolas, en un intento por restaurar su dominancia dentro del hogar.

## **1.3. Teoría criminológica**

La criminología desempeña un papel importante en el estudio de la violencia contra la mujer, pues ayuda a proporcionar conocimientos sobre la naturaleza, las causas y las consecuencias de esta problemática, se describe a continuación una de las teorías más relevantes que ha surgido cuando se considera el enfoque criminológico.

### 1.3.1. Reducción a la exposición

Esta teoría predice que conforme las mujeres tienen mejores condiciones económicas, tales como mayor participación en el mercado laboral, mayor nivel educativo y mayores ingresos, se reduce la dependencia económica hacia su pareja, así como el tiempo de exposición con la pareja violenta, en consecuencia, la violencia hacia ellas se reduce. Se sugiere además que el aumento del estatus de las mujeres y la mejora en los recursos disponibles para las víctimas de violencia doméstica, como refugios y servicios de apoyo, son factores que colaboran en la disminución de los niveles de violencia (Dugan et al. 1999).

Las intervenciones dirigidas a mejorar el acceso de las mujeres a recursos económicos y a promover su participación en el mercado laboral resultan ser estrategias efectivas para prevenir y abordar la violencia de pareja, mujeres con algún grado de propiedad de bienes, que cuentan con un empleo remunerado y que son las encargadas de sus propios ingresos, tienen menos probabilidades de experimentar violencia por parte de sus parejas (Stöckl et al. 2021)

Tabla 1.1: Teorías de la relación entre la situación económica de la mujer y la violencia de pareja

Mayores ingresos implican experimentar violencia de pareja con...			
Mayor probabilidad		Menor probabilidad	
Extractiva	La violencia es un medio para extraer dinero de su pareja, mayores ingresos de la mujer aumentan los recursos que el hombre puede extraer por medio de la violencia.	Negociación	Las mujeres con mayores oportunidades económicas tienen mayor poder de negociación, así que la amenaza de abandonar a la pareja se vuelve más creíble y experimentan menores niveles de violencia.
Recurso relativo	El poder se asocia con los ingresos y el hombre para restaurar ese poder dentro del hogar reacciona con violencia.	Dependencia marital	Mujeres con menor dependencia (mayores ingresos) tienen mayor poder de negociación y son menos propensas a permitir relaciones violentas.
Contragolpe masculino	Los hombres perciben las mejoras económicas de las mujeres como amenaza y reaccionan violentándolas, en un intento por restaurar su dominancia dentro del hogar.	Reducción a la exposición	Si las mujeres tienen mejores condiciones económicas se reduce la dependencia económica hacia su pareja, así como el tiempo de exposición con la pareja violenta y la violencia hacia ellas se reduce.

## 1.4. Evidencia empírica en México

Una forma de recibir ingresos por parte de las mujeres ocurre a través de las transferencias monetarias de programas sociales. En la mayoría de los países de bajos y medianos ingresos se ha encontrado evidencia de que las transferencias están asociadas con una reducción de la violencia de pareja. Para el caso de México, la evidencia empírica es mixta, tal como se muestra a continuación.

### 1.4.1. Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico

Angelucci (2008) estudia las determinantes de la violencia doméstica y el abuso del alcohol, mediante datos de la muestra de evaluación del programa de transferencias monetarias condicionadas Progres-a-Oportunidades se evalúa el efecto del monto de las transferencias recibidas sobre la violencia. La autora muestra que el programa disminuye el abuso de alcohol de la pareja de la mujer en un 15 % y cambia su comportamiento agresivo según el tamaño de la transferencia de efectivo recibida, la educación de los esposos y la brecha de edad conyugal.

Resulta que las transferencias pequeñas disminuyen la violencia en un 37 % para todos los hogares, mientras que las transferencias de alto monto aumentan el comportamiento agresivo de los esposos, estos hombres en particular tenían visiones tradicionales con respecto a los roles de género. Se exhibe que, si bien empoderar a las mujeres a través de programas de bienestar puede tener efectos beneficiosos como reducir el alcoholismo de sus parejas y disminuir la violencia doméstica que puedan sufrir, el riesgo de violencia puede aumentar para algunos otros hogares, en particular, en aquellos donde se brindó mayor apoyo económico a la mujer.

### 1.4.2. Public transfers and domestic violence: The roles of private information and spousal control

Es un estudio para evaluar si los programas de transferencias monetarias a las mujeres disminuyen la incidencia del abuso conyugal. Bobonis et al. (2013) usan datos de la primera versión de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH 2003) para estudiar el impacto de la recepción de transferencias del programa Oportunidades sobre la incidencia de tres tipos de violencia doméstica: física, sexual y emocional.

Los autores encuentran que las mujeres beneficiarias fueron entre 5 y 7 puntos porcentuales menos propensas de ser víctimas de violencia física que las mujeres no beneficiarias. Estos impactos pueden ser explicados como una consecuencia del incremento en el empoderamiento de las mujeres dentro del hogar. No obstante, las mujeres beneficiarias del programa fueron entre 3 y 5 puntos porcentuales más propensas a sufrir de violencia emocional que las mujeres no beneficiarias.

### 1.4.3. Domestic violence and women's earnings in Mexico

Se trata del único antecedente reciente que estudia el efecto que tiene la violencia de pareja sobre los ingresos de la mujer para el caso mexicano. Aguirre (2023) empleó datos de la ENDIREH 2006 y de la 2016 para estimar cómo la violencia de pareja afecta el nivel de ingresos recibido por las mujeres.

Cada tipo de violencia (económica, emocional y sexual) tiene asociadas distintas preguntas de la ENDIREH y sus respuestas posibles son “ninguna vez”, “una vez” y “varias veces”. Para medir la violencia Aguirre lo hace de tres maneras distintas, la primera medida que propone es crear un índice para cada tipo de violencia, el primer paso para estimar dicho índice consiste en crear una variable categórica asociada a cada pregunta, esta variable toma valores 0, 1 o 2, dependiendo de si la respuesta a la pregunta fue “ninguna vez”, “una vez” o “varias veces”, respectivamente. Después, dentro de cada tipo de violencia suma el valor de las variables correspondientes para obtener el puntaje de violencia económica, emocional y física, sin embargo, la cantidad de preguntas asociadas varía entre los tipos de violencia, de modo que el último paso es estandarizar estos puntajes para finalmente obtener tres índices (uno por cada tipo de violencia).

La segunda medida que emplea la autora es una variable dicotómica la cual asigna valor 0 si la mujer respondió “ninguna vez” a todas las preguntas asociadas al tipo de violencia y asigna valor 1 si contestó “una vez” a al menos una de las preguntas asociadas. La última medida consiste en una variable categórica donde una categoría corresponde a haber respondido con “varias veces” a al menos una pregunta asociada al tipo de violencia, una segunda categoría es si se respondió “una vez” a al menos una de las preguntas asociadas pero en ninguna respondió con “varias veces” y una última categoría para mujeres que respondieron no haber sufrido de violencia.

Para cada una de las medidas de violencia empleadas Aguirre encuentra evidencia de una relación negativa sobre los ingresos de las mujeres. En particular, cuando se considera la medida del índice de cada tipo de violencia, el efecto más grande se observa en la violencia emocional, donde resulta que el incremento de una desviación estándar en la violencia emocional reduce en promedio en 1.5 % el ingreso laboral.

## 1.5. Endogeneidad de la medida de violencia

Un aspecto que no considera Aguirre (2023) es la endogeneidad de la medida de violencia la cual proviene por las siguientes fuentes.

### 1.5.1. Simultaneidad entre ingreso y violencia

La primera fuente de endogeneidad es la simultaneidad que existe en las variables ingreso y violencia. Por una parte, de acuerdo con INEGI (2018), debido a la violencia se estima que las mujeres pierden 30 días de trabajo remunerado, lo que en consecuencia afecta sus ingresos laborales, es decir, la violencia contra la mujer afecta el nivel de ingresos que puedan percibir.

Por otro lado, el ingreso que reciben las mujeres es un factor que afecta positiva o negativamente la violencia que reciben, pues de acuerdo a la Tabla 1.1, teorías como la de contragolpe masculino o la de recurso relativo, sugieren que las mujeres son más propensas a experimentar violencia cuando tienen mayores recursos (Guarnieri y Rainer 2018; McClosey 1996); pero según teorías como la de negociación dentro del hogar o la de dependencia marital, las mujeres con mayores ingresos tienden a tener mejores niveles de negociación en la relación y, en consecuencia, tienen menos probabilidades de sufrir violencia (Baranov et al. 2020; Vyas y Watts 2009). De modo que el nivel de ingresos también afecta la probabilidad de ser víctima de violencia.

### 1.5.2. Características no observables

Otra fuente de endogeneidad son las características no observables que tienen la particularidad de afectar tanto a la violencia como al ingreso de la mujer. Ejemplo de este tipo de variables son los estereotipos y roles de género que la sociedad ha establecido; si en la relación de pareja se sigue la estructura cultural que impone a la mujer el papel de ser la encargada de las actividades dentro del hogar y designa al hombre como el encargado de proveer de recursos, entonces la mujer al no poder salir al mercado laboral a emplearse, no tiene la posibilidad de generar un ingreso propio, desarrollando naturalmente una dependencia económica que propicia el control que el esposo puede tener sobre ella, aumentando las probabilidades de violencia contra la mujer dentro del hogar (Johnson 2008).

### 1.5.3. Error de medición

La tercera fuente de endogeneidad es por error de medición en la variable de violencia, pues la información proviene del reporte de las mujeres encuestadas sobre los incidentes de violencia, de modo que puede existir una diferencia entre la realidad y lo reportado. Esta diferencia puede tener distintas fuentes de origen, ya sea por la redacción del cuestionario empleado para la encuesta, por la persona encargada de realizar las preguntas o incluso por las mujeres encuestadas.

De manera específica, las diferencias entre Aguirre (2023) y el presente trabajo son las siguientes:

- En este trabajo se emplea información de la ENDIREH 2021, siendo estos los datos más recientes representativos a nivel nacional para estimar la relación entre ingreso laboral y violencia.

- Para cuantificar la intensidad de la violencia únicamente se propone emplear el método de componentes principales, a través de este método se diferencia a la violencia en cinco diferentes tipos y se crea un índice por cada tipo.
- Se implementa el modelo de variables instrumentales con el objetivo de corregir la endogeneidad de las variables ingreso y violencia.
- De manera adicional, se hace una estimación del costo de la violencia.





## Capítulo 2

# Preeliminarios matemáticos

Para estimar la medida de la violencia se emplea un conjunto de variables por cada tipo de violencia y se considera que estas variables no deben tener el mismo peso como lo hace (Aguirre 2023), por lo tanto, se propone el método de componentes principales para determinar el peso que debe tener cada variable dentro del índice de violencia correspondiente. En este capítulo se presenta la justificación matemática del análisis de componentes principales. Además, se describen los modelos de regresión lineal y variables instrumentales, pues son los propuestos para estimar el efecto que tiene la violencia de pareja contra la mujer sobre sus ingresos.

### 2.1. Análisis de componentes principales

Lo expuesto a continuación se ha basado principalmente en Díaz-Monroy (2007) y Peña (2002). A partir de  $X_1, X_2, \dots, X_p$  variables, el análisis de componentes principales (ACP) tiene como objetivo representar adecuadamente la misma información a través de nuevas variables  $S_1, S_2, \dots, S_q$  denominadas componentes, donde  $q \leq p$ . Cada componente deberá tener media cero y son de la forma

$$S_j = \alpha_{j1}X_1 + \alpha_{j2}X_2 + \dots + \alpha_{jp}X_p \quad (2.1)$$

con  $1 \leq j \leq q$ , se debe cumplir que estas nuevas variables estén no correlacionadas entre sí, es decir,  $cov(S_i, S_l) = 0$  para  $i \neq l$ . Además, cada componente debe capturar la mayor parte posible de la variabilidad total en comparación con las componentes que le siguen. Es importante que  $cov(X_i, X_l) \neq 0$  para  $i \neq l$ , de lo contrario, al aplicar el ACP las componentes se corresponderían con cada variable por orden de magnitud en la varianza; es decir, la  $k$ -ésima componente coincide con la variable de  $k$ -ésima mayor varianza.

Para fijar ideas y sin pérdida de generalidad, considere la siguiente matriz  $\mathbb{X}$

$$\mathbb{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

la cual contiene los datos de  $n$  individuos para las  $X_i$  variables normales con media  $\bar{X}_i$  igual a cero, siendo  $n \times p$  la dimensión de  $\mathbb{X}$ , se sigue entonces que su matriz de covarianzas es

$$\mathbb{Y} = \frac{1}{n} \mathbb{X}^T \mathbb{X}$$

y como se busca obtener un subespacio de dimensión  $q$  tal que al proyectar sobre este los puntos se distorsionen cuanto menos sea posible, entonces dicho subespacio representará lo suficientemente bien la información original.

Como primer paso se busca un subespacio de dimensión uno, una línea recta  $L$  que además de contener al origen preserve la mayor información posible; dado que las variables originales tienen media cero, entonces los datos están centrados, es decir, el centro de los datos se encuentra justamente en el origen, y para que las nuevas variables mantengan una media igual a cero, los datos deben permanecer centrados, de ahí que sea necesario que  $L$  contenga al origen<sup>1</sup>. Sea  $u_1 = (u_{11}, u_{12}, \dots, u_{1p})^T$  un vector columna unitario en  $L$  y  $x_i$  el renglón  $i$  de  $\mathbb{X}$  con  $1 \leq i \leq n$ , es decir,  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ , entonces la proyección de  $x_i$  sobre  $L$  es el escalar  $s_i = \langle u_1^T, x_i \rangle$ , o bien,  $s_i = x_i u_1$ . De modo que  $s_i u_1^T$  es el vector que representa dicha proyección. Se puede ver a  $z_1 = \mathbb{X} u_1$  como la combinación lineal de variables aleatorias, por lo tanto,  $z_1$  es también una variable aleatoria, cuyas  $n$ -observaciones son  $z_{11}, z_{21}, \dots, z_{n1}$  dadas por  $z_{i1} = s_i$ , más aún,  $z_1$  tiene media  $\bar{z}_1$  igual a cero, observe que

$$\begin{aligned} \bar{z}_1 &= \frac{1}{n} \sum_{i=1}^n z_{i1} = \frac{1}{n} \sum_{i=1}^n s_i = \frac{1}{n} \sum_{i=1}^n x_i u_1 \\ &= \frac{1}{n} [x_{11} u_{11} + \dots + x_{1p} u_{1p} + x_{21} u_{11} + \dots + x_{2p} u_{1p} + \dots + x_{n1} u_{11} + \dots + x_{np} u_{1p}] \\ &= \frac{1}{n} \left[ u_{11} \left( \sum_{i=1}^n x_{i1} \right) + u_{12} \left( \sum_{i=1}^n x_{i2} \right) + \dots + u_{1p} \left( \sum_{i=1}^n x_{ip} \right) \right] \\ &= u_{11} \left( \frac{1}{n} \sum_{i=1}^n x_{i1} \right) + u_{12} \left( \frac{1}{n} \sum_{i=1}^n x_{i2} \right) + \dots + u_{1p} \left( \frac{1}{n} \sum_{i=1}^n x_{ip} \right) \end{aligned}$$

donde

$$\frac{1}{n} \sum_{i=1}^n x_{ij} = \bar{X}_j = 0$$

<sup>1</sup>En este caso las  $X_1, X_2, \dots, X_p$  variables ya cumplen que tienen media cero, sin embargo, de no ser así, primero se deben de centrar los datos.

para todo  $j = 1, 2, \dots, p$ , por la hipótesis de media cero para las  $X_j$  variables, por lo tanto

$$\bar{z}_1 = 0 \quad (2.2)$$

Al realizar la proyección  $s_i$  se puede determinar un triángulo rectángulo de catetos  $s_i$ ,  $d_i$  y con hipotenusa  $\|x_i\|$ , donde  $d_i$  no es más que la distancia entre  $x_i$  y su proyección, por lo tanto, al sumar sobre cada uno de los  $n$  triángulos se sigue que

$$\sum_{i=1}^n x_i^T x_i = \sum_{i=1}^n d_i^2 + \sum_{i=1}^n s_i^2$$

como las proyecciones  $s_i$  reflejan la información recopilada en  $L$  de cada  $x_i$  y lo que se busca es que dicha cantidad de información sea lo más grande posible, entonces se debe maximizar la suma de las  $s_i^2$ .

Por otra parte, observe que de (2.2) se sigue que

$$\text{var}(z_1) = \frac{1}{n} \sum_{i=1}^n (z_{i1} - \bar{z}_1)^2 = \frac{1}{n} \sum_{i=1}^n (s_i - \bar{z}_1)^2 = \frac{1}{n} \sum_{i=1}^n (s_i - 0)^2 = \frac{1}{n} \sum_{i=1}^n s_i^2$$

de modo que el hecho de maximizar la suma de los cuadrados de cada  $s_i$  significa que se está maximizando la varianza de  $z_1$ .

