



**”EFECTOS HETEROGÉNEOS EN SINDICALIZACIÓN,
UN ENFOQUE DESDE MACHINE LEARNING.”**

**TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN ANÁLISIS
ECONÓMICO.**

**Alumno: Leonardo Lanata Castro
Profesor Guía: Esteban Puentes**

Santiago, Enero 2023

Efectos Heterogéneos en Sindicalización, un Enfoque desde Machine Learning

Leonardo Lanata^{1,2}

¹ Facultad de Economía y Negocios, Universidad de Chile
leonardolanata@gmail.com

² Supervisor: Esteban Puentes

Abstract. *We study heterogeneity in the effect of union on wages through an observational dataset from the Encuesta de Protección Social (EPS) in Chile. The analysis uses Machine Learning to address the following associated problems: assessing treatment group positivity and covariate balance, imputing conditional average treatment effects, and interpreting imputed models. By comparing several different meta-learners models we illustrate the flexibility of these estimators in regards to predict treatments effects. We find that unionization of workers has a positive effect on wages of 3.6%, 95%-CI of [3.02%,4.18%], and that the heterogeneity in the range of [-21%,65%] is determined by industries, firm size, experience, occupations, and gender.*

Abstract. Estudiamos la heterogeneidad en el efecto de la sindicalización sobre los salarios a través de un conjunto de datos observacionales de la Encuesta de Protección Social (EPS) en Chile. El análisis utiliza *Machine Learning* para abordar los siguientes problemas: evaluar el supuesto de positividad de los grupos de tratamiento y el balance de covariables; computar efectos condicionales promedio del tratamiento e interpretar los modelos imputados. Al comparar varios modelos diferentes de *meta-learners*, ilustramos la flexibilidad de estos estimadores en lo que respecta a predecir los efectos de los tratamientos. Encontramos que la sindicalización de los trabajadores tiene un efecto positivo sobre los salarios de 3.6%, con un intervalo de confianza al 95% de [3.02%,4.18%], y que la heterogeneidad contenida en el rango de [-21%,65%] está determinada por industrias, tamaño de la firma, experiencia laboral, ocupaciones y género.

Keywords: *Machine Learning · Interpretability · Counterfactual Estimation.*

1 Introducción y motivación

El *Machine Learning* ha tenido un éxito ampliamente aceptado en la resolución de problemas de predicción. Sobre esto, hay aplicaciones que van desde el reconocimiento de imagen y voz (LeCun et al., 2015) [13] hasta el diagnóstico de enfermedades en medicina (Kononenko, 2001) [9]. En particular, el aprendizaje

supervisado va en torno a resolver algún problema de predicción, en el sentido de que produce un predicho \hat{y} a partir de un set de covariables (o *features*) x . En general, la econometría va más de buscar relaciones causales entre variables, por lo que esta premisa asume hallar un problema \hat{y} relevante. De esta manera, podríamos considerar el *Machine Learning* como una metodología válida para la resolución de ciertos problemas econométricos. Por otro lado, el atractivo de estas metodologías viene del hecho de poder descubrir patrones generalizables, de estructura compleja y con flexibilidad en el *input* de datos. Todo esto, sin caer en el problema del *overfitting*: estas herramientas pueden encontrar funciones que predican apropiadamente fuera de la muestra.

En la literatura podemos revisar varias aplicaciones de *Machine Learning* en economía. En particular, estos modelos pueden ser aplicados resolviendo preguntas tradicionales pero con nuevos conjuntos de datos. Por ejemplo, podemos medir la actividad económica utilizando imágenes satelitales o en base a titulares de las noticias con técnicas de procesamiento de lenguaje natural ³. En el caso de Chile, son muy pocas las aplicaciones documentadas. En Contreras et al., 2018 [2] estimaron varios modelos de *Machine Learning* para predecir el medio por el cual los individuos se movilizan hacia su trabajo y con sus predicciones generaron *proxies* para el tiempo de viaje de los trabajadores. En Kausel et al., 2018 [7] lograron determinar que la relación ancho-alto de la cara podría predecir el desempeño académico de los estudiantes, donde encontraron que rasgos dominantes en varones podrían tener un efecto positivo en el desempeño de asignaturas de carácter cualitativo. Recientemente, Leal et al., 2020 [12] estimaron la inflación para Chile con *Machine Learning* y encontraron que en términos predictivos, estos modelos se desempeñan de manera similar a los modelos internos del banco, pero que en términos estructurales no aportan nada debido a su dificultad interpretativa.

En particular para esta tesis, nos interesa realizar una aplicación de estas metodologías en un contexto de sindicatos. Sobre esto podemos comenzar de la base que el mercado laboral chileno ha experimentado importantes transformaciones en las últimas décadas, pero en general se comporta como un mercado competitivo (Landerretche, et al., 2013) [11]. Durante el período 1993-2009, la productividad laboral promedio en Chile creció un 2,6% en promedio, mientras que los salarios reales crecieron un 2,8%. Si se aísla un período más reciente y libre de crisis (2001-2008), la productividad laboral promedio creció un 1,66%, mientras que los salarios reales crecieron un 1,70%. Esto sugiere que los salarios en Chile se mueven aproximadamente en línea con la productividad.

De acuerdo al Informe de Actualidad Laboral de la Dirección del Trabajo, Chile tiene una tasa de sindicalización entre el 14 y 20%, que ha estado aumentando en los últimos años. La legislación laboral chilena es una de las más restrictivas en

³ Para ver varias aplicaciones en economía internacional, véase Mullainathan y Spiess, 2017 [18]

cuanto al poder de negociación sindical en el mundo. Las leyes laborales locales no hacen obligatoria la afiliación sindical para beneficiarse de los resultados de la negociación colectiva, lo que fomenta el comportamiento de los "free riders". En caso de huelga, los empleadores chilenos tienen derecho a reemplazar inmediatamente a los trabajadores si se les ha ofrecido un aumento igual a la inflación, lo que reduce significativamente el poder de negociación y hace que la sindicalización sea menos atractiva y menos efectiva en las negociaciones salariales.

La tasa de sindicalización varía significativamente entre sectores, siendo más alta en las principales industrias extractivas de exportación de Chile, como la minería, servicios públicos, salud y educación. Los servicios financieros tienen una tasa de sindicalización del 12%, y los sectores de manufactura no metálica y alimentaria tienen una tasa del 10%. Además, la tasa de sindicalización varía según el tamaño de la empresa, siendo del 54% en las grandes empresas y menos del 1% en las microempresas. Estos resultados sugieren que la sindicalización sigue siendo una característica importante del mercado laboral chileno, especialmente en sectores clave de la economía.

Según datos del Instituto Nacional de Estadísticas de Chile (INE), los trabajadores sindicalizados en promedio ganan más que los no sindicalizados. En el trimestre móvil mayo-julio de 2021, los trabajadores afiliados a sindicatos en Chile ganaron en promedio un 39,1% más que los no afiliados. Además, la brecha salarial entre ambos grupos ha venido aumentando en los últimos años, pasando de un 28,6% en 2017 a un 39,1% en 2021. Estos datos sugieren que la sindicalización puede tener un efecto positivo en los salarios de los trabajadores en Chile. Sin embargo, es importante destacar que la sindicalización puede afectar de manera diferente a distintos sectores y tipos de trabajadores, por lo que se requiere un análisis de heterogeneidad para entender completamente sus efectos en el mercado laboral chileno.

