



**UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**



**ESTIMACIÓN DEL VALOR ESTADÍSTICO DE LA VIDA EN CHILE
COMPARANDO MÉTODOS DE REGRESIÓN LINEAL Y TÉCNICAS DE
APRENDIZAJE ESTADÍSTICO**

POR

Luis Fernando Retamal Fuentes

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de
Concepción para optar al título profesional de Ingeniero Civil Industrial

Profesor Guía: Prof. Marcela Parada Contzen

Diciembre 2024
Concepción (Chile)

© 2024 Luis Fernando Retamal Fuentes

Sumario

Este estudio utiliza el método de salarios hedónicos para estimar el valor estadístico de la vida (VSL) en Chile, utilizando datos de la Encuesta de Caracterización Socioeconómica Nacional (CASEN) correspondientes a los años 2017 y 2022, así como registros de tasas de fatalidad proporcionados por la Superintendencia de Seguridad Social (SUSESO). Se emplearon diferentes métodos de estimación, incluyendo un método tradicional, un método de aprendizaje automático y dos métodos heurísticos. Los resultados indican que la mejor estimación se obtiene mediante el método de aprendizaje automático LASSO con 20 etapas, debido a su capacidad para penalizar las variables con menos relevancia y tener un mejor manejo de la multicolinealidad. El VSL obtenido fue de US \$8.86 millones. Estos resultados superan los obtenidos en estudios previos y presentan alta robustez y precisión según los análisis realizados.

Abstract

This study uses the hedonic wage method to estimate the statistical life value (VSL) in Chile, using data from the National Socioeconomic Characterization Survey (CASEN) for the years 2017 and 2022, as well as fatality rate records provided by the Superintendency of Social Security (SUSESO). Different estimation methods were employed, including a traditional method, a machine learning method, and two heuristic methods. The results indicate that the best estimate is obtained using the LASSO machine learning method with 20 stages, due to its ability to penalize irrelevant variables. The estimated VSL was US \$8.86 million. These results surpass those obtained in previous studies and exhibit high robustness and precision according to the analyses conducted.

Índice

1.	Introducción	6
1.1.	Objetivos de la memoria	7
1.1.1.	Objetivo general	7
1.1.2.	Objetivos específicos.....	7
1.2.	Alcances y limitaciones	8
1.3.	Organización del documento	8
2.	Revisión de literatura	9
2.1.	Valor estadístico de la vida usando datos del mercado laboral.....	9
2.1.1.	Estimación del valor estadístico de la vida.....	10
2.1.2.	Métodos empíricos	13
2.1.3.	Métodos de preferencia declarada	13
2.1.4.	Métodos de preferencia revelada.....	14
2.1.5.	Método de precios y salarios hedónicos	14
2.1.6.	Estimación del VSL en Chile	18
2.2.	Métodos de aprendizaje automático.....	20
2.2.1.	Etapas aprendizaje automático	20
2.2.2.	Clasificación de métodos de aprendizaje automático.....	20
2.2.3.	Aplicaciones de aprendizaje automático	21
2.2.4.	Método LASSO.....	22
2.2.5.	Aplicaciones de método LASSO.....	22
2.2.6.	Aplicación de LASSO en VSL.....	23
3.	Metodología	25
3.1.	Modelo empírico.....	25
3.2.	Método de estimación.....	25
3.2.1.	Método de Mínimos Cuadrados Ordinarios (MCO).....	25
3.2.2.	Método LASSO	26
3.2.3.	Supuesto de escasez.....	27
3.2.4.	Selección de lambda	27
3.2.5.	Otros métodos de estimación.....	29
3.3.	Test estadísticos	30
3.3.1.	Prueba F.....	30
3.3.2.	R-cuadrado	30
3.3.3.	R cuadrado ajustado	30
3.3.4.	Prueba t.....	30
3.3.5.	Valor de significación.....	31

4.	Datos	32
4.1.	Fuente de datos	32
4.2.	Creación de la muestra de estimación.....	33
4.3.	Descripción de la muestra de estimación.....	37
5.	Resultados	42
5.1.	Método OLS	42
5.1.1.	Análisis de robustez.....	46
5.2.	Método LASSO	50
5.3.	Comparación de resultados entre LASSO y métodos tradicionales	54
5.4.	Implicancia en políticas públicas	55
6.	Conclusión.....	57
7.	Referencias.....	58

Índice de tablas

Tabla 1: Resumen de la revisión de literatura	24
Tabla 2 : Tasa de accidentabilidad fatal en Chile año 2017 y 2022	34
Tabla 3: Descripción de variables	40
Tabla 4: Prueba número de variables datos 2022.....	42
Tabla 5: Resultados método OLS.....	44
Tabla 6: Estimación VLS método OLS (millones de dólares año 2017 y 2022)	45
Tabla 7: Resultados modelo OLS año 2022.....	46
Tabla 8: Análisis de multicolinealidad método OLS año 2022.....	48
Tabla 9: Primeras 5 iteraciones del método LASSO CV según datos de CASEN año 2022	50
Tabla 10: Coeficientes método LASSO según CASEN año 2022	52
Tabla 11: Valores de lambda para los diferentes métodos	53
Tabla 12: VSL mediante LASSO (millones de dólares)	54
Tabla 13: Comparación métodos de estimación año 2022.....	55

Índice de figuras

Figura 1: Densidad relativa variable LOG_WAGE	38
Figura 2: Densidad relativa de variable FATALITY_RATE.....	38
Figura 3: Cantidad de encuestados por región del país	39
Figura 4: Salario promedio por nivel educacional	39
Figura 5: Evolución de los coeficientes en el modelo LASSO	51

1. Introducción

La evaluación del costo de una vacuna, las regulaciones laborales, la asignación de recursos e incluso la preparación ante un desastre natural son razones clave para medir el valor de las vidas humanas con precisión (Ministerio de Desarrollo, 2017). Una herramienta útil en este proceso es el valor estadístico de la vida (abreviado habitualmente como VSL, por sus siglas en inglés). Este método estima los beneficios obtenidos por la disminución del riesgo de muerte o, el costo asociado con la pérdida de una vida humana calculado a partir de la cantidad de dinero máxima que las personas están dispuestas a pagar para disminuir dicho riesgo (COFEMER, 2013). El VSL permite a las autoridades competentes evaluar de manera cuantitativa las intervenciones en seguridad, medio ambiente y salud, justificando eficientemente cada una de las inversiones (Ministerio de Desarrollo Social, 2017).

Existen diversos métodos para estimar el VSL. Uno de los métodos más tradicionales es el de salarios hedónicos, que utiliza datos del mercado laboral considerando salarios y niveles de riesgos en los puestos de trabajo para estimar el trade-off entre ingresos y riesgos (Viscusi, 1993). Este método utiliza las elecciones observadas de los agentes económicos para revelar información sobre su disposición a pagar marginalmente por reducir los riesgos de fatalidad (Evans & Taylor, 2020).

Algunas de las limitaciones que existen para el modelo de salarios hedónicos son las limitaciones de los datos. Esto se debe a que los datos disponibles para describir las características de los trabajadores y los atributos laborales no son exhaustivos, lo que significa que los modelos de salarios hedónicos suelen sufrir un sesgo por omisión de variables (Evans & Taylor, 2020). Una de las formas recientes y prácticas para analizar grandes cantidades de datos es el aprendizaje automático. Esto corresponde a técnicas para analizar datos que implica enseñar a las máquinas a responder preguntas o localizar hallazgos, y luego probar esos resultados para determinar su precisión (Awad, 2023). Con el uso de técnicas de aprendizaje automático se puede analizar y predecir los sectores de demanda de empleo para el futuro (Alsayed & Awad, 2023). Según algunos estudios como Motamedi (2022), una de las técnicas de aprendizaje automático más exitosas para acelerar la extracción de las mejores características de grandes conjuntos de datos es el uso del operador de contracción y selección mínima absoluta (LASSO, por sus siglas en inglés).

En la presente memoria de título se exploran los diversos métodos de estimación del valor estadístico de la vida y se implementan técnicas de aprendizaje automático. El pilar central de esta investigación es la aplicación del método LASSO junto con enfoques de machine learning para mejorar la precisión y la eficiencia en el análisis. Este informe incluye la recolección y preparación de datos, la aplicación del método LASSO, junto a la evaluación de su desempeño. Finalmente se realiza una comparación de los métodos tradicionales frente al método propuesto.

En esta memoria se contribuye a la innovación en políticas públicas y en particular a la estimación del VSL. Superando limitaciones de métodos tradicionales y ofreciendo una idea innovadora para el análisis de grandes volúmenes de datos. Desde una perspectiva científica mejora las estimaciones y enriquece la literatura con nuevas metodologías en este campo. Profesionalmente, optimiza la toma de decisiones en áreas tan críticas como la seguridad y salud, y proporciona una base sólida para la capacitación de nuevos profesionales en el uso de técnicas avanzadas de análisis de datos.

Para esta memoria se utilizaron datos de la encuesta de caracterización socioeconómica nacional (CASEN) y datos de la Asociación Chilena de Seguridad (ACHS). Luego se prepararon los datos y se utilizó el método de salarios hedónicos para estimar el valor estadístico de la vida. Luego se formulará y aplicará el método LASSO para la estimación del VSL. Se evaluarán los resultados obtenidos y se compararon ambos métodos.

1.1. Objetivos de la memoria

1.1.1. Objetivo general

Comparar la estimación del valor estadístico de una vida (VSL) para Chile utilizando métodos de estimación LASSO y métodos tradicionales de salarios hedónicos de regresión lineal múltiple.

1.1.2. Objetivos específicos

Estimar el VSL utilizando el método de salarios hedónicos para datos de Chile y analizar el rendimiento de la estimación, en términos de precisión y robustez.

Aplicar el método de estimación LASSO a los mismos datos para calcular el VSL, adaptando el método a las características específicas de un conjunto de datos.

Comparar los resultados obtenidos mediante LASSO con los resultados de métodos tradicionales, destacando las diferencias en la precisión, interpretabilidad y aplicabilidad de cada enfoque.

1.2. Alcances y limitaciones

Este estudio se enfoca en la estimación del VSL utilizando el método de salarios hedónicos en Chile con información observada del mercado laboral y aplicando el método LASSO en comparación con métodos tradicionales utilizando datos de la encuesta CASEN del período comprendido entre los años 2017 a 2022.

Algunas de las limitaciones del estudio se dieron al momento de elaborar la muestra y variables para los modelos. En primer lugar, dado que las encuestas CASEN 2017 y CASEN 2022 no tenían el mismo nombre para las preguntas de ciertos datos, entonces se utilizaron diferentes archivos do-file según sea el año de la encuesta analizada. Otro problema se presentó al momento de ejecutar los modelos ya que muchas variables tenían archivos faltantes por lo que esas observaciones no se consideraban en el análisis. Esto implicó una notable disminución de la muestra analizada. Además, al existir pocos estudios para el caso chileno y la diferencia de estimación entre los pocos que hay, entonces se debió analizar bastante literatura para comprender mejor la forma de estimación.

1.3. Organización del documento

El resto del documento se organiza de la siguiente forma. En el capítulo 2 se hace una revisión exhaustiva a la literatura, se revisan documentos y artículos donde se estima el VSL, se observan y comparan otros métodos ya realizados. En el capítulo 3 se describen los métodos de estimación utilizados y las técnicas aplicadas en el estudio. En el capítulo 4 se da una descripción de los datos utilizados. En el capítulo 5 se presentan y analizan los resultados obtenidos a partir de los métodos aplicados. En el capítulo 6 se concluyen los hallazgos del estudio.

2. Revisión de literatura

2.1. Valor estadístico de la vida usando datos del mercado laboral

El valor estadístico de la vida representa cuánto estaría dispuesta a gastar la sociedad para evitar una muerte no identificada (Andersson, 2020). Al ser la vida de una persona un valor no transable en el mercado es necesario recurrir a métodos indirectos de valorización. Uno de los enfoques más utilizados es el análisis del mercado laboral (Gentry, 2016), que se basa en el método de salarios hedónicos. Este método estima el VSL observando cómo los trabajadores son compensados por asumir riesgos adicionales en su lugar de trabajo.

El informe de seguridad vial de la OMS indica que 1.35 millones de personas mueren en el mundo por accidentes de tránsito (World Health Organization, 2018). Estos accidentes son un problema que los gobiernos y las autoridades locales intentan mitigar, lo que plantea la interrogante de por qué no se busca eliminar completamente las muertes o reducirlas a cero. Sin embargo, esto tiene un costo asociado, el cual debe ser analizado por las autoridades y ver si los beneficios obtenidos son mayores a los costos. Hay varias formas de medir este costo, uno de ellos es el análisis costo-beneficio (Wim Wijnen, 2009). En este contexto se debe asignar un valor monetario a salvar vidas, además este análisis permite orientar la asignación de recursos entre sectores y para políticas con objetivos múltiples. Otra forma de medir el costo asociado a eliminar las muertes o reducirlas es el análisis costo-efectividad. Este análisis “permite una asignación de recursos más eficiente al identificar las políticas de seguridad más rentables” (Andersson, 2020).

La disposición a pagar por seguridad vial y por ende por menos muertes tiene dos tipos de métodos de medición, uno de ellos es el método de preferencia revelada y otro es el método de preferencia declarada (Wim Wijnen, 2009). Los métodos de preferencia revelada valoran las reducciones de riesgo basándose en el comportamiento real, por ejemplo, cuánto dinero gastan realmente las personas en medidas de seguridad. Mientras que los métodos de preferencia declarada utilizan cuestionarios en los que se pregunta directamente a personas cuánto están dispuestos a pagar. Siendo los métodos de preferencia revelada menos aplicables a la realidad, esto se debe a que muchos de los implementos comprados para la seguridad como airbag o cinturones de seguridad vienen incluidos en el automóvil. Por ende, los métodos de preferencia declarada son más utilizados ya que no requieren de gran cantidad de

datos y se pueden utilizar para otros fines como medición de tiempos de viaje. Entre los métodos más usados de preferencia declarada se encuentra el método de valoración contingente, el cual consiste en pedir a individuos de una muestra representativa de la población en riesgo que declaren su disposición a pagar.

Los valores de la disposición a pagar (WTP, por sus siglas en inglés) pretenden reflejar las preferencias, percepciones y actitudes hacia el riesgo de los afectados por las decisiones en las que se van a utilizar los valores (De Blaeij, 2003). Este riesgo puede medirse como la probabilidad de un accidente fatal cada 100.000 habitantes. El valor estadístico de la vida se puede considerar como una medida de la tasa marginal de sustitución entre riesgo de mortalidad y riqueza, y se calcula comúnmente a partir de la disposición a pagar de un individuo por pequeñas reducciones en su riesgo de muerte (Robinson et al., 2018). El nombre dado “valor estadístico de la vida” puede sonar un poco controversial para algunas personas (Cameron, 2010), creyendo que se refiere a dar valor a salvar una vida identificada.