Nótese además lo siguiente

$$\text{var}(z_1) = \frac{1}{n} z_1^T z_1$$

de donde se sigue que

$$\text{var}(z_1) = \frac{1}{n} (\mathbb{X}u_1)^T \mathbb{X}u_1 = \frac{1}{n} u_1^T \mathbb{X}^T \mathbb{X}u_1 = u_1^T \mathbb{Y}u_1$$

El problema entonces es maximizar la varianza de  $z_1$  sujeto al hecho de que  $u_1$  es un vector unitario ( $u_1^T u_1 = 1$ ). Para resolverlo se emplean multiplicadores de Lagrange. La función a maximizar es

$$u_1^T \mathbb{Y}u_1 - \lambda_1 (u_1^T u_1 - 1) \quad (2.3)$$

Para  $i, j = 1, 2, \dots, p$ , sea  $y_{ij}$  la entrada de la matriz  $\mathbb{Y}$  en el renglón  $i$  y la columna  $j$ , observe que

$$u_1^T \mathbb{Y}u_1 = \sum_{i=1}^p \sum_{j=1}^p u_{1i} y_{ij} u_{1j}$$

entonces, derivando término a término, para  $k = 1, 2, \dots, p$ , se tiene que

$$\frac{\partial}{\partial u_{1k}} (u_1^T \mathbb{Y}u_1) = \frac{\partial}{\partial u_{1k}} \left( \sum_{i=1}^p \sum_{j=1}^p u_{1i} y_{ij} u_{1j} \right)$$

donde los casos de interés son cuando  $j = k$  y cuando  $i = k$ , de modo que

$$\frac{\partial}{\partial u_{1k}} \left( \sum_{i=1}^p \sum_{j=1}^p u_{1i} y_{ij} u_{1j} \right) = \sum_{i=1}^p u_{1i} y_{ik} + \sum_{j=1}^p y_{kj} u_{1j} = \sum_{i=1}^p y_{ik} u_{1i} + \sum_{j=1}^p y_{kj} u_{1j}$$

dado que  $\mathbb{Y}$  es una matriz simétrica, entonces  $y_{ij} = y_{ji}$ , y como los índices de las sumas son arbitrarios, entonces

$$\begin{aligned} \frac{\partial}{\partial u_{1k}} \left( \sum_{i=1}^p \sum_{j=1}^p u_{1i} y_{ij} u_{1j} \right) &= \sum_{i=1}^p y_{ik} u_{1i} + \sum_{i=1}^p y_{ki} u_{1i} \\ &= 2 \sum_{i=1}^p y_{ki} u_{1i} \end{aligned}$$

por lo tanto

$$\frac{\partial}{\partial u_1} (u_1^T \mathbb{Y} u_1) = 2\mathbb{Y} u_1$$

Entonces, al derivar la función (2.3) respecto a  $u_1$  e igualar a cero se obtiene

$$2\mathbb{Y} u_1 = 2\lambda_1 u_1$$

o bien,

$$\mathbb{Y} u_1 = \lambda_1 u_1$$

de modo que  $u_1$  es eigenvector de  $\mathbb{Y}$  cuyo eigenvalor asociado es  $\lambda_1$ .

Se puede notar que

$$u_1^T \mathbb{Y} u_1 = u_1^T \lambda_1 u_1 = \lambda_1 u_1^T u_1 = \lambda_1 \quad (2.4)$$

de modo que la varianza de  $z_1$  es  $\lambda_1$ ; al ser esta la cantidad que se quiere maximizar,  $\lambda_1$  será el mayor eigenvalor de la matriz de covarianzas  $\mathbb{Y}$ , donde  $u_1$  además de ser el eigenvector asociado a  $\lambda_1$ , también es el vector que define los coeficientes de cada variable en  $S_1$  (la primera componente), es decir,  $\alpha_{1j} = u_{1j}$  con  $j = 1, \dots, p$ .

Para un subespacio de dimensión 2, el mejor plano que ajusta las  $p$  variables es el generado por  $u_1$  y  $u_2$ , donde  $u_2$  además de ser unitario, se pide que sea ortogonal a  $u_1$  para que se cumpla que la primera componente no esté correlacionada con la segunda componente. Ahora, lo que se quiere es maximizar la varianza de  $z_2 = \mathbb{X} u_2$ , donde

$$\text{var}(z_2) = u_2^T \mathbb{Y} u_2$$

Por lo tanto, resolviendo esto como un problema de optimización mediante multiplicadores de Lagrange, bajo las condiciones  $u_2^T u_2 = 1$  y  $u_2^T u_1 = 0$ , la función que se busca maximizar es:

$$u_2^T \mathbb{Y} u_2 - \lambda_2 (u_2^T u_2 - 1) - \phi (u_2^T u_1)$$

donde  $\lambda_2$  y  $\phi$  son los multiplicadores de Lagrange correspondientes. Derivando respecto a  $u_2$  e igualando a cero, se tiene

$$2\mathbb{Y}u_2 - 2\lambda_2 u_2 - \phi u_1 = 0 \quad (2.5)$$

multiplicando por  $u_1^T$  por la izquierda

$$2u_1^T \mathbb{Y}u_2 - 2\lambda_2 u_1^T u_2 - \phi u_1^T u_1 = 0 \quad (2.6)$$

de la igualdad (2.4) se sigue que  $u_1^T \mathbb{Y} = \lambda_1 u_1^T$ , por lo tanto se tiene que

$$2u_1^T \mathbb{Y}u_2 - 2\lambda_2 u_1^T u_2 - \phi u_1^T u_1 = 2\lambda_1 u_1^T u_2 - 2\lambda_2 u_1^T u_2 - \phi u_1^T u_1$$

al ser  $u_1$  ortogonal a  $u_2$ , entonces

$$2\lambda_1 u_1^T u_2 = 2\lambda_1 \cdot 0 = 0$$

de manera similar

$$2\lambda_2 u_1^T u_2 = 2\lambda_2 \cdot 0 = 0$$

y además  $u_1$  es unitario, de modo que

$$\phi u_1^T u_1 = \phi \cdot 1 = \phi$$

entonces, de la ecuación (2.6) se tiene que

$$\phi = 0$$

por lo tanto, volviendo a la ecuación (2.5), se concluye que

$$\mathbb{Y}u_2 = \lambda_2 u_2$$

de modo que  $\lambda_2$  es el segundo eigenvalor más grande de  $\mathbb{Y}$  cuyo eigenvector asociado es  $u_2$ , con lo que  $\{u_1, u_2\}$  generan al subespacio de dimensión dos que mejor ajusta a los datos.

Extendiendo lo anterior con un algoritmo análogo, es como se obtienen las componentes que generen al subespacio de dimensión  $q$  que se busca. Como se ha visto, las  $q$  componentes están determinadas por los eigenvectores asociados a los  $q$  eigenvalores más grandes de la matriz de covarianzas de  $\mathbb{X}$  (matriz  $\mathbb{Y}$ ). En general, si  $u_j = (u_{j1}, u_{j2}, \dots, u_{jp})$  es el eigenvector asociado al  $j$ -ésimo eigenvalor más grande de  $\mathbb{Y}$ , para  $i = 1, \dots, p$ , se tiene que

$$u_{ji} = \alpha_{ji}$$

en la ecuación (2.1).

## 2.2. Regresión lineal

Cuando se tiene un muestreo aleatorio de una población y es de interés conocer la relación entre dos o más variables, una herramienta bastante útil es la regresión lineal. El objetivo de la regresión lineal es poder estimar los parámetros de una ecuación que ayude a predecir una variable, que se denotará como variable de respuesta o variable dependiente, a partir de una o más variables denominadas variables explicativas. Además, permite identificar si los cambios en la o las variables explicativas tienen un efecto sobre la variable dependiente<sup>2</sup>.

Para ilustrar la idea de la regresión lineal se presenta la figura 2.1 donde se considera el caso de una variable dependiente  $y$  con una variable explicativa  $x$ . Las observaciones de los pares ordenados  $(x, y)$ , provenientes del muestreo aleatorio, se colocan en el plano. Lo que hará la regresión lineal es calcular la recta que “mejor se ajuste” a los datos que se tienen (recta color rojo), así, para cada  $x_i$  se tendrá un valor ajustado  $\hat{y}_i$ , y con “mejor se ajuste” se hace referencia a que la diferencia  $y_i - \hat{y}_i$  en promedio sea 0.

Nombrando ahora a la variable explicativa como  $x_1$ , bajo el supuesto de que una variable aleatoria  $y$  proviene de una distribución normal, entonces el mejor modelo para predecir a esta  $y$ , dados los valores conocidos de  $x_1$  es

$$y = E(y|x_1) + u$$

donde  $u$  es una variable aleatoria que se denominará término de error, el cual representa a aquellos factores no observables que pueden afectar a la variable  $y$ . Se supone además que  $u$  tiene una distribución normal con media 0 y varianza constante  $\sigma^2$ .

En el modelo más sencillo de regresión lineal se considera únicamente a una variable explicativa, como en la figura (2.1); a este se le conoce como un modelo de regresión lineal simple. Al agregar  $k - 1$  variables explicativas o también llamadas variables de control, y bajo los supuestos anteriores, se obtiene el modelo de regresión lineal múltiple que luce así

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad (2.7)$$

donde  $y$  es la variable dependiente o de respuesta,  $\beta_0$  es un término constante,  $x_i$  son cada una de las variables independientes o explicativas (con  $i = 1, \dots, k$ ),  $\beta_i$  es el coeficiente asociado a la variable  $x_i$  y  $u$  es el término de error. De manera equivalente, se puede escribir el modelo como

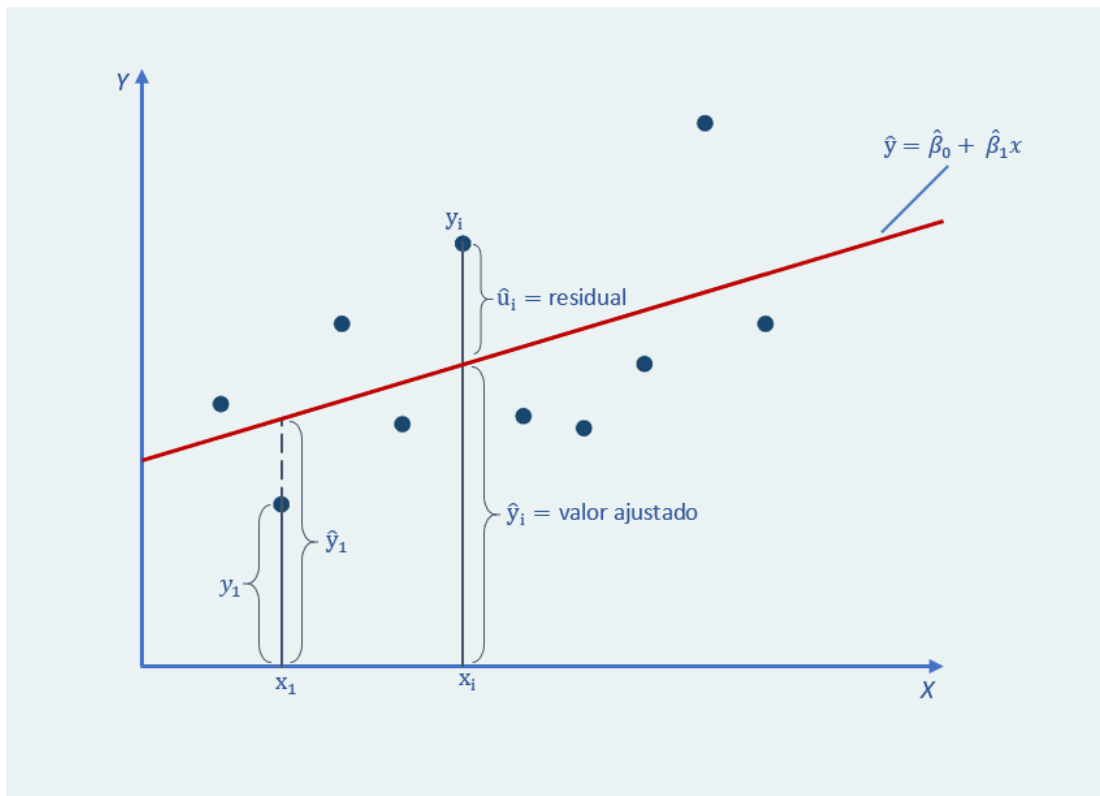
$$y = \beta_0 + \beta_1 x_1 + \alpha X + u$$

el cual resulta conveniente para distinguir a las variables de control, variables contenidas en el vector  $X$ , de la variable explicativa de interés, sin embargo, para fines de este capítulo se recurrirá principalmente al modelo escrito en la forma (2.7).

---

<sup>2</sup>Esta sección, al igual que la siguiente (Variables instrumentales), se basan en lo expuesto en el libro de Wooldridge (2010).

Figura 2.1: Regresión lineal simple



Fuente: elaboración propia.

Para obtener estimadores confiables de  $\beta_0$  y  $\beta_i$  con los datos que se están manejando, se tiene que restringir la relación entre el término de error y las variables explicativas. Se establece entonces el siguiente supuesto sobre  $u$

$$E(u) = E(u|x_1, x_2, \dots, x_k) = 0 \quad (2.8)$$

con lo que en promedio, los factores observables no afectan a  $y$ , además, si  $x = (x_1, x_2, \dots, x_k)$  se supondrá que

$$\text{cov}(x, u) = 0 \quad (2.9)$$

de modo que todo factor en  $u$  debe estar no correlacionado con  $x$ , se dice entonces que  $x_i$  es una variable explicativa exógena, mientras que si por alguna razón  $x_i$  está correlacionada con  $u$ , se le denomina variable explicativa endógena .

Al tomar el valor esperado de  $y$  dado  $x$  se tiene

$$\begin{aligned} E(y|x) &= E(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u|x) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \end{aligned}$$

observe que este valor esperado es una función lineal de  $x$ , lo que nos dice que al aumentar  $x_i$  en una unidad, el valor esperado de  $y$  se modifica en  $\beta_i$  unidades.

Mediante el método de mínimos cuadrados ordinarios (MCO) se calculan los estimadores  $\hat{\beta}_i$  para cada coeficiente  $\beta_i$  (con  $i = 0, \dots, k$ ) de la ecuación (2.7). Se define ahora al valor ajustado como

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

de modo que el residual, que es la diferencia entre el valor real y el valor ajustado, se expresa como

$$\hat{u} = y - \hat{y}$$

El método de MCO elige aquellas estimaciones que minimizan los residuales cuadrados, por lo tanto, haciendo uso de una muestra de  $n$  observaciones, la suma

$$\sum_{j=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \dots - \hat{\beta}_k x_{jk})^2 \quad (2.10)$$

es lo más pequeña posible al tomar simultáneamente los estimadores  $\hat{\beta}_i$ .

Al igual que en un problema de cálculo donde se desea obtener el mínimo, se deriva la ecuación (2.10) respecto a cada parámetro  $\hat{\beta}_i$  y se iguala a cero, obteniendo las siguientes  $k+1$  ecuaciones

$$\left. \begin{aligned} \sum_{j=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \dots - \hat{\beta}_k x_{jk}) &= 0 \\ \sum_{j=1}^n x_{j1} (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \dots - \hat{\beta}_k x_{jk}) &= 0 \\ &\vdots \\ \sum_{j=1}^n x_{jk} (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \dots - \hat{\beta}_k x_{jk}) &= 0 \end{aligned} \right\} \quad (2.11)$$

Observe que si se define  $v_j = (1, x_{j1}, \dots, x_{jk})$  y  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)^T$ , entonces el sistema anterior se puede escribir como

$$\sum_{j=1}^n v_j^T (y_j - v_j \hat{\beta}) \equiv 0$$

lo cual es equivalente a

$$(\mathbf{X}^T \mathbf{X}) \hat{\beta} = \mathbf{X}^T Y$$

donde  $v_j$  es la  $j$ -ésima fila de la matriz  $\mathbf{X}$  y  $Y = (y_1, \dots, y_n)^T$ . Así, finalmente se tiene que el vector  $\hat{\beta}$  esta determinado de la siguiente forma



$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y \quad (2.12)$$

Para que  $\hat{\beta}$  este bien definido es crucial que la matriz  $(\mathbf{X}^T \mathbf{X})$  sea invertible, lo cual requiere que las columnas de  $\mathbf{X}$  sean linealmente independientes, de modo que otro supuesto fuerte en la regresión lineal será que en la muestra no exista relación lineal entre las variables independientes y que además ninguna de éstas sea constante.

En el Apéndice A se muestra a detalle que

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \sum_{i=1}^k \hat{\beta}_i \bar{x}_i \\ \frac{cov(x_1, y)}{var(x_1)} \\ \frac{cov(x_2, y)}{var(x_2)} \\ \vdots \\ \frac{cov(x_k, y)}{var(x_k)} \end{pmatrix} \quad (2.13)$$

Escribiendo el modelo de regresión lineal en su forma matricial como

$$Y = \mathbf{X}\beta + U \quad (2.14)$$

donde el vector columna  $U$  esta formado por los términos de error asociados a cada  $y_i$ , se tiene lo siguiente

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(Y) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{X}\beta + U) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{X}\beta + E(U)] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T [\mathbf{X}\beta] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X})\beta \\ &= \beta \end{aligned}$$

de modo que  $\hat{\beta}$  es un estimador insesgado.

Volviendo al caso particular de una regresión simple (una única variable explicativa)

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + u$$

observe que

$$\begin{aligned}
 cov(y, x) &= cov(\hat{\beta}_0 + \hat{\beta}_1 x + u, x) \\
 &= cov(\hat{\beta}_0, x) + cov(\hat{\beta}_1 x, x) + cov(u, x) \\
 &= \hat{\beta}_1 cov(x, x) + cov(u, x) \\
 &= \hat{\beta}_1 var(x) + cov(u, x)
 \end{aligned}$$

por lo tanto

$$\hat{\beta}_1 = \frac{cov(y, x) - cov(u, x)}{var(x)} \quad (2.15)$$

y gracias al supuesto de exogeneidad de  $x$  indicado en (2.9), la igualdad (2.15) es consistente con (2.13) para el caso  $k = 1$ , por lo tanto  $\hat{\beta}_1$  se mantiene como estimador insesgado.