Así, este trabajo nos enfocaremos en aplicar herramientas de aprendizaje automático para estudiar la heterogeneidad en los efectos de la sindicalización. De hecho, en términos teóricos, las metodologías de *Machine Learning* sobresalen en la superación de las conocidas limitaciones de los métodos tradicionales que emplea la econometría para resolver este tipo de problemas. Por ejemplo, los métodos de *matching* no suelen tener un buen desempeño cuando el conjunto de covariables es de varias dimensiones (Rubin y Thomas, 1996) [22]. Los modelos lineales generalizados no son lo suficientemente flexibles para descubrir interacciones variables y tendencias no lineales. Además, los métodos basados en *propensity* tienen problemas de alta-varianza en la estimación (Lee et al., 2011) [14]. Por el contrario, el aprendizaje supervisado ha demostrado ser útil para descubrir patrones en datos de alta dimensionalidad (Lecun et al., 2015) [13], aproximar funciones complejas y en el intercambio (*trade-off*) entre sesgo y varianza (Swaminathan y Joachims, 2015) [24].

En este estudio observacional basado en datos de la Encuesta de Protección Social de Chile, aplicamos estimadores de dos etapas de la familia de modelos *Meta-Learners*, diseñados para caracterizar la heterogeneidad en el efecto de la sindicalización sobre el salario de los trabajadores. La heterogeneidad final se computará empleando el modelo de mejor desempeño. Con esto en consideración, este estudio es relevante y aporta a la literatura en el sentido de aplicar y caracterizar un marco metodológico para la evaluación detallada de programas en Chile, utilizando *Machine Learning*.

En la sección 2 describiremos nuestro problema y revisaremos la metodología de estimación. En la sección 3 revisaremos los resultados e interpretaremos el modelo. Por último en la sección 4 concluiremos nuestro análisis.

2 Metodología y Datos

2.1 Descripción del problema y los datos

Estudiamos el efecto de los sindicatos en los salarios de los trabajadores, basados en datos de la Encuesta de Protección Social (EPS) de Chile. Para la edición de 2002-2009, la muestra comprende un total de 3462 observaciones. Esta base de datos cuenta con información a nivel individuo, de su salario, experiencia laboral, industria y tamaño de la firma en la que trabaja, si es que este tiene algún tipo de limitación física y entre otras. La afiliación a un sindicato (desde ahora el tratamiento) es representada por una variable binaria $T \in \{0, 1\}$, mientras el *outcome*, $Y \in R$ representa al logaritmo del salario por hora. Los trabajadores son observados a través de las covariables X_0 que corresponde a la experiencia en años; X_1 que representa categóricamente el nivel educacional, siendo el primero nivel menor a 12 años, luego exactamente 12 y para más de 12 años. X_2 , X_3 , X_4 y X_5 son variables dicotómicas que identifican respectivamente a trabajadores de mediana habilidad, si tienen alguna restricción física, su género y si son trabajadores del sector privado o no. X_6 corresponde al tamaño de la empresa, comenzando el primer tramo entre 10 y 49 trabajadores, luego entre 50 y 199, para finalmente clasificar a las firmas mas grandes con más de 500. X_7 describe las posibles ocupaciones (profesionales, técnicos, administrativos, venta y servicios, etc). Por último X_8 describe la industria del trabajador, ya sea manufactura, minería, construcción, etc. Por conveniencia, dejamos a $X = [X_1, X_2, \dots, X_8]^T$ representar el conjunto de covariables observables para los trabajadores de la muestra. Así, dejamos a $\{x_i, t_i, y_i\}$ denotar la observación correspondiente para el individuo $i \in \{1, \dots, n\}$. Los grupos del tratamiento observados los denotaremos como G_0 para el control y G_1 para los tratados y se definen por $G_z = \{i \in \{1, \dots, n\} : t_i = z\}$. El conjunto de datos observables lo denotamos $\mathcal{D} = \{(x_1, z_1, y_1), \dots, (x_n, z_n, y_n)\}$ y la densidad de las variables como $p(X, T, Y)$.

A continuación, presentamos las estadísticas descriptivas para la muestra descrita, segmentando por ocupaciones y afiliación al sindicato. La Tabla 2.1 presenta estadísticas resumidas combinadas en nuestra muestra. Lo primero a destacar

es la tasa creciente de sindicalización desde 2004 hasta 2009. Lo segundo a destacar es la alta tasa de sindicalización, superior al 20% en ambas submuestras, lo que es alto en comparación con otros conjuntos de datos. La estadística descriptiva también muestra que los salarios son más altos en promedio en el sector sindicalizado. Además, se observa una clara relación positiva entre el estatus sindical y el tamaño de la empresa y el sector público. La experiencia promedio también es ligeramente más alta en el sector sindicalizado. Se tiene que la distribución salarial del sector sindicalizado está desplazada hacia la derecha del sector no sindicalizado. Aunque este desplazamiento está altamente sesgado y no se puede asumir que sea más que la correlación ya mencionada, el tamaño del desplazamiento sugiere que hay diferencias significativas entre los dos grupos que se mantendrán incluso después de controlar por la endogeneidad del estatus sindical.

Variables	Todas las Ocupaciones				Ocupaciones de mediana habilidad			
	Sindicalizado		No Afiliado		Sindicalizado		No Afiliado	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
log(salario/hora)	7.450	0.633	7.180	0.631	7.347	0.521	7.152	0.525
Género	0.617	0.487	0.708	0.455	0.626	0.485	0.692	0.462
Restricción Física	0.046	0.209	0.025	0.158	0.026	0.160	0.027	0.162
Educación: 12 años	0.571	0.496	0.533	0.499	0.779	0.416	0.685	0.465
Educación >12 años	0.274	0.446	0.171	0.377	0.121	0.327	0.108	0.311
Experiencia	28.584	10.673	29.385	11.775	26.889	10.090	27.924	10.872
Experiencia al cuadrado	930.578	670.795	1,001.989	779.552	824.321	586.691	897.672	680.454
Trabajador del sector Privado r	0.623	0.485	0.875	0.331	0.737	0.442	0.910	0.286
Tamaño de la Firma: 10-49	0.155	0.362	0.322	0.468	0.147	0.355	0.315	0.465
Tamaño de la Firma: 50-199	0.234	0.424	0.216	0.412	0.189	0.393	0.204	0.404
Tamaño de la Firma: 200+	0.587	0.493	0.255	0.436	0.642	0.481	0.245	0.431
<i>Ocupaciones</i>								
Profesionales	0.137	0.344	0.081	0.273	0.000	0.000	0.000	0.000
Técnicos	0.128	0.334	0.068	0.252	0.158	0.366	0.090	0.286
Dependientes	0.140	0.347	0.135	0.341	0.179	0.384	0.204	0.404
Servicios y Retail	0.128	0.334	0.128	0.335	0.189	0.393	0.200	0.400
Agricultura y pesca capacitada	0.033	0.180	0.065	0.247	0.026	0.160	0.052	0.222
Artesanía y similares	0.119	0.324	0.165	0.371	0.174	0.380	0.243	0.429
Operación de planta y maquinaria	0.176	0.382	0.150	0.358	0.274	0.447	0.211	0.409
Sin capacitación	0.125	0.331	0.194	0.396	0.000	0.000	0.000	0.000
<i>Industrias</i>								
Minería	0.052	0.222	0.007	0.085	0.074	0.262	0.007	0.082
Manufactura	0.140	0.347	0.153	0.360	0.205	0.405	0.211	0.409
Utilidades	0.018	0.134	0.010	0.098	0.026	0.160	0.013	0.115
Construcción	0.043	0.202	0.119	0.324	0.032	0.175	0.112	0.316
Comercio mayorista, minorista y hotelero	0.112	0.316	0.188	0.391	0.168	0.375	0.265	0.442
Transporte, almacenamiento y comunicaciones	0.055	0.228	0.080	0.271	0.053	0.224	0.112	0.316
Servicios Financieros	0.067	0.250	0.079	0.270	0.074	0.262	0.076	0.266
Servicios Personales	0.453	0.499	0.219	0.414	0.311	0.464	0.142	0.349
Ratio de Afiliación a Sindicato 2002	16.2				18.0			
Ratio de Afiliación a Sindicato 2004	20.5				23.5			
Ratio de Afiliación a Sindicato 2006	25.6				27.4			
Ratio de Afiliación a Sindicato 2009	28.5				29.9			
N ^a de individuos	1,154				635			

Nota: Estadísticas desde la EPS 2009.