2.1.1. Estimación del valor estadístico de la vida

El modelo teórico del VSL se basa en la idea que cada individuo actuará como maximizador de la utilidad esperada (Jones-Lee, 1974). Esto quiere decir que si tuviéramos la probabilidad p de supervivencia y $u(w)$ represente las utilidades del estado de riqueza (w). Luego sean los estados de permanecer vivo ($s=a$) o morir ($s=d$) entonces si cada individuo elige la decisión que maximiza la utilidad esperada esto se representaría en la ecuación (1).

$$EU(w, p) = (1 - p)u_d(w) + pu_a(w) \quad (1)$$

En la ecuación (1) la expresión $EU(w, p)$ representa la utilidad asociada a la riqueza w , luego $u_a(w)$ es la utilidad condicionada a la ocurrencia de la decisión vida, mientras que $u_d(w)$ representa utilidad condicionada a la ocurrencia de la decisión muerte. Las expresiones $u_a(w)$ y $u_d(w)$ son funciones únicas y al menos dos veces diferenciables (Jones-Lee, 1974). La expresión “ p ” en la ecuación representa la probabilidad de permanecer vivo. Una suposición que se debe considerar es que el individuo quiere más riqueza antes que menos riqueza además tiene aversión o rechazo al riesgo por ende se deben considerar las derivadas de la función $u_a(w)$ tal y como se presenta la ecuación (2) y en la ecuación (3).

$$\frac{du_a(w)}{dw} > 0 \quad (2)$$

$$\frac{d^2u_a(w)}{dw^2} < 0 \quad (3)$$

La ecuación (2) muestra que la utilidad $u_a(w)$ es una función creciente respecto a la riqueza w . Por lo tanto, mientras aumenta la riqueza la utilidad también se incrementa. La ecuación (3) muestra que hay rendimientos decrecientes de la utilidad respecto a la riqueza. Esto quiere decir que mientras la riqueza aumenta la tasa a la que aumenta la utilidad empieza a disminuir. Esto muestra la aversión al riesgo donde los individuos prefieren evitar riesgos significativos, dado que los beneficios marginales de la riqueza disminuyen a medida que acumulan más. Otro aspecto a considerar es que para un determinado nivel de riqueza el individuo preferiría el estado de vida antes que el estado de muerte tal y como muestra la ecuación (4).

$$u_a(w) > u_d(w) \quad (4)$$

Por último, se introduce una hipótesis bastante fuerte la cual podría ignorarse en otros análisis, pero servirá para este. Se supondrá que el individuo muestra una sensibilidad más pronunciada a las variaciones de la riqueza o del riesgo financiero si espera vivir en lugar de morir durante el período en curso. El excedente compensatorio o disposición a pagar (WTP), y el excedente equivalente o disposición a aceptar (WTA), para un cambio en el riesgo de mortalidad Δp (Freeman, 2014), se puede derivar de la ecuación (1). Definiendo EU_0 según la ecuación (1) y $C(\varepsilon)$ como la disposición a pagar para la reducción del riesgo ε , entonces $C(\varepsilon)$ se obtiene de la ecuación (5).

$$(p + \varepsilon)u_a(w - C(\varepsilon)) + (1 - p - \varepsilon)u_d(w - C(\varepsilon)) = EU_0 \quad (5)$$

Y de forma similar si dejamos que $P(\varepsilon)$ denote la WTA, la cual refleja la cantidad mínima que una persona está dispuesta a aceptar como compensación por un aumento en el nivel de riesgo. Luego el nivel de riesgo viene definido por ε por lo tanto $P(\varepsilon)$ se puede determinar según la ecuación (6).

$$(p + \varepsilon)u_a(w + P(\varepsilon)) + (1 - p - \varepsilon)u_d(w + P(\varepsilon)) = EU_0 \quad (6)$$

De las ecuaciones (5) y (6) se desprende que mientras mayor sea el cambio en el riesgo, mayor será la WTP o la WTA. Sin embargo, se espera que el riesgo ε sea pequeño y la WTA

y WTP casi iguales y proporcionales a ε (Hammitt, 2000). Como se mencionó anteriormente el VSL es la media poblacional de la tasa marginal de sustitución entre la riqueza y el riesgo de mortalidad por ende puede tomarse el límite de la WTP o WTA cuando $\varepsilon \cong 0$. Dado lo anterior la estimación del VSL se muestra en la ecuación (7).

$$VSL = -\left.\frac{dw}{dp}\right|_{EU \text{ constant}} = \frac{u_a(w) - u_d(w)}{p \frac{du_a(w)}{dw} + (1-p) \frac{du_d(w)}{dw}} \quad (7)$$

En la ecuación (7) el término $-dw/dp$ indica que se calculará la derivada de la riqueza w con respecto a la probabilidad de muerte p manteniendo la utilidad EU constante. Esto muestra cuánto se debe cambiar la riqueza de una persona para compensar un cambio infinitesimal en la probabilidad de muerte, de manera tal, que su utilidad esperada permaneciera constante. Dados los supuestos para las ecuaciones (2) y (3) entonces el VSL es siempre positivo. La ecuación (7) es definida de forma similar por otros autores tales como Viscusi (2003), con ligeros cambios en la terminología utilizada, sin embargo, expresa lo mismo. En muchas aplicaciones empíricas la ecuación (7) no es aplicable directamente puesto que requiere muchos datos. Simplemente se preguntaría a los encuestados cuál es su disposición a pagar para un pequeño cambio en el riesgo (Δp) y el VSL en este caso sería dado por la ecuación (8).

$$VSL = \frac{WTP}{(\Delta p)} \quad (8)$$

En la ecuación (8) la disposición a pagar es proporcional al cambio en el riesgo, pero debe interpretarse solo como una aproximación ya que la verdadera relación es casi proporcional. Esto quiere decir que la WTP puede entregar solo una aproximación del VSL, ya que no siempre refleja perfectamente las preferencias de los individuos (Hammitt, 2000).

Esta forma de medición del VSL fue utilizada en un estudio en la India (Bharti et al., 2022), donde se consultó a 401 motociclistas de Patna. En este caso se consideró el pago por un casco con mayor previsión de lesiones. Se consideraron otros factores como un casco básico y distintas formas de pago como tarjeta o incluso dinero de los propios encuestados. Además, se consideraron medidas necesarias como cierto nivel educacional para mejorar el rango de respuestas de acuerdo con las necesidades de la entrevista. La reducción del riesgo de muerte, según las estimaciones, era del 50%, lo que implicaba que la disposición a pagar por un casco

mejorado estaba orientada a disminuir las muertes de 15 por cada cien mil motociclistas a solo 7,5 muertes. El VSL correspondiente a esta disposición a pagar se presenta en la ecuación (9).

$$VSL = \frac{346.13}{7.5 \times 10^{-5}} = 46.15 \times 10^5 \quad (9)$$

En la ecuación (9) se muestra que el cambio de riesgo o la variación del riesgo es 7.5/100,000 dada la disminución de 7.5 muertes por cada 100 mil personas. El numerador presenta la disposición a pagar de 346.13 rupias que fue el valor promedio que respondió el total de motociclistas. Finalmente, la estimación del VSL es de 46.15×10^5 rupias lo que equivale a 61,810 dólares. Luego a través del método indirecto de elección declarada se obtuvo un VSL de 76,600 rupias lo que equivale a unos 1,030 dólares. En otro estudio realizado en U.S.A. (Mon et al., 2018), el resultado fue de 135,712 dólares lo cual indica que el VSL es mucho menor en la India. Esta diferencia de más de 70,000 dólares entre el método directo en India y el caso de U.S.A puede deberse a elementos clave como el costo de vida, cultura, percepción del riesgo entre otros factores.

2.1.2. Métodos empíricos

Como se indicó anteriormente, no existen precios de mercado específicos para calcular el valor estadístico de la vida. Por ello, los analistas deben emplear técnicas que utilicen datos fuera del ámbito comercial, como escenarios hipotéticos. Estas técnicas se agrupan en dos tipos: los métodos de preferencia revelada, basados en datos observados en el mercado, y los métodos de preferencia declarada, que recurren a situaciones hipotéticas para captar las preferencias individuales (Andersson, 2020).

2.1.3. Métodos de preferencia declarada

El método de preferencias declaradas se trata de encuestas que crean situaciones o mercados hipotéticos y se le pide a cierto grupo de personas que declaren sus preferencias ante estos mercados. Entre los principales métodos se encuentran el método de valoración contingente y los experimentos de elección (Remoundou et al., 2009). Los participantes se enfrentan a escenarios sin consecuencias reales. Además, requiere una buena explicación y entendimiento de los encuestados ya que deben responder como si fuera una situación real. Incluso se puede indicar, tal y como se señala en algunos artículos (Carson, 2014). que las

preferencias de la encuesta hipotética pueden influir en alguna decisión de política del mundo real. A pesar de las tantas problemáticas que pueden presentar los métodos de preferencia declarada, son los métodos que más información pueden entregar ante distintos escenarios debido al alto control que tiene el analista que crea la encuesta. Incluso pueden obtenerse datos de preferencias que no existen en el mercado actualmente.

2.1.4. Métodos de preferencia revelada

Estos métodos se basan en el hecho de que los individuos en los mercados toman decisiones que revelan hasta qué punto están dispuestos a intercambiar otros bienes (Andersson, 2020). Dentro de las técnicas no monetarias estos métodos implican observar el comportamiento o analizar documentos de forma tal que se pueda determinar indirectamente las preferencias de las personas. Cuando se utilizan métodos que involucran precios de mercado la disposición a pagar refleja el límite inferior de las preferencias de la persona (Andersson, 2020), esto se muestra en la ecuación (10) donde Δp refleja un pequeño cambio en el riesgo.

$$WTP < VSL \cdot \Delta p \quad (10)$$

Uno de los principales métodos de preferencia revelada a partir de mercados laborales es el método de los salarios hedónicos.

2.1.5. Método de precios y salarios hedónicos

Los precios hedónicos se definen como ‘los precios implícitos de los atributos’, y se revelan a los agentes económicos a partir de los precios observados de productos diferenciados, y de las cantidades específicas de características asociadas con ellos (Rosen, 1974). Este método fue utilizado para estimar el valor de la seguridad de los coches utilizando datos del mercado estadounidense resultando un valor de vida estadística de US \$3,375,000 dólares (Atkinson, 1990). Además, tiene otras aplicaciones como el análisis de los trabajadores que requieren una compensación por aceptar trabajos con mayor riesgo (Gentry, 2016). Incluso se utilizó para estudios ambientales como por ejemplo para demostrar que la calidad del agua tiene un efecto significativo en el valor de las propiedades a lo largo de la bahía de Chesapeake (Leggett & Bockstael, 2000). En esta última aplicación se determinaron los beneficios potenciales de una mejora en la calidad del agua e intentaron incluir variables omitidas por otros estudios similares. Finalmente llegaron a la conclusión que los propietarios de viviendas frente al mar tienen una disposición positiva a pagar por la reducción de las emisiones de

gases de efecto invernadero.

La formalización del modelo de precios hedónicos se planteó en el artículo “Precios hedónicos y mercado implícito” de Rosen (1974). En este artículo se plantea que dado un mercado competitivo con individuos maximizadores de utilidad y empresas maximizadoras de beneficios. Dado el precio P de un producto Q (q_1, q_2, \dots, q_k), donde el vector tiene los atributos del bien dado por la ecuación (11).

$$P = P(q_1, q_2, \dots, q_k) \quad (11)$$

Luego en el equilibrio la disposición a pagar por un atributo concreto (q_k) será dado por la ecuación (12).

$$MWTP_{q_k} = \frac{\partial P(Q)}{\partial q_k} \quad (12)$$

La ecuación (12) muestra el precio implícito, que viene dado por la derivada parcial de la función de precio hedónico con respecto a dicho atributo. Esta ecuación estima el efecto sobre el precio de las variaciones en el atributo de interés, controlando al mismo tiempo otros atributos que influyen en el precio. Una aplicación de la ecuación (12) es en la seguridad de un vehículo si q_k representa una característica de seguridad del coche que reduce el riesgo de muerte en caso de accidente entonces MWTP puede interpretarse como el VSL (Andersson, 2020). Sin embargo, el acceso a un atributo que represente esta característica puede no resultar posible, por ende, se accede a estadísticas de accidentes y se observa la relación con precios de las características de seguridad.

Suponiendo que q_k , F y D denotan el riesgo real, la tasa de accidentes mortales y un vector de características de los accidentes respectivamente. Si F es una función tanto de q_k como de D , luego $F=g(q_k, D)$ y la relación entre F y q_k es monótona entonces $q_k=g^{-1}(F, D)$. Por lo tanto, en lugar de estimar la función de precios con el verdadero riesgo presentado en la ecuación (13), donde q_k puede no estar disponible entonces se muestra en la ecuación (14).

$$P = P(q_k, Q_{-k}) \quad (13)$$

$$P = P(F, D, Q_{-k}) \quad (14)$$

Por lo tanto, el VSL se estima según la ecuación (15). En este caso el signo menos se incluye

para garantizar que el VSL es positivo ya que P es una función del riesgo y no de la seguridad.

$$VSL = - \frac{\partial(F, D, Q_{-k})}{\partial F} \quad (15)$$

Según el mismo autor (Andersson, 2020) esta forma de la ecuación es poco aplicable en la vida real además que se describe en función de la seguridad vial, y las formas funcionales más comunes son semilogarítmicas y log-lineal.

El concepto de salarios hedónicos se puede ilustrar de una forma práctica. Lo primero a considerar es que los trabajos para cualquier persona son en función de sus preferencias salariales y del riesgo que quieran tomar (Viscusi W. K., 2003). Si $U(w)$ representa la utilidad de un trabajador sano con salario w . Además, sea $V(w)$ la utilidad de un trabajador lesionado con un salario w . Comúnmente la compensación que tienen los trabajadores luego de un accidente es función de su trabajo. Se supone que los trabajadores prefieren estar sanos antes que lesionados o enfermos [$U(w) > V(w)$] y que la utilidad marginal de la renta es positiva [$U'(w) > 0, V'(w) > 0$]. Además, los trabajadores eligen las combinaciones de salario-riesgo que maximicen su utilidad esperada, entonces la utilidad esperada del trabajador se representa por Z y se muestra en la ecuación (16).

$$Z = (1 - p)U(w) + pV(w) \quad (16)$$

$$\frac{dw}{dp} = - \frac{Z_p}{Z_w} = \frac{U(w) - V(w)}{(1 - p)U'(w) + pV'(w)} > 0 \quad (17)$$

Luego el equilibrio a lo largo de la curva de la ecuación (16) está dado por la ecuación (17). En esta ecuación el primer término indica cuánto debe cambiar el salario para compensar un cambio en el riesgo. Luego Z_p podría interpretarse como la disminución en la utilidad debido a un aumento en la probabilidad de un evento riesgoso, como una lesión. Z_w podría interpretarse como el incremento en la utilidad de un trabajador derivado de un aumento en el salario. La letra p indica la probabilidad de estar lesionado. $U'(w)$ representa cómo cambia la utilidad marginal de un trabajador sano con respecto a un cambio en el salario. $V'(w)$ representa el cambio en la utilidad marginal de un trabajador lesionado con respecto a un cambio en el salario. Finalmente, todo debe ser mayor a 0 ya que si la probabilidad de sufrir un daño o estar lesionado aumenta entonces por definición el salario debe aumentar.

En la ecuación (17) se plantea que la compensación entre salario y riesgo es, en consecuencia,

igual a la diferencia en los niveles de utilidad en los dos estados dividida por la utilidad marginal esperada del ingreso (Viscusi W. K., 1993). Y esta compensación es la que utilizan los economistas para calcular el VSL utilizando los coeficientes de una función de salarios hedónicos.

La ecuación canónica de salarios hedónicos utilizada en los cálculos del VSL es presentada en la ecuación (18).

$$\ln(w_{ijk}) = \alpha_1 \pi_{ijk} + X_{ijk} \gamma + u_{ijk} \quad (18)$$

En la ecuación (18) para un trabajador i en la industria j y cuya ocupación es k , el $\ln(w_{ijk})$ es el logaritmo natural o neperiano de la tasa salarial por hora, π representa la tasa de fatalidad relacionada con el trabajo. X es un vector que contiene tanto variables demográficas como variables de características de trabajo. Finalmente, u_{ijk} es un término de error que puede exhibir heterocedasticidad condicional y dentro de la autocorrelación del riesgo de fatalidad, lo cual debe reflejarse en los errores estándar calculados de los coeficientes (Kniesner & Leeth, 2010).

Un ejemplo de utilización de la ecuación de salarios hedónicos es en el artículo llamado “Una aplicación del Método de Salarios Hedónicos” (Riera Font et al., 2007). En este artículo se estima el valor del VSL a partir de una amplia base de datos del mercado de trabajo español. En esta base de datos se incluye la ocupación y la rama de actividad para los periodos del año 1997 al año 2000 de trabajadores españoles. En ese artículo se adoptó la forma funcional semilogarítmica para la ecuación de los salarios hedónicos la cual se presenta en la ecuación (19).

$$\ln(w_i) = \beta_0 + \beta'_x X + \beta'_r R + u_i \quad (19)$$

En la ecuación (19) w_i representa el salario para cada individuo i , β_0 es la constante, X recoge los factores determinantes del salario correspondientes a las características individuales y del puesto de trabajo, β_x es el vector de parámetros de las variables de la matriz X . Para esta ecuación R representa los índices de riesgo mortal y no mortal incorporados al modelo, β_r son los coeficientes estimados para los índices de riesgo y u_i es el término de error (Riera Font et al., 2007).