Cabe mencionar que debido a que MCO minimiza a (2.10), resulta que se está minimizando también la varianza de  $\hat{u}$ , pues observe que

$$\begin{aligned}
 var(\hat{u}) &= E(\hat{u}^2) - E(\hat{u})^2 \\
 &= E(\hat{u}^2) \\
 &= \frac{1}{n} \sum_{j=1}^n \hat{u}_j^2 \\
 &= \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2 \\
 &= \frac{1}{n} \sum_{j=1}^n (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \dots - \hat{\beta}_k x_{jk})^2
 \end{aligned}$$

donde se ha utilizado que

$$E(\hat{u}) = 0$$

hecho que se sigue directamente de la primera ecuación del sistema (2.11), pues  $\hat{u} = y - \hat{y}$ , y a su vez  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$ , con lo que

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^n \hat{u}_i = 0 \quad (2.16)$$

De (2.16) se sigue que

$$\bar{y} = \bar{\hat{y}} \quad (2.17)$$

donde  $\bar{y}$  es la media de  $y$  y  $\bar{\hat{y}}$  la media de  $\hat{y}$ . Por otra parte, por definición

$$cov(\hat{y}, \hat{u}) = \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(\hat{u}_i - \bar{\hat{u}})$$

donde  $\bar{\hat{u}}$  es la media de  $\hat{u}$ , la cual por (2.16) es cero y con (2.17) se tiene que

$$\begin{aligned}
 cov(\hat{y}, \hat{u}) &= \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})(\hat{u}_i - \bar{\hat{u}}) \\
 &= \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})\hat{u}_i \\
 &= \frac{1}{n-1} \left[ \sum_{i=1}^n \hat{y}_i \hat{u}_i - \sum_{i=1}^n \bar{\hat{y}} \hat{u}_i \right] \\
 &= \frac{1}{n-1} \left[ \sum_{i=1}^n \hat{y}_i \hat{u}_i - \bar{\hat{y}} \sum_{i=1}^n \hat{u}_i \right]
 \end{aligned}$$

por (2.16), entonces

$$\begin{aligned}
 cov(\hat{y}, \hat{u}) &= \frac{1}{n-1} \left[ \sum_{i=1}^n \hat{y}_i \hat{u}_i - \bar{\hat{y}} \cdot 0 \right] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \hat{y}_i \hat{u}_i \\
 &= \frac{1}{n-1} \sum_{i=1}^n (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}) \hat{u}_i \\
 &= \frac{1}{n-1} \left[ \hat{\beta}_0 \sum_{i=1}^n \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^n x_{i1} \hat{u}_i + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} \hat{u}_i \right]
 \end{aligned}$$

y de acuerdo a (2.11) se tiene que

$$\sum_{i=1}^n x_{ij} \hat{u}_i = 0$$

para todo  $j = 1, 2, \dots, k$ , por lo tanto

$$cov(\hat{y}, \hat{u}) = 0 \tag{2.18}$$

Observe ahora que

$$\begin{aligned}
 \text{var}(y) &= \frac{1}{n-1} \sum_{i=1}^n [y_i - \bar{y}_i]^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n [y_i - \bar{y}_i + \hat{y}_i - \hat{y}_i]^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}_i)]^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n [\hat{u}_i + (\hat{y}_i - \bar{y}_i)]^2 \\
 &= \frac{1}{n-1} \left[ \sum_{i=1}^n \hat{u}_i^2 + 2\hat{u}_i(\hat{y}_i - \bar{y}_i) + (\hat{y}_i - \bar{y}_i)^2 \right]
 \end{aligned}$$

y como

$$\text{cov}(\hat{y}, \hat{u}) = \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)(\hat{u}_i)$$

que de acuerdo a (2.18) es cero, entonces

$$\begin{aligned}
 \text{var}(y) &= \frac{1}{n-1} \left[ \sum_{i=1}^n \hat{u}_i^2 + (\hat{y}_i - \bar{y}_i)^2 \right] \\
 &= \frac{1}{n-1} \sum_{i=1}^n \hat{u}_i^2 + \frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2
 \end{aligned}$$

y por las igualdades (2.16) y (2.17), se tiene que

$$\text{var}(y) = \text{var}(\hat{u}) + \text{var}(\hat{y})$$

lo que indica que la varianza total de  $y$  se puede descomponer como la varianza explicada por  $\hat{y}$  y la explicada por  $\hat{u}$ . Por lo tanto, el hecho de haber minimizado la varianza de  $\hat{u}$  implica que se ha maximizado la varianza de  $\hat{y}$ .

Note además que

$$\begin{aligned}
 \text{cov}(y, \hat{y}) &= \text{cov}(\hat{y} + \hat{u}, \hat{y}) \\
 &= \text{cov}(\hat{y}, \hat{y}) + \text{cov}(\hat{u}, \hat{y}) \\
 &= \text{var}(\hat{y}) + \text{cov}(\hat{y}, \hat{u})
 \end{aligned}$$

por la ecuación (2.18) se sigue entonces que

$$\text{cov}(y, \hat{y}) = \text{var}(\hat{y})$$

por lo tanto, como se ha maximizado la varianza de  $\hat{y}$ , resulta que también se maximizó la covarianza entre  $y$  y  $\hat{y}$ .

### 2.3. Variables instrumentales

Una de las razones por las que una de las variables explicativas que se está considerando en el modelo (2.14) resulta no ser exógena es cuando existe simultaneidad entre dicha variable y la variable dependiente, como es el caso de las variables violencia e ingreso. Bajo esta situación, realizar una estimación con una regresión lineal generaría un efecto estimado sesgado, pues al violarse el supuesto de exogeneidad la ecuación (2.15) no es consistente con el estimador mostrado en (2.13) y entonces no hay certeza de que  $\hat{\beta}$  sea un estimador insesgado. Por lo tanto, para corregir la endogeneidad se implementa el método de variables instrumentales, el cual consiste en estimar las siguientes ecuaciones por medio de una regresión de mínimos cuadrados en dos etapas

$$\begin{aligned} y_1 &= \alpha_0 + \alpha Z + \gamma X + \epsilon \\ y &= \beta_0 + \beta_1 \hat{y}_1 + \theta X + u \end{aligned}$$

a continuación se explicará el significado de cada uno de los elementos que conforman ambas ecuaciones.

Para fijar ideas considere el modelo

$$y = \beta_0 + \beta_1 y_1 + \theta_1 x_1 + \dots + \theta_k x_k + v \quad (2.19)$$

donde  $y$  es la variable dependiente,  $v$  es el término de error con el supuesto de que tiene distribución normal con media cero y varianza constante  $\sigma^2$ ,  $\beta_0$  es una constante,  $\beta_1$  es el coeficiente de la variable explicativa  $y_1$  (la cual se sospecha que está correlacionada con  $v$ ) y  $\theta_i$  es el coeficiente asociado a la variable de control  $x_i$  ( $i = 1, \dots, k$ ), a estas  $x_i$  se les suele llamar también *instrumentos incluidos*.

Lo siguiente es considerar un conjunto de variables observables denominadas *variables instrumentales*, *instrumentos excluidos* o simplemente *instrumentos*. Sin pérdida de generalidad se supondrá que se emplean  $r$  instrumentos y se define  $Z = (x_{k+1}, \dots, x_{k+r})$  el vector de instrumentos excluidos. Estos instrumentos no van a ser variables arbitrarias, pues además de no estar presentes en el modelo (2.19), se debe cumplir que  $Z$  sea exógena en (2.19) y que sea relevante, es decir, es necesario que los instrumentos satisfagan la particularidad de no afectar a la variable dependiente  $y$ , pero sí a la variable explicativa endógena  $y_1$ , en términos matemáticos esto se escribe como

$$\text{cov}(Z, v) = 0 \quad \text{y} \quad \text{cov}(Z, y_1) \neq 0$$

Sea  $X = (x_1, \dots, x_k)$ ,  $Z$  al igual que  $X$  no están correlacionadas con  $v$ , de modo que cualquier combinación lineal de éstas tampoco estará correlacionada con  $v$ , sin embargo, hay infinitas posibilidades de formar dicha combinación lineal, no obstante, observe que al escribir a  $y_1$  como

función lineal de las variables exógenas y un término de error, se tiene una regresión lineal de la forma

$$y_1 = \alpha_0 + \alpha Z + \gamma X + \epsilon \quad (2.20)$$

donde  $\alpha$  y  $\gamma$  son vectores de coeficientes desconocidos y  $\alpha_0$  es una constante, mientras que  $\epsilon \sim N(0, \sigma^2)$ ,  $E(\epsilon) = 0$  y  $cov((X, Z), \epsilon) = 0$ . Al estimar esta regresión por MCO se obtiene la combinación lineal

$$\hat{y}_1 = \hat{\alpha}_0 + \hat{\alpha}Z + \hat{\gamma}X \quad (2.21)$$

y como se vió en la sección anterior, MCO maximiza la covarianza entre  $y_1$  y  $\hat{y}_1$ , por lo tanto  $\hat{y}_1$  no solo es una variable instrumental válida para  $y_1$ , sino que es la mejor, pues es la variable que mayor correlación tiene con  $y$ .

Hasta aquí llega la primera etapa del método, que consiste en emplear la muestra para hacer la regresión (2.20), de modo que se obtienen los valores ajustados mostrados en (2.21). La segunda etapa consiste en estimar ahora los coeficientes de la ecuación

$$y = \beta_0 + \beta_1 \hat{y}_1 + \theta X + u \quad (2.22)$$

donde  $\theta$  es el vector que contiene a cada uno de los coeficientes  $\theta_i$ ,  $u \sim N(0, \sigma^2)$ ,  $E(u) = 0$  y  $cov((\hat{y}_1, X), u) = 0$ . Se realiza entonces una regresión lineal por MCO tomando a  $y$  como variable dependiente y a  $\hat{y}_1$  y  $X$  como variables explicativas.

### 2.3.1. Validez de los instrumentos

Recuerde que los instrumentos en  $Z$  deben estar relacionados con la variable explicativa endógena  $y_1$ , pero no deben afectar a la variable dependiente  $y$ , es decir, los instrumentos cumplen las propiedades de relevancia y exogeneidad. La relevancia de  $Z$  asegura que  $y_1$  puede ser explicada a partir de los instrumentos empleados; sin relevancia, la estimación del efecto que tiene  $y_1$  sobre  $y$  será muy débil y poco fiable. Para examinar que los instrumentos excluidos satisfagan esta propiedad, se emplean dos pruebas de hipótesis, una prueba  $F$  en la primera etapa con la hipótesis nula de que en conjunto los instrumentos no tienen relación significativa con la variable endógena y la prueba de Kleibergen-Paap rank Wald cuya hipótesis nula establece que la variable endógena no puede ser descrita a partir de los instrumentos. Rechazar la hipótesis nula de cada prueba brinda evidencia de que los instrumentos excluidos son relevantes (Kleibergen y Paap 2006; Stock y Yogo 2002).

Por su parte, la exogeneidad asegura que los instrumentos no están correlacionados con el término de error  $u$ ; de no cumplirse la propiedad de exogeneidad, la estimación estará sesgada porque  $Z$  estaría capturando efectos no observables que considera la variable  $u$ . Como herramienta para evaluar dicha propiedad se emplea la prueba de Sargan - Hansen, en esta se plantea la hipótesis nula de que los instrumentos excluidos están no correlacionados con el término de error  $u$  y

están correctamente excluidos de la ecuación estimada en la segunda etapa, no poder rechazar esta hipótesis brinda validez de las variables empleadas como instrumentos en la modelación (Hayashi 2000). En conclusión, verificar que se cumplen ambas propiedades permite contar con evidencia de que los coeficientes estimados son insesgados y consistentes, los detalles de las tres pruebas de hipótesis mencionadas se describen en el Apéndice B.





## Capítulo 3

# Datos y método

Este capítulo se organiza de la siguiente manera, primero se habla acerca de la ENDIREH, de la organización de su base de datos y de cómo se ha delimitado la muestra de estudio para el análisis realizado en este trabajo. Después, se describe la construcción de los índices para cuantificar la intensidad de la violencia. Posteriormente se describe la construcción de las variables implementadas para el modelado y enseguida se presenta la estadística descriptiva sobre la muestra de estudio. Finalmente, se explican los modelos propuestos para analizar la relación entre ingresos y violencia.

### 3.1. Muestra de estudio

Los datos utilizados para este análisis provienen de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) 2021, ésta forma parte de una colección estadística que brinda información sobre la situación de violencia contra las mujeres en México. Dicha encuesta es levantada por el Instituto Nacional de Estadística y Geografía (INEGI) y tuvo su origen en 2003, tras su éxito se tuvo una segunda edición en 2006 y a partir de entonces se genera cada cinco años, teniendo a la encuesta de 2021 como la más reciente de la serie.

La ENDIREH ofrece información sobre las distintas experiencias de violencia que han enfrentado las mujeres seleccionadas<sup>1</sup> de 15 años y más en cinco distintos ámbitos de su vida: escolar, laboral, comunitario, familiar y de pareja. La encuesta también recopila información sobre las personas agresoras, así como los lugares donde ocurrieron las situaciones de violencia y variables complementarias para profundizar en el entendimiento de esta problemática de la violencia contra la mujer, tales como los antecedentes de violencia en la familia de las mujeres y de sus parejas.

Toda la información captada de la encuesta se distribuye en cuatro diferentes bases de datos:

---

<sup>1</sup>La ENDIREH selecciona a una de las mujeres de 15 años y más por vivienda para verificar su situación conyugal actual e indagar sobre sus experiencias de violencia.

- **TVIV**: informa sobre las características básicas de la vivienda, los servicios y bienes de los que esta dispone, además del número de personas que residen habitualmente ahí y los grupos de personas que comparten los gastos para alimentación.
- **TSDem**: almacena las características sociodemográficas, económicas y culturales de cada una de las personas residentes de la vivienda.
- **TB\_SEC\_III**: permite identificar a las mujeres de 15 años y más que fueron seleccionadas para verificar su situación conyugal actual.
- **TB\_SEC\_IVaVD**: contiene toda la información asociada a la situación de violencia de las mujeres de 15 años y más que resultaron seleccionadas.

Dado que este trabajo consiste en analizar la violencia de pareja y su relación con el ingreso laboral, se restringe la muestra de estudio a mujeres trabajadoras de 15 años y más que están casadas o viven en unión libre y cuya pareja actual es residente en el hogar; con estas especificaciones la muestra contiene 24,549 registros que, al considerar el factor de expansión, representan a un total de 9,537,807 mujeres casadas o unidas.

## 3.2. Estimación de los índices de violencia

Con el fin de indagar, de manera desglosada y con un lenguaje comprensible para las mujeres entrevistadas, sobre los actos de los que son víctimas, la ENDIREH presenta preguntas por bloques enfocadas a un tipo de violencia en específico. En la Tabla 3.1 se presentan las preguntas de interés que se emplean para construir los índices de violencia, estas se refieren a los hechos ocurridos en los últimos doce meses relativo al momento de aplicación de la encuesta. Las respuestas se encuentran en escala tipo Likert, de modo que para cada pregunta se crea una variable categórica a la cual se le asigna el valor cero si la respuesta sobre la ocurrencia del evento fue “no ocurrió”, se le asigna el valor uno si la respuesta fue “una vez” y, si la mujer encuestada respondió “muchas veces” o “pocas veces” se asigna el valor dos.

En este trabajo se considera que las respuestas sobre la violencia no deben tener el mismo peso como lo hace Aguirre (2023). Por ejemplo, no parece ser lo que mismo cuando una mujer responde que una vez su pareja “la ha amenazada con algún arma (cuchillo, navaja, pistola o rifle) o con quemarla”, al hecho de que una vez su pareja “le ha destruido, tirado o escondido cosas”, dado que estas preguntas se encuentran categorizadas en el mismo tipo de violencia no deberían ponderar de la misma forma dentro del índice de violencia. Por lo tanto, para crear los índices de violencia se emplea el método de componentes principales visto en el Capítulo 2, gracias a este método se logra identificar cinco tipos de violencia: económica, emocional, acoso, física y sexual. En la Tabla 3.1, además de las preguntas que engloban cada uno de los índices construídos, se presenta también el peso (cargas factoriales) que tiene cada respuesta en la combinación lineal.

Tabla 3.1: Variables que incluye cada uno de los índices de violencia y sus respectivas cargas factoriales.

Relación actual- De octubre de 2020 a la fecha, ¿esto ha ocurrido...	Cargas factoriales
<b>Violencia Económica</b>	
1.- le ha prohibido trabajar o estudiar?	0.202
2.- se ha adueñado o le ha quitado bienes (terrenos, casas, departamento, coche, etcétera)?	0.133
3.- se ha gastado el dinero que se necesita para la casa?	0.299
4.- no ha cumplido con dar el gasto o ha amenazado con no darlo?	0.289
5.- aunque tenga dinero ha sido codo o tacaño con los gastos de la casa?	0.310
6.- le ha reclamado por cómo gasta usted el dinero?	0.279
<b>Violencia Emocional</b>	
7.- la ha avergonzado, ofendido, menospreciado o humillado (le ha dicho que es fea o la ha comparado con otras mujeres)?	0.242
8.- la ha ignorado, no la toma en cuenta o no le brinda cariño?	0.250
9.- le ha dicho que usted lo engaña?	0.174
10.- le ha hecho sentir miedo?	0.130
11.- la ha amenazado con dejarla/abandonarla, dañarla, quitarle a los(as) hijos(as) o correrla de la casa?	0.203
12.- le ha dejado de hablar?	0.262
13.- ha hecho que los hijos(as) o parientes se pongan en su contra?	0.098
14.- se ha enojado mucho porque no está listo el quehacer, porque la comida no está como él quiere o cree que usted no cumplió con sus obligaciones?	0.206
<b>Acoso</b>	
15.- la ha encerrado, le ha prohibido salir o que la visiten?	0.230
16.- la ha vigilado, espiado, la ha seguido cuando sale de su casa o se le aparece de manera sorpresiva?	0.216
17.- la ha amenazado con algún arma (cuchillo, navaja, pistola o rifle) o con quemarla?	0.426
18.- la ha amenazado con matarla, matarse él o matar a los niños(as)?	0.405
19.- le ha destruido, tirado o escondido cosas de usted o del hogar?	0.251
<b>Violencia Física</b>	
20.- la ha empujado o le ha jalado el cabello?	0.379
21.- la ha pateado?	0.278
22.- le ha aventado algún objeto?	0.270

23.-	la ha golpeado con el puño o con algún objeto?	0.394
24.-	la ha tratado de ahorcar o asfixiar?	0.229
Violencia Sexual		
25.-	le ha exigido con amenazas o chantajes tener relaciones sexuales, aunque usted no quiera?	0.400
26.-	cuando tienen relaciones sexuales la ha obligado a hacer cosas que a usted no le gustan?	0.432
27.-	ha usado su fuerza física para obligarla a tener relaciones sexuales?	0.422

Nota: cada línea indica la pregunta que le hicieron a la mujer. Para cada pregunta se creó una variable a la cual se le asigna valor uno si la respuesta fue “una vez”, se le asigna valor dos si la respuesta fue “muchas veces” o “pocas veces” y se asigna valor cero si la respuesta fue “no ocurrió”. Las cargas factoriales corresponden a los pesos estimados mediante el método de componentes principales para realizar un índice por cada tipo de violencia de pareja, para ello se utilizó el software Stata, versión 14.

Fuente: elaboración propia con información de la ENDIREH 2021.

### 3.3. Variables para el modelado

En busca de obtener información más clara y concisa, la ENDIREH ha tenido cambios en el cuestionario de algunas de sus ediciones, no obstante, también se han perdido métricas relevantes. En la primera edición (ENDIREH 2003) para recopilar datos acerca de los ingresos, se consideraba la pregunta “Habitual o regularmente, ¿cuántas horas trabaja cada día de la semana?”; en ediciones posteriores se ha omitido, de modo que al no contar con el número de horas trabajadas, para el análisis sólo es posible identificar el ingreso mensual que reciben las mujeres debido al trabajo realizado. Para estimar este ingreso mensual se emplean las preguntas “¿Cuánto gana o recibe usted por su trabajo?” y “¿Cada cuándo?”, tal como lo indica la Tabla 3.2.