Con el problema descrito, podemos adoptar el enfoque de modelos causales de Neyman-Rubin (Rubin, 2005) [21], y denotar por $Y(0), Y(1)$ los *outcomes* potenciales correspondientes a la intervención $T = 0$ y $T = 1$ respectivamente.

El objetivo de este estudio es caracterizar la heterogeneidad en el efecto del tratamiento $Y(1) - Y(0)$ a través de las variables observables X que recién definimos. Todas estas variables corresponden a posibles fuentes de heterogeneidad en el tratamiento (ver por ejemplo (Landerretche et al., 2013) [11]). Como es bien sabido, este efecto no es identificable sin supuestos adicionales dado que cada trabajador es observado solamente en uno de los grupos de tratamiento. En cambio, estimaremos el *efecto condicional promedio del tratamiento* (*CATE* por sus siglas en inglés) con respecto a las covariables observadas X .

$$\tau(x) = E[Y(1) - Y(0) \mid X = x] \quad (1)$$

La ecuación 1 es identificable a partir de datos observacionales bajo el supuesto estándar de ignorabilidad,

$$Y(1), Y(0) \perp T \mid X,$$

consistencia,

$$Y = TY(1) + (1 - T)Y(0)$$

Y positividad,

$$\forall x : p(T = 0 \mid X = x) > 0 \iff p(T = 1 \mid X = x) > 0$$

El *CATE* condicionado al conjunto completo de covariables X es lo más cerca que estaremos de estimar el efecto del tratamiento para un trabajador individual. Sin embargo, para el diseño y asignación de políticas, es poco frecuente -y necesario- tener este nivel de resolución. Más adelante, siguiendo a Johansson (2018), [6] estimaremos, además, los efectos condicionales con respecto a las variables individuales de X . Con esto, podremos obtener el efecto promedio marginal estratificado por industria, ocupación, o cualquier sub-muestra o función de X .

2.2 Marco Teórico y Metodológico

En esta sección primero expondremos a grandes rasgos el marco unificado de meta-algoritmos (*meta-learners*) propuesto en Künzel et al., 2019 [10] que comprende el estado de arte en cuanto a estimación de heterogeneidad en efectos de un tratamiento, a través de aprendizaje supervisado. Luego, revisaremos el *pipeline* estándar de *machine learning*. Sobre esto, describiremos el proceso de división de la base de datos en la sub-muestra de entrenamiento y validación, computaremos la estimación de los *outcomes* potenciales y caracterizaremos la heterogeneidad en los *CATE*.

Meta-Learners

Existe un creciente interés por estimar y analizar efectos heterogéneos de un tratamiento en estudios tanto experimental y observacionales. Siguiendo a Künzel et al., 2019 [10] describiremos algunos meta-algoritmos que pueden sacarle ventaja a cualquier método de aprendizaje supervisado o de regresión para estimar

los *CATE* en un contexto de asignación binaria del tratamiento. La idea central de los meta-algoritmos es descomponer la estimación de los *CATE* en varios sub-problemas de regresión que pueden ser resueltos con cualquier regresión o método de aprendizaje supervisado.

El meta-algoritmo más común para estimar la heterogeneidad en el efecto del tratamiento comprende dos pasos. Primero, emplea un algoritmo de aprendizaje de base (por ejemplo, regresiones lineales o métodos basados en árboles de decisión) para estimar los *outcomes* condicionales para individuos bajo control y aquellos bajo tratamiento. Segundo, en la muestra de validación, toma la diferencia entre estas estimaciones para computar los *CATE*. En general, nos referimos a este mecanismo de estimar por separado como *T-learner*, "T" siendo una abreviación de *Two-Steps* en inglés. En la figura 1 se describe un diagrama con el proceso de entrenamiento y predicción de este algoritmo.

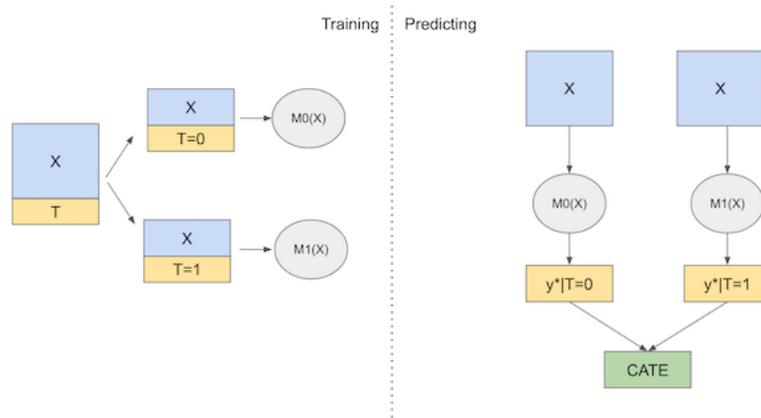


Fig. 1: *T-Learner Diagram*

Estrechamente relacionado con el *T-learner* está la idea de estimar el resultado usando todas las covariables (o *features*) y el indicador de tratamiento T , pero sin darle a éste último un rol fundamental en las predicciones. En este caso, el *CATE* predicho para un trabajador individual es entonces la diferencia entre los valores predichos cuando el indicador de asignación de tratamiento cambia, con todas las demás variables fijas. Nos referimos a este meta-algoritmo como *S-learner*, ya que utiliza un único estimador. En la figura 2 podemos ver su diagrama.

Por último, hay un modelo de dos etapas llamado *X-learner*, que en su primera etapa es idéntico al *T-learner*, luego en la segunda imputamos el efecto del tratamiento para control y tratamiento usando los modelos de la primera. En el último paso calibra los modelos de la segunda etapa a través de un modelo de *propensity score*. En otras palabras, se favorece el modelo que fue entrenado utilizando más datos. En la figura 3 está el diagrama resumen.