Luego para determinar el VSL se utiliza la ecuación (20). Obteniéndose finalmente para el

periodo de 1997 a 2000 un valor de VSL de 2.69 millones de euros.

$$VSL = \frac{\partial(w_i)}{\partial(R_i)} \cdot 1000 \cdot (\bar{w}) \cdot \frac{\partial(w_i)}{\partial(R_i)} \cdot 1000 \cdot (\bar{w}) \cdot \beta_r \cdot 1000 \quad (20)$$

2.1.6. Estimación del VSL en Chile

Para países en vías de desarrollo no hay mucha experiencia en la aplicación de métodos para estimar el valor estadístico de la vida (Cifuentes, 1999). Y de las pocas veces que se llega a estimar se utiliza el enfoque del capital humano de forma poco coherente (Rizzi & Ortúzar, 2003).

Para el caso de Chile hay varios estudios que intentan una aproximación del valor estadístico de la vida, un ejemplo de ello es el artículo “The value of a statistical life in Chile”, donde se estima el VSL según datos de la CASEN. Los resultados dieron un valor de US \$4,625,958 (Parada-Contzen et al., 2012). Este valor resultó evidentemente menor a la situación de países desarrollados pero superior al estimado por estudios anteriores para la región.

Otro estudio (Mardones & Riquelme, 2018) estimó el valor estadístico de la vida utilizando el método de salarios hedónicos y lo extrapolo a otros países de la región. Los datos se obtuvieron de la CASEN 2013 y de la Superintendencia de Seguridad Social. Los resultados dieron un valor del VSL de US \$3.7 millones. Luego se extrapolaron para otros países de la región usando el PIB per cápita y varían desde US \$0.01 a US \$5.2 millones.

En el estudio de Rizzi & Ortúzar (2003) se utilizó el método de preferencias declaradas con el fin de evaluar el valor de una vida estadística para las autopistas interurbanas chilenas. Se presentaban rutas de un viaje hipotético. Las personas optan por rutas más seguras lo que se veía reflejado en la variable riesgo. Se observaron diferentes rangos de valor para el VSL según factores y tipos de personas. Como aquellas con comportamiento lexicográfico (priorizan ciertos factores de forma absoluta) donde el VSL varía desde US \$650,000 a US \$1,300,000. Es necesario recalcar que a diferencia del estudio de Parada-Contzen (2012) y Mardones & Riquelme (2018), el estudio de Rizzi y Ortuzar es bastante antiguo por ende el VSL estimado es bastante menor.

En el estudio de Cifuentes (1999) se preguntó directamente la disposición a pagar por una reducción de cantidad de muertes en varios rangos de tiempo. Esta encuesta fue aplicada a

28 estudiantes de la carrera de Ingeniería de la Universidad Católica de Chile y se obtuvo el valor de VSL de US \$413,703 por la reducción de muertes de 1 en 1000 por los próximos 10 años. Luego para el caso de 5 en 1000 para los próximos 10 años se redujo el VSL a US \$310,277. Al final del estudio se compara el valor obtenido con los métodos comunes de esa época (método del capital humano), y se obtuvo un valor 5 veces mayor ya que en general los resultados de esa época daban un VSL de US \$66,000.

En el estudio de Iragüen & Ortúzar (2004) se analizó y diseñó un experimento de preferencia declarada (SP, por sus siglas en inglés) para estimar la disposición a pagar (DAP) para reducir el riesgo de accidentes fatales en áreas urbanas. Se presentaron diferentes rutas de acuerdo al tiempo de viaje, costo y número de accidentes fatales de automóviles por año. En este estudio se obtuvo un valor de VSL de US \$274,500. Pero dados los intervalos de confianza muy grandes se recomienda según los autores considerar el valor más bajo para el VSL que corresponde a US \$125,000.

Hojman, Ortuzar, & Rizzi (2005) configuraron una encuesta a través de una página web en la que los individuos debían elegir entre dos rutas para un viaje hipotético entre dos ciudades. Las variables consideradas en el estudio fueron: tiempo de viaje, tasa de peaje, víctimas mortales al año y víctimas gravemente heridas al año. Se estableció un nivel base y cada una de las opciones tenía variaciones respecto a ese nivel. Por ende habían opciones con 20 minutos menos de viaje pero con mayor costo de peaje, adicionalmente tenían 4 víctimas mortales menos. Al final el VSL obtenido fue de US \$300,000.

El Instituto de Sociología de la Universidad Católica de Chile realizó un informe (GreenLabUC, 2014) en el que estimaba el valor estadístico de la vida en dos contextos distintos. Uno de ellos estaba relacionado con el riesgo asociado al tráfico, diseñaron una encuesta de elección declarada que fue respondida por una muestra aleatoria estratificada de hogares del área metropolitana de Santiago. Se utilizaron varios modelos obteniendo un VSL entre 3.47 a 4.65 millones de dólares. En otro estudio (Rizzi & De La Mazza, 2017) consideran que el valor del VSL óptimo a utilizar para ese estudio era de 4 millones de dólares aproximadamente.

La OCDE también plantea un valor de VSL para ser utilizado en Chile. Esto se presenta en su informe (OECD, 2014) donde mencionan los valores para varios países como Austria,

Canadá y Finlandia. La estimación se realizó en base a las muertes por contaminación ambiental. Para el caso de Chile estima que para 2005 el VSL fue de 1.5 millones de dólares y para 2010 fue de 1.9 millones de dólares.

2.2. Métodos de aprendizaje automático

Ante la gran cantidad de datos disponibles en la actualidad se hace necesario obtener métodos y técnicas más eficientes que puedan organizarlos y comprenderlos para un mejor manejo de la información. Aquellos sistemas capaces de aprender tanto de los datos como de los cambios en los datos para resolver problemas se denominan métodos de aprendizaje automático o machine learning (Farhat et al., 2020). Algunos autores lo describen como el “estudio científico de algoritmos y modelos estadísticos que utilizan los sistemas informáticos para realizar una tarea específica sin estar programados explícitamente” (Batta, 2020).

2.2.1. Etapas aprendizaje automático

El aprendizaje automático tiene 6 etapas independiente del algoritmo utilizado (Alzubi et al., 2018). Estas 6 etapas son recopilación y preparación de datos, selección de características, elección del algoritmo, selección de modelos y parámetros, entrenamiento y evaluación del desempeño. En general se inicia por la recopilación de los datos luego se escogen aquellas características que no son relevantes para el aprendizaje y se eliminan, luego se escoge el mejor algoritmo para el problema presente. Posteriormente se realiza una intervención inicial para establecer los valores más apropiados de varios parámetros. Se entrena el modelo con algunos de los datos disponibles y finalmente el modelo debe probarse con datos no vistos para evaluar cuánto se ha aprendido utilizando varios parámetros de desempeño como precisión, exactitud y recuperación.

2.2.2. Clasificación de métodos de aprendizaje automático

El aprendizaje automático se puede clasificar según el enfoque utilizado para el proceso de aprendizaje. Podemos encontrar cuatro categorías principales: aprendizaje supervisado, no supervisado, semisupervisado y de refuerzo. Según Farhat (2020) en los métodos de aprendizaje supervisado tenemos un conjunto de datos de entrenamiento en los que conocemos la estructura y el resultado. Tomamos estos datos y entrenamos un modelo de aprendizaje automático para que pueda comprender patrones. Una vez que se ha entrenado el

modelo, podemos usarlo para predecir resultados de datos en los que se desconocen los resultados. Los métodos de aprendizaje no supervisado aprenden la estructura de los datos en sí, sin necesidad de un etiquetado previo encontrando patrones. El aprendizaje semisupervisado proporciona un marco poderoso para aprovechar los datos no etiquetados cuando las etiquetas son limitadas o costosas de obtener. Los métodos por refuerzo sirven cuando sabemos lo que buscamos, pero no sabemos cómo conseguirlo. Existen otros enfoques o paradigmas (Alzubi et al., 2018) como por ejemplo aprendizaje evolutivo, aprendizaje en conjunto, red neuronal artificial, aprendizaje basado en instancias, algoritmos de reducción de dimensionalidad y aprendizaje híbrido.

2.2.3. Aplicaciones de aprendizaje automático

Hoy en día son múltiples las aplicaciones de machine learning en el diario vivir. Por ejemplo, para la filtración de mensajes de spam ya sea en mensajes directos o correos electrónicos. El modelo basado en el aprendizaje automático buscará, comparará y se basará en los correos spam anteriores. Otra aplicación es en la conducción de vehículos autónomos donde se aplican modelos de aprendizaje automático.

En el ámbito del mercado laboral es ventajoso combinar el aprendizaje automático y los datos del mercado laboral para los demandantes de empleo, las empresas y los responsables de la toma de decisiones (Alsayed & Awad, 2023). Sirviendo como predictor de empleos con mayor demanda el uso de machine learning. “En un mercado laboral interactúan empleados y empresarios, estos últimos compiten por los empleados más calificados, mientras que los empleados compiten por los puestos de trabajo más satisfactorios. La inteligencia del mercado laboral es la actividad encargada de analizar los datos del mercado laboral utilizando tecnologías de aprendizaje automático. El aprendizaje automático se utiliza para examinar, analizar y predecir el comportamiento futuro del mercado laboral” (Alsayed & Awad, 2023).

En el artículo de Orozco-Castañeda (2024) se busca desarrollar un indicador basado en aprendizaje automático para el mercado laboral colombiano. Empleando máquinas de vectores soporte para modelos de regresión y redes neuronales para pronosticar las tasas mensuales de empleo y desempleo. Inician con un análisis de componentes principales para desarrollar indicadores de la condición del mercado laboral colombiano. Luego hicieron un modelo autorregresivo (AR) para comparar el rendimiento de los modelos de aprendizaje

automático. Posteriormente se aplicaron máquinas de vectores soporte para regresión y redes neuronales para proyectar las tasas mensuales de ocupación total y desempleo de Colombia. Se concluye que las redes neuronales superan en general la predicción de las tasas de desempleo en comparación con modelos autorregresivos y con las máquinas de vectores.

2.2.4. Método LASSO

En estadística, LASSO (least absolute shrinkage and selection operator) es un método de análisis de regresión que selecciona las variables y las regula para evitar el sobreajuste y busca una mejor precisión en la predicción e interpretabilidad de los resultados (Wikipedia contributors, 2024). Este método se basa en una investigación geofísica de 1986; sin embargo, se popularizó 10 años después por el estadístico Robert Tibshirani (Wikipedia, 2024). Es un caso particular de la regresión de mínimos cuadrados penalizada con una función de penalización L1 (Muthukrishnan & Rohini, 2016).

Según la descripción en el artículo “¿What is Lasso regression?” de IBM (2024): LASSO es un método popularmente usado en aprendizaje automático (machine learning) para datos de alta dimensionalidad, ya que facilita la selección de características. Para ello, añade un término de penalización a la suma residual de cuadrados (RSS), que luego se multiplica por el parámetro de regularización (lambda o λ). Este parámetro de regularización controla la cantidad de regularización aplicada. Los valores más altos de lambda aumentan la penalización, reduciendo más coeficientes hacia cero, lo que reduce la importancia de algunas de las características del modelo (o las elimina por completo), lo que da a una selección automática de características.

Entre las ventajas de utilizar LASSO está que mejora tanto la precisión de la predicción como la interpretabilidad del modelo al combinar las buenas cualidades de la regresión y la selección de subconjuntos. Además, reduce la variabilidad de las estimaciones al reducir algunos de los coeficientes exactamente a cero.

2.2.5. Aplicaciones de método LASSO

LASSO puede equilibrar el sesgo y la varianza reduciendo algunos coeficientes a cero (IBM, 2024). En una aplicación práctica de este método al mercado laboral se utiliza para analizar la brecha salarial entre hombres y mujeres, incorporando datos de 1.7 millones de trabajadores en Suiza (Strittmatter & Wunsch, 2021). En este caso se utilizaron datos de

personas sujetas a diferentes grados de regulación de mercado laboral y se detectaron problemas como la falta de hombres respecto a la cantidad de mujeres (Strittmatter & Wunsch, 2021, pág. 4). Se compararon los resultados respecto a otros métodos como la descomposición de BlinderOaxaca (BO) que es utilizada para analizar y cuantificar las causas detrás de las diferencias salariales entre dos grupos como hombres y mujeres (Vicéns Otero, 2012). Obteniendo que algunos resultados de BO resultan ser engañosos, ya que los métodos más minuciosos muestran brechas salariales hasta un 50% más bajas.

2.2.6. Aplicación de LASSO en VSL

Como ya se mencionó la estimación del valor estadístico de la vida (VSL) es crucial en la toma de decisiones en políticas públicas y otros ámbitos. Sin embargo, esta estimación puede verse afectada por la inclusión de numerosas variables predictoras, y algunas pueden no ser relevantes (Viscusi W. K., 2003). LASSO ayuda a seleccionar correctamente las variables (Tibshirani, 1996). La penalización de LASSO, controlado por lambda, evita el sobreajuste al reducir la complejidad del modelo (IBM, 2024). Esto es útil cuando hay muchas variables predictoras y algunas pueden estar correlacionadas (UCSF Knowledge Hub, 2024). Además, con LASSO se mejora la interpretabilidad de los resultados al existir menos variables implicadas (IBM, 2024).

Tabla 1: Resumen de la revisión de literatura

Autores (año)	Objetivo	Método de estimación	Variable dependiente	País	Hallazgo
Alsayed & Awad (2023)	Presentar un marco que utiliza técnicas de aprendizaje automático para analizar las tendencias del mercado laboral	Análisis de clustering, modelado de temas combinación de resultados	Demanda futura de trabajos, sectores o habilidades	Bahréin	El uso de aprendizaje automático mejora la predicción de tendencias laborales y demanda de habilidades. Esto facilita la toma de decisiones para empleadores y buscadores de empleo.
Atkinson (1990)	Estimar el valor de una vida estadística usando datos del mercado de automóviles.	Regresión hedónica	Costo de uso del automóvil	EE. UU.	El valor de la vida estadística es de \$3.35 millones de dólares en 1986.
GreenLabUC (2014)	Estimar el valor estadístico de la vida en Chile	Método de preferencias declaradas	Disposición a pagar (WTP)	Chile	El valor de la vida estadística es de \$4 millones de dólares para Chile
Mardones & Riquelme (2018)	Obtener el valor de una vida estadística en Chile	Método de salarios hedónicos	Logaritmo natural del salario	Chile	El VSL es de US \$3.7 millones para Chile
Mon, Jomnonkwao et al. (2018)	Determinar los costos vinculados con la reducción del riesgo de mortalidad en accidentes de tráfico	Método de valoración contingente	Disposición a pagar	Myanmar	El VSL estimado es de \$135,712 dólares
Orozco-Castañeda et al. (2024)	Desarrollar nuevo indicador para el mercado laboral colombiano usando técnicas de machine learning	Máquinas de soporte vectorial para la regresión y redes neuronales	Tasa de ocupación y tasa de desempleo	Reino Unido	Los modelos de machine learning, especialmente las redes neuronales, superan a los enfoques autorregresivos tradicionales en la predicción de las tasas de empleo
Rosen (1974)	Formular una teoría de precios hedónicos	Regresión hedónica y estimación simultánea de funciones de oferta y demanda	Precio del producto	EE. UU.	Los precios hedónicos revelan cómo los consumidores valoran los atributos de los productos, permitiendo estimar precios implícitos por cada característica
Wim Wijnen (2009)	Analizar el uso del Valor de una Vida Estadística en análisis costo-beneficio de seguridad vial.	Método de disposición a pagar	Valor de una Vida Estadística	Países Bajos	El valor de una vida estadística en los Países Bajos se estimó en 2.2 millones de euros

Fuente: Elaboración propia

3. Metodología

3.1. Modelo empírico

El método de salarios hedónicos es utilizado para estimar el VSL. Esto con el fin de analizar la relación entre el salario y diversas características laborales, demográficas y de riesgo. La especificación del modelo se basa en la premisa de que el salario de un individuo refleja no solo su productividad, sino también las características del entorno laboral, incluidos los riesgos a los que está expuesto. En este informe se analizan 3 métodos distintos para determinar el VSL. El primero de ellos es el método de mínimos cuadrados, método utilizado con mayor frecuencia para este tipo de análisis según Evans & Taylor (2020). Posteriormente se aplica el modelo de estimador de mínimos cuadrados en dos etapas propuesto por Lee (2003). En este método se estiman en primer lugar los valores de los riesgos de fatalidad y no fatalidad para posteriormente aplicar el modelo de salarios hedónicos. Finalmente se aplican métodos heurísticos.