Para el modelado se han considerado algunas variables de control que incluyen datos de la mujer tales como su edad, escolaridad, si tiene hijos y si habla alguna lengua indígena; además de datos sobre el hogar como el tamaño, el tipo de localidad de la residencia, el nivel socioeconómico y el grado de acceso a servicios de telecomunicaciones, siendo estos dos últimos un par de índices creados también a través del método de componentes principales. De igual manera, en la Tabla 3.2 se detalla la construcción de estas variables de control.

Tabla 3.2: Construcción de variables para el modelado.

Variable	Fuente de información	Construcción
Dependiente		
Ingresos mensuales	TB_SEC_IVaVD	Se obtiene a partir de las preguntas “¿cuánto gana o recibe usted por su trabajo?” y “¿Cada cuándo?”.

Independientes		
Tamaño del hogar	TSDem	Variable creada a partir de diferenciar a los individuos mediante el número de identificación de la vivienda y el número de hogar de la seleccionada.
Edad de la mujer	TSDem	Variable obtenida de la pregunta “¿Cuántos años cumplidos tiene (NOMBRE)?”
Años de escolaridad	TSDem	De la pregunta “¿Hasta qué año o grado aprobó (NOMBRE) en la escuela?” la encuesta genera las variables NIV y GRA que registran el nivel de escolaridad (primaria, secundaria, etc.) y el último año aprobado, respectivamente; con estas variables se contruye el total de años de escolaridad del individuo.
1 = Pertenece a un área urbana	TB_SEC_IVaVD	Indicadora con valor uno si en la encuesta se identifica a la mujer como parte de un dominio urbano o complemento urbano y cero si pertenece a un dominio rural.
1 = Tiene al menos un hijo	TB_SEC_IVaVD	Variable dicotómica con valor uno si en la pregunta “En total, ¿cuántas hijas e hijos nacidos vivos ha tenido?” la respuesta es un número entre 1 y 25, con valor cero si la respuesta fue “Ninguna(o)”.
1 = Habla lengua indígena	TSDem	Variable indicadora con valor uno si en la pregunta “¿(NOMBRE) habla algún dialecto o lengua indígena?” la respuesta es “Sí”, cero si respondió “No”.
Nivel socioeconómico del hogar	TVIV	Es un índice creado mediante componentes principales, se incluyen preguntas sobre condiciones de la vivienda, tales como el tipo de piso de la vivienda y si la vivienda cuenta con radio, televisión, refrigerador, lavadora, automóvil, agua entubada dentro de la vivienda y drenaje conectado a la red pública.
Servicios de telecomunicaciones	TVIV	Índice creado utilizando el método de componentes principales, se incluyen datos acerca de si la vivienda cuenta con teléfono fijo, computadora y servicio de internet.

Nota: TB\_SEC\_IVaVD es la base de datos que recopila información de la encuesta realizada a las mujeres seleccionadas de 15 años y más. TSDem se refiere a la base de datos con información de las características sociodemográficas de residentes de la vivienda. TVIV hace referencia a la base de datos que recopila la información de las características de la vivienda y hogares en la vivienda. Todas las bases antes mencionadas forman parte de los datos recopilados de la ENDIREH (2021).

### 3.4. Estadística descriptiva de la muestra

Para ver cómo es que cada tipo de violencia afecta de manera diferenciada a las mujeres, se presenta la Tabla 3.3, que muestra los valores medios de los índices de violencia estimados para ciertos grupos poblacionales. De acuerdo a la construcción de los índices de violencia, una mayor

magnitud es señal de mayor nivel de violencia. Se observa que las mujeres en los rangos de 15 a 29 años y 40 a 49 años son quienes sufren en mayor medida de violencia emocional, además de que aquellas mujeres de 15 a 29 años son también las principales víctimas de violencia física y acoso, por su parte, las mujeres en el rango de edad de 40 - 49 años son además quienes sufren en mayor medida violencia económica y sexual. Por otro lado, las mujeres de 60 años y más son las que experimentan en menor medida de todos los tipos de violencia.

Referente a la escolaridad, aquellas mujeres con estudios de secundaria que corresponden al 31.57 % de la muestra son las que padecen en mayor medida cuatro de los cinco tipos de violencia, estas son económica, emocional, sexual y acoso, sólo las mujeres con estudios de a lo más primaria son las que experimentan en mayor grado la violencia física. Por lo que corresponde al tipo de localidad de residencia, las mujeres en las zonas rurales son más propensas a someterse a violencia física, sexual y acoso, mientras que las mujeres de zonas urbanas padecen en mayor medida de violencia económica y emocional.

En cuanto al tener o no hijo(s), aquellas mujeres que no son madres sufren en menor medida todos los tipos de violencia. Con respecto al nivel socioeconómico, se encuentra que las mujeres de hogares muy pobres (nivel socioeconómico muy bajo) son quienes sufren mayormente los cinco tipos de violencia considerados. Esto evidencia que la violencia se encuentra presente en mayor grado en grupos poblacionales con determinadas características.

Tabla 3.3: Composición de la muestra y valores medios de los distintos tipos de índices de violencia por grupos poblacionales.

	Composición		Valores medios			
	(%)	Violencia Económica	Violencia Emocional	Acoso	Violencia Física	Violencia Sexual
Edades (en años)						
15 a 29	19.54	0.01	0.03	0.03	0.03	-0.01
30 a 39	25.42	0.00	0.01	0.02	-0.01	-0.01
40 a 49	21.72	0.03	0.03	0.02	0.00	0.04
50 a 59	17.17	0.00	-0.01	-0.01	0.00	0.00
60 y más	16.15	-0.09	-0.11	-0.04	-0.01	-0.06
Escolaridad						
A lo más primaria	27.91	0.01	-0.02	0.02	0.05	0.01
Secundaria	31.57	0.02	0.03	0.04	0.02	0.02
Bachillerato	20.48	-0.02	0.01	-0.01	-0.03	-0.02
Profesional o posgrado	20.04	-0.05	-0.03	-0.04	-0.08	-0.04
Tipo de localidad						
Área Rural	27.07	-0.02	-0.04	0.02	0.01	0.01

Área Urbana	72.93	0.00	0.01	0.00	0.00	-0.01
Hijo(s)						
No tiene	7.67	-0.10	-0.10	-0.01	-0.06	-0.05
Tiene al menos uno	92.33	0.00	0.01	0.01	0.00	0.00
Nivel Socioeconómico del Hogar						
Muy bajo	25.01	0.05	0.03	0.05	0.08	0.03
Bajo	28.18	-0.01	0.01	0.00	0.00	-0.01
Medio	22.59	-0.02	-0.01	-0.02	-0.03	-0.03
Alto	24.22	-0.04	-0.03	-0.02	-0.05	-0.01

Notas: En la categoría de Primaria en la sección de escolaridad se han agrupado a aquellas mujeres que a lo más terminaron la primaria, mientras que para las otras categorías (Secundaria, Bachillerato, Profesional y Posgrado) se han agrupado a mujeres que cursaron al menos un grado del nivel correspondiente. Las categorías que se presentan en la sección de nivel socioeconómico del hogar corresponden a los cuartiles de la distribución del índice. Utilizando el software Stata, versión 14, se estiman los valores medios de los índices de violencia para ciertos grupos poblacionales. La columna Composición muestra la composición de la muestra de estudio. Los índices están contruidos de tal forma que una mayor magnitud es señal de mayor nivel de violencia.

Fuente: elaboración propia con información de la ENDIREH 2021.

### 3.5. Modelos

Un modelo que estudie la relación entre los ingresos de pareja y cada tipo de violencia de pareja es el que se muestra a continuación:

$$y_i = \beta_0 + \beta_1 IV_i^j + \alpha X_i + u_i \quad (3.1)$$

donde  $IV_i^j$  es el índice de violencia del tipo  $j$  recibido por la mujer  $i$ , con  $j = \{\text{Económica, Emocional, Acoso, Física, Sexual}\}$  y  $X_i$  almacena variables de control, pues es un vector fila de características observables tales como escolaridad de la mujer, su edad y el cuadrado de su edad, nivel socioeconómico y tamaño del hogar, grado de acceso a servicio de telecomunicaciones e indicadores del tipo de localidad de residencia (rural o urbana), si la mujer tiene al menos un hijo y si la mujer habla alguna lengua indígena. Por su parte,  $u_i$  es un término de error, el cual es una variable aleatoria que tiene como supuestos una distribución normal con media cero, varianza constante y que su covarianza tanto con el vector  $X_i$  como con el índice de violencia  $IV_i^j$  sean cero. De modo que la variable dependiente  $y_i$ , que no es más que el logaritmo del ingreso mensual de la mujer  $i$ , está determinada por  $IV_i^j$ , por las variables de control  $X_i$  y por los factores no observables que considera  $u_i$ , siendo  $\beta_1$  el parámetro de interés, pues este coeficiente representa el cambio en los ingresos por cambio unitario en la medida de la violencia del tipo  $j$  cuando las variables de control permanecen constantes.

No obstante, por lo visto en el Capítulo 1, existe la posibilidad de simultaneidad entre las variables ingreso y violencia porque el nivel de los ingresos puede afectar el grado de violencia hacia las mujeres (Angelucci, 2008; Bobonis et al., 2013) y, el nivel de violencia puede afectar los

ingresos recibidos (Aguirre 2023). Por lo tanto, se está violando el supuesto sobre  $u_i$  de que su covarianza con el índice de violencia  $IV_i^j$  sea cero, así que emplear el método de regresión lineal múltiple llevaría a que los coeficientes estimados del modelo sean sesgados e inconsistentes.

Para corregir la endogeneidad de las variables de ingreso y violencia se emplea el método de variables instrumentales descrito en el Capítulo 2, donde el principal reto para poder implementarlo es encontrar un vector de instrumentos válidos, variables relacionadas con los niveles de violencia que no tengan un efecto directo sobre el ingreso de la mujer. Alonso-Borrego y Carrasco (2017) analizan la relación simultánea entre la condición de empleo y el riesgo de violencia doméstica para el caso de España, para ello emplean como instrumento la información geográfica sobre el empleo y las tasas de desempleo por género y edad. Usando datos de Turquía, Dildar (2021) toma en cuenta la endogeneidad entre el empleo y la violencia empleando como instrumento al promedio de la situación laboral de las mujeres por provincia. Por su parte, Fajardo-Gonzalez (2021) emplea como instrumento una indicadora de que el esposo fue maltratado o golpeado regularmente por sus padres o padrastros para estudiar la relación entre la violencia y la probabilidad de emplearse para el caso de mujeres de Colombia.

Los instrumentos que aquí se emplean son los antecedentes de violencia durante la infancia tanto de la mujer como de su pareja, además de los apoyos con los que cuenta la mujer ajeno a su pareja. Específicamente se consideran las siguientes variables: una indicadora de que la pareja de la mujer sufría violencia familiar durante su infancia, una indicadora de que durante la infancia de la pareja de la mujer su papá le pegaba a su mamá, un índice de los antecedentes de violencia durante la infancia de la mujer y un índice que mide el nivel de apoyo que recibe la mujer independiente del recibido por parte de su pareja, la construcción de estas variables se detalla en la Tabla 3.4.

Los antecedentes de violencia son importantes pues ocurre una transmisión intergeneracional de la violencia (Knaul y Ramírez 2003; Martin et al. 2002). Mientras que si la mujer puede recibir apoyo de terceros, ya sea por parte de familiares, amigos o vecinos, hace que no dependa al cien por ciento de su pareja, con lo que es factible la amenaza de abandonar el hogar y en consecuencia tiene efectos sobre el nivel de violencia. La hipótesis es que los antecedentes de violencia durante la infancia propician ejercer, para el caso de la pareja, y tolerar, para el caso de la mujer, la violencia en la etapa adulta y el nivel de apoyo recibido fuera del hogar afectan el nivel de violencia, pero una vez controlando por características individuales y del hogar, estos antecedentes y apoyos no tienen relación sobre el nivel de ingresos recibidos.

De modo que las dos ecuaciones a estimar con el método de variables instrumentales son las siguientes

$$IV_i^j = \alpha_0 + \alpha Z_i + \gamma X_i + \epsilon_i \quad (3.2)$$

$$y_i = \beta_0 + \beta_1 \hat{IV}_i^j + \theta X_i + u_i \quad (3.3)$$



donde  $Z_i$  es el vector de los instrumentos excluidos que se han considerado para la implementación del método y  $\epsilon_i$  al igual que  $u_i$  satisfacen los mismos supuestos que en la Sección 2.3. Todas las estimaciones incluyen los pesos muestrales que se especifican en la encuesta para la mujer elegida dentro del hogar.

En el siguiente capítulo se presentan los resultados de la estimación del modelo sin considerar endogeneidad de las variables y los resultados por variables instrumentales, así como las respectivas pruebas de la validez de los instrumentos.

Tabla 3.4: Construcción de los instrumentos.

Variable	Construcción
1 = La pareja de la mujer sufría violencia familiar durante la infancia	Variable dicotómica que vale uno si se contestó “de vez en cuando” o “seguido” a la pregunta “Cuando su pareja o esposo era niño (hasta antes de cumplir 15 años), ¿le pegaban o insultaban en su casa...”, con valor cero si la respuesta fue “No le pegaban ni lo insultaban”.
1 = Durante la infancia de la pareja de la mujer, su papá le pegaba a su mamá	Variable indicadora que vale uno si se contestó “Sí le pegaba” a la pregunta “¿Sabe si cuando su pareja o esposo era niño (hasta antes de cumplir 15 años), a la mamá de él le pegaba su marido?” y vale cero si respondió “No le pegaba”.
Antecedentes de violencia durante la infancia de la mujer	Es un índice que se forma con el método de componentes principales, incluye preguntas relacionadas con situaciones que pudieron ocurrirle a la mujer durante su infancia (hasta antes de cumplir 15 años), entre dichas situaciones están si le tocaron sus partes íntimas o la obligaron a tocar las partes íntimas de otra persona sin su consentimiento, si la obligaron a mostrar sus partes íntimas y/o a mirar las partes íntimas de otra persona, si la obligaron a mirar escenas o actos sexuales o pornográficos, si intentaron forzarla a tener relaciones sexuales, si la obligaron a tener relaciones sexuales bajo amenazas o usando la fuerza y si la obligaron a realizar actos sexuales a cambio de dinero o regalos.
Ajeno a su pareja, la mujer cuenta con apoyo para el hogar	A partir de la cuestión “En caso de que usted necesitara ayuda, sin incluir a su esposo o pareja, ¿a quién o quiénes recurriría para...” se consideran las preguntas “cuidar un rato a sus hijas/hijos cuando tiene alguna emergencia o se enferman?”, “hacer alguna tarea o labor?” y “cuando usted se enferma?”, se crea una variable para cada pregunta, asignando valor uno si la respuesta fue “Vecina(o)”, “Amiga(o)”, “Compañera(o)”, “Algún familiar” u “Otra persona”, con estas tres variables y mediante el método de componentes principales se crea este índice.

Nota: la información necesaria para la creación de estas cuatro variables se obtiene de la base de datos que recopila información de la encuesta realizada a las mujeres seleccionadas de 15 años y más, correspondiente a la ENDIREH (2021).

## Capítulo 4

# Resultados

En este capítulo se presenta la interpretación de las estimaciones obtenidas del modelo de regresión lineal, correspondiente al caso donde no se considera la endogeneidad de las variables de ingreso y violencia. Posteriormente, se exhiben los resultados de la primera y segunda etapa del modelo de variables instrumentales y las pruebas de hipótesis de validez de los instrumentos implementados. Al final del capítulo se realiza una estimación del costo agregado de la violencia.

### 4.1. Omitiendo la endogeneidad de la violencia: Modelo de regresión lineal

La estimación del efecto de la violencia sobre el ingreso laboral sin corregir por la endogeneidad de la violencia se presenta en la Tabla 4.1, se trata de los coeficientes estimados de la ecuación (3.1) para cada tipo de violencia. En la columna (1) se presenta el coeficiente estimado del modelo que considera la violencia económica como una variable explicativa; seguido de la violencia emocional, después el acoso, posteriormente la violencia física y finalmente la columna (5) corresponde al modelo que toma a la violencia sexual como variable explicativa.

Se observa que el aumento de una desviación estándar (DE) en la violencia económica y en la violencia emocional reduce, en promedio, el ingreso laboral mensual de la mujer en 1.1% y 1.2%, respectivamente. Los efectos estimados obtenidos son parecidos en magnitud a los encontrados por Aguirre (2023), pues encontró que al aumentar una DE en la violencia económica y en la violencia emocional se reduce el ingreso laboral en promedio en 0.9% y 1.5%, respectivamente. Para los demás tipos de violencia considerados no se encuentra un efecto estadísticamente significativo sobre el ingreso laboral. En cuanto a las variables de control incluidas en la modelación se observan comportamientos usuales, es decir, aquellas mujeres de hogares más pequeños, con mayor edad, con más años de escolaridad, que viven en zonas urbanas, sin hijos, con mayor nivel socioeconómico y con servicios de telecomunicación son las que obtienen en promedio mayores ingresos laborales.

Tabla 4.1: Coeficientes estimados de la relación ingresos-violencia sin corregir por la endogeneidad de la violencia.