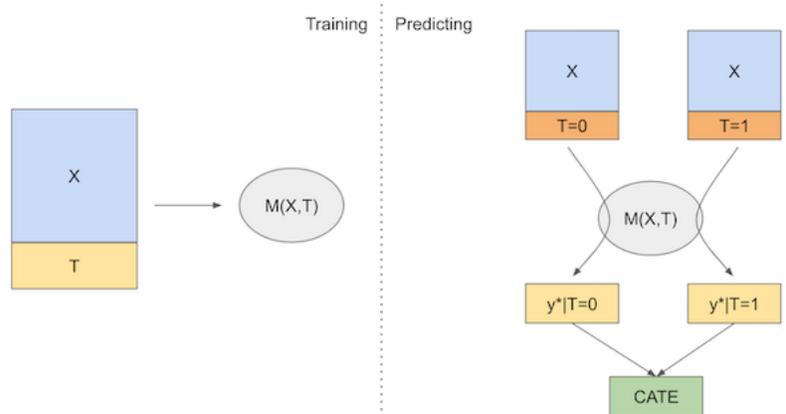


Fig. 2: *S-Learner Diagram*

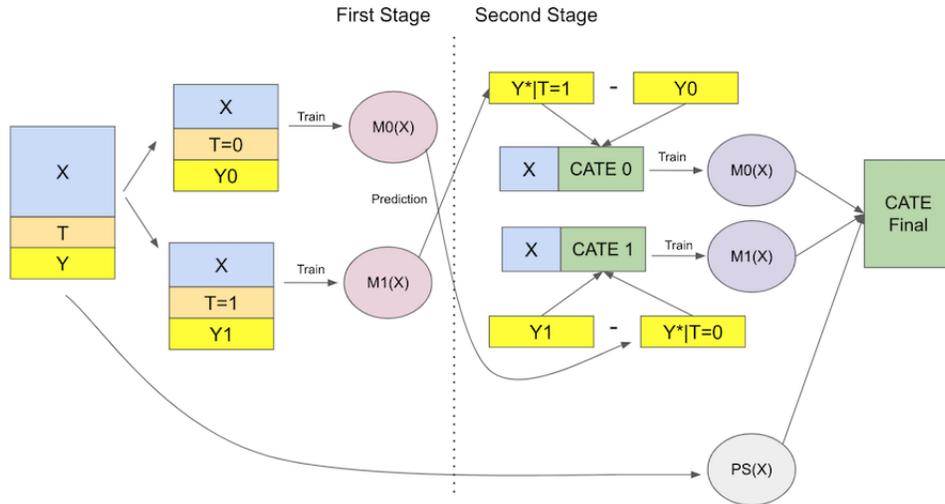


Fig. 3: *X-Learner Diagram*

Cabe destacar por cierto, que no todos los métodos que apuntan a capturar la heterogeneidad de los efectos del tratamiento se encasillan dentro de la clase de meta-algoritmos. Por ejemplo, algunos investigadores analizan la heterogeneidad por estimar los efectos promedio del tratamiento para subgrupos significativos (ver Hansen y Bowers, 2009 [5]). Otro ejemplo es utilizando *causal forests* (ver Wager y Athey, 2017 [26]). A pesar de que teóricamente el desempeño de los *X-learners* debiese ser superior a cualquier metodología supervisada. Puede que dada la naturaleza del problema, la calibración de hiper-parámetros y la calidad de los datos disponibles, se tenga que empíricamente los modelos se comporten mejor o peor en determinadas circunstancias. En la siguiente sección describiremos el proceso de regularización en *Machine Learning*. Luego, empleando *Grid Search* ajustaremos los mejores hiper-parámetros para los meta-algoritmos, de

manera de obtener las mejores predicciones posibles y evitar el problema de *overfitting*.

Revisión de la Metodología

La flexibilidad de los estimadores de *Machine Learning* crea al mismo tiempo oportunidades y desafíos. Por ejemplo, es típico de que el número de parámetros del modelo que mejor se ajuste a un determinado problema de predicción, exceda el número de grados de libertad disponibles en la muestra. De hecho, en la literatura un apropiado método de regularización es un estándar requerido para mitigar el problema de *overfitting*. Más aún, la gran mayoría de modelos alcanzan su mejor ajuste solamente después de una configuración minuciosa de los hiper-parámetros que controlan el *trade-off* entre sesgo y varianza⁴. En particular, para la elección de los parámetros empleamos un algoritmo de fuerza bruta *Grid Search*. Por último, con el fin de validar un modelo correctamente ajustado, es una práctica estándar en *Machine Learning* realizar un proceso de división de la base de datos. A continuación, aplicaremos este *pipeline* para la estimación de los *CATE*, procediendo a través de los siguientes pasos.

1. Dividiremos la base de datos \mathcal{D} en dos particiones, un *set de entrenamiento* \mathcal{D}_t y un *set de validación* \mathcal{D}_v , respectivamente para el ajuste de parámetros y selección del modelo.
2. Ajustaremos los estimadores f_0, f_1 de los *outcomes* potenciales $Y(0)$ e $Y(1)$ en \mathcal{D}_t y a través de un algoritmo de *Grid Search* con optimización bayesiana, escogemos los hiper-parámetros basados en el error obtenido en \mathcal{D}_v .
3. Computaremos el *CATE*, $\hat{\tau}_i := f_1(x_i) - f_0(x_i)$, para cada trabajador en \mathcal{D} y ajustaremos un modelo interpretable a través del enfoque unificado *SHAP* (Lee, et al., 2017) [15] que describiremos más adelante.

Este *pipeline* nos permite encontrar los estimadores (tipo *caja negra*) con el mejor ajuste posible en el paso 2 sin tener en cuenta la interpretabilidad de sus predicciones. Al ajustar un modelo más simple e interpretable a los efectos imputados en el paso 3, podemos explicar las predicciones del modelo más complejo en términos de cantidades conocidas. Este procedimiento es particularmente adecuado cuando el tratamiento es una función más simple que la respuesta y también nos permite controlar lo granular del enfoque con el que

⁴ La regularización de los modelos de *Machine Learning* por lo general es un proceso de prueba y error empírica, en donde se suele automatizar el proceso de optimización de parámetros a través de algoritmos. Recientes trabajos han sometido a simulaciones exhaustivas distintos tipos de algoritmos de búsqueda tales como *Grid Search*, *Random Search* y *Genetic Algorithm* (Liashchynskiy, 2020) [16] donde en la mayoría de aplicaciones suele destacar *Grid Search* para efectos predictivos, a pesar de que en términos computacionales puede ser menos eficiente en comparación a lo demás algoritmos mencionados.

estudiaremos la heterogeneidad.

Los datos de la EPS tienen una naturaleza de varios niveles; los trabajadores se agrupan en empresas e industrias y cada nivel está asociado con su propio conjunto de covariables. En este sentido, la literatura es rica con estudios de efectos causales en entornos multinivel, ver, por ejemplo, Gelman y Hill (2006) [3]. Sin embargo, este estudio está dirigido principalmente al análisis de los efectos de alto nivel (por ejemplo, nivel-industria), en intervenciones en materia de bajo nivel (por ejemplo, a nivel-trabajador) y en el aumento de la volatilidad que viene con tal análisis. En las siguientes secciones describiremos cada paso de nuestra metodología en detalle.

Paso 1. División de \mathcal{D} : Para permitir la estimación imparcial del error de predicción y seleccionar parámetros de ajuste, dividimos el conjunto de datos \mathcal{D} en dos partes con el 70% de los datos utilizados para el conjunto de entrenamiento \mathcal{D}_t y el 30% para el conjunto de validación \mathcal{D}_v . Para prevenir *overfitting*, siguiendo a Johansson (2018) [6] equilibramos \mathcal{D}_t y \mathcal{D}_v construyendo un gran número de divisiones distribuidas de manera uniforme y seleccionando la que minimiza la distancia euclidiana entre el primer y segundo momento de todas las covariables. Aumentamos la influencia de la variable de tratamiento T por un factor de 10 en esta comparación para asegurar que los grupos de tratamiento se dividen uniformemente entre \mathcal{D}_t y \mathcal{D}_v .

Paso 2. Estimación de *outcomes* potenciales: El efecto de tratamiento promedio condicional es la diferencia entre los *outcomes* potenciales esperados. Aquí denotados μ_0 y μ_1 . Bajo ignorabilidad escrito con respecto a X (ver arriba), tenemos que:

$$\mu_z(x) := E[Y(t) | X = x] = E[Y | X = x, T = t] \text{ para } t \in \{0, 1\}$$

y por lo tanto, $\tau(x) = \mu_1(x) - \mu_0(x)$. Una ruta directa para estimar τ es independientemente ajustar los condicionales $E[Y | X = x, T = t]$ para cada valor de $t \in \{0, 1\}$ y computar su diferencia. Esto ha denominado recientemente el enfoque de meta-algoritmos para distinguirlo de otros paradigmas de aprendizaje (Künzel et al., 2017) [10]. A continuación, cubriremos brevemente la teoría que motiva este método y señalar algunas de sus deficiencias. Para estudiar la heterogeneidad, consideramos varios estimadores "meta" tomando en cuenta distintas bases, tales como algoritmos *Gradient Boosting*, regresiones *Ridge Lasso*, *Random Forest* y *Generalized Forest* ⁵.