3.2. Método de estimación

Para la estimación del VSL, se implementa el método de mínimos cuadrados ordinarios y el método de aprendizaje automático LASSO.

3.2.1. Método de Mínimos Cuadrados Ordinarios (MCO)

Sea “y” denominada variable dependiente o variable explicada, y “x” denominada variable independiente o variable explicativa (Wooldridge, 2013). El elemento u represente el valor residual. Luego la ecuación que representa el modelo de regresión lineal de un modelo que incluye ambas variables este se muestra en la ecuación (21).

$$y = \beta_0 + \beta_1 x + u \quad (21)$$

La regresión lineal múltiple es un modelo que puede capturar la relación lineal entre múltiples variables y características. La fórmula general del modelo de regresión lineal múltiple se presenta en la ecuación (22). Luego esta ecuación se puede representar también por notación vectorial en la ecuación (23).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + u \quad (22)$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \cdot \\ x_n^T \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_n \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdot \\ u_n \end{bmatrix} \quad (23)$$

Para resolver la ecuación (22) existen varios métodos entre ellos se encuentra el método de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés). Este método minimiza la suma de los cuadrados de los residuos. Su desglose y explicación matemática es bastante extensa, pero puede resumirse en la ecuación (24).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (24)$$

En el caso de esta investigación se aplica este modelo considerando el logaritmo natural del salario (LOG_WAGE) como variable dependiente. Este enfoque estima la relación entre el salario y las variables independientes que describen las características laborales, demográficas y de riesgo.

3.2.2. Método LASSO

El método LASSO se utiliza para la predicción y selección de modelos y como componente de estimadores para realizar inferencias (StataCorp LLC, 2024). En Stata existen diversas opciones para aplicar este método los cuales sirven para inferencias, estimación y selección de modelos. Además, dependerá el tipo de datos la opción a escoger ya sea regresión lineal, logística o de Poisson para modelar un resultado continuo, binario o de conteo. En un comienzo se pueden incluir todas las características que uno cree que deberían ir en el modelo. Supongamos que “y” es la variable dependiente y las variables x_1, x_2 hasta x_p un conjunto de variables independientes. El número de variables puede ser incluso mayor al número de observaciones. Este método en Stata se puede usar para predicción, selección de modelo o inferencia. Por predicción, se refiere a la predicción del valor de un resultado condicionado a un gran conjunto de posibles regresores, ya sea dentro como fuera de la muestra. Por el contrario, se entiende la selección de variables que se correlacionan bien con el resultado en un conjunto de datos y probar si esas mismas variables predicen bien el resultado en otros conjuntos de datos. Por inferencia se entiende a la inferencia para interpretar y dar significado a los coeficientes del modelo ajustado. Para el caso particular de esta investigación es necesaria la inferencia puesto que se debe obtener un

modelo para determinar cómo las variaciones en el riesgo laboral (u otros factores relevantes) afectan los salarios.

LASSO no selecciona necesariamente las variables que aparecen en el modelo verdadero, pero sí selecciona un conjunto de variables que están correlacionadas con ellas. LASSO busca la solución al modelo de la ecuación (25).

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (25)$$

La búsqueda de la solución se realiza minimizando la expresión de la ecuación (26). Donde el primer término indica el error de predicción dentro de la muestra. Luego el segundo término dado por lambda y la sumatoria indica una penalización que aumenta de valor cuanto más complejo es el modelo (StataCorp LLC, 2024). Este término ayuda a omitir las variables. Si se minimiza la expresión (26) con respecto a β_j y λ la solución dará $\lambda=0$. Esto muestra un modelo con máxima complejidad. LASSO minimiza para valores dados de λ . Luego elige la mejor de esas soluciones en función de otro criterio como el error de estimación fuera de la muestra.

$$\frac{1}{2N} (\mathbf{y} - \mathbf{X}\beta')' (\mathbf{y} - \mathbf{X}\beta') + \lambda \sum_{j=1}^p |\beta_j| \quad (26)$$

3.2.3. Supuesto de escasez

Cuando se utiliza LASSO para predicción se debe suponer que el modelo real tiene pocas variables en relación con la cantidad de datos presentes. Según Tibshirani & Wasserman (2017) la mejor selección de subconjuntos y estimadores de LASSO tienen una propiedad en especial: sus soluciones son dispersas. Esto quiere decir que habrán muchos valores de β_j con valor 0.

3.2.4. Selección de lambda

El método LASSO ofrece varias formas de obtener lambda (λ) una de estas formas es a través de validación cruzada (CV, por sus siglas en inglés), otra forma es LASSO adaptativo, estimador plugin o criterio de información bayesiano.

3.2.4.1. Validación cruzada

En el método de validación cruzada (llamado CV, por sus siglas en inglés). En primer lugar, se divide aleatoriamente el conjunto de datos completo en una cierta cantidad de grupos $K >$

1. La cantidad típica de valores es 5 o 10. Se fija un grupo como el conjunto de prueba y se designan los $K - 1$ grupos restantes como el conjunto de entrenamiento. Luego se aplica el lazo a los datos de entrenamiento para un rango de diferentes valores t , y se usa cada modelo ajustado para predecir las respuestas en el conjunto de prueba, registrando los errores de predicción cuadráticos medios para cada valor de λ (límite del criterio LASSO). Este proceso se repite un total de K veces, con cada uno de los K grupos teniendo la oportunidad de desempeñar el papel de los datos de prueba, con los $K - 1$ grupos restantes utilizados como datos de entrenamiento. Luego estas K estimaciones del error de predicción se promedian para cada valor de λ , lo que produce una curva de error de validación cruzada.

3.2.4.2. LASSO adaptativo

Zou (2006) propuso el LASSO adaptativo como un medio para ajustar modelos más dispersos que el LASSO común. En primer lugar se estima un valor piloto de β a través de validación cruzada. Luego se resuelve mediante la expresión (27). En este caso el término $w_j = 1/|\beta_j|^p$. Una de las ventajas de este método es el uso del valor piloto. Otra ventaja es que se puede rescatar el modelo original bajo ciertas condiciones.

$$\min \left\{ \frac{1}{2} \|y - X\beta\|_2^2 + \lambda \sum_{j=1}^p w_j |\beta_j| \right\} \quad (27)$$

3.2.4.3. Estimador plugin

Uno de los métodos heurísticos es el estimador plugin. En este caso se utiliza un valor de λ lo suficientemente grande para dominar el ruido inherente a la estimación de los coeficientes cuando las cargas de penalización κ_j están en sus niveles óptimos (StataCorp, 2023). El valor de λ es dado por la ecuación (28). Lo que se muestra es la probabilidad de cómo se elige λ^* , el parámetro de regularización, asegurando que sea suficientemente grande para satisfacer ciertas condiciones de probabilidad relacionadas con los pesos κ_j y las contribuciones h_j de cada variable. Es necesario remarcar que “ $\xrightarrow{N \rightarrow \infty}$ ” indica que, conforme N (el tamaño de la muestra) tiende a infinito, la probabilidad converge a 1, lo que significa que el evento ocurre casi con certeza para muestras suficientemente grandes. En este caso c es una constante y además la puntuación j del estimador no penalizado en los coeficientes verdaderos β_0 es dada en la expresión (28). La puntuación del estimador j -ésimo no

penalizado en los coeficientes verdaderos β_0 es dada por la expresión (29).

$$P\left(\lambda^* \geq c \max_{1 \leq j \leq p} \left| \frac{1}{N} \sum_{i=1}^N h_j(y_i, x_i \beta'_0) \right| \right) \rightarrow_N 1 \quad (28)$$

$$\frac{1}{N} \sum_{i=1}^N h_j(y_i, x_i \beta'_0) \quad (29)$$

Según Belloni & Chernozhukov (2011) el valor adecuado de c es 1.1, luego N es el tamaño de la muestra, además p es el número de covariables candidatas en el modelo, entonces la probabilidad de no eliminar una variable cuando tiene un coeficiente de cero es dada por la ecuación (30). Finalmente el valor óptimo de λ es dado por la ecuación (31). En esta ecuación $\hat{\sigma}$ es un estimador de la varianza del término de error.

$$\gamma = 0.1 / \ln [\max \{p, N\}] \quad (30)$$

$$\lambda_{homocedástico} = \frac{c \hat{\sigma}}{\sqrt{N}} \Phi^{-1} \left(1 - \frac{\gamma}{2p} \right) \quad (31)$$

3.2.5. Otros métodos de estimación

3.2.5.1. Modelo de Estimadores de Mínimos Cuadrados en Dos Etapas (2SLS)

Otro método, es el 2SLS, particularmente útil en contextos de análisis donde las variables de riesgo pueden estar correlacionadas con la variable dependiente. O incluso en casos donde pareciera que la teoría no encaja con la realidad ya que en la regresión no se explica la variable dependiente por las variables independientes. O en casos en que la variable independiente termina siendo endógena o está determinada conjuntamente a la variable dependiente. El modelo 2SLS presupone que existen instrumentos, que están correlacionados con los predictores problemáticos, pero no con el término de error (IBM, 2024).

3.2.5.2. Modelo de Heckman

El modelo de Heckman, también llamado método de Heckit, es un método para estimar modelos de regresión que sufren el sesgo de selección de la muestra. En el libro Wooldridge (2013) se describe como un “procedimiento econométrico utilizado para corregir el sesgo de selección de la muestra debido a truncamiento fortuito u otra forma de falta de datos no aleatoria”.

3.3. Test estadísticos

3.3.1. Prueba F

Se utiliza la Prueba F para evaluar si el modelo global es significativo, es decir, si al menos una de las variables independientes tiene una relación significativa con la variable dependiente. La prueba F se utiliza en el análisis de regresión para probar si la hipótesis de que todos los parámetros del modelo son cero es cierta (Kissell & Poserina, 2017). Dada la hipótesis nula que todos los coeficientes en una regresión son nulos, luego la hipótesis alternativa dice que al menos alguno de los coeficientes es distinto de 0 tal como se muestra en la expresión (33).

$$H_0: \beta_1 = \beta_2 = \dots = \beta_j = 0 \quad \text{para algún } j \quad (32)$$

$$H_1: \beta_j \neq 0 \quad \text{para algún } j \quad (33)$$

La Prueba F expresa la relación entre la varianza explicada y la varianza no explicada. En general si $F > 2.5$ entonces se rechaza la hipótesis nula, y hay algún parámetro β_j distinto de 0.

3.3.2. R-cuadrado

R-cuadrado también llamado coeficiente de determinación es una medida de la precisión global del modelo. Se mide en una escala de 0 a 1. Un valor de 1 indica que el modelo es preciso, y el modelo predice perfectamente los valores en el campo objetivo. Un valor de 0 indica un modelo que no tiene valor predictivo. En el mundo real, R^2 se sitúa entre estos valores.

3.3.3. R cuadrado ajustado

R^2 ajustado es una medida corregida de bondad de ajuste (precisión de modelo) para los modelos lineales. Identifica el porcentaje de varianza en el campo de destino que se explica por la entrada o las entradas.

3.3.4. Prueba t

La prueba t se utiliza para comprobar la hipótesis nula de que el valor de un parámetro es cero. Si se cumpliera esta hipótesis indicaría que el factor explicativo seleccionado no tiene ningún poder predictivo en una ecuación de regresión. En la ecuación (34) y (35) se muestran las hipótesis. En general si el estadístico t es mayor a 2 entonces se rechaza la hipótesis nula.

$$H_0: \beta_j = 0 \quad (34)$$

$$H_1: \beta_j \neq 0 \quad (35)$$

3.3.5. Valor de significación

El valor de significación indica la probabilidad de que un resultado se haya producido por azar, si el valor de significación es inferior al nivel de significación (por defecto, 0.05), se considera que la prueba es estadísticamente significativa.

4. Datos

4.1. Fuente de datos

Los datos para esta investigación provienen de diversas fuentes. La primera fuente de datos es el Boletín Estadístico de Seguridad Social de diversos años. El cual se encuentra disponibles en la página web de la Superintendencia de Seguridad Social (SUSESO). Este boletín contiene información anual de variables referentes a los regímenes previsionales y asistenciales que fiscaliza la Superintendencia de Seguridad Social. Entre algunas de las estadísticas presentes en el boletín está el número promedio mensual de trabajadores protegidos por el seguro de la ley N° 16.744, este establece normas sobre accidentes del trabajo y enfermedades profesionales. Además, muestra el número de accidentes de trabajo, de trayecto y de enfermedades profesionales. Por último, se muestra la tasa de accidentabilidad según actividad económica y mutualidad a la que pertenecía el trabajador. Dentro de esto se muestra la información de 3 mutualidades: Asociación Chilena de Seguridad (ACHS), Mutual de Seguridad de la Cámara Chilena de la Construcción (MUSEG), y por último del Instituto de Seguridad del Trabajo (IST). Para la presente investigación solo se considerará la información de la ACHS para poder compararla con los datos de otras fuentes. Este boletín contiene una información bastante amplia, en el año 2022 se analizaron 7,178,925 trabajadores de los cuales 2,826,657 pertenecen a la Asociación Chilena de Seguridad. Siendo 1,586,278 hombres y 1,240,378 mujeres y la mayoría pertenece a la región Metropolitana de Santiago.

La segunda fuente de datos corresponde a la Encuesta de Caracterización Socioeconómica Nacional (CASEN) la cual como su página web lo indica tiene por objetivo “conocer periódicamente la situación de los hogares y de la población” y “evaluar el impacto de la política social” (Ministerio de Desarrollo Social y Familia, 2023). Esta encuesta tiene por objeto de estudio los hogares que habitan las viviendas particulares ocupadas que se ubican en el territorio nacional. Se desarrolla desde 1990 y tiene periodicidad bianual o trianual. Para el año 2022 se encuestaron 72.056 hogares de 335 comunas de las 16 regiones del país (Ministerio de Desarrollo Social y Familia, 2023).

Para el año 2022 según información de la ACHS las industrias que tuvieron mayor tasa de accidentes fueron la pesca, la industria manufacturera y la industria de actividades

inmobiliarias (Superintendencia de Seguridad Social, 2022). Abarcando en total el 40% de los accidentes registrados en ese año. En la tabla 2 se muestra un resumen del detalle para ese año y para cada industria. Se puede apreciar que el mayor número de muertes se presenta en la industria del transporte y de la agricultura.

4.2. Creación de la muestra de estimación

Para la creación de la muestra de estimación se consideraron diversas características de los entrevistados entre las cuales se encuentra el salario, el tipo de trabajo, número de hijos menores a 6 años entre otros. Según Taylor (2003) no es fácil decidir qué variables o características se deben incluir en la función de salarios hedónicos. Según Kniesner (2010) entre las variables típicas que se deben integrar en este tipo de análisis está la edad, edad al cuadrado, estado civil y la región de residencia. Por lo que se creó una variable llamada “AGE” que considera el cuadrado de la edad de los entrevistados. Para cada una de las regiones a las que pertenecen los entrevistados se creó una variable dummy que se puede capturar posteriormente en el modelo. Esto incluye a cada una de las 16 regiones del país. Para el estado civil se creó una variable dummy para indicar si la persona está casada o no.