	(1)	(2)	(3)	(4)	(5)
Violencia Económica	-0.0112** (0.0054)				
Violencia Emocional		-0.0115*** (0.0042)			
Acoso			0.0040 (0.0071)		
Violencia Física				0.0012 (0.0050)	
Violencia Sexual					-0.0013 (0.0054)
Tamaño del hogar	-0.0212*** (0.0058)	-0.0213*** (0.0058)	-0.0215*** (0.0059)	-0.0215*** (0.0059)	-0.0215*** (0.0059)
Edad de la mujer	0.0185*** (0.0045)	0.0185*** (0.0045)	0.0183*** (0.0045)	0.0183*** (0.0045)	0.0184*** (0.0045)
Edad de la mujer al cuadrado	-0.0002*** (0.0001)	-0.0002*** (0.0001)	-0.0002*** (0.0001)	-0.0002*** (0.0001)	-0.0002*** (0.0001)
Años de escolaridad	0.0669*** (0.0029)	0.0669*** (0.0029)	0.0670*** (0.0029)	0.0670*** (0.0029)	0.0670*** (0.0030)
1 = Pertenece a un área urbana	0.120*** (0.0295)	0.120*** (0.0294)	0.119*** (0.0295)	0.119*** (0.0296)	0.119*** (0.0295)
1 = Tiene al menos un hijo	-0.149*** (0.0275)	-0.148*** (0.0275)	-0.150*** (0.0275)	-0.150*** (0.0276)	-0.150*** (0.0276)
1 = Habla lengua indígena	-0.062 (0.0445)	-0.062 (0.0445)	-0.0611 (0.0449)	-0.0609 (0.0450)	-0.0609 (0.0450)
Nivel socioeconómico del hogar	0.152*** (0.0145)	0.153*** (0.0146)	0.153*** (0.0146)	0.153*** (0.0147)	0.153*** (0.0146)
Servicios de telecomunicaciones	0.0980*** (0.0124)	0.0984*** (0.0123)	0.0988*** (0.0123)	0.0989*** (0.0123)	0.0989*** (0.0123)
Constante	7.471*** (0.1040)	7.473*** (0.1040)	7.476*** (0.1050)	7.476*** (0.1050)	7.476*** (0.1050)
R-cuadrado	0.274	0.274	0.274	0.274	0.274
Observaciones	24,549	24,549	24,549	24,546	24,546

Nota: se presentan los coeficientes estimados de la ecuación (3.1). Los errores estándar se reportan entre paréntesis. Coeficiente significativo al \*\*\*1%, \*\*5%, \*10%. Empleando el software Stata, versión 14, los modelos se estimaron con errores estándar robustos y agrupados por estado.

Fuente: elaboración propia con información de la ENDIREH 2021.

## 4.2. Corrigiendo por la endogeneidad de la violencia: Modelo con variables instrumentales

Se emplea el método de variables instrumentales descrito por las ecuaciones (3.2) y (3.3) para abordar la problemática de la endogeneidad de las variables de ingreso y violencia. En la primera etapa del método se describe a cada tipo de violencia a partir de los instrumentos que se han considerado para la modelación, esto se logra estimando los coeficientes de la ecuación (3.2). Dichos coeficientes se presentan en la Tabla 4.2, donde en cada columna se presenta el modelo para cada tipo de violencia, es decir, se pueden observar aquellos factores que están asociados con la violencia recibida por las mujeres que se encuentran en alguna situación de relación de pareja.

De la Tabla 4.2 se puede observar que el tamaño del hogar tiene un efecto significativo únicamente en el modelo que considera al índice de violencia económica como variable dependiente, se indica que al aumentar un integrante en el tamaño del hogar el nivel de violencia económica aumenta en 0.02 DE. El incremento de la edad de la mujer se asocia con mayores niveles de violencia económica, sexual y acoso, siendo los incrementos de 0.01, 0.02 y 0.01 DE, respectivamente. Por otra parte, los años de escolaridad de la mujer presentan efectos estadísticamente significativos solamente en la violencia sexual; la relación entre estas variables es negativa, es decir, las mujeres con mayor cantidad de años de escolaridad tienen, en promedio, un nivel de violencia sexual más bajo. El tipo de localidad afecta tanto a la violencia económica como a la violencia emocional, se observa que aquellas mujeres cuya vivienda pertenece a una zona urbana sufren en mayor medida de violencia económica y emocional que las mujeres pertenecientes a zonas rurales.

Se observa también que las mujeres que tienen al menos un hijo padecen en mayor grado de violencia económica, emocional, física y acoso que aquellas mujeres que no tienen hijos. En cuanto a las mujeres hablantes de alguna lengua indígena, únicamente se tienen efectos significativos en un tipo de violencia; en promedio, padecen un menor nivel de violencia económica que las mujeres que no hablan lengua indígena. Respecto a aquellas mujeres de niveles socioeconómicos más altos, sufren en menor medida de violencia económica, emocional, física y acoso. Finalmente, las mujeres con un mayor acceso a servicios de telecomunicaciones padecen menores niveles de violencia económica, emocional, física y sexual.

En términos de los instrumentos excluidos, se observan efectos significativos que asocian de manera positiva a los antecedentes de violencia con el nivel de violencia. Si la pareja de la mujer sufría de violencia familiar hasta antes de los 15 años se estima un aumento promedio de 0.21 DE en el nivel de violencia económica que sufre la mujer, en 0.29 DE en la violencia emocional, 0.14 DE en la violencia física y en 0.04 DE en la violencia sexual, sin embargo, para el acoso no se obtiene un efecto estadísticamente significativo. Mientras que el hecho de que durante la infancia de la pareja de la mujer, a su mamá le pegaba su marido se asocia con efectos significativos que corresponden a incrementos promedio para todos los tipos de violencia considerados, estos

aumentos van desde 0.09 DE, correspondiente al acoso, hasta 0.28 DE en el caso de la violencia emocional. De modo que si la pareja de la mujer era víctima de violencia familiar cuando era un niño o si su mamá experimentó violencia de pareja, se incrementan los niveles de violencia que la pareja provoca a la mujer, es decir, existe una transmisión intergeneracional de la violencia; de hecho, esto concuerda con lo encontrado por Martin et al. (2002) y Knaul y Ramírez (2003).

Para los antecedentes de violencia durante la infancia de la mujer se encuentran efectos estadísticamente significativos en todos los tipos de violencia que se han considerado. El efecto más grande se presenta en el modelo que considera la violencia emocional como variable dependiente, se tiene que este tipo de violencia aumenta en promedio en 0.41 DE al aumentar una DE en los antecedentes de violencia de la mujer durante la infancia. Mientras que el efecto más pequeño se observa en el modelo correspondiente a la violencia sexual, con una magnitud de 0.16 DE. Lo cual ratifica el patrón de que las mujeres que fueron víctimas de violencia durante su etapa infantil siguen sufriendo violencia en su etapa adulta (Knaul y Ramírez 2003).

Finalmente, con respecto al apoyo para el hogar que la mujer recibe, ajeno a su pareja, únicamente se encuentra efecto estadísticamente significativo en la violencia física, la cual se reduce en promedio en 0.04 DE al incrementar en una DE los apoyos que recibe la mujer.

Tabla 4.2: Coeficientes estimados de la primera etapa del método de variables instrumentales

	Violencia Económica	Violencia Emocional	Acoso	Violencia Física	Violencia Sexual
<b>Instrumentos excluidos</b>					
1 = La pareja de la mujer sufría violencia familiar durante la infancia	0.212*** (0.0461)	0.291*** (0.0363)	-0.0138 (0.0230)	0.138*** (0.0261)	0.0403** (0.0164)
1 = Durante la infancia de la pareja de la mujer, su papá le pegaba a su mamá	0.211*** (0.0587)	0.276*** (0.0420)	0.0894** (0.0354)	0.180*** (0.0434)	0.0952** (0.0387)
Antecedentes de violencia durante la infancia de la mujer	0.334*** (0.0337)	0.414*** (0.0528)	0.186*** (0.0390)	0.279*** (0.0458)	0.156*** (0.0549)
Ajeno a su pareja, la mujer cuenta con apoyo para el hogar	-0.0139 (0.0106)	-0.0078 (0.0134)	-0.0137 (0.0106)	-0.0406*** (0.0152)	0.0129 (0.0169)
<b>Instrumentos incluidos</b>					
Tamaño del hogar	0.0169** (0.0068)	0.0117 (0.0092)	-0.0112 (0.0090)	0.0021 (0.0041)	0.0068 (0.0078)
Edad de la mujer	0.0103* (0.0053)	0.0025 (0.0047)	0.0119*** (0.0040)	-0.0022 (0.0067)	0.0174*** (0.0038)
Edad de la mujer al cuadrado	-0.00014** (0.0001)	-0.0001 (0.0001)	-0.00017*** (0.0000)	0.0000 (0.0001)	-0.0002*** (0.0000)
Años de escolaridad	-0.00315 (0.0037)	-0.00243 (0.0044)	-0.0050 (0.0036)	-0.0059 (0.0047)	-0.0091** (0.0043)
1 = Pertenece a un área urbana	0.116** (0.0499)	0.0781* (0.0424)	0.0194 (0.0404)	0.0324 (0.0456)	0.0082 (0.0458)
1 = Tiene al menos un hijo	0.115*** (0.0257)	0.110** (0.0450)	0.0690*** (0.0234)	0.0511* (0.0281)	0.0226 (0.0142)
1 = Habla lengua indígena	-0.154***	-0.1310	0.0306	-0.0442	-0.0637

	(0.0530)	(0.0872)	(0.0715)	(0.0529)	(0.0747)
Nivel socioeconómico del hogar	-0.0911***	-0.0394*	-0.0672**	-0.0503**	-0.0442
	(0.0272)	(0.0206)	(0.0325)	(0.0253)	(0.0405)
Servicios de telecomunicaciones	-0.0610***	-0.0326**	-0.0097	-0.0489***	0.00745
	(0.0137)	(0.0127)	(0.0190)	(0.0143)	(0.0190)
Constante	-0.478***	-0.327***	-0.192**	-0.0890	-0.345***
	(0.1180)	(0.0969)	(0.0974)	(0.1150)	(0.0995)

Nota: se presentan los coeficientes estimados de la ecuación (3.2). Los errores estándar se reportan entre paréntesis. Coeficiente significativo al \*\*\*1 %, \*\*5 %, \*10 %. Empleando el software Stata, versión 14, los modelos se estimaron con errores estándar robustos y agrupados por estado.

Fuente: elaboración propia con información de la ENDIREH 2021.

Una vez depurado cada índice de violencia de su correlación con  $u_i$  en (3.1), es decir, ya que se ha corregido la endogeneidad de la violencia, se procede a la segunda etapa del método de variables instrumentales, determinada por la estimación de los coeficientes de la ecuación (3.3), dichas estimaciones se presentan en la Tabla 4.3. A diferencia del modelo donde no se corrige la endogeneidad de la violencia, en este modelo se obtienen efectos estadísticamente significativos para la mayoría de los tipos de violencia.

Se estima que el incremento en una DE de la violencia económica en promedio reduce en 8.4 % los ingresos mensuales, mientras que el incremento de una DE de la violencia emocional reduce en promedio en 6.6 % el ingreso laboral. Para el modelo de la violencia física se estima que el incremento de una DE en esta, reduce en promedio en 10.5 % el ingreso mensual. Por su parte, el aumento de una DE en la violencia sexual reduce el ingreso mensual en 17.4 %. Finalmente, es el modelo del acoso el único donde no se tiene un efecto estadísticamente significativo sobre el ingreso de la mujer.

Tabla 4.3: Coeficientes estimados de la relación ingresos-violencia corrigiendo por la endogeneidad de la violencia.

	(1)	(2)	(3)	(4)	(5)
Violencia Económica	-0.0839** (0.0359)				
Violencia Emocional		-0.0655** (0.0289)			
Acoso			-0.1260 (0.0796)		
Violencia Física				-0.105** (0.0448)	
Violencia Sexual					-0.174* (0.0899)
Tamaño del hogar	-0.0239*** (0.0060)	-0.0245*** (0.0064)	-0.0270*** (0.0071)	-0.0252*** (0.0058)	-0.0241*** (0.0070)

Edad de la mujer	0.0226*** (0.0044)	0.0219*** (0.0044)	0.0229*** (0.0051)	0.0215*** (0.0042)	0.0246*** (0.0052)
Edad de la mujer al cuadrado	-0.00028*** (0.0001)	-0.00027*** (0.0001)	-0.00028*** (0.0001)	-0.00026*** (0.0000)	-0.0003*** (0.0001)
Años de escolaridad	0.0680*** (0.0035)	0.0681*** (0.0035)	0.0678*** (0.0036)	0.0676*** (0.0036)	0.0668*** (0.0036)
1 = Pertenece a un área urbana	0.0885** (0.0355)	0.0838** (0.0345)	0.0810** (0.0346)	0.0825** (0.0349)	0.0797** (0.0340)
1 = Tiene al menos un hijo	-0.135*** (0.0241)	-0.138*** (0.0240)	-0.137*** (0.0242)	-0.141*** (0.0245)	-0.140*** (0.0251)
1 = Habla lengua indígena	-0.0886* (0.0506)	-0.0842* (0.0492)	-0.0727 (0.0559)	-0.0800 (0.0542)	-0.0876* (0.0517)
Nivel socioeconómico del hogar	0.154*** (0.0163)	0.159*** (0.0164)	0.154*** (0.0167)	0.156*** (0.0162)	0.155*** (0.0197)
Servicios de telecomunicaciones	0.101*** (0.0110)	0.104*** (0.0111)	0.105*** (0.0116)	0.101*** (0.0122)	0.107*** (0.0120)
Constante	7.414*** (0.1220)	7.433*** (0.1230)	7.425*** (0.1310)	7.445*** (0.1260)	7.392*** (0.1330)
R-cuadrado	0.2690	0.2720	0.2500	0.2610	0.2340
Observaciones	16,729	16,729	16,729	16,727	16,727

Nota: se presentan los coeficientes estimados de la segunda etapa del método de variables instrumentales, es decir, corresponden a los coeficientes estimados de la ecuación (3.3). Los errores estándar se reportan entre paréntesis. Coeficiente significativo al \*\*\*1 %, \*\*5 %, \*10 %. Empleando el software Stata, versión 14, los modelos se estimaron con errores estándar robustos y agrupados por estado.

Fuente: elaboración propia con información de la ENDIREH 2021.

Como se vió en el Capítulo 2, en el método de variables instrumentales es necesario verificar que los instrumentos satisfacen la particularidad de afectar el nivel de violencia (propiedad de relevancia) pero no el nivel de ingresos recibidos (propiedad de exogeneidad) para tener fundamentos de que se han estimado coeficientes insesgados y consistentes. La Tabla 4.4 muestra las tres pruebas de hipótesis que se realizan, de acuerdo con la sección 2.3, para verificar la validez de los instrumentos, en ella se presenta el valor de cada estadístico de prueba y sus respectivos valores  $p$ .

En la Tabla 4.4 se observa que en la prueba  $F$  en la primera etapa, la hipótesis nula que establece que en conjunto los instrumentos no tienen relación significativa con los respectivos índices de violencia se rechaza al 1 % de significancia para todos los tipos de violencia. De manera similar, para todos los tipos de violencia se rechaza la hipótesis nula de la prueba de Kleibergen - Paap rank Wald al 1 % de significancia, esta hipótesis nula establece que la variable endógena no puede ser descrita a partir de los instrumentos. Por lo tanto, las variables empleadas que se han considerado como instrumentos resultan satisfacer la propiedad de relevancia.

En cuanto a la prueba de Sargan - Hansen para verificar la exogeneidad de los instrumentos,



se establece la hipótesis nula de que los instrumentos excluidos no están correlacionados con el término de error  $u_i$  y fueron correctamente excluidos de la ecuación (3.3), se puede observar que en el caso de la violencia física esta hipótesis nula no se puede rechazar al 1% de significancia, mientras que para el resto de los tipos de violencia esta hipótesis no se puede rechazar al 5% de significancia. De modo que no hay evidencias que pongan en duda la validez de los instrumentos.

Tabla 4.4: Pruebas de hipótesis sobre la validez de los instrumentos para cada uno de los modelos estimados.

	Violencia Económica	Violencia Emocional	Acoso	Violencia Física	Violencia Sexual
<b>Pruebas de relevancia</b>					
$F$ en la primera etapa	115.983 (0.000)	243.393 (0.000)	5.841 (0.001)	49.198 (0.000)	15.610 (0.000)
Prueba de Kleibergen - Paap rank Wald	21.402 (0.000)	22.789 (0.000)	14.951 (0.005)	19.675 (0.001)	13.777 (0.008)
<b>Prueba de exogeneidad</b>					
Prueba de Sargan - Hansen	7.548 (0.056)	7.512 (0.057)	7.602 (0.055)	8.113 (0.044)	5.859 (0.119)

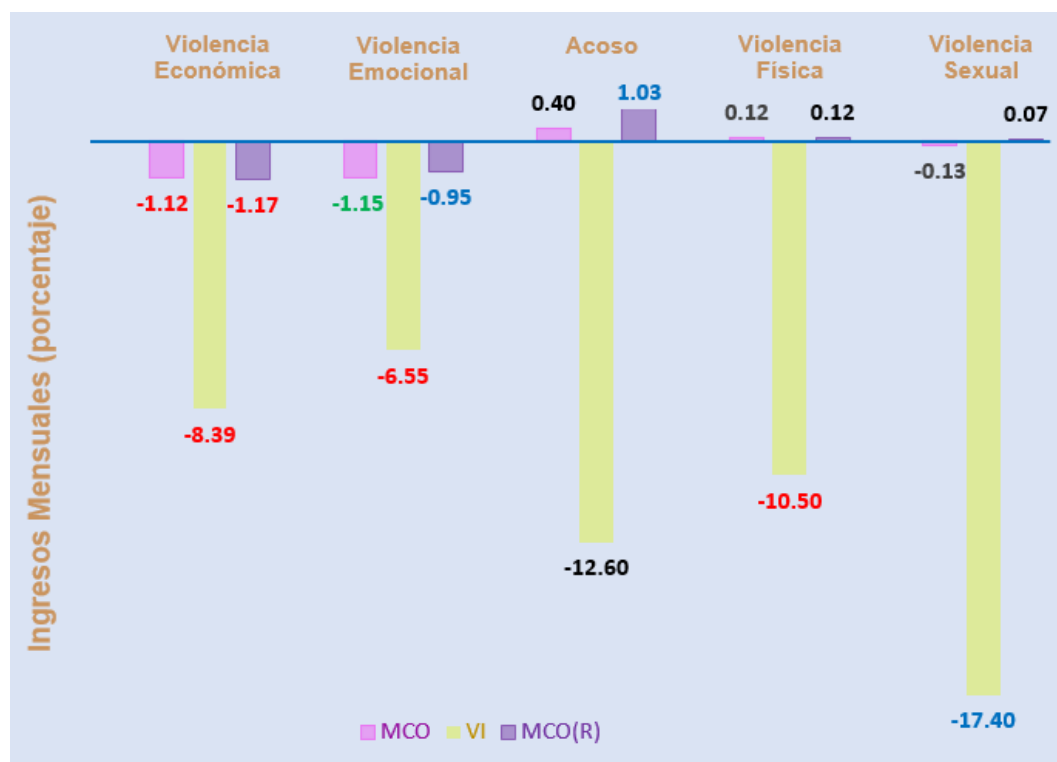
Nota: se presenta el valor de cada estadístico de prueba estimado en el software Stata, versión 14. Entre paréntesis se muestran los valores  $p$ .