⁵ Este tipo de algoritmos difieren de los tradicionales basados en árboles de decisión, básicamente en que emplean un kernel ponderado, más no la versión clásica para así mitigar la *maldición de la dimensionalidad*, ver (Athey et al., 2019). [1]

Aproximamos μ_0, μ_1 usando las hipótesis f_0, f_1 y medimos su calidad por alguna métrica de precisión. Los riesgos condicionales de grupo empíricos y esperados se definen de la siguiente manera:

$$\underbrace{\mathcal{R}_t(f_t) := E[(\mu_t(x) - f_z(x))^2 \mid T = t]}_{\text{Riesgo esperado condicional de grupo}} \quad (2)$$

$$\underbrace{\hat{\mathcal{R}}_t(f_t) := \frac{1}{|G_t|} \sum_{i \in G_t} (f(x_i; \theta) - y_i)^2}_{\text{Riesgo empírico condicional de grupo}} \quad (3)$$

Nunca observamos μ_t directamente, sino que los algoritmos aprenden de observaciones ruidosas de y . La teoría del aprendizaje estadístico ayuda a resolver este problema al acotar el riesgo esperado en términos de su contra parte empírica y una medida de la complejidad de la función (Vapnik, 1999) [19]. Para distintas hipótesis en una clase \mathcal{F} con una medida de complejidad particular $\mathcal{C}_{\mathcal{F}}(\delta, n)$ con dependencia logarítmica en n , se cumple con una probabilidad mayor que $1 - \delta$, lo siguiente

$$\forall f_t \in \mathcal{F} : \mathcal{R}_t(f_t) \leq \hat{\mathcal{R}}_t(f_t) + \frac{\mathcal{C}_{\mathcal{F}}(\delta, n)}{\sqrt{n}} - \sigma_Y^2,$$

donde σ_Y^2 es un límite de la varianza esperada en Y (ver Johansson et al, 2018 [6] para una derivación completa). Esta clase de límites ilustra el equilibrio entre sesgo-varianza que es típico en *Machine Learning* y motiva el uso de regularización para controlar la complejidad del modelo. Como se mencionó, en nuestra investigación, consideramos varios modelos *meta-learners* que estiman cada resultado potencial de forma independiente utilizando la minimización empírica del riesgo regularizada, resolviendo el siguiente problema.

$$f_t = \arg \min_{f(\cdot; \theta) \in \mathcal{F}} \hat{\mathcal{R}}_t(f(x; \theta)) + \lambda r(\theta) \quad (4)$$

Aquí $f(x; \theta)$ es una función parametrizada por θ y $r(\theta)$ es un modelo regularizador de parámetros como lo son los penalizadores l_1 (*Lasso*) y l_2 (*Ridge*).

A estas alturas cabe destacar una limitación que tienen los meta-algoritmos. En particular, se tiene que estos algoritmos no comparten información entre los diferentes *outcomes* potenciales. En problemas donde $Y(0)$ es una función más compleja que el efecto τ en sí mismo, los *meta-learners* implican un desperdicio en términos de poder estadístico (Künzel et al., 2017; Nie y Wager, 2017) [10] [20]. Como alternativa se podría adoptar un enfoque de redes neuronales *TAR-Net* de Shalit et al., 2017 [23] donde la arquitectura planteada permite aprender una representación de ambos grupos de tratamiento de manera conjunta, pero predecir potenciales *outcomes* separadamente. Esto tiene la ventaja de compartir información entre grupos de tratamiento en el aprendizaje del *outcome* promedio, pero permite flexibilidad en la aproximación del efecto.

Ahora considerando que, las métricas de riesgo y el problema de minimización descrito, están definidos con respecto a asignaciones observables del tratamiento, para estimar el *CATE*, queremos que nuestras estimaciones de los *outcomes* potenciales, de igual forma, sean precisas para las asignaciones del contrafactual. En otras palabras, queremos que el riesgo en toda la muestra,

$$\mathcal{R}(f_t) := E[(\mu_t(x) - f_z(x))^2]$$

sea pequeño. Cuando los grupos de tratamiento $p(X | T = 0)$ y $p(X | T = 1)$ están desbalanceados, el riesgo esperado dentro de un grupo de tratamiento puede no ser representativo del riesgo en la población completa. Este es otro inconveniente de los estimadores *meta-learners*, dado que no se ajustan a esta posible discrepancia.

En un trabajo reciente, Shalit et al., 2017 [23] caracterizan la diferencia entre $\mathcal{R}(f_t)$ y $\mathcal{R}_t(f_t)$ y acotaron el error en las estimaciones de *CATE* utilizando una métrica de distancia entre los grupos de tratamiento. En particular consideran la familia de distancias *IPM* (*integral probability metric*), resultando en la siguiente relación entre el riesgo del grupo completo y del grupo de tratamiento.

$$\mathcal{R}(f_t) \leq \mathcal{R}_t(f_t) + IPM_G(p(X | T = 0), p(X | T = 1)) \quad (5)$$

En Shalit et al., 2017 [23] los autores utilizaron la distancia media máxima basada en *kernel* (Ver Greton et al., 2012) [4] para regularizar y equilibrar las representaciones aprendidas por la arquitectura *TARNet*, minimizando el límite superior en el riesgo del *CATE*. En Johansson, 2018 [6] el autor utilizó esta misma arquitectura comparándola con meta-algoritmos, donde en general el desempeño fue bastante similar entre los modelos debido a un buen balance de las covariables. En particular, queda como una arista de investigación a futuro aplicar estas metodologías en nuestro problema, sin embargo, como veremos en secciones más adelante, el balance de nuestras covariables es ideal, teniendo que las diferencias entre las distribuciones condicionales del tratamiento, son pequeñas.

Cabe destacar que en este caso, para la ajustar los estimadores, empleamos un *Grid-Search* con optimización bayesiana para la elección de los hiperparámetros. Este tipo de *Grid-Search* utiliza un modelo probabilístico para seleccionar las mejores combinaciones y es más eficiente que los de tipo exhaustivo y aleatorio, ya que utiliza información previa para seleccionar las combinaciones de hiperparámetros más prometedoras.

Paso 3. Caracterización de la heterogeneidad en los *CATE*: Después de ajustar los modelos f_0, f_1 para cada outcome potencial, el *CATE* se computa para cada trabajador como $\hat{\tau}_i = f_1(x_i) - f_0(x_i)$. A diferencia de los regresores lineales, las predicciones de la mayoría de los estimadores de *Machine Learning* son difíciles de interpretar directamente a través de los parámetros del modelo. Por

esta razón, a menudo estos modelos son considerados métodos *black box* (Lipton, 2016) [17]. Sin embargo, en el estudio de la heterogeneidad, es crucial caracterizar para qué trabajadores el efecto de una intervención es bajo y para quienes es alto. Para lograr esto, adoptamos la práctica común de la interpretación *post-hoc*: ajustar un modelo más simple e interpretable $h \in \mathcal{H}$ a los efectos imputados $\{\hat{\tau}_i\}$.

Para esto seguiremos el enfoque unificado *SHAP* (ver Lee, et al., 2017) [15] donde se describe que $h(x_i)$ puede ser una función de un solo atributo, como por ejemplo el tamaño de la empresa donde trabaja el individuo, promediando efectivamente sobre otras covariables. Esta suele ser una buena manera de descubrir tendencias globales en los datos, pero descuida las interacciones significativas entre las variables, así como el modelo lineal. Como alternativa más flexible, realizamos una etapa intermedia de detección de la heterogeneidad ajustando modelos *Causal Forest* para así encontrar las variables que mayor heterogeneidad inducen al modelo, con todo lo demás constante (ver Johansson, 2018 [6]).

3 Resultados

A continuación, presentamos los resultados de nuestro análisis.