Al analizar las demás variables que se podían incluir se recurre a la literatura. Por ejemplo, la presencia de niños pequeños puede tener un mayor impacto en la participación laboral y los salarios de las mujeres. Cukrowska-Torzewska (2020) menciona que “las madres tienden a percibir salarios más bajos porque eligen empleos que, por término medio, pagan menos, pero son más compatibles con el cuidado de los hijos”. En otro estudio Looze (2017) menciona que “los niños en edad preescolar reducen la movilidad laboral de las mujeres, que está asociada al crecimiento salarial”. Por lo anterior se estableció la variable “CHILDREN” para definir si la persona tiene 6 años o menos. Luego se estableció una variable que indica la cantidad de niños menores de 6 años llamada “childsix” en el hogar. Finalmente se estableció la variable que busca aquellos hogares con mujeres mayores a 15 años y donde existen niños menores a 6 años en casa llamada “CHILDSIX”. Al analizar si el tamaño de la empresa donde se desempeña el trabajador es una variable relevante para el análisis de salarios hedónicos se recurrió a la literatura donde hay diversas opiniones. Mientras Lanoie (1995) dice que “el tamaño de la empresa puede afectar las primas salariales en el análisis de riesgo salarial”, hay otros autores como Leigh (1987) que mencionan que la inclusión o

Tabla 2 : Tasa de accidentabilidad fatal en Chile año 2017 y 2022

Industria	Accidentes fatales		Tasa de accidentes fatales (1/10.000 trabajadores)		Accidentes no fatales		Tasa de accidentes no fatales (1/100 trabajadores)	
	2017	2022	2017	2022	2017	2022	2017	2022
Agricultura, ganadería, caza y silvicultura	36	10	1.053	0.544	14,368	6239	4.20	3.39
Pesca	6	4	1.549	1.305	1,556	1197	4.02	3.91
Explotación de minas y canteras	4	2	0.683	0.565	730	340	1.25	0.96
Industrias Manufactureras	23	4	0.455	0.149	23,757	9830	4.70	3.66
Suministro de electricidad, gas y agua	1	1	0.333	0.550	485	224	1.61	1.23
Construcción	44	6	0.745	0.307	24,345	7242	4.12	3.70
Comercio, reparación de vehículos y otros	8	4	0.112	0.104	26,381	8953	3.68	2.33
Hoteles y restaurantes	6	0	0.275	0.000	10,235	4201	4.70	3.39
Transporte, almacenamiento y comunicaciones	57	17	1.613	0.947	16,291	5133	4.61	2.86
Intermediación financiera	3	0	0.168	0.000	2,093	699	1.17	0.91
Actividades inmobiliarias, empresariales y de alquiler	20	2	0.257	0.041	20,727	9412	2.66	1.95
Administración pública y defensa	4	1	0.105	0.044	9,292	4222	2.44	1.84
Enseñanza	0	1	0.000	0.040	7,049	4184	1.85	1.69
Servicios sociales y de salud	2	0	0.108	0.000	4,263	2391	2.30	1.75
Otras actividades de servicios comunitarios	7	2	0.281	0.110	7,398	4353	2.97	2.39
Hogares privados con servicio doméstico	0	1	0.000	0.424	859	353	2.00	1.50
Organizaciones y órganos extraterritoriales	0	0	0.000	0.000	13	0	1.84	0.00

Fuente: Superintendencia de Seguridad Social

exclusión de este tipo de variable no tiene efecto alguno. Finalmente se creó una variable que contiene el tamaño de la empresa en la que trabaja cada encuestado.

Para la información de nivel educacional se crearon variables dummy para capturar de mejor forma los datos. Se incluye a personas que no tuvieron educación formal, hasta aquellas personas que terminaron o están en educación superior. Para la encuesta CASEN 2022 existen dos variables que engloban esta información que son las variables e6a_no_asiste y la variable e6a_asiste. Pero se incluyeron dentro de una misma categoría aquellas personas que llegaron hasta un nivel educacional y las personas que están actualmente en ese nivel. Por ejemplo, una persona que llegó a estudiar hasta la educación básica (e6a_no_asiste=7) y una persona que actualmente se encuentra en ese nivel (e6a_asiste=7) se registran en la misma categoría. En total son 5 categorías. Para el año 2017 en cambio la variable e6a incluye las personas que estudiaron en el pasado o están actualmente en cierto nivel educacional.

Otro aspecto a considerar es la cantidad de personas en el hogar. Además, se considera si la persona tiene contrato de trabajo.

Además, se categoriza según el tipo de trabajador o rol que tiene en una empresa, siendo la variable o15 la que define esto y toma valores desde 1 correspondiente a patrón(a) o empleador(a) hasta el valor 9 que corresponde a persona no remunerada.

Morgan (2015) muestra que una de las variables necesarias de incluir en el análisis es el tipo de trabajo que tengan según la cantidad de horas que trabajan las personas. Por ello indica que se debe controlar ciertas variables como la jornada completa frente a la jornada parcial. Para esto se crea la variable FULLTIME que indica que una persona tiene contrato de horario completo.

Como medida de riesgo no existe una variable que pueda definirse de forma estricta para todos los casos, por ende, se recurre al número de accidentes fatales como una aproximación al verdadero riesgo inherente (Andersson H. , 2005). Estos datos se extraen desde la página web de la Superintendencia de Seguridad Social (SUSESO). Los cuales están disponibles en boletines informativos, la información de accidentes se presenta en forma de tasas. Estas se calculan dividiendo el total de accidentes del trabajo fatales ocurridos a los trabajadores protegidos por el número de trabajadores de cada rubro. Para el modelo de salarios hedónicos se trabaja con la variable FATALITY_RATE que denota esta tasa. Además, se incluye la tasa

de accidentes no fatales denominada “NON_FATALITY_RATE”.

Las variables utilizadas se listan a continuación según su respectiva categoría:

Salario: se incluye el logaritmo natural del salario como variable dependiente (*LOG_WAGE*).

Empresa: se incluyen variables según el tamaño de la empresa (*FIRMSIZE1*, *FIRMSIZE2*, *FIRMSIZE3*, *FIRMSIZE4*, *FIRMSIZE5*, *FIRMSIZE6*). Adicionalmente se agrega una variable que indica si tiene contrato de trabajo la persona (*SIGNED_CONTRACT*, *NOT_SIGNED_IN*). Además, se incorpora una variable dummy según el tipo de contrato que tiene la persona (*FULLTIME*, *PART_TIME*). Posteriormente se incorpora una variable para definir el tipo de trabajo del encuestado, ya sea del área pública, privada o del hogar o una persona sin trabajo dentro de hogar (*FREELANCE*, *PUBLIC_EMPLOYEE*, *HOUSEHOLD_EMPLOYEE*, *FFAA*, *UNPAID_FAMILY*). Por último, se incluye una variable que corresponde al tamaño promedio de otras empresas del rubro sin incluir la del encuestado (*AVG_OTHERS_FIRMS*).

Formación: se incluyen variables dummy de acuerdo con el nivel educacional de la persona entrevistada siendo el primer nivel sin educación, luego las personas que llegaron a educación básica hasta las personas con estudios superiores o más (*WITHOUT_EDUCATION*, *EDUCATION_1*, *EDUCATION_2*, *EDUCATION_3*, *EDUCATION_4*).

Individual: una variable dummy que determina si es hombre o mujer (*WOMAN*), una variable continua que expresa el cuadrado de la edad de la persona (*AGE*) y otra variable para los ingresos no laborales (*NON_LABOR_INCOME*).

Hogar: una variable que indica si hay niños menores de 6 años, una variable para determinar el número de niños menores de 6 años que hay en cada casa y si hay mujeres mayores de 15 años que los puedan cuidar (*CHILD_SIX*), y una variable que muestra la cantidad de personas en el hogar (*HOUSEHOLD_SIZE*). Una variable que indica el estado civil del encuestado (*MARRIED*), adicionalmente una variable que indica los años de escolaridad del esposo/a (*EDUCATION_SPOUSE*), y se agrega una variable que indica si el esposo/a trabaja (*JOBSPOUSE*).

Geográfica: una variable que determina la región de cada entrevistado (*TARAPACA*, *ANTOFAGASTA*, *ATACAMA*, *COQUIMBO*, *VALPARAISO*, *OHIGGINS*, *MAULE*, *BIOBIO*,

ARAUCANIA, LAGOS, AYSEN, MAGALLANES, METROPOLITANA, RIOS, ARICA, NUBLE).

Riesgo: una variable que incluye la tasa de personas fallecidas por industria (*FATALITY_RATE*), otra variable que incluye la tasa de personas accidentadas no fallecidas por industria (*NON_FATALITY_RATE*).

4.3. Descripción de la muestra de estimación

Los datos de la CASEN 2022 muestran más de 200 mil datos de personas entrevistadas. Sin embargo, en el proceso de limpieza de datos, se ha reducido significativamente el número de observaciones, de más de 200,000 a poco más de 83,000 personas. Esta reducción se debe principalmente a la exclusión de individuos muy jóvenes, que generalmente no están en el mercado laboral, y personas muy ancianas que no perciben ingresos salariales o no se encuentran trabajando en la actualidad. La limpieza de la muestra contribuye a enfocar el análisis en individuos con mayor probabilidad de estar en el mercado laboral activo, lo cual es crucial para estudiar el impacto de factores laborales sobre los salarios. En la tabla 3 se encuentra una descripción de las variables según la encuesta CASEN 2022.

En la figura 1 se muestra el histograma de densidad de la variable *LOG_WAGE*. La distribución del logaritmo de los salarios muestra una alta concentración alrededor del promedio, con una asimetría a la derecha, lo que indica que existen algunos individuos con salarios significativamente más altos que el resto de la población. La desviación estándar de 0.6495 refleja una variabilidad considerable en los ingresos. Este fenómeno puede ser explicado por factores como la diversidad en los niveles educativos, las ocupaciones, el riesgo laboral, el tipo de contrato laboral y las disparidades salariales entre géneros. Estos hallazgos sugieren que, aunque el ingreso promedio es un buen indicador de la situación general, la dispersión de los salarios es amplia.

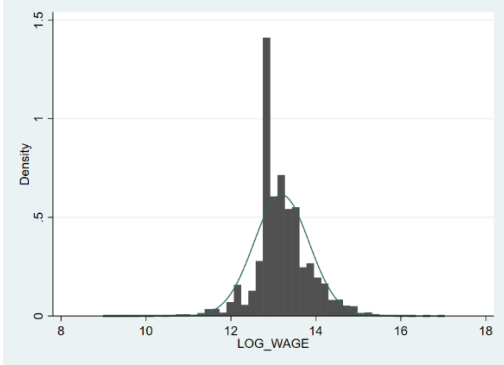
En la figura 2 se puede apreciar la tasa de fatalidad. Aunque la tasa promedio de accidentes fatales es baja (0.0000457), el rango muestra una variación considerable. Lo cual justifica la implementación de políticas sectoriales de seguridad laboral más estrictas y mayor fiscalización en sectores de alto riesgo como transporte y agricultura.

En la Figura 3 se presenta la cantidad de personas encuestadas en cada región del país, y se observa una variabilidad significativa entre las distintas regiones. Las regiones más pobladas,

como la Región Metropolitana y Biobío, albergan una proporción considerable de la población nacional, mientras que las regiones más al sur, como Aysén y Magallanes, tienen una menor cantidad de habitantes. Las disparidades regionales en la distribución laboral sugieren la necesidad de políticas diferenciadas para fomentar el empleo en regiones menos densas y mejorar la conectividad laboral.

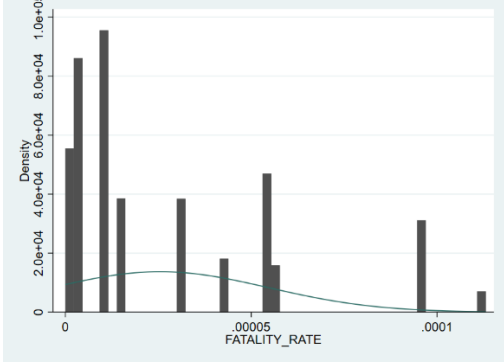
En la figura 4 se observa la variación de los sueldos promedios según el nivel educacional de los entrevistados. Los resultados indican que existe una clara tendencia a que los salarios aumenten a medida que se incrementa el nivel educativo. Las personas sin educación formal tienen un salario promedio significativamente más bajo (375,656 CLP) en comparación con aquellos con educación universitaria o de posgrado (1,111,403 CLP). Este patrón sugiere que la educación superior tiene un impacto positivo en los ingresos laborales, lo cual es consistente con la teoría económica que asocia una mayor educación con mayores habilidades y productividad. Posiblemente políticas educativas que promuevan la capacitación técnica o profesional pueden mejorar las oportunidades salariales y reducir desigualdades.

Figura 1: Densidad relativa variable LOG_WAGE



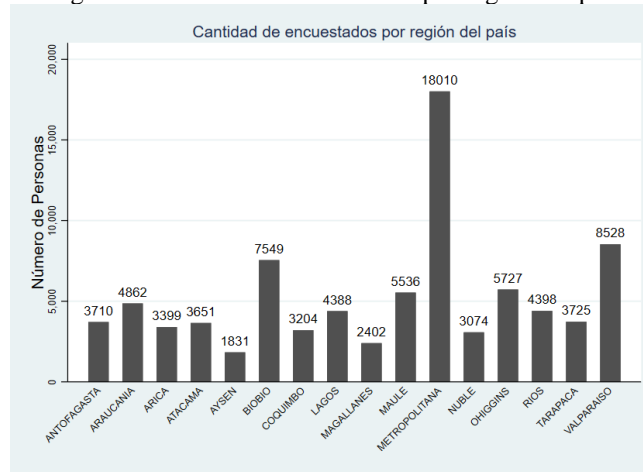
(a) Distribución del logaritmo del salario calculado con datos del período 2022.
Fuente: Elaboración propia basada en datos de CASEN 2022

Figura 2: Densidad relativa de variable FATALITY_RATE



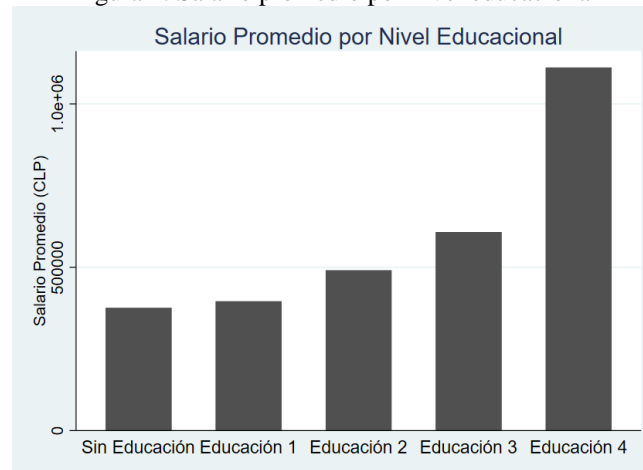
(a) Distribución de la tasa de fatalidad calculada con datos del período 2022.
Fuente: Elaboración propia basada en datos de CASEN 2022

Figura 3: Cantidad de encuestados por región del país



(a) Distribución de la tasa de fatalidad calculada con datos del período 2022.
Fuente: Elaboración propia basada en datos de CASEN 2022

Figura 4: Salario promedio por nivel educacional



(a) Distribución de la tasa de fatalidad calculada con datos del período 2022.
Fuente: Elaboración propia basada en datos de CASEN 2022

Tabla 3: Descripción de variables

Variable	Obs	Promedio	Desviación estándar	Mínimo	Máximo
LOG_WAGE	58,208	13.1920	0.6495	8.9871	17.0343
FATALITY_RATE	83,994	0.0000255	0.0000291	0	0.0001131
TARAPACA	83,994	0.0443	0.2059	0	1
ANTOFAGASTA	83,994	0.0442	0.2055	0	1
ATACAMA	83,994	0.0435	0.2039	0	1
COQUIMBO	83,994	0.0381	0.1915	0	1
VALPARAISO	83,994	0.1015	0.3020	0	1
OHIGGINS	83,994	0.0682	0.2521	0	1
MAULE	83,994	0.0659	0.2481	0	1
BIOBIO	83,994	0.0899	0.2860	0	1
ARAUCANIA	83,994	0.0579	0.2335	0	1
LAGOS	83,994	0.0522	0.2225	0	1
AYSEN	83,994	0.0218	0.1460	0	1
MAGALLANES	83,994	0.0286	0.1667	0	1
RIOS	83,994	0.0524	0.2228	0	1
ARICA	83,994	0.0405	0.1971	0	1
NUBLE	83,994	0.0366	0.1878	0	1
METROPOLITANA	83,994	0.2144	0.4104	0	1
WITHOUT_EDUCATION	83,994	0.0053	0.0726	0	1
EDUCATION_1	83,994	0.1745	0.3796	0	1
EDUCATION_2	83,994	0.4481	0.4973	0	1
EDUCATION_3	83,994	0.1277	0.3337	0	1
EDUCATION_4	83,994	0.2445	0.4298	0	1
WOMAN	83,994	0.4396	0.4963	0	1
AGE	83,994	2086.0800	1281.9900	225	8464
CHILD_SIX	83,975	0.0981	0.3378	0	4
AVG_OTHERS_FIRMS	83,994	85.3031	71.0074	0	621.81
NON_FATALITY_RATE	83,994	0.0254	0.0085	0	0.0391
FULLTIME	83,994	0.6202	0.4853	0	1
SIGNED_CONTRACT	83,994	0.5995	0.4900	0	1
NOT_SIGNED_IN	83,994	0.0069	0.0829	0	1
MARRIED	83,994	0.3098	0.4624	0	1
EDUCATION_SPOUSE	64,872	8.4521	6.4375	0	29
HOUSEHOLD_SIZE	83,994	0.0257	0.1582	0	1
NON_LABOR_INCOME	83,994	62579.3300	101901.3000	0	1,849,188
FIRMSIZE1	83,994	0.2244	0.4172	0	1
FIRMSIZE2	83,994	0.1537	0.3606	0	1
FIRMSIZE3	83,994	0.0623	0.2417	0	1
FIRMSIZE4	83,994	0.1615	0.3680	0	1
FIRMSIZE5	83,994	0.1402	0.3472	0	1

(continuación) Tabla 3: Descripción de variables

Variable	Obs	Promedio	Desviación estándar	Mínimo	Máximo
JOBSCOUSE	83,994	0.3200	0.4665	0	1
FREELANCE	83,994	0.2849	0.4514	0	1
PUBLIC_EMPLOYEE	83,994	0.1368	0.3436	0	1
FREELANCE	83,994	0.2849	0.4514	0	1
PUBLIC_EMPLOYEE	83,994	0.1368	0.3436	0	1
HOUSEHOLD_EMPLOYEE	83,994	0.0257	0.1582	0	1
PART_TIME	83,994	0.0751	0.2635	0	1
FFAA	83,994	0.0072	0.0844	0	1
UNPAID_FAMILY	83,994	0.0045	0.0668	0	1

Fuente: Encuesta de Caracterización Socioeconómica Nacional 2022

5. Resultados

5.1. Método OLS

En primer lugar, se realizaron múltiples pruebas para saber qué variables debían incluirse en el modelo. En un comienzo se incluyeron pocas variables para luego ir agregando más de acuerdo con valores estadísticos como R^2 . En la tabla 4 se muestran los valores de R^2 , el número de observaciones y el número de variables que se fueron probando.