Fuente: elaboración propia con información de la ENDIREH 2021.

De las Tablas 4.1 y 4.3 se puede observar que hay una notoria diferencia en las observaciones consideradas entre los modelos de regresión lineal múltiple por MCO y variables instrumentales, esto debido a que no todos los individuos en la muestra responden las preguntas que se utilizan para las variables que se emplean como instrumentos. En la Figura 4.1 se ilustra una comparativa de los efectos obtenidos, se muestra lo estimado por variables instrumentales y lo hecho por MCO tanto lo correspondiente a la muestra original como también una estimación para el modelo restringido a las mismas observaciones que considera el modelo de variables instrumentales.

Se puede observar que el efecto estimado de la regresión lineal por MCO en realidad estaba siendo subestimado, pues la magnitud del efecto estimado al corregir la endogeneidad de las variables ingreso y violencia, correspondiente al modelo de variables instrumentales, es mucho mayor. Otro punto importante es exhibir la inconsistencia del modelo de regresión lineal, pues resulta que al restringir la muestra ahora se obtiene un efecto significativo en el modelo que considera al acoso como variable explicativa, se observa que el aumento de una DE de este tipo de violencia en promedio incrementa en 1.03% el ingreso mensual de la mujer.

Figura 4.1: Efecto de la violencia de pareja sobre el ingreso mensual.



Nota: los efectos mostrados en negro no son estadísticamente significativos, en verde son significativos al 1%, en rojo al 5% y en azul al 10%. MCO indica las estimaciones cuando no se corrige la endogeneidad de la violencia. VI representa las estimaciones corrigiendo la endogeneidad de la violencia. MCO(R) muestra las estimaciones del modelo sin corregir la endogeneidad de la violencia pero estimado sobre la muestra restringida a las mismas observaciones que considera el modelo de variables instrumentales.

Fuente: elaboración propia.

### 4.3. Estimación del costo de la violencia

Con motivo del Día Internacional de la Eliminación de la Violencia contra la Mujer, el INEGI presenta de manera anual un comunicado de prensa con información estadística sobre la situación de violencia que sufren las mujeres en México. En INEGI (2018), con datos de la ENDIREH 2016 se da a conocer una primera estimación del costo de la violencia, de acuerdo a los días perdidos que fueron causados por algún tipo de violencia que experimentó la mujer y que como consecuencia se ve limitada a desempeñar sus labores. De un total de 401,050 mujeres se estimó que en promedio cada mujer perdió prácticamente un mes de trabajo, que en términos agregados equivale al 0.01% del PIB<sup>1</sup>.

<sup>1</sup>Producto Interno Bruto correspondiente al año 2016

En INEGI (2018) el costo se estima con la pregunta “A causa de los problemas con su esposo, ¿de octubre de 2015 a la fecha, aproximadamente cuántos días dejó de acudir a trabajar?”, sin embargo, en el presente trabajo se sigue una metodología distinta para obtener el costo de la violencia en términos de días de trabajo perdidos. De acuerdo al modelo de variables instrumentales, se estimó que hay una reducción del 8.4% en el salario mensual al aumentar una DE en la violencia emocional, lo que quiere decir que en promedio al mes la mujer pierde 2.52 días laborales a causa de este tipo de violencia que le ejerce su pareja, por lo tanto, al año pierde en promedio 30.2 días de trabajo. Por su parte, la violencia emocional es la que tiene en promedio el menor número de días de trabajo perdidos, pues su reducción del 6.6% en el salario mensual se refleja en una pérdida promedio de 23.6 días de trabajo al año. La violencia física al reducir en promedio en 10.5% los ingresos, implica más de un mes de trabajo perdido, mientras que la violencia sexual, que es la violencia que en particular afecta en mayor magnitud los ingresos mensuales, se traduce en poco más de dos meses de trabajo. Estos resultados se muestran en la Tabla 4.5, se puede observar que la fila correspondiente al acoso se encuentra vacía, esto se debe a que el acoso al no tener un efecto estadísticamente significativo sobre el salario, tampoco lo tiene para el número de días de trabajo perdidos.

Tabla 4.5: Días de trabajo perdidos anualmente.

	Reducción en el salario mensual (%)	Días de trabajo perdidos al año
Violencia económica	8.39	30.2
Violencia emocional	6.55	23.6
Acoso	-	-
Violencia física	10.5	37.8
Violencia sexual	17.4	62.6

Fuente: elaboración propia.

De manera adicional, es posible hacer una estimación del costo agregado que tiene cada tipo de violencia, tal como se muestra en la Tabla 4.6. Debido a que se están empleando las estimaciones del modelo de variables instrumentales, se considera que la muestra se restringe a las observaciones que contribuyen en el modelo de variables instrumentales correspondiente, sobre esta muestra de estudio ya restringida se obtiene un ingreso mensual promedio y de acuerdo al porcentaje de la reducción en el salario mensual, se puede obtener la estimación de la pérdida anual de cada mujer. Al multiplicarlo por el total de las mujeres de esta muestra se obtiene el costo total de cada tipo de violencia de pareja. Se puede observar que en el caso de la violencia sexual, que implica en promedio una reducción del 17.4% del salario mensual, se traduce a que en promedio, cada mujer sufre una pérdida anual de 16,517 pesos en sus ingresos laborales, lo

que en términos agregados equivale al 0.60% del PIB<sup>2</sup>.

Tabla 4.6: Costos de la violencia de pareja.

	Mujeres que viven con su pareja	Salario promedio (\$)	Pérdida anual por mujer (\$)	Costo de la violencia de pareja (\$)	Costo respecto al PIB (%)
Violencia económica	6,480,336	7,910.35	7,964.14	51,610,318,662.37	0.29
Violencia emocional	6,480,336	7,910.35	6,217.54	40,291,726,726.88	0.23
Acoso	6,480,336	7,910.35	-	-	-
Violencia física	6,479,804	7,910.52	9,967.26	64,585,868,278.53	0.36
Violencia sexual	6,479,804	7,910.52	16,517.17	107,028,010,290.14	0.60

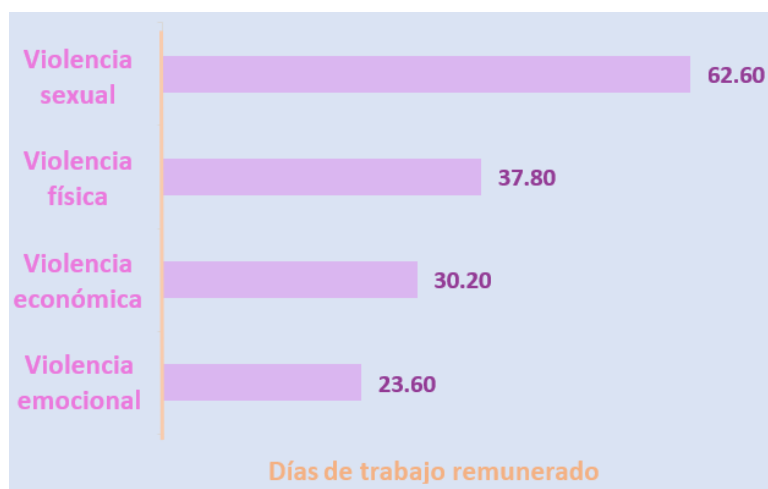
Nota: la muestra de análisis corresponde a las observaciones consideradas para el modelo de variables instrumentales, sobre esta muestra se estima un salario mensual promedio, precios constantes de 2021. Se considera el PIB de México en el año 2021 de 17.811 billones de pesos mexicanos.

Fuente: elaboración propia.

De manera gráfica, en la Figura 4.2 se muestran las estimaciones del costo de la violencia en términos de días de trabajo remunerado perdidos anualmente, mientras que la Figura 4.3 ilustra el costo de la violencia en términos agregados como porcentaje que representa respecto al PIB del año 2021.

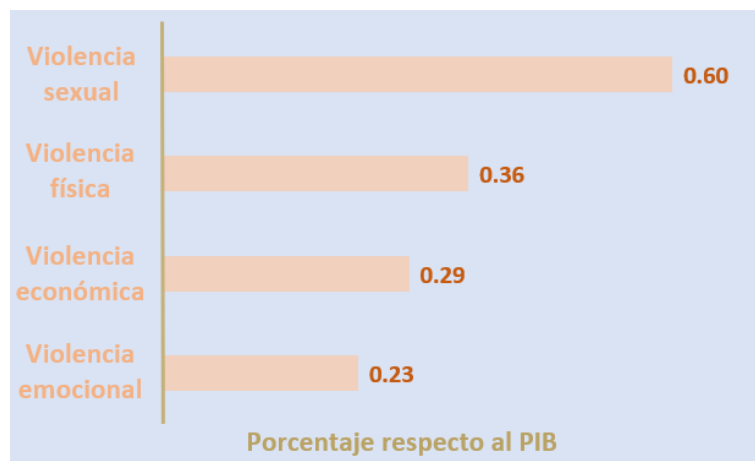
<sup>2</sup>Producto Interno Bruto de 17.811 billones de pesos mexicanos, correspondiente al año 2021.

Figura 4.2: Días de trabajo remunerado perdidos al año.



Fuente: elaboración propia.

Figura 4.3: Costo de la violencia respecto al PIB.



Fuente: elaboración propia.



# Conclusiones

Cuando se analiza el efecto que tiene la situación económica de las mujeres sobre la probabilidad de ser violentadas por su pareja la conclusión es ambigua, resulta que el hecho de empoderar a la mujer no parece solucionar el problema de la violencia de la que pueda ser víctima, pues así como la mejora de su condición económica puede reducir la exposición a relaciones violentas, también puede incrementar la tendencia de ser violentadas. Esta vertiente ha sido bastante estudiada, sin embargo la situación recíproca no, es decir, el estudio del efecto que tiene la violencia contra la mujer sobre su economía, más aún, Aguirre (2023) es apenas el primer análisis de esta problemática para el caso de México. El presente documento tiene como objetivo colaborar en el estudio de dicha problemática tan relevante y que requiere de mayor atención de la que se le ha prestado hasta el momento.

En este trabajo se dimensionaron los efectos que tienen los distintos tipos de violencia sobre el ingreso laboral mensual de las mujeres que se encuentran en una relación de pareja, ya sea casadas o en unión libre. Las principales aportaciones de este trabajo son: en primer lugar, se emplearon los datos más recientes representativos a nivel nacional para estimar la relación entre ingreso laboral y violencia; se implementó el método de componentes principales para cuantificar la intensidad de la violencia, diferenciada en cinco tipos (económica, emocional, acoso, física y sexual); se hizo la observación de la existencia de endogeneidad en las variables ingreso y violencia, problema que requiere un tratamiento especial al momento del modelado y resulta que hay significativas diferencias al considerar este asunto y al evadirlo. Finalmente, a partir de los resultados obtenidos y en términos del número de días perdidos anualmente a causa de la violencia, se hizo una estimación del costo de la violencia y se mostró en términos agregados el porcentaje del PIB que representa dicho costo.

Los efectos estimados al no corregir la endogeneidad de la violencia van de -1.15% a 0.40% y son estadísticamente significativos solamente para dos tipos de violencia, mientras que al corregir la endogeneidad se obtienen efectos que van de -17.40% a -8.55% y que son estadísticamente significativos para casi todos los tipos de violencia. Exhibiendo así la gran importancia de tomar en cuenta este problema al momento de realizar un análisis de la relación entre la violencia de pareja contra la mujer y sus ingresos.

Cuando no se corrige la endogeneidad de las variables violencia e ingreso, el análisis de la relación violencia-ingresos se efectúa mediante un modelo de regresión lineal múltiple y se encuentra que tanto la violencia emocional como la económica afectan negativamente los ingresos mensuales de aquellas mujeres que tienen una relación de pareja, la violencia económica implica una disminución en promedio en 1.12 % de sus ingresos, mientras que la económica los reduce en 1.15 %, sin embargo, para los demás tipos de violencia no se encuentran efectos estadísticamente significativos.

Por otra parte, al corregir la endogeneidad, los resultados muestran que la violencia de pareja de tipo económica, emocional, física y sexual reducen en promedio los ingresos laborales de las mujeres casadas o en unión libre. Siendo la violencia sexual la que presenta el efecto más alarmante, pues ésta disminuye en promedio en 17.4 % el ingreso mensual de las mujeres, de manera concreta, si una mujer gana el ingreso promedio<sup>3</sup>, este ingreso se disminuye en \$1,376 pesos<sup>4</sup> al mes a causa de la violencia sexual ejercida por parte de su pareja, con esto se estima que en promedio una mujer pierde poco más de dos meses de trabajo remunerado al año y en términos agregados equivale al 0.60 % del PIB<sup>5</sup>. El segundo efecto más grande se observa en la violencia física, seguido de la violencia económica y la emocional.

La inconsistencia que muestra el modelo cuando no se corrige la endogeneidad es un punto a destacar, pues cuando se restringe la muestra se obtiene que en promedio el ingreso de la mujer aumenta en 1.03 % al incrementar en una DE el acoso, este aumento podría ser explicado porque la mujer al ser hostigada en su hogar a través de un ambiente hostil y rodeada de amenazas prefiere estar fuera del mismo, siendo el trabajo un posible escape, quizá presentando horas o incluso días extra, sin embargo aquí se presenta una limitante de trabajo, pues no hay datos en la ENDIREH que ayuden a verificar esta suposición.

Otra limitación de trabajo es que no es posible identificar el mecanismo por el cual la violencia reduce el ingreso laboral, una posible explicación sería una menor productividad laboral o incluso puede que el ingreso se ve afectado debido a ausencias laborales para que la mujer atienda cuestiones de salud. Sin embargo, la ENDIREH no proporciona información que ayude a rastrear posibles causas de la reducción en los ingresos, a pesar de que su primera edición indagaba sobre las horas trabajadas, en las siguientes ediciones no fue así, contar con este dato, con los días trabajados o tener una desagregación del ingreso laboral puede ayudar para diferenciar el mecanismo por el que ocurre la disminución del ingreso laboral. Si se tuviera información acerca del número de horas o días trabajados y se encuentra que el aumento de los niveles de violencia disminuye la intensidad del trabajo, resultaría entonces que el ingreso mensual de la mujer disminuye a causa de las ausencias laborales. Mientras que la desagregación del ingreso laboral de las mujeres por concepto como salario, bonos de productividad o ingreso por negocios propios, permitiría estimar

<sup>3</sup>A partir de los 16,727 registros que representan a 6,479,804 mujeres de la muestra de estudio restringida, se estima un salario mensual promedio de 7,910.52 pesos mexicanos.

<sup>4</sup>Precios constantes de 2021.

<sup>5</sup>PIB correspondiente al año 2021.



los modelos para cada tipo de concepto e identificar cuál de éstos se está reduciendo, si se observa que no hay una reducción en los salarios, pero sí en los bonos de productividad o en los ingresos por negocios propios se tendría evidencia de que el efecto se debe a una baja en la productividad.

La violencia contra las mujeres es un grave problema a nivel internacional que afecta tanto las vidas como la salud de las mujeres (World Health Organization 2021). Además de la pérdida o daño económico, la violencia de pareja genera en las mujeres problemas en la salud mental, trastornos depresivos, ansiedad, estrés postraumático, suicidio, trastornos gastrointestinales y ginecológicos (Plazaola-Castaño y Ruiz-Pérez 2004; Rivara et al. 2019). Cuando una mujer es violentada, se presentan repercusiones en el capital humano de sus hijos a través de la salud, pues las mujeres que sufren violencia suelen asistir a menos citas prenatales, los hijos son más propensos de tener diarrea y menor peso y la probabilidad de que reciban vacunas es menor (Agüero 2013). También existen consecuencias sociales, tales como la transmisión intergeneracional de la violencia que se hace presente en aquellos hogares donde la mujer es violentada por su cónyuge (Safranoff y Tiravassi 2018).

La violencia de pareja genera consecuencias negativas a corto, mediano y largo plazo, afectando tanto a las mujeres, como a los niños y las familias (World Health Organization 2021). Por ello, es crucial no solo profundizar en esta problemática, sino también dirigir las políticas públicas hacia la prevención, atención, sanción y eliminación de cualquier forma de violencia dirigida contra las mujeres, tales como: promover campañas de divulgación donde se dimensionen los daños que genera la violencia de pareja y se exhorte a las mujeres a dirigirse a solicitar ayuda en cuanto sea requerida; dar mayor difusión a los distintos organismos con los que México cuenta para atender la violencia contra la mujer, pues organismos como la Comisión Nacional para Prevenir y Erradicar la Violencia contra las Mujeres (CONAVIM) o el Centro de Atención a la Violencia Intrafamiliar (CAVI) no suelen ser tan conocidas por las mujeres en situación de violencia; implementar programas para mejorar la capacitación de policías, jueces y personal médico en el manejo de casos de violencia de género; agilizar los procesos de denuncia que por lo regular resultan ser complicados y las víctimas prefieren no continuar con el caso; desarrollar protocolos adecuados por parte de las autoridades para el trato con las víctimas, algunas víctimas experimentan traumas al tener que narrar repetidamente su experiencia o se enfrentan a autoridades que constantemente cuestionan su papel de víctima; implementar mecanismos que aseguren la protección y el bienestar tanto de la mujer como de sus hijos y/o personas dependientes de ella, con el fin de que las víctimas de violencia no duden en levantar una denuncia.



# Apéndice A

## Desarrollo del estimador $\hat{\beta}$

En la descripción del método de Regresión Lineal en el Capítulo 2, se llegó a la igualdad (2.12):

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

sin embargo, en la práctica se busca una forma de  $\hat{\beta}$  que hable más acerca de los datos. Se muestra a continuación cómo se puede escribir el vector  $\hat{\beta}$  en términos de las características observables del modelo, a esto se le conoce como *identificación del modelo*.

Recordar que

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ 1 & x_{31} & x_{32} & \dots & x_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

Sea  $S = \mathbf{X}^T \mathbf{X}$ ,  $S$  cumple las siguientes propiedades:

- 1)  $S$  es una matriz simétrica.
- 2)  $S$  es una matriz positiva definida.

La primera propiedad no hace falta probarla, pues es trivial por la definición de  $S$ . La prueba de la segunda propiedad se muestra a continuación.

*Demostración.*  $S$  es positiva definida.