3.1 Balance de Covariables

Primero, investigamos si se cumple el supuesto de positividad comparando las estadísticas de las covariables de los grupos de tratamiento y control. En la figura 4, visualizamos las distribuciones marginales de cada covariable. Por otro lado, en la figura 5 observamos una proyección *2D t-SNE* de todo el conjunto de covariables (Maaten y Hinton, 2008) [25]. La diferencia observada entre las distribuciones marginales de los dos grupos de tratamiento es en general pequeña. También la proyección no lineal *t-SNE* revela una muy baja diferencia entre los grupos de tratamiento. Cuanto menor sea el desequilibrio entre grupos de tratamiento, más cerca está nuestro problema de poder resolverse con aprendizaje supervisado estándar. Dicho de otra manera, el ratio de densidad $p(T = 1 | X)/p(T = 0 | X)$ es cercano a 1 y la distancia *IPM* entre distribuciones condicionales es pequeña (ver 5). Por lo tanto, no se espera que una arquitectura como la propuesta por Shalit et al., 2017 [23] tenga un gran efecto en los resultados (Johansson, 2018) [6].

3.2 Estimación de los outcomes potenciales

En nuestro análisis, comparamos los tres meta-algoritmos presentados en el apartado metodológico. En principio, la primera etapa de cada uno de los modelos *off-the-shelf* estuvo basada en algoritmos de aprendizaje de *Gradient Boosting*, Regresiones *Ridge Lasso*, *Random Forest* y *Generalized Forest*. Para cada

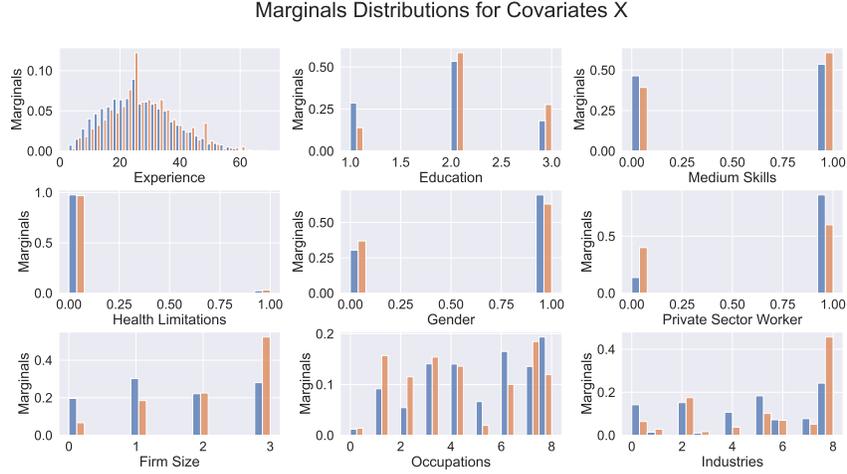


Fig. 4: Marginales de las covariables X . El color azul corresponden a los trabajadores del grupo de control, mientras que el naranja a los de tratamiento (sindicalizados).



Fig. 5: Proyección $2D$ t -SNE de las covariables X mostrando baja diferencia estadística entre los grupos de tratamiento.

familia de estimadores, ajustamos modelos de ambos *outcomes* potenciales en el conjunto de entrenamiento \mathcal{D}_t y la selección de parámetros (y modelos de primera etapa) a través *Grid Search* empleando el R^2 como métrica de validación en el conjunto \mathcal{D}_v . Del proceso de búsqueda, el algoritmo de primera etapa ganador fue un *Random Forest*, modelo del que computamos nuestra estimación final (con sus hiper-parámetros respectivos ya ajustados). Para estimar incertidumbre en las predicciones del modelo, realizamos *Bag of Little Bootstrapping* a nivel trabajador del conjunto de entrenamiento (Kleiner et al., 2012) [8] ajustando los modelos a cada muestra del *bootstrap*. Además, comparamos los meta-algoritmos con modelos de *Propensity Score Matching* y MCO Ajustados, metodologías clásicas de econometría.

Estimator	ATE
Naive	0.24
T-Learner	0.0707 [0.0514 - 0.09]
S-Learner	0.036 [0.0302 - 0.0418]
X-Learner	0.0304 [0.0175 - 0.0432]
PS Match	0.1085 [0.0461 - 0.1708]
OLS	0.0984 [0.0441 - 0.1526]

Table 1: Estimación del ATE para los distintos modelos y sus intervalos de confianza.

En la tabla 1, mostramos la estimación del efecto de tratamiento promedio (ATE) de cada modelo y los intervalos de confianza del 95% basados en el *bootstrap* empírico para los algoritmos *meta-learners*. Además, reportamos la estimación *naive* del ATE , es decir, la diferencia entre los *outcomes* promedios observados en los dos grupos de tratamiento. De los resultados, observamos que los *meta-learners* cuentan con estimaciones similares, mientras que las metodologías clásicas arrojan valores dispares. En particular, el algoritmo *S-Learner* es aquel con mayor poder predictivo y reporta un valor del ATE igual a 3.6%, lo que es un resultado robusto de acuerdo a magnitudes encontradas anteriormente en la literatura. Landerretche, et al., 2013 [11], empleando un modelo dinámico de dos etapas sobre la misma EPS, encontraron un ATE de 4.3% en salarios producto de la sindicalización. Más aún, el *X-Learner* alcanza este valor de ATE dentro de su intervalo de confianza empírico. Finalmente, en la figura 6 vemos las estimaciones de los $CATE$ marginales a cada una de las covariables del problema. En general se aprecia bastante heterogeneidad en las primas salariales, pudiendo tener valores negativos en algunas industrias y positivos en otras. Similar ocurre con la variable categórica de ocupación y tamaño de la empresa.

Con esto, dependiendo del tipo de análisis y la cantidad de datos, podría ser pertinente emplear los meta-algoritmos para corregir los sesgos de selección y tener efectos causales. Esto como alternativa a las metodologías dinámicas o de Heckmann. Sobre este último punto cabe destacar que aún no se prueban las propiedades asintóticas de los meta-algoritmos (a diferencia de muchas de las metodologías adoptadas por la econometría tradicional) lo que podría suponer una limitación al momento de dar validez externa a las estimaciones del posible efecto causal.

3.3 Heterogeneidad en efectos causales

En esta sección examinamos más a fondo el $CATE$ para cada trabajador imputado por el modelo de mejor ajuste, en este caso el *S-Learner* basado en *Random Forest*. En la sección anterior vimos una marcada heterogeneidad en la distintas covariables del problema, teniéndose en promedio un efecto del 3.6% sobre los salarios producto de la sindicalización.