Tabla 4: Prueba número de variables datos 2022

Número de variables	Observaciones	R^2	Prueba F	Prob > F
2	58,208	0.0012	72.59	0
3	58,208	0.0018	52.69	0
5	58,208	0.0076	111.70	0
20	58,208	0.2041	785.48	0
44	45,591	0.5104	1130.63	0

En la tabla 4 se aprecia que a medida que aumenta el número de variables va aumentando el valor de R^2 . Esto significa que aumenta la proporción de la variabilidad de la variable dependiente que es explicada por las variables independientes del modelo. Aunque se debe tener cuidado ya que por definición al aumentar la cantidad de variables R^2 aumentará o se mantendrá constante.

Luego se escoge la cantidad de 44 variables explicativas para el modelo. Se inicia con el modelo de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés). Se obtiene el valor del sueldo promedio de \$624,630 y luego se determina el VSL. Dando un valor de 8.93 millones de dólares para el año 2022, considerando el valor del dólar a \$859 según lo indicado por el Banco Central. Para el año 2017 se obtiene un VSL de 2.75 millones de dólares según el modelo OLS con el precio del dólar a \$615.

De la tabla 6 se puede apreciar que el valor del VSL obtenido para 2022 es alto considerando análisis anteriores ya que en 2010 se obtuvieron valores de 1.5 millones de dólares y en 2018 de 3.7 millones de dólares como se mencionó anteriormente. Además, es necesario fijarse en los valores de los coeficientes de la regresión de la tabla 5. Dado que la mayoría de las regiones del país tienen valores negativos para ambos años, y la variable omitida en la regresión fue la Región Metropolitana, entonces los coeficientes deben interpretarse en relación con la variable omitida. Por ejemplo, la región de Tarapacá tiene un valor de -0.078 por ende en promedio los sueldos son un 7.6% más bajos en relación con la región Metropolitana. Para el caso del nivel de educación la variable omitida es el nivel profesional.

Por ende, si vemos los coeficientes mientras más bajo sea el nivel educacional más bajo será el salario en relación con los profesionales. Para el caso de “sin educación” dado el coeficiente -0.614 en promedio las personas tienen un 45.8% menos de sueldo en relación con las personas con trabajos de formación profesional, esto se determina según la ecuación (36). Para el sexo de la persona, la variable omitida es hombre, luego el coeficiente -0.161 para mujeres muestra que en promedio tienen un sueldo 14.9% más bajo que los hombres.

$$(e^{-0.614} - 1) * 100\% = -0.458\% \quad (36)$$

Tabla 5: Resultados método OLS

Variable	2022		2017	
	Coficiente	Error estándar	Coficiente	Error estándar
FATALITY_RATE	1023.0940	83.9410	297.2190	265.1580
TARAPACA	-0.0780	0.0110	-0.1550	0.0490
ANTOFAGASTA	0.0380	0.0110	-0.1400	0.0530
ATACAMA	-0.0350	0.0110	-0.1440	0.0600
COQUIMBO	-0.0840	0.0120	-0.0580	0.0540
VALPARAISO	-0.1330	0.0080	-0.4340	0.0390
OHIGGINS	-0.1050	0.0090	-0.2810	0.0410
MAULE	-0.1560	0.0090	-0.3210	0.0430
BIOBIO	-0.1400	0.0080	-0.0840	0.0380
ARAUCANIA	-0.1780	0.0100	-0.1910	0.0450
LAGOS	-0.0830	0.0100	-0.3600	0.0460
AYSEN	0.0330	0.0160	-0.2700	0.0650
MAGALLANES	0.0630	0.0130	-0.2060	0.0540
RIOS	-0.1490	0.0100	-0.2600	0.0500
ARICA	-0.1050	0.0120	-0.1040	0.0600
NUBLE	-0.1540	0.0120	-0.6150	0.0570
WITHOUT_EDUCATION	-0.6140	0.0350	-0.6670	0.1280
EDUCATION_1	-0.6320	0.0080	-0.3940	0.0410
EDUCATION_2	-0.5150	0.0060	-0.4370	0.0320
EDUCATION_3	-0.3870	0.0070	-0.3290	0.0310
WOMAN	-0.1610	0.0050	-0.2620	0.0250
AGE	0.0000	0.0000	0.0000	0.0000
CHILD_SIX	-0.0100	0.0070	0.1030	0.0300
AVG_OTHERS_FIRMS	0.0010	0.0000	0.0000	0.0000
NON_FATALITY_RATE	-4.1670	0.3030	1.9600	1.2140
FULLTIME	0.0970	0.0150	0.0480	0.0370
SIGNED_CONTRACT	0.2400	0.0070	-0.0360	0.0280
NOT_SIGNED_IN	0.1160	0.0240	-0.0200	0.0750
MARRIED	0.0350	0.0050	0.1120	0.0220
EDUCATION_SPOUSE	0.0210	0.0000	0.0200	0.0030
HOUSEHOLD_SIZE	0.0030	0.0020	-0.0620	0.0070
NON_LABOR_INCOME	0.0000	0.0000	0.0000	0.0000
FIRMSIZE1	-0.2400	0.0140	-0.2290	0.0570
FIRMSIZE2	-0.1900	0.0080	-0.1340	0.0370
FIRMSIZE3	-0.1560	0.0090	-0.0520	0.0410
FIRMSIZE4	-0.1130	0.0060	0.0540	0.0270
FIRMSIZE5	-0.0500	0.0060	0.0860	0.0280
JOBSPOUSE	-0.0420	0.0060	-0.7790	0.0590
FREELANCE	omitida	omitida	omitida	omitida
PUBLIC_EMPLOYEE	0.0320	0.0060	0.0050	0.0300

(continuación) Tabla 5: Resultados método OLS

Variable	Coficiente	Error estándar	Coficiente	Error estándar
HOUSEHOLD_EMPLOYEE	-0.0010	0.0180	-0.0430	0.0730
PART_TIME	-0.4830	0.0160	omitida	omitida
FFAA	0.1690	0.0200	-0.0970	0.0780
UNPAID_FAMILY	omitida	omitida	omitida	omitida
_cons	13.3450	0.0200	13.0680	0.0930

Fuente: Elaboración propia con base en resultados obtenidos utilizando Stata

Tabla 6: Estimación VLS método OLS (millones de dólares año 2017 y 2022)

Año	2017	2022
VLS	2.75	8.93

Fuente: Elaboración propia con base en resultados obtenidos utilizando Stata

5.1.1. Análisis de robustez

Dados los resultados de la tabla 7 se observa una muestra numerosa lo cual es importante para asegurar que los resultados sean estadísticamente significativos. Luego al ver el valor de la prueba F con un valor de 1130,63 implica que el modelo es significativamente diferente de un modelo sin variables explicativas. Además, la probabilidad es de 0,000 por lo que se puede rechazar la hipótesis nula que todos los coeficientes son iguales a cero. Con un valor de R^2 de 0.510 implica que el modelo explica el 51% de la variabilidad total de la variable dependiente (en este caso LOG_WAGE). Si bien no es un valor muy alto, puede ser razonable pues en el modelo de salarios hedónicos hay otros factores no modelados que pueden influir en el salario. Dado que la diferencia entre el R-cuadrado y el R-cuadrado ajustado es pequeña, esto sugiere que el modelo no está sobre ajustado (overfitted) y que las variables explicativas son relevantes. El valor del error cuadrático medio (Root MSE, por sus siglas en inglés), sugiere que las predicciones del modelo tienen una desviación estándar de aproximadamente 0.44 unidades respecto a los valores reales. En la tabla 8 se muestra el Factor de Inflación de la Varianza (VIF, por sus siglas en inglés), este índice mide cuánto se incrementa la varianza de un coeficiente de la regresión debido a la colinealidad de una variable independiente con las otras variables del modelo (PennState , 2024).

Tabla 7: Resultados modelo OLS año 2022

	SS	df	MS	Número de obs	45.591
Modelo	9538.064	42	227.096	F(42,45548)	1130.63
Residuos	9148.675	45,548	0.200	R^2	0.510
Total	18686.74	45,590	0.409	Root MSE	0.44817

Fuente: Elaboración propia en base a resultados de Stata

En otras palabras, el VIF indica la extensión en que las estimaciones de los coeficientes son afectadas por la correlación entre las variables independientes. Cuando el valor del VIF es cercano a 1 indica que no hay colinealidad entre las variables, es decir, las variables independientes no están correlacionadas entre sí.

Cuando los resultados del VIF de las variables están entre 1 y 5 indica cierta correlación entre las variables. Finalmente, cuando es superior a 10 puede ser un indicio de que la variable está altamente correlacionada con otras variables, lo que puede generar inestabilidad en los coeficientes estimados. En la tabla 8 se observan los resultados del análisis de

multicolinealidad para el modelo OLS. En este caso las variables que tienen un VIF más alto son aquellas relacionadas con el tipo de jornada laboral, estas variables podrían tener correlación con el resto de las variables, pero al ser cercanas a 5 no indica problemas de multicolinealidad. Por otro lado, las variables `WITHOUT_EDUCATION` y `NON_SIGNED_IN` tienen el VIF más bajo por ende no están altamente correlacionadas con el resto de las variables. El promedio de VIF es 1,54 lo que es un valor bajo y sugiere que, en general, no existe multicolinealidad grave en el modelo. Las variables están razonablemente correlacionadas, pero no de manera que afecten gravemente la estabilidad de los coeficientes en el modelo.

Tabla 8: Análisis de multicolinealidad método OLS año 2022

Variable	VIF	1/VIF
PART_TIME	5.23	0.19121
FULLTIME	5.17	0.193387
EDUCATION_SPOUSE	2.06	0.485454
FIRMSIZE1	1.93	0.518007
HOUSEHOLD_EMPLOYEE	1.92	0.521502
EDUCATION_1	1.88	0.532968
EDUCATION_2	1.78	0.561178
JOB_SPOUSE	1.74	0.574186
AGE	1.65	0.604708
NON_FATALITY_RATE	1.65	0.607887
MARRIED	1.54	0.648179
WOMAN	1.47	0.6812
AVG_OTHERS_FIRMS	1.41	0.706807
FATALITY_RATE	1.38	0.725322
FIRMSIZE2	1.35	0.740102
EDUCATION_3	1.34	0.746701
VALPARAISO	1.34	0.749045
SIGNED_CONTRACT	1.33	0.750706
BIOBIO	1.33	0.754421
FIRMSIZE4	1.32	0.75533
PUBLIC_EMPLOYEE	1.3	0.771666
FIRMSIZE5	1.28	0.78412
OHIGGINS	1.27	0.788358
MAULE	1.26	0.7938
CHILD_SIX	1.25	0.801917
LAGOS	1.21	0.827042
ARAUCANIA	1.21	0.828825
NON_LABOR_INCOME	1.19	0.838238
RIOS	1.19	0.839248
HOUSEHOLD_SIZE	1.19	0.843822
FIRMSIZE3	1.19	0.843867
ATACAMA	1.18	0.848447
ANTOFAGASTA	1.17	0.851527
TARAPACA	1.16	0.859699
NUBLE	1.15	0.869384
ARICA	1.15	0.871073
COQUIMBO	1.14	0.875322
MAGALLANES	1.12	0.889019
AYSEN	1.08	0.928483
FFAA	1.07	0.932209
NOT_SIGNED_IN	1.06	0.940881

(continuación) Tabla 8: Análisis de multicolinealidad método OLS año 2022

Variable	VIF	1/VIF
WITHOUT_EDUCATION	1.04	0.958709
PROMEDIO	1.54	

Fuente: Elaboración propia en base a resultados de Stata

5.2. Método LASSO

En primer lugar, se inicia un ajuste a los datos usando validación cruzada (CV, por sus siglas en inglés) de 10 pliegues o folds. Esto significa que los datos se dividen en 10 partes, y el modelo se entrena 10 veces. En cada iteración, el modelo usa 9 pliegues para entrenar y 1 pliegue para evaluar su desempeño. El valor de lambda varía en cada iteración iniciando desde 0.2447354 como se muestra en la tabla 9. El número de coeficientes distintos de cero, es decir, la cantidad de variables incluidas en el modelo varía desde 0 hasta 41. El proceso se realiza mediante validación cruzada con 100 lambdas y 10 folds, lambda fue seleccionado por el criterio de detención `stop()` y fue el número 71. El valor de lambda es el parámetro de regularización en el modelo LASSO, que controla la cantidad de penalización sobre los coeficientes y en este caso dio 0.0003634 como se aprecia en la tabla 11. A medida que el valor de lambda aumenta, más coeficientes se reducen a cero, lo que da lugar a un modelo más sencillo, en la figura 5 se muestra la evolución de los coeficientes para el análisis del año 2022. Entre cada iteración del modelo se presenta el valor de CVF que corresponde al desempeño del modelo. Un CVF más bajo es mejor, ya que indica que el modelo se ajusta mejor a los datos sin sobre ajustarse, en este caso se inicia con un valor de 0.409 y se termina con un valor de 0.201.

Tabla 9: Primeras 5 iteraciones del método LASSO CV según datos de CASEN año 2022

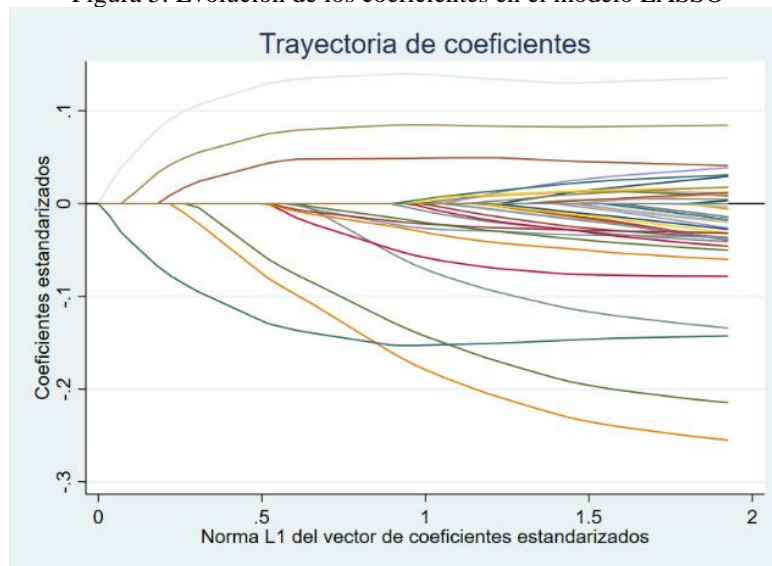
Grid	Value Lambda	Nonzero coef.	CVF	Folds
1	0.2447354	0	0.4096848	1...5....10
2	0.2229938	2	0.3945014	1...5....10
3	0.2031836	2	0.3791053	1...5....10
4	0.1851334	3	0.3635685	1...5....10
5	0.1686866	3	0.3503314	1...5....10

Fuente: Elaboración propia en base a resultados del uso de Stata

En el eje X de la figura 5 se muestra la suma de los valores absolutos de los coeficientes penalizados (la L-norma) yendo de 0 a 2. Cada línea traza el coeficiente penalizado para una de las covariables estandarizadas del modelo. Se observa que a medida que la norma L1 aumenta, más coeficientes se activan (se vuelven diferentes de cero), mientras que a medida que L1 disminuye más coeficientes se vuelven 0. Esto se debe a la penalización aplicada a cada una de las variables. En el gráfico, las líneas que se activan primero (se alejan de cero) a medida que disminuye λ representan las variables más relevantes para el modelo. Estas variables son seleccionadas antes porque tienen un mayor impacto en la variable dependiente.