De acuerdo al Criterio de Sylvester, una matriz simétrica  $A$  es positiva definida si el determinante de cada submatriz principal de  $A$  es positivo.

Observe que para las  $k$  variables explicativas en (2.7) la media está dada por

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

con  $j = 1, \dots, k$ , por lo tanto, la matriz  $S$  luce de la siguiente forma

$$S = \begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 & \cdots & n\bar{x}_k \\ n\bar{x}_1 & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ n\bar{x}_2 & \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n\bar{x}_k & \sum_{i=1}^n x_{i1}x_{ik} & \sum_{i=1}^n x_{i2}x_{ik} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{pmatrix}$$

Si las entradas de  $S$  se etiquetan de la siguiente manera

$$S = \begin{pmatrix} s_{00} & s_{01} & \cdots & s_{0k} \\ s_{01} & s_{11} & \cdots & s_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{0k} & s_{1k} & \cdots & s_{kk} \end{pmatrix}$$

entonces las  $k + 1$  submatrices principales de  $S$  están dadas por

$$S_t = \begin{pmatrix} s_{00} & s_{01} & \cdots & s_{0f} \\ s_{01} & s_{11} & \cdots & s_{1f} \\ \vdots & \vdots & \ddots & \vdots \\ s_{0f} & s_{1f} & \cdots & s_{ff} \end{pmatrix}$$

con  $t = 0, 1, 2, \dots, k$  y  $f = 0, 1, \dots, t$ .

Es claro que la primera submatriz principal tiene determinante positivo, pues

$$S_0 = s_{00} = n > 0$$

para las siguientes  $k$  submatrices es conveniente obtener el determinante por el método de eliminación de Gauss. Como primer paso, con un múltiplo de la primera fila se obtendrán ceros en la

primera columna de  $S_t$ , es decir:

$$\begin{aligned}
 S_t &= \begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 & \cdots & n\bar{x}_t \\ n\bar{x}_1 & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{it} \\ n\bar{x}_2 & \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 & \cdots & \sum_{i=1}^n x_{i2}x_{it} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n\bar{x}_t & \sum_{i=1}^n x_{i1}x_{it} & \sum_{i=1}^n x_{i2}x_{it} & \cdots & \sum_{i=1}^n x_{it}^2 \end{pmatrix} \\
 &\sim \begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 & \cdots & n\bar{x}_t \\ 0 & \sum_{i=1}^n x_{i1}^2 - n\bar{x}_1^2 & \sum_{i=1}^n x_{i1}x_{i2} - n\bar{x}_1\bar{x}_2 & \cdots & \sum_{i=1}^n x_{i1}x_{it} - n\bar{x}_1\bar{x}_t \\ 0 & \sum_{i=1}^n x_{i1}x_{i2} - n\bar{x}_1\bar{x}_2 & \sum_{i=1}^n x_{i2}^2 - n\bar{x}_2^2 & \cdots & \sum_{i=1}^n x_{i2}x_{it} - n\bar{x}_2\bar{x}_t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sum_{i=1}^n x_{i1}x_{it} - n\bar{x}_1\bar{x}_t & \sum_{i=1}^n x_{i2}x_{it} - n\bar{x}_2\bar{x}_t & \cdots & \sum_{i=1}^n x_{it}^2 - n\bar{x}_t^2 \end{pmatrix} \quad (\text{A.1})
 \end{aligned}$$

ahora bien, como la varianza muestral de  $x_j$  (para  $j = 1, 2, \dots, k$ ) esta dada por

$$s_j^2 = \text{var}(x_j) = \frac{1}{n-1} \left[ \sum_{i=1}^n x_{ij}^2 - n\bar{x}_j^2 \right]$$

se sigue que

$$s_j^2(n-1) = \sum_{i=1}^n x_{ij}^2 - n\bar{x}_j^2 \quad (\text{A.2})$$

Además, la covarianza muestral entre  $x_j$  y  $x_l$  (con  $j \neq l$ ) es

$$\text{cov}(x_j, x_l) = \frac{1}{n-1} \left[ \sum_{i=1}^n x_{ij}x_{il} - n\bar{x}_j\bar{x}_l \right]$$

entonces se tiene la siguiente igualdad

$$\text{cov}(x_j, x_l)(n-1) = \sum_{i=1}^n x_{ij}x_{il} - n\bar{x}_j\bar{x}_l \quad (\text{A.3})$$

De modo que la matriz en (A.1) se puede escribir en términos de varianzas y covarianzas muestrales de las variables explicativas de la siguiente manera

$$\begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 & \cdots & n\bar{x}_t \\ 0 & s_1^2(n-1) & cov(x_1, x_2)(n-1) & \cdots & cov(x_1, x_t)(n-1) \\ 0 & cov(x_1, x_2)(n-1) & s_2^2(n-1) & \cdots & cov(x_2, x_t)(n-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & cov(x_1, x_t)(n-1) & cov(x_2, x_t)(n-1) & \cdots & s_t^2(n-1) \end{pmatrix}$$

y como cada  $x_j$  es independiente entonces

$$cov(x_i, x_j) = 0 \quad (\text{A.4})$$

para todo  $i \neq j$ , con lo que la matriz anterior queda como

$$\begin{pmatrix} n & n\bar{x}_1 & n\bar{x}_2 & \cdots & n\bar{x}_t \\ 0 & s_1^2(n-1) & 0 & \cdots & 0 \\ 0 & 0 & s_2^2(n-1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & s_t^2(n-1) \end{pmatrix}$$

la cual ya es una matriz triangular, obteniendo entonces que

$$det(S_t) = n(n-1)^t \prod_{p=1}^t s_p^2$$

donde cada variable  $x_j$  tiene una varianza no nula, por lo tanto el determinante de  $S_t$  es estrictamente mayor a cero. Concluyendo así que, al ser  $S$  una matriz simétrica y tal que el determinante de cada una de sus submatrices principales es positivo, entonces  $S$  es una matriz positiva definida.

□

La matriz  $S$  es difícil de operar tal cual está definida, no obstante, mediante la Factorización de Cholesky, que se enuncia a continuación, se puede obtener una matriz equivalente a  $S$  y que sea más sencilla de manejar.

**Factorización de Cholesky.** Si  $A$  es una matriz real, simétrica y positiva definida de dimensión  $d$ , entonces existe una única matriz  $Q$  tal que

$$A = Q^T Q$$

donde  $Q$  es una matriz triangular superior con entradas positivas en su diagonal.

Más aún, las entradas de  $Q$  se obtienen de la siguiente manera

$$\begin{aligned} q_{ii} &= \left[ a_{ii} - \sum_{k=1}^{i-1} q_{ki}^2 \right]^{\frac{1}{2}} & i = 1, 2, \dots, d \\ q_{ij} &= \frac{1}{q_{ii}} \left[ s_{ij} - \sum_{k=1}^{i-1} q_{ki} q_{kj} \right] & i < j \\ q_{ij} &= 0 & j < i \end{aligned}$$

Al ser  $S$  una matriz simétrica y positiva definida, entonces admite una descomposición como producto de matrices triangulares. Sea  $Q$  de la forma

$$Q = \begin{pmatrix} q_{00} & q_{01} & \cdots & q_{0k} \\ & q_{11} & \cdots & q_{1k} \\ & & \ddots & \vdots \\ & & & q_{kk} \end{pmatrix}$$

de acuerdo a la factorización de Cholesky, se tiene que  $q_{00} = \sqrt{n}$ , mientras que

$$\begin{aligned} q_{0j} &= \frac{n\bar{x}_j}{\sqrt{n}} \\ &= \bar{x}_j \sqrt{n} \end{aligned}$$

para  $j = 1, \dots, k$ . Con  $q_{11}$  se tiene lo siguiente

$$\begin{aligned} q_{11} &= \left[ \sum_{i=1}^n x_{i1}^2 - \left( \frac{n\bar{x}_1}{\sqrt{n}} \right)^2 \right]^{\frac{1}{2}} \\ &= \left( \sum_{i=1}^n x_{i1}^2 - n\bar{x}_1^2 \right)^{\frac{1}{2}} \end{aligned}$$

por (A.2) se sigue que

$$\begin{aligned} q_{11} &= \left( \sum_{i=1}^n x_{i1}^2 - n\bar{x}_1^2 \right)^{\frac{1}{2}} \\ &= \sqrt{s_1^2(n-1)} \end{aligned}$$

Para  $j = 2, \dots, k$ , las entradas  $q_{1j}$  estan dadas por

$$\begin{aligned} q_{1j} &= \frac{1}{q_{11}} [s_{1j} - q_{01}q_{0j}] \\ &= \frac{1}{q_{11}} \left[ \sum_{i=1}^n x_{i1}x_{ij} - (\bar{x}_1\sqrt{n}) (\bar{x}_j\sqrt{n}) \right] \\ &= \frac{1}{q_{11}} \left( \sum_{i=1}^n x_{i1}x_{ij} - n\bar{x}_1\bar{x}_j \right) \end{aligned}$$

por (A.3) se tiene entonces que

$$\begin{aligned} q_{1j} &= \frac{1}{q_{11}} \left( \sum_{i=1}^n x_{i1}x_{ij} - n\bar{x}_1\bar{x}_j \right) \\ &= \frac{\text{cov}(x_1, x_j)(n-1)}{q_{11}} \end{aligned}$$

y por (A.4) se sigue que para  $j = 2, \dots, k$

$$q_{1j} = \frac{0 \cdot (n-1)}{q_{11}} = 0$$

Hasta el momento  $Q$  luce de la siguiente forma

$$Q = \begin{pmatrix} \sqrt{n} & \bar{x}_1\sqrt{n} & \bar{x}_2\sqrt{n} & \cdots & \bar{x}_k\sqrt{n} \\ & \sqrt{s_1^2(n-1)} & 0 & \cdots & 0 \\ & & q_{22} & \cdots & q_{2k} \\ & & & \ddots & \vdots \\ & & & & q_{kk} \end{pmatrix}$$

y a partir de este punto el cálculo de las entradas restantes se simplifica, de manera analítica, observe que para obtener el elemento  $j$  de la diagonal se consideran únicamente a  $s_{jj}$  y los elementos de la  $j$ -ésima columna de  $Q$ , pronto se convencerá de que a excepción del mismo  $q_{jj}$  y  $q_{0j}$ , el resto de elementos de dicha columna son cero.



Note que para el caso  $j = 2$  se tiene

$$\begin{aligned}
 q_{22} &= \left[ s_{22} - \sum_{p=0}^1 q_{p2}^2 \right]^{\frac{1}{2}} \\
 &= [s_{22} - (q_{02}^2 + q_{12}^2)]^{\frac{1}{2}} \\
 &= [s_{22} - (q_{02}^2 + 0)]^{\frac{1}{2}} \\
 &= \left[ \sum_{i=1}^n x_{i2}^2 - (\bar{x}_2 \sqrt{n})^2 \right]^{\frac{1}{2}} \\
 &= \left( \sum_{i=1}^n x_{i2}^2 - n\bar{x}_2^2 \right)^{\frac{1}{2}}
 \end{aligned}$$

y de la igualdad (A.2) se sigue que

$$q_{22} = \left( \sum_{i=1}^n x_{i2}^2 - n\bar{x}_2^2 \right)^{\frac{1}{2}} = \sqrt{s_2^2(n-1)}$$

para los elementos restantes de esta fila se tiene que

$$\begin{aligned}
 q_{2j} &= \frac{1}{q_{22}} \left[ s_{2j} - \sum_{p=0}^1 q_{p2}q_{pj} \right] & j = 3, 4, \dots, k \\
 &= \frac{1}{q_{22}} [s_{2j} - q_{02}q_{0j} - q_{12}q_{1j}] \\
 &= \frac{1}{q_{22}} [s_{2j} - q_{02}q_{0j} - 0] \\
 &= \frac{1}{q_{22}} [s_{2j} - q_{02}q_{0j}] \\
 &= \frac{1}{q_{22}} \left[ \sum_{i=1}^n x_{i2}x_{ij} - (\bar{x}_2 \sqrt{n}) (\bar{x}_j \sqrt{n}) \right] \\
 &= \frac{1}{q_{22}} \left( \sum_{i=1}^n x_{i2}x_{ij} - n\bar{x}_2\bar{x}_j \right)
 \end{aligned}$$

y por (A.3), entonces

$$q_{2j} = \frac{\text{cov}(x_2, x_j)(n-1)}{q_{22}}$$

además, por (A.4), se tiene que

$$q_{2j} = \frac{\text{cov}(x_2, x_j)(n-1)}{q_{22}} = \frac{0 \cdot (n-1)}{q_{22}} = 0$$

De modo que la fila que se ha calculado está formada por ceros fuera de la diagonal, y gracias a esta forma que presentan las filas creadas, ocurre lo mismo para el resto de filas. Observe que cuando se quiere obtener al elemento  $q_{ab}$  ( $a < b$ ), la fila de entradas  $q_{0j}$  (con  $j = 0, 1, \dots, k$ ) es la única que aporta elementos distintos de cero, entonces los sumandos en

$$\sum_{p=0}^{a-1} q_{pa} q_{pb}$$

serán cero a excepción del primero, dado por

$$\begin{aligned} q_{0a} q_{0b} &= (\bar{x}_a \sqrt{n})(\bar{x}_b \sqrt{n}) \\ &= n \bar{x}_a \bar{x}_b \end{aligned}$$

por lo tanto

$$\begin{aligned} q_{ab} &= \frac{1}{q_{aa}} \left[ s_{ab} - \sum_{p=0}^{a-1} q_{pa} q_{pb} \right] \\ &= \frac{1}{q_{aa}} \left( \sum_{i=1}^n x_{ia} x_{ib} - n \bar{x}_a \bar{x}_b \right) \\ &= \frac{\text{cov}(x_a, x_b)(n-1)}{q_{aa}} \\ &= \frac{0 \cdot (n-1)}{q_{aa}} \\ &= 0 \end{aligned}$$

Volviendo a los elementos de la diagonal, para la entrada  $q_{jj}$  se requiere de los primeros  $j-1$  elemento de la columna a la que pertenece  $q_{jj}$ , sin embargo, estos elementos forman parte de las  $j-1$  filas creadas previamente, las cuales se ha mostrado contienen ceros fuera de la diagonal, a excepción de la primera fila, por lo tanto, de

$$q_{jj} = \left[ s_{jj} - \sum_{p=0}^{j-1} q_{pj}^2 \right]^{\frac{1}{2}} \quad j = 2, 3, \dots, k$$

se sigue que

$$\begin{aligned}
 q_{jj} &= \left[ s_{jj} - \sum_{p=0}^{j-1} q_{pj}^2 \right]^{\frac{1}{2}} \\
 &= \left[ s_{jj} - (q_{0j}^2 + q_{1j}^2 + q_{2j}^2 + \dots + q_{(j-1)j}^2) \right]^{\frac{1}{2}} \\
 &= \left[ s_{jj} - (q_{0j}^2 + 0 + 0 + \dots + 0) \right]^{\frac{1}{2}} \\
 &= \left[ \sum_{i=1}^n x_{ij}^2 - (\bar{x}_j \sqrt{n})^2 \right]^{\frac{1}{2}} \\
 &= \left( \sum_{i=1}^n x_{ij}^2 - n\bar{x}_j^2 \right)^{\frac{1}{2}}
 \end{aligned}$$

nuevamente recurriendo a la igualdad (A.2), se sigue que

$$q_{jj} = \sqrt{s_j^2(n-1)}$$

con lo que  $Q$  queda de la siguiente forma

$$Q = \begin{pmatrix} \sqrt{n} & \bar{x}_1 \sqrt{n} & \bar{x}_2 \sqrt{n} & \dots & \bar{x}_k \sqrt{n} \\ & \sqrt{s_1^2(n-1)} & 0 & \dots & 0 \\ & & \sqrt{s_2^2(n-1)} & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & \sqrt{s_k^2(n-1)} \end{pmatrix}$$

Es sencillo ver que la matriz inversa de  $Q$  es

$$Q^{-1} = R = \begin{pmatrix} \frac{1}{\sqrt{n}} & -\frac{\bar{x}_1}{\sqrt{s_1^2(n-1)}} & -\frac{\bar{x}_2}{\sqrt{s_2^2(n-1)}} & \dots & -\frac{\bar{x}_k}{\sqrt{s_k^2(n-1)}} \\ & \frac{1}{\sqrt{s_1^2(n-1)}} & 0 & \dots & 0 \\ & & \frac{1}{\sqrt{s_2^2(n-1)}} & \dots & 0 \\ & & & \ddots & \vdots \\ & & & & \frac{1}{\sqrt{s_k^2(n-1)}} \end{pmatrix}$$

Recordar que

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (\text{A.5})$$

y como  $\mathbf{X}^T \mathbf{X} = S$  entonces

$$(\mathbf{X}^T \mathbf{X})^{-1} = S^{-1} = (Q^T Q)^{-1} = R R^T$$

donde  $RR^T$  es una matriz mucho más sencilla de calcular que  $(\mathbf{X}^T\mathbf{X})^{-1}$ , como se muestra a continuación

$$RR^T = \begin{pmatrix} \left(\frac{1}{n} + \frac{1}{n-1} \sum_{i=1}^n \frac{\bar{x}_i}{s_i^2}\right) & -\frac{\bar{x}_1}{s_1^2(n-1)} & -\frac{\bar{x}_2}{s_2^2(n-1)} & \cdots & -\frac{\bar{x}_k}{s_k^2(n-1)} \\ -\frac{\bar{x}_1}{s_1^2(n-1)} & \frac{1}{s_1^2(n-1)} & 0 & \cdots & 0 \\ -\frac{\bar{x}_2}{s_1^2(n-1)} & 0 & \frac{1}{s_2^2(n-1)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\bar{x}_k}{s_1^2(n-1)} & 0 & 0 & \cdots & \frac{1}{s_k^2(n-1)} \end{pmatrix}$$

de (A.5) se sigue que

$$\hat{\beta} = (RR^T)(\mathbf{X}^T Y)$$

como  $Y = (y_1, y_2, \dots, y_n)^T$ , se tiene entonces que

$$\hat{\beta} = \begin{pmatrix} \left(\frac{1}{n} + \frac{1}{n-1} \sum_{i=1}^n \frac{\bar{x}_i}{s_i^2}\right) & -\frac{\bar{x}_1}{s_1^2(n-1)} & -\frac{\bar{x}_2}{s_2^2(n-1)} & \cdots & -\frac{\bar{x}_k}{s_k^2(n-1)} \\ -\frac{\bar{x}_1}{s_1^2(n-1)} & \frac{1}{s_1^2(n-1)} & 0 & \cdots & 0 \\ -\frac{\bar{x}_2}{s_1^2(n-1)} & 0 & \frac{1}{s_2^2(n-1)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\bar{x}_k}{s_1^2(n-1)} & 0 & 0 & \cdots & \frac{1}{s_k^2(n-1)} \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_{i1}y_i \\ \sum_{i=1}^n x_{i2}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}$$

donde a excepción de la primera entrada de  $\hat{\beta}$ , a la cual se denominará  $\hat{\beta}_0$ , los demás términos son de la forma

$$\begin{aligned} \hat{\beta}_j &= \frac{\sum_{i=1}^n x_{ij}y_i}{s_j^2(n-1)} - \frac{n\bar{x}_j\bar{y}}{s_j^2(n-1)} \\ &= \frac{1}{s_j^2(n-1)} \left[ \sum_{i=1}^n x_{ij}y_i - n\bar{x}_j\bar{y} \right] \\ &= \frac{1}{s_j^2(n-1)} [\text{cov}(x_j, y)(n-1)] \\ &= \frac{\text{cov}(x_j, y)}{\text{var}(x_j)} \end{aligned}$$