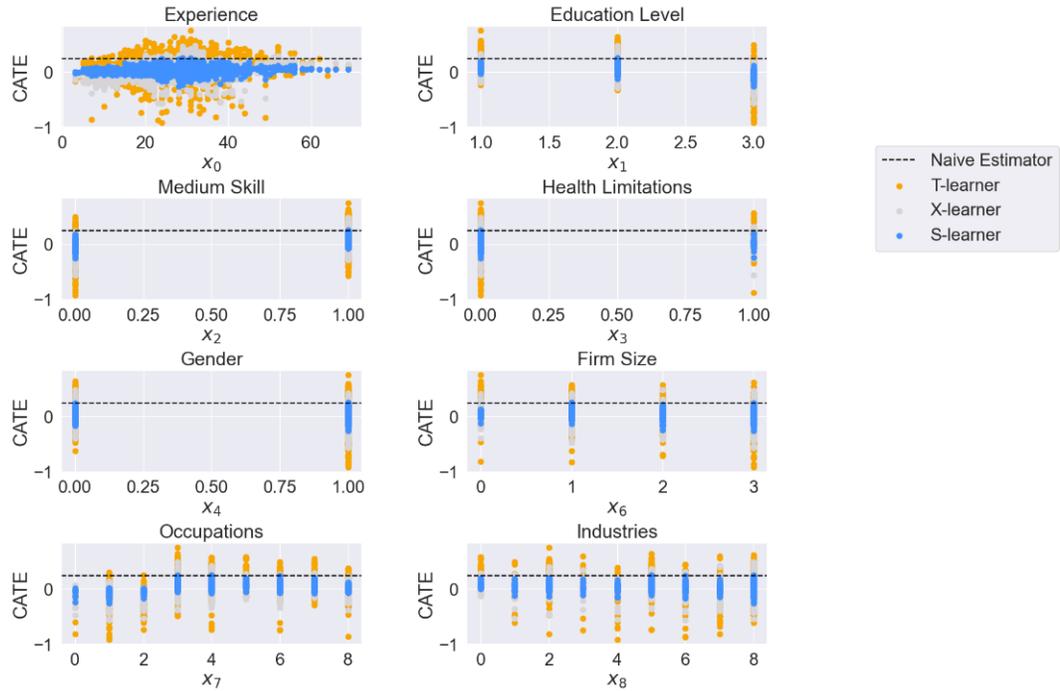


Fig. 6: Heterogeneidad en efectos causales para los distintos *meta-learners*.

Luego, para descubrir los canalizadores de la heterogeneidad, siguiendo a Johanson, 2018 [6] ajustamos un estimador de *Causal Forest* a los efectos imputados para así estimar la importancia de cada covariable, es decir, la frecuencia con la que se utiliza para dividir los nodos de un árbol. De acuerdo a esta metodología las covariables más importantes del *Causal Forest* fueron X_8 (industria), X_0 (experiencia), X_7 (ocupaciones), X_6 (tamaño de la firma) y X_4 (sexo del trabajador). En la figura 7 estratificamos el *CATE* imputado con respecto a estas covariables.

Según lo que se observa, cuando el trabajador tiene relativamente poca experiencia laboral, el premio salarial estaría contenido entre -30% y 3%, lo que sería menor que el *ATE* manteniéndose así hasta cumplidos 15 años. A los 40 años de experiencia, se observa que el premio por riesgo es el máximo en torno a 60%, para luego converger al valor medio. En el caso de profesionales, técnicos y administrativos, el premio sería menor que el *ATE*, teniéndose magnitudes entre -5 y -15% , mientras que para el resto de ocupaciones sería mayor, con niveles en torno al 20%. Por industria, se observa que solamente en manufacturas el efecto es menor al *ATE* con primas salariales de -5%, en minería se tiene el valor máximo entre las industrias con primas salariales cercanas al 40% y por su parte, servicios tiene valores bastante cercanos a la media. A nivel de género, se aprecia una brecha salarial de aproximadamente 8% en el tratamiento. De esta manera confirmamos la hipótesis de este estudio que buscaba identificar

patrones no lineales en la heterogeneidad de los efectos causales en salarios dada la sindicalización de los trabajadores.

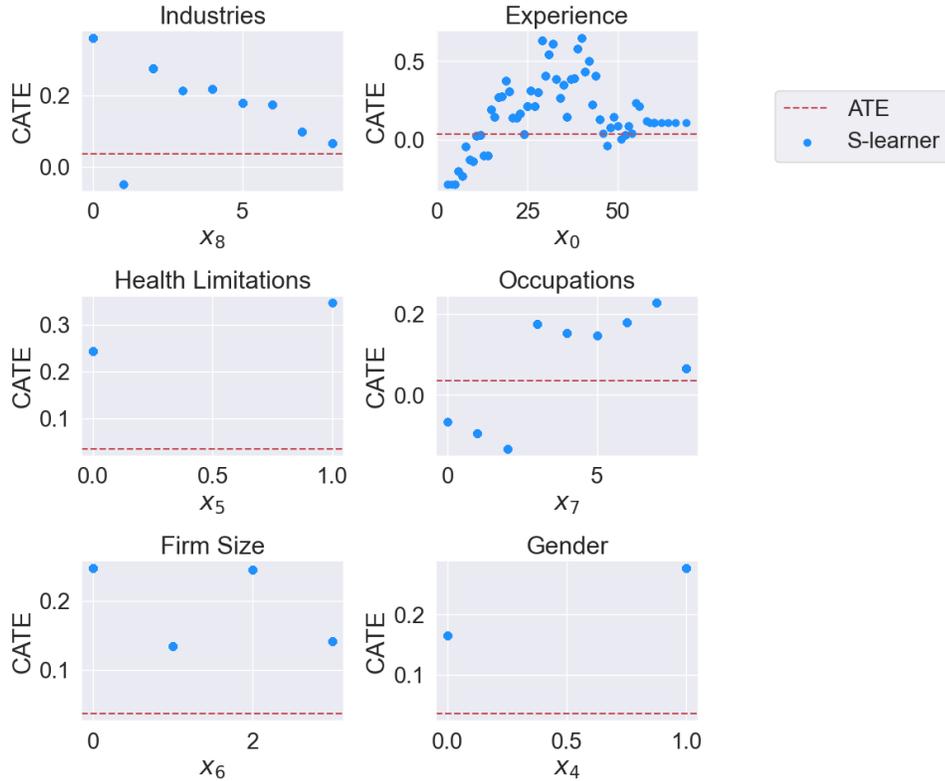


Fig. 7: Heterogeneidad en efectos causales estimada usando *S-Learner*.

3.4 Interpretabilidad del Modelo

En este trabajo sacrificamos la facilidad explicativa de los modelos lineales en virtud de la precisión en nuestras estimaciones y en poder mostrar patrones no lineales en los efectos del tratamiento. Aún así, nos gustaría comprender los determinantes del modelo y poder entender qué lleva al modelo a realizar una determinada predicción, más allá de encontrar la variables que inducen heterogeneidad.

La creciente tensión entre la precisión y la interpretabilidad de las predicciones de los modelos de *Machine Learning* ha motivado el desarrollo de métodos que ayuden a los usuarios a interpretar las predicciones. El marco *SHAP* propuesto por Lee y Lundberg, 2017 [15] identifica la clase de métodos de importancia

aditiva de las covariables y muestra que hay una única solución de teoría de juegos en esta clase que se adhiere a las propiedades deseables de consistencia, ausencia y precisión local. Esta metodología busca computar un modelo aditivo sencillo e interpretable h que hace *mapping* del modelo predictivo real. El hilo de unificación metodológica que teje *SHAP* a través de la literatura es una señal alentadora de que los principios comunes sobre la interpretación de modelos pueden llevar al desarrollo de nuevos métodos futuros.

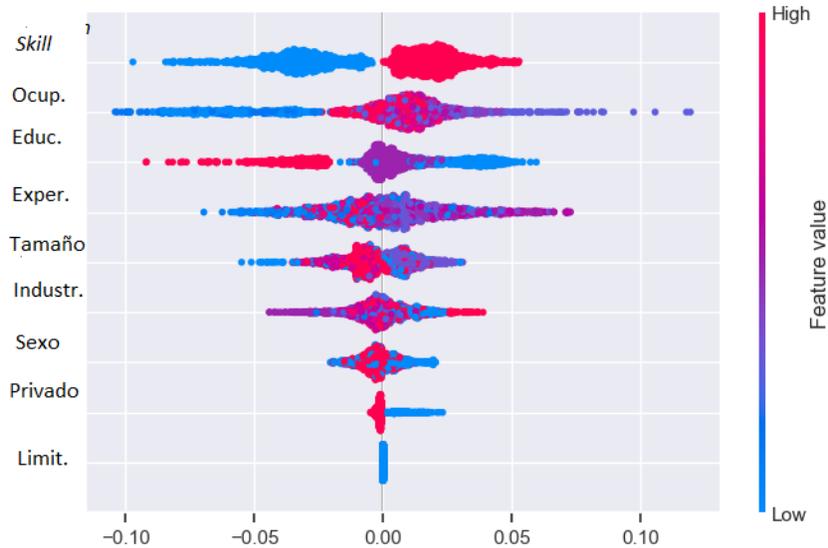


Fig. 8: *Shapley Additive Explanation Values (SHAP)*. Se ilustra el impacto que tiene cada variable en el *outcome*. Para esto se realiza un *mapping* de las covariables de manera de poder explicar el modelo adicionando cada una de éstas. Los valores coloreados hacia *high* representan un alto impacto de dicha variable sobre el resultado del modelo, mientras que aquellos cargados hacia la dirección contraria *low* explicarían en baja proporción las predicciones del modelo.