Para este modelo LASSO la cantidad de observaciones fue 45,591 y las variables seleccionadas fueron 41 de 44.

Figura 5: Evolución de los coeficientes en el modelo LASSO



Fuente: Elaboración propia basado en resultados de Stata

Los resultados del modelo indican un R-cuadrado (R^2) de 0,5094 lo cual indica que el modelo explica aproximadamente el 51% de la variabilidad en la variable dependiente. En otras palabras, el 51% de la variabilidad en la variable dependiente (LOG_WAGE) puede ser explicado por las variables independientes incluidas en el modelo. Los coeficientes estimados para el modelo se encuentran presentes en la tabla 10.

Dados los resultados de los coeficientes estimados en la tabla 5 y la tabla 10, se observa que los resultados son cercanos al menos para los coeficientes de FATALITY_RATE y la constante del método OLS y LASSO CV. Por lo que los valores de VSL estimados serán similares para el año 2022 mediante estos métodos, esto podría darse porque el valor de lambda (el parámetro de regularización) es muy pequeño. Por lo tanto, se busca utilizar otro método llamado LASSO adaptativo. Este método se aplica cuando algunas variables debieran recibir una penalización especial y no igual a las demás. El valor de lambda seleccionado fue de 0.0003061 para este método y los coeficientes estimados están en la tabla 10. Para optimizar la selección de las variables y la penalización para cada una de ellas se puede aumentar el número de pasos o etapas en el proceso adaptativo. En este caso se inició con 2 etapas y luego se aumentó a 20 etapas que es la cantidad recomendable, con esto el modelo

logra mayor refinamiento al penalizar más variables irrelevantes, estabilizando el resultado.

Tabla 10: Coeficientes método LASSO según CASEN año 2022

Variable	LASSO CV	LASSO adaptativo 2 etapas	LASSO adaptativo 20 pasos	Estimador Plugin heterocedástico	Estimador Plugin homocedástico
FATALITY_RATE	1022.478	1022.478	1023.261	933.256	937.741
TARAPACA	-0.0780	-0.0780	-0.0780	omitida	omitida
ANTOFAGASTA	0.0380	0.0380	0.0380	0.0690	0.0690
ATACAMA	-0.0350	-0.0350	-0.0340	omitida	omitida
COQUIMBO	-0.0840	-0.0840	-0.0840	omitida	omitida
VALPARAISO	-0.1330	-0.1330	-0.1330	-0.1020	-0.1020
OHIGGINS	-0.1050	-0.1050	-0.1050	-0.0730	-0.0740
MAULE	-0.1560	-0.1560	-0.1570	-0.1250	-0.1250
BIOBIO	-0.1400	-0.1400	-0.1410	-0.1100	-0.1090
ARAUCANIA	-0.1780	-0.1780	-0.1780	-0.1460	-0.1460
LAGOS	-0.0830	-0.0830	-0.0840	omitida	omitida
AYSEN	0.0330	0.0330	0.0320	0.0640	0.0640
MAGALLANES	0.0630	0.0630	0.0620	0.0940	0.0940
RIOS	-0.1490	-0.1490	-0.1500	-0.1180	-0.1180
ARICA	-0.1050	-0.1050	-0.1050	-0.0740	-0.0730
NUBLE	-0.1540	-0.1540	-0.1540	-0.1220	-0.1220
WITHOUT_EDUCATION	-0.6140	-0.6140	-0.6140	-0.6220	-0.6200
EDUCATION_1	-0.6320	-0.6320	-0.6320	-0.6380	-0.6380
EDUCATION_2	-0.5150	-0.5150	-0.5140	-0.5190	-0.5190
EDUCATION_3	-0.3870	-0.3870	-0.3860	-0.3890	-0.3890
WOMAN	-0.1610	-0.1610	-0.1630	-0.1630	-0.1630
AGE	0.0000	0.0000	0.0000	0.0000	0.0000
CHILD_SIX	-0.0100	-0.0100	omitida	omitida	omitida
AVG_OTHERS_FIRMS	0.0010	0.0010	0.0010	0.0010	0.0010
NON_FATALITY_RATE	-4.1630	-4.1640	-4.1680	-4.0390	-4.0360
FULLTIME	0.0970	0.0970	0.0970	0.1010	0.1000
SIGNED_CONTRACT	0.2400	0.2400	0.2400	0.2310	0.2390
NOT_SIGNED_IN	0.1160	0.1160	0.1160	omitida	0.1170
MARRIED	0.0350	0.0350	0.0360	0.0360	0.0360
EDUCATION_SPOUSE	0.0210	0.0210	0.0210	0.0210	0.0210
HOUSEHOLD_SIZE	0.0030	0.0030	omitida	omitida	omitida
NON_LABOR_INCOME	0.0000	0.0000	0.0000	0.0000	0.0000
FIRMSIZE1	-0.2400	-0.2400	-0.2400	-0.2430	-0.2410
FIRMSIZE2	-0.1900	-0.1900	-0.1910	-0.1970	-0.1950
FIRMSIZE3	-0.1560	-0.1560	-0.1570	-0.1620	-0.1620
FIRMSIZE4	-0.1130	-0.1130	-0.1130	-0.1170	-0.1170
FIRMSIZE5	-0.0500	-0.0500	-0.0500	-0.0540	-0.0540

(continuación) Tabla 10: Coeficientes método LASSO según CASEN año 2022

Variable	LASSO CV	LASSO adaptativo 2 etapas	LASSO adaptativo 20 pasos	Estimador Plugin heterocedástico	Estimador Plugin homocedástico
JOBSCOPE	-0.042	-0.042	-0.043	-0.043	-0.042
FREELANCE	omitida	omitida	omitida	omitida	omitida
PUBLIC_EMPLOYEE	0.032	0.032	0.032	0.026	0.026
PART_TIME	-0.483	-0.483	-0.484	-0.48	-0.479
FFAA	0.169	0.169	0.169	0.16	0.161
UNPAID_FAMILY	omitida	omitida	omitida	omitida	omitida
_cons	13.345	13.345	13.356	13.332	13.325

Fuente: Elaboración propia en base a resultados del uso de Stata

Como se aprecia en la tabla 12 no hay diferencias en la estimación del VSL entre las dos formas de LASSO adaptativo. Luego se aplicaron otros métodos adicionales llamados estimador plugin heterocedástico y estimador plugin homocedástico que corresponden a métodos heurísticos.

Tabla 11: Valores de lambda para los diferentes métodos

Método	ID	lambda	Noncero coef	R ²	error
Validación cruzada	71	0.0003634	41	0.5093	0.2011222
LASSO adaptativo 2 pasos (ridge)	198	0.0002782	41	0.5094	0.2010986
LASSO adaptativo 20 pasos (ridge)	1897	0.0003061	38	0.5093	0.201114
Estimador plugin heterocedástico	*1	0.0190234	34	0.5016	
Estimador plugin homocedástico	**1	0.0085367	35	0.5013	

Fuente: Elaboración propia en base a resultados del uso de Stata

Notas: * El valor de lambda se obtuvo mediante por la fórmula plugin asumiendo errores heterocedásticos

** El valor de lambda se obtuvo mediante por la fórmula plugin asumiendo errores homocedásticos

Los valores de los coeficientes resultantes están en la tabla 10 y el valor del VSL obtenido a partir de estos métodos están en la tabla 12. En el caso de los métodos heurísticos el valor de lambda es escogido por una fórmula y no sufre modificaciones a lo largo de las iteraciones. Al comparar los valores obtenidos de VSL en la tabla 6 y la tabla 12 se puede observar una disminución en la estimación, esto se debe a la selección de las variables, la regularización y la optimización de cada método. Además, los métodos como el estimador plugin homocedástico tienden a reducir el valor requerido debido a la simplificación que implica asumir una varianza constante.

En la tabla 12 también se observan los valores obtenidos para el VSL del año 2017. Hubo diferencia entre los valores obtenidos por el método LASSO y el método OLS de la tabla 6.

Luego hubo una disminución en el valor para el método LASSO adaptativo en 20 pasos y con los métodos heurísticos no se pudo determinar el VSL puesto que excluían a la variable FATALITY_RATE del modelo.

Tabla 12: VSL mediante LASSO (millones de dólares)

	2017	2022
LASSO validación cruzada (CV)	2.57	8.66
LASSO adaptativo 2 pasos	2.29	8.86
LASSO adaptativo 20 pasos	2.54	8.86
Estimador plugin heterocedástico		3.85
Estimador plugin homocedástico		3.69

Fuente: Elaboración propia en base a resultados de Stata

5.3. Comparación de resultados entre LASSO y métodos tradicionales

En este estudio se utilizaron diferentes métodos para la obtención del VSL para Chile, utilizando datos del año 2017 y 2022. Uno de ellos fue el método OLS y otro de aprendizaje automático denominado LASSO. Además, se incorporaron métodos heurísticos.

Al aplicar el método tradicional para la CASEN del año 2022 se obtuvo un VSL de 8.93 dando un resultado convincente respecto a estudios anteriores. Además, este método produce un ajuste más directo a los datos al no aplicar una penalización, aunque esto puede traer consigo problemas de multicolinealidad. Por otro lado, al aplicar el método LASSO se observa que tres variables son eliminadas para el método con 20 etapas, lo que puede mejorar la precisión predictiva al reducir el sobreajuste. Otra diferencia es la interpretabilidad. Esto se debe a que OLS ofrece una interpretación más directa en la influencia de cada variable en el logaritmo del salario. Por otro lado, LASSO tiende a iniciar con todos los coeficientes de las variables con valor 0, y luego algunas variables se mantienen así, de forma tal que el modelo sea más parsimonioso, pero más difícil de interpretar. En este caso además existen muchas variables y muchas observaciones por lo que el método LASSO suele ser mejor para medir efectos de multicolinealidad. Esto se demuestra en la tabla 8 donde algunas de las variables tenían valores altos de VIF.

En la tabla 13 se muestra una comparación general de los métodos aplicados para la encuesta CASEN 2022 y se puede concluir que tienen un rendimiento similar a pesar de la forma de estimación. Los valores de R^2 oscilan entre 0.5013 y 0.510, lo que indica que todos los métodos tienen un poder predictivo comparable para estimar el Valor Estadístico de la Vida

(VSL). Por otro lado, los valores de VSL obtenidos son bastante consistentes para los métodos no heurísticos. Oscilan entre 8.66 y 8.93 millones de dólares lo que podría indicar robustez en los resultados a pesar de los diferentes enfoques de estimación. Otro aspecto a considerar es el tiempo de ejecución ya que métodos como OLS tienen un tiempo bajísimo de ejecución en comparación con LASSO de 20 etapas. Luego si el tiempo de ejecución fuera un factor relevante se debería escoger este método. Al analizar posteriormente los resultados de los métodos heurísticos tienen baja robustez y son poco convincentes debido a la diferencia de estimación respecto a los demás métodos.

Tabla 13: Comparación métodos de estimación año 2022

Método	Ventajas	Desventajas	Número de variables	Rendimiento (R ²)	VSL (millones de dólares)	Tiempo de ejecución
OLS	Fácil interpretación	Sensible a multicolinealidad	42	0.51	8.93	1 segundos
LASSO CV	Falta de regularización	Poca o nula penalización	41	0.5093	8.66	19 segundos
LASSO 2 etapas	Reduce Sobreajuste	Pérdida de interpretabilidad	41	0.5094	8.86	25 segundos
LASSO 20 etapas	Elimina variables irrelevantes	Mayor complejidad computacional	38	0.5093	8.86	235 segundos
Estimadores plugin	Fácil aplicabilidad	Poca robustez	34-35	0.5013-0.5016	3.69-3.85	4 segundos

Fuente: Elaboración propia en base a resultados de Stata

Si la interpretabilidad y la rapidez son cruciales, OLS es una buena opción para la estimación del VSL, pero hay que tener en cuenta la sensibilidad a la multicolinealidad. Si se busca una reducción de sobreajuste y una selección más precisa de variables, los métodos LASSO (CV, 2 etapas o 20 etapas) pueden ser más apropiados. Sin embargo, si el tiempo de ejecución es una preocupación importante, LASSO CV y LASSO de 2 etapas ofrecen un buen equilibrio entre precisión y eficiencia.

5.4. Implicancia en políticas públicas

Al comparar las diferentes metodologías utilizadas en este estudio, se obtuvo una estimación más precisa y robusta del valor estadístico de la vida. Este resultado es fundamental para la formulación de políticas públicas enfocadas en la seguridad laboral y la protección de los trabajadores. Gracias a estas mejoras, es posible justificar una mayor inversión en medidas de seguridad en diversas industrias y una mejor prevención de accidentes laborales futuros.

En comparación, un estudio previo aplicado en Chile estimó el VSL utilizando el método de preferencias declaradas con datos de 2014 (Ministerio de Desarrollo Social, 2017). Aquel análisis se basó en una muestra mucho más pequeña, con solo 1,200 personas, limitada a la Región Metropolitana, mientras que este estudio abarca una muestra representativa a nivel nacional, con más de 80,000 observaciones. Además, las encuestas de preferencias declaradas están sujetas a diversos sesgos cognitivos y de comportamiento, como la sobreestimación del riesgo, la deseabilidad social y la inconsistencia en las respuestas. Estos factores afectan la validez de las estimaciones.

Por el contrario, los métodos automáticos utilizados en este trabajo, como LASSO adaptativo, ofrecen varias ventajas clave. Permiten manejar grandes volúmenes de datos y seleccionar automáticamente las variables más relevantes, eliminando aquellas que presentan colinealidad o carecen de variabilidad significativa. Este enfoque garantiza una mayor precisión en la identificación de los factores que determinan los riesgos laborales y los salarios, lo que es esencial para diseñar políticas públicas más focalizadas, como regulaciones específicas por sector o incentivos para industrias de alto riesgo.

Además, los métodos utilizados responden al comportamiento real de los trabajadores, reflejando mejor su disposición a aceptar riesgos y aumentando la validez externa de los resultados. Una vez disponible la base de datos, los modelos automáticos permiten realizar análisis de manera rápida y repetitiva, lo que facilita actualizaciones constantes y simulaciones de escenarios futuros. A diferencia del método de preferencias declaradas, que implica costos elevados y un proceso complejo de diseño y validación, el enfoque de este estudio proporciona una evaluación dinámica y sostenible de políticas.

En conclusión, el análisis comparativo de metodologías llevado a cabo en este estudio fortalece la capacidad para desarrollar políticas públicas basadas en evidencia empírica robusta, mejorando la eficiencia y precisión en la asignación de recursos para la seguridad laboral y la protección de los trabajadores.

6. Conclusión

En el presente estudio, se estimó el Valor Estadístico de la Vida (VSL) utilizando diversos métodos de estimación, incluyendo métodos tradicionales como mínimos cuadrados ordinarios (OLS), enfoques automatizados como el método LASSO y sus variaciones, así como métodos heurísticos. Los resultados obtenidos muestran que el VSL estimado mediante el método tradicional fue de US \$8.93 millones, mientras que el estimado mediante LASSO con 20 etapas fue de US \$8.86 millones.