Para fines prácticos,  $\hat{\beta}_0$  se escribirá en términos de los estimadores  $\hat{\beta}_j$  ( $j = 1, \dots, k$ ) y las medias de las variables  $x_j$  y  $y$ . A partir del modelo de regresión múltiple

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

se puede despejar al término de error  $u$

$$u = y - \beta_0 - \beta_1 x_1 - \dots - \beta_k x_k$$

y como  $E(u) = 0$ , entonces se llega a la siguiente ecuación

$$E(y - \beta_0 - \beta_1 x_1 - \dots - \beta_k x_k) = 0 \quad (\text{A.6})$$

recurriendo a la muestra de datos, se toman los estimadores  $\hat{\beta}_j$  de modo que satisfagan la contraparte muestral de (A.6), es decir

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

se sigue que

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_k \bar{x}_k = 0$$

donde  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , así, finalmente se obtiene una expresión para  $\hat{\beta}_0$ :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_k \bar{x}_k$$

por lo tanto

$$\hat{\beta} = \begin{pmatrix} \bar{y} - \sum_{i=1}^k \hat{\beta}_i \bar{x}_i \\ \frac{\text{cov}(x_1, y)}{\text{var}(x_1)} \\ \frac{\text{cov}(x_2, y)}{\text{var}(x_2)} \\ \vdots \\ \frac{\text{cov}(x_k, y)}{\text{var}(x_k)} \end{pmatrix}$$



## Apéndice B

# Pruebas de relevancia y exogeneidad de los instrumentos

Lo expuesto en este apéndice se basa en la fundamentación teórica que implementa el software Stata, versión 14, el cual toma como referencia lo hecho por Stock y Yogo (2002) para la Prueba  $F$ , Kleibergen y Paap (2006) para la prueba de Kleibergen-Paap rank Wald y Hayashi (2000) para la prueba de Sargan - Hansen. Cuando se emplea el método de variables instrumentales, descrito por las ecuaciones

$$\begin{aligned}y_1 &= \alpha_0 + \alpha Z + \gamma X + \epsilon \\y &= \beta_0 + \beta_1 \hat{y}_1 + \theta X + u\end{aligned}$$

verificar que las variables empleadas como instrumentos excluidos cumplen las propiedades de relevancia y de exogeneidad brinda evidencia de que se han estimado coeficientes consistentes e insesgados, se describen a continuación las tres pruebas de hipótesis que se emplean para esto, las dos primeras para verificar la relevancia de los instrumentos y la tercera para evaluar que se cumpla la propiedad de exogeneidad de los instrumentos.

### Prueba $F$

La prueba  $F$  sirve para probar si un conjunto de variables explicativas no tiene efecto sobre la variable dependiente, esta prueba se emplea en la primera etapa de la estimación de los coeficientes del modelo de variables instrumentales y se establece la hipótesis nula de que no hay relación significativa entre los instrumentos y la variable endógena, lo cual se puede escribir de la siguiente manera

$$H_0 : \alpha \equiv 0$$

mientras que la hipótesis alterna es que al menos uno de los coeficientes es distinto de cero, es decir, si  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ , entonces

$$H_a : \alpha_j \neq 0 \quad \text{para al menos un } \alpha_j \text{ de } \alpha.$$

El estadístico  $F$  que se emplea se define como

$$F = \frac{\frac{SC}{k+r}}{\frac{SRC}{n-k-r-1}} \quad (\text{B.1})$$

donde  $SC$  se define como

$$SC = \sum_{i=1}^n (\hat{y}_{i1} - \bar{y}_1)^2$$

con  $\bar{y}_1 = \frac{1}{n} \sum_{i=1}^n \hat{y}_{i1}$ . Por su parte  $SRC$  es la suma de los residuales cuadrados dada por

$$SRC = \sum_{i=1}^n (y_{i1} - \hat{y}_{i1})^2$$

Se puede probar que el estadístico  $F$  (B.1) se obtiene del cociente de dos variables aleatorias independientes  $ji$ -cuadradas, cada una dividida entre sus grados de libertad correspondientes. La variable aleatoria  $ji$ -cuadrada del numerador tiene  $k + r$  grados de libertad, mientras que la variable aleatoria  $ji$ -cuadrada del denominador tiene  $n - k - r - 1$  grados de libertad. Por definición, esto es una variable aleatoria con distribución  $F$  con  $(k + r, n - k - r - 1)$  grados de libertad, lo cual se suele escribir como  $F \sim F_{k+r, n-k-r-1}$ .

Una vez estimado  $F$ , se define un nivel de significancia del  $c\%$  y si  $q$  es el percentil  $100 - c$  en la distribución  $F_{k+r, n-k-r-1}$ , entonces la hipótesis nula se rechaza a favor de la hipótesis alterna al nivel de significancia elegido si se satisface la siguiente desigualdad

$$F > q$$

Si no se puede rechazar la hipótesis nula, no hay evidencia de que alguno de los instrumentos excluidos ayude a explicar a  $y_1$ , mientras que al rechazarla, se dice que  $x_{k+1}, x_{k+2}, \dots, x_{k+r}$  son conjuntamente significativas al nivel de significancia correspondiente. De modo que, rechazar la hipótesis nula en la prueba  $F$  de la primera etapa indica que los instrumentos son relevantes y tienen una correlación significativa con la variable endógena, de lo contrario, no rechazar la hipótesis nula indica que los instrumentos pueden no ser suficientemente fuertes, es decir, no tienen una correlación significativa con la variable endógena.



### Prueba de Kleibergen-Paap rank Wald

Para evaluar que los instrumentos sean relevantes, se emplea la prueba de Kleibergen-Paap rank Wald, la cual es una prueba de subidentificación y de identificación débil. La subidentificación se refiere a que no existen suficientes instrumentos excluidos para identificar<sup>1</sup> los coeficientes de la ecuación (2.22), mientras que la identificación débil ocurre cuando los instrumentos excluidos están correlacionados con la variable endógena, pero esta correlación es muy baja.

La prueba establece la hipótesis nula de que la variable endógena no puede ser descrita a partir de los instrumentos, lo cual es equivalente a

$$H_0 : \alpha \equiv 0$$

donde  $\alpha$  es el vector de coeficientes para los instrumentos en la ecuación de la primera etapa (2.20). Mientras que la hipótesis alterna de la prueba es que al menos uno de los instrumentos es relevante para explicar la variable endógena, es decir, con  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_r)$ ,

$$H_a : \alpha_j \neq 0 \quad \text{para al menos un } \alpha_j$$

El estadístico para esta prueba está dado por

$$LM = n \frac{v^2}{1 + v}$$

donde  $n$  son las observaciones de la muestra y  $v = \|\alpha\| = \left( \sum_{j=1}^r \alpha_j^2 \right)^{\frac{1}{2}}$ .

Se asume que el estadístico tiene una distribución  $\chi^2$  con  $r$  grados de libertad. El rechazo de la hipótesis nula indica entonces que el modelo está correctamente identificado y por lo tanto, los instrumentos son relevantes.

### Prueba de Sargan - Hansen

La prueba de Sargan-Hansen es una prueba de restricciones de sobreidentificación; en el contexto de variables instrumentales, una ecuación se dice sobreidentificada cuando hay más instrumentos excluidos de los necesarios para identificar los coeficientes de la ecuación (2.22).

La hipótesis nula establece que los instrumentos son válidos, es decir, no correlacionados con el término de error  $u$ , y que los instrumentos están correctamente excluidos de la ecuación estimada en la segunda etapa.

---

<sup>1</sup>Entiendase a "identificar" como: escribir los coeficientes estimados en términos de varianzas y covarianzas de los instrumentos.

Por su parte, la hipótesis alterna sugiere que los instrumentos no son válidos en el modelo. Esto puede deberse a varias razones, tales como la presencia de correlación entre los instrumentos y los errores del modelo, o la inclusión de instrumentos irrelevantes que no capturan completamente la variabilidad de las variables endógenas.

El estadístico para la prueba de Sargan-Hansen es el estadístico de Sargan, definido como

$$J = nR^2$$

donde  $n$  son las observaciones de la muestra y  $R^2$  es el coeficiente de determinación ( $R$ -cuadrado) que se obtiene a partir de una regresión lineal que considera como variable dependiente a  $\hat{u}$ , residual del modelo de variables instrumentales (el valor de  $y$  real menos el valor de  $y$  predicho), y como variables explicativas al conjunto de instrumentos tanto incluidos como excluidos.

El estadístico  $J$  de la prueba tiene una distribución  $\chi^2$  con  $(r - 1)$  grados de libertad. Del rechazo de la hipótesis nula se concluye que por lo menos algunos de los instrumentos no son exógenos, de modo que se pone en duda la validez de las variables empleadas como instrumentos en la modelación.

# Bibliografía

- Abramsky, T., Lees, S., Stöckl, H., Harvey, S., Kapinga, I., Ranganathan, M., Mshana, G., & Kapiga, S. (2019). Women's income and risk of intimate partner violence: Secondary findings from the MAISHA cluster randomised trial in North-Western Tanzania. *BMC Public Health*, *19*, 1-15. <https://doi.org/10.1186/s12889-019-7454-1>
- Agüero, J. (2013). Causal Estimates of the Intangible Costs of Violence Against Women in Latin America and the Caribbean. <https://doi.org/10.2139/ssrn.2367687>
- Aguirre, E. (2023). Violencia doméstica e ingresos laborales de las mujeres en México. *Estudios Económicos de El Colegio de México*, *38*, 143-165. <https://doi.org/10.24201/ee.v38i1.438>
- Alonso-Borrego, C., & Carrasco, R. (2017). Employment and the risk of domestic violence: does the breadwinner's gender matter? *Applied Economics*, *49*. <https://doi.org/10.1080/00036846.2017.1299103>
- Angelucci, M. (2008). Love on the rocks: Domestic violence and alcohol abuse in rural Mexico. *B.E. Journal of Economic Analysis and Policy*, *8*. <https://doi.org/10.2202/1935-1682.1766>
- Arceo-Gómez, E. O., & Campos-Vázquez, R. M. (2014). Evolución de la brecha salarial de género en México. *El Trimestre Económico*, *81*, 619-653. <https://doi.org/https://doi.org/10.20430/ete.v81i323.125>
- Baranov, V., Cameron, L., Suarez, D. C., & Thibout, C. (2020). Theoretical Underpinnings and Meta-analysis of the Effects of Cash Transfers on Intimate Partner Violence in Low- and Middle-Income Countries. <https://doi.org/10.1080/00220388.2020.1762859>
- Becker, G. S. (1993). *A Treatise on the Family*. Harvard University.
- Bobonis, G. J., González-Brenes, M., & Castro, R. (2013). Public transfers and domestic violence: The roles of private information and spousal control. *American Economic Journal: Economic Policy*, *5*, 179-205. <https://doi.org/10.1257/pol.5.1.179>
- Contreras, J. M. (2008). La legitimidad social de la violencia contra las mujeres en la pareja. Un estudio cualitativo con varones en la ciudad de México. <https://biblioteca.clacso.edu.ar/Mexico/crim-unam/20100329121733/Estudiosobreculturagenero.pdf>

- Covarrubias, A. (2018). Poder, normas sociales y desigualdad de las mujeres en el hogar. *Nósis. Revista de Ciencias Sociales y Humanidades*, 27, 140-158. <https://doi.org/10.20983/noesis.2018.1.7>
- Díaz-Monroy, L. G. (2007). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia.
- Dildar, Y. (2021). Is Economic Empowerment a Protective Factor Against Intimate Partner Violence? Evidence from Turkey. *European Journal of Development Research*, 33, 1695-1728. <https://doi.org/10.1057/s41287-020-00311-x>
- Dugan, L., Nagin, D. S., & Rosenfeld, R. (1999). Explaining the Decline in Intimate Partner Homicide. *Homicide Studies*, 3, 187-214. <https://doi.org/10.1177/1088767999003003001>
- Eggers del Campo, I., & Steinert, J. I. (2022). The Effect of Female Economic Empowerment Interventions on the Risk of Intimate Partner Violence: A Systematic Review and Meta-Analysis. <https://doi.org/10.1177/1524838020976088>
- Fajardo-Gonzalez, J. (2021). Domestic violence, decision-making power, and female employment in Colombia. *Review of Economics of the Household*, 19. <https://doi.org/10.1007/s11150-020-09491-1>
- Guarnieri, E., & Rainer, H. (2018). *Female Empowerment and Male Backlash*. <https://doi.org/10.2139/ssrn.3198483>
- Hayashi, F. (2000). *Econometrics*. Princeton: Princeton University Press.
- INEGI. (2018). Estadísticas a propósito del Día Internacional de la Eliminación de la Violencia contra la Mujer (25 de noviembre)/ Datos nacionales. Comunicado de prensa Núm. 588/18. [https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2018/violencia2018\\_nal.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2018/violencia2018_nal.pdf)
- INEGI. (2021). Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) 2021. <https://www.inegi.org.mx/programas/endireh/2021/>
- INEGI. (2022a). Encuesta nacional sobre dinámica de las relaciones en los hogares (ENDIREH) 2021. Comunicado de Prensa, Núm. 485/22. [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2022/endireh/Endireh2021\\_Nal.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2022/endireh/Endireh2021_Nal.pdf)
- INEGI. (2022b). Sistema Integrado de Estadísticas de Violencia contra las Mujeres. <https://sc.inegi.org.mx/SIESVIM1/paginas/consultas/tablero.jsf>
- Johnson, M. P. (2008). *A Typology of Domestic Violence: Intimate Terrorism, Violent Resistance, and Situational Couple Violence*. Northeastern University Press.
- Kibris, A., & Nelson, P. (2022). Female income generation and intimate partner violence: Evidence from a representative survey in Turkey. *Journal of International Development*, 1-16. <https://doi.org/10.1002/jid.3713>

- 
- Kimmel, M. S. (2000). *The Gendered Society*. Oxford University Press.
- Kleibergen, F., & Paap, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, *133*, 97-126.
- Knaul, F. M., & Ramírez, M. Á. (2003). El impacto de la violencia intrafamiliar en la probabilidad de violencia intergeneracional, la progresión escolar y el mercado laboral en México. <https://ssrn.com/abstract=2053836>
- Martin, S. L., Moracco, K. E., Garro, J., Tsuia, A. O., Kupper, L. L., Chase, J. L., & Campbell, J. C. (2002). Domestic violence across generations: Findings from northern India. *International Journal of Epidemiology*, *31*, 560-572. <https://doi.org/10.1093/ije/31.3.560>
- McCloskey, L. (1996). Socioeconomic and coercive power within the family. *Gender and Society*, *10*, 449-463. <https://doi.org/10.1177/089124396010004006>
- Peña, D. (2002). *Análisis de Datos Multivariantes*. McGraw-Hill.
- Plazaola-Castaño, J., & Ruiz-Pérez, I. (2004). Violencia contra la mujer en la pareja y consecuencias en la salud física y psíquica. *Medicina Clínica*, *122*, 461-467. [https://doi.org/https://doi.org/10.1016/S0025-7753\(04\)74273-6](https://doi.org/https://doi.org/10.1016/S0025-7753(04)74273-6)
- Rao, V., & Bloch, F. (2002). Terror as a Bargaining Instrument a Case Study of Dowry Violence in Rural India. *American Economic Review*, *92*, 1029-1043.
- Rivara, F., Adhia, A., Lyons, V., Massey, A., Mills, B., Morgan, E., Simckes, M., & Rowhani-Rahbar, A. (2019). The effects of violence on health. *Health Affairs*, *38*. <https://doi.org/10.1377/hlthaff.2019.00480>
- Safranoff, A., & Tiravassi, A. (2018). La transmisión intergeneracional de la violencia. *Banco Interamericano de Desarrollo*. <https://webimages.iadb.org/publications/spanish/document/La-transmisi%C3%B3n-intergeneracional-de-la-violencia-Testimonios-desde-la-c%C3%A1rcel.pdf>
- Stock, J. H., & Yogo, M. (2002). *Testing for Weak Instruments in Linear IV Regression* (Working Paper N.º 284). National Bureau of Economic Research. <https://doi.org/10.3386/t0284>
- Stöckl, H., Hassan, A., Ranganathan, M., & Hatcher, A. M. (2021). Economic empowerment and intimate partner violence: a secondary data analysis of the cross-sectional Demographic Health Surveys in Sub-Saharan Africa. *BMC Women's Health*. <https://doi.org/https://doi.org/10.1186/s12905-021-01363-9>
- Torres-Jacome, E. Y., & Fernand-Desfrancois, P. G. (2021). Influencia de la violencia de género e intrafamiliar hacia la mujer en el desempeño laboral en Quito 2020. *Eruditus*, *2*, 9-28. <https://doi.org/https://doi.org/10.35290/re.v2n3.2021.451>
- Vyas, S., & Watts, C. (2009). How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published

evidence. *Journal of International Development*, 21, 577-602. <https://doi.org/10.1002/jid.1500>

Wooldridge, J. M. (2010). *Introducción a la econometría. Un enfoque moderno*. Cengage Learning Editores.

World Health Organization. (2021). Violence against women prevalence estimates, 2018: global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women. *World Report on Violence and Health*. <https://iris.who.int/bitstream/handle/10665/341337/9789240022256-eng.pdf?sequence=1>