En la figura 8 encontramos los valores *SHAP* para nuestro caso. De acuerdo a lo observado, podemos destacar que los valores más bajo de la *mapping-covariable* "mediana habilidad" (*skill*) disminuyen la prima salarial predicha dado el tratamiento, mientras que valores altos la aumentan. En el caso de la ocupación solamente tenemos claro que cuando la ocupación del individuo es profesional, técnico o administrativo, esta variable no aumenta el *outcome* dado el tratamiento. Por último, la última relación importante es respecto cuando la escolaridad es baja, en este caso las primas salariales tienden a disminuir.

4 Conclusiones

El *Machine Learning* ofrece una amplia gama de herramientas para la aproximación de funciones y proporciona garantías teóricas para la estimación estadística en caso de, por ejemplo, errores de especificación del modelo. Esto, convierte al aprendizaje supervisado en un marco adecuado para la estimación de los efectos causales no lineales, desequilibrados o con datos de alta dimensionalidad. Sin embargo, la flexibilidad viene con un precio, muchos métodos vienen con parámetros de ajuste que son difíciles de establecer para estimación causal y requieren de mucho ejercicio empírico de prueba y error. La interpretabilidad de los modelos sufre también. Si bien, se han hecho progresos de forma independiente en cada uno de estos problemas, como el *Grid Search* y los valores *SHAP*, aún no ha surgido un conjunto estandarizado de herramientas que sea amplio consenso en la literatura de *Machine Learning*.

Para este trabajo propusimos una metodología para descubrir patrones complejos de heterogeneidad en el efecto de un tratamiento. Además, estos métodos son fácilmente reproducibles con cualquier otro programa. En particular, en el análisis de los datos de la EPS, la metodología que propusimos parece adecuada para estudiar positividad, *outcomes* potenciales y heterogeneidad en los efectos causales. El modelo que mejor ajustaba las predicciones fue el *S-Learner* basado en *Random Forest*, el cual nos dio como resultado un *ATE* igual a 3.6%, mientras que la heterogeneidad del efecto en la prima salarial está bien contenida entre $[-0.2, 0, 6]$. Descubrimos varios patrones de heterogeneidad, destacando los efectos dispares entre industrias, donde las primas salariales por sindicalización serían relativamente más bajas en manufactura y servicios. La experiencia del trabajador también reporta heterogeneidad; se observa que cuando se tiene poca experiencia el premio salarial es menor al *ATE*, esto hasta aproximadamente los 15 años de experiencia. Sin embargo, el análisis también abre algunas preguntas metodológicas. La naturaleza multinivel de las covariable no se tiene en cuenta en los modelos *meta-learners*. Por otro lado, la regularización de modelos aplicados a datos de nivel múltiples ha sido considerablemente menos estudiada que para datos de un solo nivel (Johansson, 2018) [6]. Adicionalmente, como es señalado por varios autores (Künzel et al., 2017; Nie y Wager, 2017) [10] [20], los meta-algoritmos pueden desperdiciar poder estadístico al no comparar los grupos de tratamiento. Esto correspondería a una de las razones por las que enfoques *TARNet* o de redes neuronales podrían tener una pequeña ventaja en términos predictivos y sería un buen ejercicio a futuro replicar este estudio con dichas metodologías. Por último las propiedades asintóticas deseables no se han comprobado para este tipo de metodologías.

References

1. Athey, S., T.J.W.S.: Generalized Random Forest. arXiv, 1864.03423(3). (2019)

2. Contreras D., Hojman D., M.M.R.P.S.N.: The impact of Commuting Time over Educational Achievement, A Machine Learning Approach. Serie de Documentos de Trabajo. Departamento de Economía - Universidad de Chile. (2018)
3. Gelman, A., H.J.: Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press. (2006)
4. Gretton, A., B.K.J.M.S.B.S.A.: A kernel Two-Sample Test. Journal of Machine Learning Research. 12(5) : 103-128. (2012)
5. Hansen, J.: Attributing Effects to a Cluster Randomized get out the Vote Campaign. Journal of the American Statistical Association, 104(487): 873-885. (2009)
6. Johansson, F.: Machine Learning Analysis of Heterogeneity in the Effect of Student Mindset Interventions. Atlantic Causal Inference Conference, arXiv, 1811.0597(1). (2018)
7. Kausel E., Ventura S., V.M.D.D.V.F.: Does Facial Structure Predict Academic Performance?. Personality and Individual Differences, 0191-8869(129): 1-5. (2018)
8. Kleiner, A., T.A.S.P.J.M.: The Big Data Bootstrap. arXiv, 1206.6415(1). (2012)
9. Kononenko, I.: Machine Learning for Medical Diagnosis history, State of the Art and Perspective. Artificial Intelligence in Medicine, 23(1): 89-109. (2001)
10. Künzel, S., J.S.B.P.Y.B.: Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning. arXiv: 1706.0346(6). (2019)
11. Landerretche, O., L.N.P.E.: The Union Effect on Wages in Chile. A Two-Stage Approach Using Panel Data. CEIS, 27(2): 164-191. (2013)
12. Leal F., M.C., E, Z.: Inflation Forecast in Chile with Machine Learning Methods. Banco Central de Chile. Documentos de trabajo n° 860. (2020)
13. LeCun, Y, B.Y.H.G.: Deep Learning. Nature, 521(3): 436-444. (2015)
14. Lee, B., L.J.S.E.: Weight Trimming and Propensity Score Weighting. PLoS ONE 6(3). (2011)
15. Lee, S., L.S.: A unified Approach to Interpreting Model Predictions. arXiv, 1705.07874(2). (2017)
16. Liashchynskyi, P., L.P.: Grid Search, Random Search, Genetic Algorithm. arXiv, 1912.06059(1). (2020)
17. Lipton, Z.: The Mythos of Model Interpretability. arXiv, 1606.03490(3). (2016)
18. Mullainathan, S.: Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 2(31): 87-106. (2017)
19. Naumovich, V.: An overview of Statistical Learning Theory on Neural Networks. IEEE Transactions on Neural Networks(5): 10. (1999)
20. Nie, X., W.S.: Learning Objectives for Treatment Effect Estimation. arXiv, 1403.03452(2). (2017)
21. Rubin, D.: Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. Journal of the American Statistical Association, 100(469): 322-331. (2005)
22. Rubin, D., N.T.: Matching using Estimated Propensity Scores: Relating Theory to Practice. International Biometric Society, 52(1): 249-264. (1996)
23. Shalit, U., J.F.S.D.: Estimating Individual Treatment Effect: Generalization Bounds and Algorithms. arXiv, 1606.03976(5). (2017)
24. Swaminathan, A., J.T.: Counterfactual Risk Minimization: Learning from Logged Bandit Feedback. arXiv, 1502.02362(2). (2015)
25. Van Der Maaten, L., H.G.: Visualizing Data using T-SNE. Journal of Machine Learning. 102(11) : 47-74. (2008)
26. Wager, S., A.S.: Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. arXiv, 1510.04342(4). (2017)