



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

ESTIMACIONES Y DESCOMPOSICIONES DE LOS RETORNOS A LA EDUCACIÓN:
EVIDENCIA DEL MERCADO LABORAL CHILENO EN LÍNEA

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN ECONOMÍA APLICADA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

RONALD LEBLEBICI GARO

PROFESOR GUÍA:
BENJAMÍN VILLENA ROLDÁN

MIEMBROS DE LA COMISIÓN:
ALEJANDRA MIZALA SALCES
MAURICIO TEJADA GUZMÁN

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN ECONOMÍA APLICADA; INGENIERO CIVIL INDUSTRIAL
POR: RONALD LEBLEBICI GARO
FECHA: 2019
PROF. GUÍA: BENJAMÍN VILLENA ROLDÁN

ESTIMACIONES Y DESCOMPOSICIONES DE LOS RETORNOS A LA EDUCACIÓN: EVIDENCIA DEL MERCADO LABORAL CHILENO EN LÍNEA

Se usa información de una bolsa de trabajo en línea (www.trabajando.com) para estimar y descomponer los retornos salariales de los niveles educacionales en Chile. La disponibilidad de información que se tiene parece ser nacional e internacionalmente inusual, pues se tiene la distribución completa de salarios ofrecidos en avisos. El enfoque está en la demanda de trabajo, por lo que los retornos estimados representan las valoraciones de los empleadores. De este modo, se pueden evadir problemas de endogeneidad frecuentes al observar datos de encuestas sin información sobre habilidades de los trabajadores. Adicionalmente, se comparan datos del sitio web con información de la Encuesta Suplementaria de Ingresos, que realiza anualmente el Instituto Nacional de Estadísticas de Chile. Se concluye que la bolsa de empleo concentra sus vacantes en la Región Metropolitana y algunas industrias. Además hay diferencias importantes entre los retornos estimados con ambas bases de datos. De todos modos, se puede usar estos datos para representar parte importante del mercado laboral chileno.

Las estimaciones realizadas contemplan efectos fijos de empleadores y cargos de trabajo usados para analizar diferencias de políticas salariales de empresas y rentabilidades de cargos. También permiten descomponer el retorno “bruto” de la educación en tres componentes según su presunto origen: (1) el acceso a empresas con políticas salariales más o menos generosas, (2) el acceso a cargos con remuneraciones relativamente altas o bajas y (3) las valoraciones que tienen los empleadores por la educación de los postulantes (retorno “neto” o “intrínseco”). Los resultados arrojados por estimaciones kernel no-paramétricas, estimaciones de Mincer y descomposiciones de Gelbach permiten concluir que:

1. El mercado laboral en línea valora considerablemente la educación de los postulantes. Niveles educativos superiores están asociados a salarios ampliamente mayores.
2. Parte importante de los retornos a la educación se explican por accesos a cargos y empresas. El retorno bruto de la educación se compone principalmente por el denominado retorno intrínseco y, en segundo lugar, por el acceso a cargos mejor remunerados. El acceso a empresas, a pesar de quedar en tercer lugar, sigue teniendo un rol importante.
3. En general, un mayor nivel educacional es requerido en vacantes publicadas por empresas que poseen políticas salariales más generosas y cuyos cargos ofrecidos tienden a tener remuneraciones relativamente mayores. Esto no implica que los cargos y empresas tengan siempre efectos positivos/negativos para los retornos asociados a niveles de educación altos/bajos.
4. Las políticas salariales de las firmas son más homogéneas al observar cargos con bajos requerimientos educacionales. Análogamente, las rentabilidades salariales de los cargos son más homogéneas para cargos con mayores requerimientos educacionales.

A Rosete y Umit, por enseñarme a darlo todo por lo que uno ama y sueña.

Agradecimientos

Agradezco profundamente a mis padres Rosete y Umit, quienes han luchado incansablemente por mis hermanas y por mí. Se han entregado por completo. Nos han destinado tiempo y energías que muchas veces no tenían para darnos lo que han creído mejor. De ustedes aprendí a amar, respetar, soñar, perseverar, ser responsable y tener coraje para afrontar los desafíos que nos pone la vida. Sin ustedes no sería quien soy ni estaría donde estoy. Gracias a Catalina y Carolina por aguantarme, apoyarme y quererme a pesar de todos mis defectos. Su compañía ha sido emocional y prácticamente muy importante para mí en todos estos años. Sepan que me enorgullese tener hermanas a quienes admirar, de quienes aprender y a quienes querer.