La diferencia observada entre ambos métodos puede atribuirse a la penalización aplicada por LASSO, que tiende a reducir la magnitud de los coeficientes menos relevantes y, en consecuencia, ajusta la estimación final. Esta penalización ha implicado una disminución en el VSL estimado, lo cual es consistente con las expectativas teóricas del modelo de LASSO, que prioriza la simplificación y generalización del modelo sobre el ajuste perfecto a los datos de entrenamiento.

En términos generales, los métodos de estimación automáticos como LASSO ofrecen una alternativa robusta al método tradicional, especialmente en presencia de multicolinealidad y cuando se trabaja con un gran número de variables independientes. Sin embargo, la penalización aplicada por LASSO también conlleva una disminución en el VSL, lo que resalta la importancia de ajustar cuidadosamente los parámetros del modelo según el contexto y los objetivos del análisis.

Este estudio resalta las ventajas y limitaciones de cada enfoque, subrayando que, mientras que los métodos tradicionales como OLS son más fáciles de interpretar, los métodos de aprendizaje automático como LASSO ofrecen ventajas en términos de manejo de la multicolinealidad y la selección de variables relevantes, aunque con una ligera reducción en la estimación final del VSL.

7. Referencias

- Alsayed, N., & Awad, W. (2023). A framework for Labor Market Analysis using Machine Learning., (págs. 1-5). Manama, Bahrain. doi:<https://doi.org/10.1109/ITIKD56332.2023.10099838>
- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of Physics: Conference Series*, 1142. IOP Publishing Ltd. doi:<https://doi.org/10.1088/1742-6596/1142/1/012012>
- Andersson, H. (2005). The value of safety as revealed in the Swedish car market: An application of the hedonic pricing approach. *Journal of Risk and Uncertainty*, 30(3), 211–239. doi:<https://doi.org/10.1007/s11166-005-1154-1>
- Andersson, H. (2020). The value of a statistical life. *Advances in Transport Policy and Planning*, 6, 75-99. doi:<https://doi.org/10.1016/bs.atpp.2020.07.004>
- Atkinson, S. E. (1990). The valuation of risks to life: Evidence from the market for automobiles. *Review of Economics and Statistics*, 72(1), 133–136.
- Awad, N. A. (2023). A framework for Labor Market Analysis using Machine Learning. 2023 *International Conference on IT Innovation and Knowledge Discovery (ITIKD)*, (págs. 1-5). Manama, Bahrain. doi:<https://doi.org/10.1109/ITIKD56332.2023>
- Batta, M. (2020). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR)*, 9(1), 381-386. doi:<https://doi.org/10.21275/ART20203995>
- Belloni, A., & Chernozhukov, V. (2011). *High dimensional sparse econometric models: An introduction*. (P. Alguier, E. Gautier, & G. Stoltz, Edits.) Berlín: Springer.
- Bharti, S., Bandyopadhyaya, R., & Raju, N. K. (December de 2022). Estimation of Willingness to Pay and Value of Statistical Life for Road Crash Fatality Reduction for Motorcyclists: A Case Study of Patna, India. *J. Inst. Eng. India Ser. A*, 103(4), 1315–1323. doi: <https://doi.org/10.1007/s40030-022-00680-y>
- Cameron, T. (2010). Euthanizing the value of a statistical life. *Review of Environmental Economics and Policy*, 4(2), 161-178.
- Carson, R. T. (2014). Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. *Journal of the Association of Environmental and Resource Economists*, 1(1/2), 171-207. doi:<https://doi.org/10.1086/676450>
- Cheng, X. V. (2019). Evaluation of cultural ecosystem services: A review of methods. *Ecosystem Services*, 37. doi:<https://doi.org/10.1016/j.ecoser.2019.100925>
- Cifuentes, L., Prieto, J., & Escobari, J. (1999). Valuing Mortality Risk Reductions at Present and at an Advanced Age: Preliminary Results from a Contingent Valuation Study in Chile. *XVII Latin American Meeting of the Econometric Society*. Cancun, Mexico.
- COFEMER. (2013). *Guía para evaluar el impacto de la regulación - Vol. I. Métodos y Metodologías*. México: COFEMER, SE, y APEC.
- Cukrowska-Torzewska, E., & Matysiak, A. (2020). The motherhood wage penalty: A meta-analysis. *Social Science Research*, 88–89, 102416. doi:<https://doi.org/10.1016/j.ssresearch.2020.102416>.
- De Blaeij, A. F. (2003). The value of statistical life in road safety: a meta-analysis. *Journal of Risk and Uncertainty*, 27(3), 251-277. doi:[https://doi.org/10.1016/S0001-4575\(02\)00105-7](https://doi.org/10.1016/S0001-4575(02)00105-7)
- Evans, M., & Taylor, L. (2020). Using revealed preference methods to estimate the value of reduced mortality risk: best practice recommendations for the hedonic wage model. *Review of Environmental Economics and Policy*, 14(2), 282-301. doi:<https://doi.org/10.1093/reep/reaa006>
- EViews. (2024). *EViews 8 User Guide*. Obtenido de EViews: https://www.eviews.com/EViews8/ev8echeck_n.html
- Farhat, R., Mourali, Y., Jemni, M., & Ezzedine, H. (2020). An overview of machine learning technologies and their use in e-learning. *2020 International Multi-Conference on:*

- “*Organization of Knowledge and Advanced Technologies*” (OCTA). (págs. 1-4). IEEE. doi:<https://doi.org/10.1109/OCTA49274.2020.9151758>
- Freeman, A. M. (2014). *The Measurement of Environmental and Resource Values: Theory and Methods* (3rd Edition ed.). New York: Routledge. doi:<https://doi.org/10.4324/9781315780917>
- Gentry, E. V. (2016). The fatality and morbidity components of the value of statistical life. *J. Health Econ.*, 46, 90–99.
- GreenLabUC. (2014). *Estimación del valor de la vida estadística asociado a contaminación atmosférica y accidentes de tránsito*. Santiago: Instituto de Sociología, Facultad de Ciencias Sociales, Pontificia Universidad Católica de Chile y DICTUC S.A. Obtenido de <https://greenlab.uc.cl/wp-content/uploads/2016/04/2014-VVE-Informe-Final-23Dic2014-1125-casos.pdf>
- Hammit, J. (2000). Evaluating contingent valuation on environmental health risks: the proportionality test. *Journal of the Association of Environmental and Resource Economists*, 20(1), 14-19.
- Hojman, P., Ortuzar, J., & Rizzi, L. (2005). On the joint valuation of averting fatal and severe injuries in highway accidents. *Journal of Safety Research*, 36(4), 377-386. doi: <https://doi.org/10.1016/j.jsr.2005.07.003>
- IBM. (30 de junio de 2020). *Linear regression from scratch*. Obtenido de IBM Developer: <https://developer.ibm.com/articles/linear-regression-from-scratch/>
- IBM. (16 de enero de 2024). *IBM*. Obtenido de <https://www.ibm.com/topics/lasso-regression>
- IBM. (2024). *Two-stage least squares regression model*. Obtenido de IBM: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-two-stage-least-squares-model>
- Iragüen, P., & Ortúzar, J. (2004). Willingness-to-pay for reducing fatal accident risk in urban areas: an Internet-based Web page stated preference survey. *Accident Analysis & Prevention*, 36(4), 513-524. doi:[https://doi.org/10.1016/S0001-4575\(03\)00057-5](https://doi.org/10.1016/S0001-4575(03)00057-5)
- Jagsi, R., Bekelman, J., Chen, A., Chen, R., & Hoffman, K. (2014). Considerations for Observational Research Using Large Data Sets in Radiation Oncology. *International Journal of Radiation Oncology Biology Physics*, 90(1), 11-24. doi:<https://doi.org/10.1016/j.ijrobp.2014.05.013>
- Jones-Lee, M. (1974). The value of changes in the probability of death or injury. *Journal of Political Economy*, 82(4), 835-849.
- Kissell, R., & Poserina, J. (2017). *Optimal sports math, statistics, and fantasy*. Academic Press. doi:<https://doi.org/10.1016/B978-0-12-805163-4.00002-5>
- Kniesner, T. J., Viscusi, W. K., & Christopher, W. (2010). *The Value of a Statistical Life: Evidence from Panel Data*. Syracuse, NY: Syracuse University, Center for Policy Research. Obtenido de <https://surface.syr.edu/cpr/44>
- Kniesner, T., & Leeth, J. (2010). *Hedonic wage equilibrium: Theory, evidence and policy*. Institute for the Study of Labor (IZA), Bonn, Alemania. Obtenido de <https://docs.iza.org/dp5076.pdf>
- Lanoie, P., Pedro, C., & Latour, R. (1995). The value of a statistical life: A comparison of two approaches. *J Risk Uncertainty*, 10, 235–257. doi:<https://doi.org/10.1007/BF01207553>
- Lee, L.-f. (2003). Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews*, 22(4), 307-335.
- Leggett, C., & Bockstael, N. (2000). Evidence of the Effects of Water Quality on Residential Land Prices. *Journal of Environmental Economics and Management*, 39(2), 121-144. doi:<https://doi.org/10.1006/jeem.1999.1096>
- Leigh, J. (1987). Gender, firm size, industry, and estimates of the value-of-life. *Journal of Health Economics*, 6(3), 255-273. doi:[https://doi.org/10.1016/0167-6296\(87\)90012-9](https://doi.org/10.1016/0167-6296(87)90012-9)
- Looze, J. (2017). Why do(n't) they leave?: Motherhood and women's job mobility. *Social Science*

- Research*, 65, 47-59. doi:<https://doi.org/10.1016/j.ssresearch.2017.03.004>
- Mardones, C., & Riquelme, M. (2018). Estimation of the Value of Statistical Life in Chile and Extrapolation to Other Latin American Countries. *Latin American Research Review*, 53(4), 815–830. doi:10.25222/larr.61
- Ministerio de Desarrollo Social y Familia. (2023). *Pobreza por ingresos, pobreza multidimensional y distribución de los ingresos*. Santiago: Ministerio de Desarrollo Social y Familia. Obtenido de <https://www.desarrollosocialyfamilia.gob.cl>
- Ministerio de Desarrollo Social y Familia. (s.f.). *Observatorio Social*. Recuperado el 06 de octubre de 2024, de <https://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen>
- Ministerio de Desarrollo, S. (2017). *Estimación del valor de la vida: Estadística en Chile a través del enfoque de disposición a pagar*. Obtenido de <https://sni.gob.cl/storage/docs/Minuta%20actualiza%20VVE%20Final%20052017.pdf>
- Mon, E. E., Jomnonkwao, S., Khampirat, B., Satiennam, W., & Ratanavaraha, V. (2018). Willingness to pay for mortality risk reduction for traffic accidents in Myanmar. *Accident Analysis & Prevention*, 118, 18-28. doi:<https://doi.org/10.1016/j.aap.2018.05.018>
- Morgan, F. (2015). The motherhood penalty and its impact of the career decisions of working women. doi: <https://doi.org/10.13140/RG.2.1.3070.9288>
- Motamedi, F., Pérez-Sánchez, H., Mehridehnavi, A., Fassihi, A., & Ghasemi, F. (2022). Accelerating Big Data Analysis through LASSO-Random Forest Algorithm in QSAR Studies. *Bioinformatics*, 38(2), 469-475. doi:<https://doi.org/10.1093/bioinformatics/btab659>
- Muthukrishnan, R., & Rohini, R. (2016). A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)* (págs. 18-20). Coimbatore, India: IEEE. doi:<https://doi.org/10.1109/ICACA.2016.7887916>
- OECD. (2014). *The Cost of Air Pollution: Health Impacts of Road Transport*. París: OECD Publishing. doi:<https://doi.org/10.1787/9789264210448-en>
- Orozco-Castañeda, J., Sierra-Suárez, L., & Vidal, P. (2024). Labor market forecasting in unprecedented times: A machine learning approach. *Bulletin of Economic Research*, 1–23. doi:<https://doi.org/10.1111/boer.12451>
- Parada-Contzen, M., Riquelme-Won, A., & Vasquez-Lavin, F. (2012). The value of a statistical life in Chile.
- PennState , U. (03 de 12 de 2024). *Detecting multicollinearity using variance inflation factors*. Obtenido de <https://online.stat.psu.edu/stat462/node/180/>
- Remoundou, K., Koundouri, P., Kontogianni, A., Nunes, P. A., & Skourtos, M. (2009). Valuation of natural marine ecosystems: An economic perspective. *Environmental Science & Policy*, 12(7), 1040-1051. doi:<https://doi.org/10.1016/j.envsci.2009.06.006>
- Riera Font, A., Ripoll Penalva, A. M., & Mateu Sbert, J. (2007). Estimación del valor estadístico de la vida en España: Una aplicación del Método de Salarios Hedónicos. *Hacienda Pública Española / Revista de Economía Pública*, 181(2), 29-48.
- Rizzi, L. I., & De La Mazza, C. (2017). The external costs of private versus public road transport in the Metropolitan Area of Santiago, Chile. *Transportation Research Part A: Policy and Practice*, 98, 123-140. doi:<https://doi.org/10.1016/j.tra.2017.02.002>
- Rizzi, L., & Ortúzar, J. (2003). Stated preference in the valuation of interurban road safety. *Accident Analysis & Prevention*, 35(1), 9-22. doi:[https://doi.org/10.1016/S0001-4575\(01\)00082-3](https://doi.org/10.1016/S0001-4575(01)00082-3)
- Robinson, L., Hammitt, J., & O’Keeffe, L. (2018). *Valuing Mortality Risk Reductions in Global Benefit-Cost Analysis*. Guidelines for Benefit-Cost Analysis Project.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34-55.
- StataCorp. (2023). *Stata 18 Lasso Reference Manual*. College Station, TX: Stata Press.
- StataCorp LLC. (14 de noviembre de 2024). *Lasso model selection and prediction in Stata*. Obtenido de Stata: <https://www.stata.com/features/overview/lasso-model-selection->

- prediction/
- Strittmatter, A., & Wunsch, C. (2021). *The gender pay gap revisited with big data: Do methodological choices matter?* doi:<https://doi.org/10.48550/arXiv.2102.09207>
- Superintendencia de Seguridad Social. (2022). *Boletín Estadístico de Seguridad Social 2022*. Santiago.
- Taylor, L. (2003). The Hedonic Method. En P. Champ, K. Boyle, & T. Brown (Edits.), *A Primer on Nonmarket Valuation* (Vol. 3). Springer, Dordrecht. doi:https://doi.org/10.1007/978-94-007-0826-6_10
- Tibshirani, R., & Wasserman, L. (2017). *Sparsity, the Lasso, and Friends*. Statistical Machine Learning.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288. Obtenido de <https://www.jstor.org/stable/2346178>
- UCSF Knowledge Hub. (2024). *UCSF Knowledge Hub*. Obtenido de <https://khrc.ucsf.edu/lasso-correlated-variable-selection>
- Vicéns Otero, J. (2012). *Descomposición Oaxaca-Blinder en Modelos Lineales y No Lineales*.
- Viscusi, W. K. (1993). The value of risks to life and health. *Journal of Economic Literature*, 31(4), 1912–1946. doi:<http://www.jstor.org/stable/2728331>
- Viscusi, W. K. (2003). *The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World*. Cambridge, MA: National Bureau of Economic Research. Obtenido de <http://www.nber.org/papers/w9487>
- Viscusi, W., & Aldy, J. (2007). Labor market estimates of the senior discount for the value of statistical life. *Journal of Environmental Economics and Management*, 53(3), 377-392. doi:<https://doi.org/10.1016/j.jeem.2006.10.005>
- Wikipedia, c. (2024). *Wikipedia, The Free Encyclopedia*. Obtenido de [https://en.wikipedia.org/w/index.php?title=Lasso_\(statistics\)&oldid=1229890899](https://en.wikipedia.org/w/index.php?title=Lasso_(statistics)&oldid=1229890899)
- Wim Wijnen, P. W. (2009). Valuation of road safety effects in cost–benefit analysis. *Evaluation and Program Planning*, 32(4), 326-331. doi:<https://doi.org/10.1016/j.evalprogplan.2009.06.015>
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5 ed.). Mason, OH: Cengage Learning.
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western Cengage Learning.
- World Health Organization. (2018). *Global status report on road safety 2018*. Geneva: World Health Organization. Obtenido de <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>
- Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429. doi:<https://doi.org/10.1198/016214506000000735>