Agradezco a mis abuelos Ángela, Elías, Jacqueline y John; quienes llegaron desde el otro lado del planeta, superaron una serie de dificultades y peligros, trabajaron arduamente y –por sobretodo– educaron. Sin el sacrificio que ustedes hicieron sería poco probable ser un profesional con valores el día de hoy. Las imágenes y palabras de quienes no están me han acompañado e inspirado hasta el día de hoy. Agradezco particularmente y felicito a mi abuela Jacqueline, quien está con nosotros y de quien soy nieto mayor y primer egresado de la universidad. Agradezco enormemente al resto de mi familia. Tengo la suerte de tener a mi alrededor gente maravillosa con la que he compartido hermosas experiencias de vida. A pesar de que no puedo nombrar a todas y todos, quiero que sepan que no me he olvidado de ustedes y, más aún, han estado presentes en mí siempre.

No imagino cómo habrían sido estos últimos 7 años sin la compañía de Bárbara. Sin duda eres una de las personas más importantes en mi vida y tenerte a mi lado en los buenos y malos momentos ha sido una fortuna. Gracias por tu confianza plena, por tu apoyo incondicional y por lo mucho que has alegrado mis días desde que nos conocimos.

Carlos, Cristian y Felipe son los mejores amigos que pude haber tenido en la universidad. Agradezco la amistad y colaboración que hemos tenido estos años, y espero de corazón que siempre las sigamos teniendo. Del mismo modo agradezco a todas las buenas amistades que he desarrollado en Osorno y Santiago. Su compañía ha sido fundamental para sentir que tengo una gran familia en cada ciudad.

Agradezco a los profesores Benjamín y Elton. Ha sido un honor aprender de ustedes, trabajar juntos y –en definitiva– tener como mentores a dos excelentes personas y profesionales.

Finalmente, agradezco a mis camaradas activistas por su compañerismo, su trabajo y su auténtico deseo de justicia y respeto. Las esperanzas de quienes nacen privados de sus libertades y sus vidas son ustedes. Los animales nos necesitan. ¡Nunca bajemos los brazos!

Tabla de Contenido

Introducción	1
1. Datos	8
1.1. Trabajando.com (TBN)	8
1.1.1. Presentación	8
1.1.2. Muestra del estudio	9
1.2. Encuesta Suplementaria de Ingresos (ESI)	10
1.2.1. Presentación de la base de datos	10
1.2.2. Muestra del estudio	11
1.3. Comparación TBN-ESI y variables de interés	12
1.3.1. Resumen	12
1.3.2. Niveles educacionales	13
1.3.3. Salarios	15
1.3.4. Regiones	17
1.3.5. Experiencia	19
1.3.6. Industrias	19
1.3.7. Tamaño de las empresas	20
1.3.8. Retornos de la educación sobre los salarios	23
1.3.9. Empresas (TBN)	27
1.3.10. Palabras presentes en las descripciones de los cargos (TBN)	28
2. Distribuciones de efectos fijos: análisis exploratorio	30
2.1. Políticas salariales de las empresas	30
2.2. Rentabilidades salariales de los cargos	36
3. Metodología	41
3.1. Notación	41
3.1.1. Niveles educacionales	42
3.1.2. Experiencia	42
3.1.3. Años	42
3.1.4. Empresas	42
3.1.5. Cargos	43
3.2. Estimaciones de Mincer	43
3.3. Descomposición de Gelbach	45
3.3.1. Contexto de origen	45
3.3.2. Descripción teórica	45

3.3.3. Descripción práctica	47
4. Resultados	49
4.1. Estimaciones de Mincer	49
4.2. Descomposición de Gelbach	52
4.2.1. Educación media científico humanista	53
4.2.2. Educación media técnico profesional	55
4.2.3. Educación superior técnico profesional	56
4.2.4. Educación universitaria	56
5. Discusión	57
Conclusión	61
Anexos	62
Bibliografía	65

Índice de Tablas

1.1.	Distribución de jornadas requeridas por vacantes del sitio web.	10
1.2.	Salario mínimo bruto legal en Chile a partir del 1 de julio de 2009.	11
1.3.	Distribución educacional de avisos y vacantes TBN y flujo (ponderado) ESI.	14
1.4.	Estadísticos de distribuciones empíricas de salarios reales por hora, por nivel educacional, de avisos y vacantes TBN y flujo (ponderado) ESI.	18
1.5.	Distribución regional de avisos y vacantes TBN y flujo (ponderado) ESI.	19
1.6.	Distribución por sectores industriales de avisos y vacantes TBN y flujo (ponderado) ESI.	20
1.7.	Distribución por tamaños de empresas de avisos y vacantes TBN y flujo (ponderado) ESI.	22
1.8.	Retornos estimados de los distintos niveles educacionales, obtenidos mediante una regresión que combina datos de ESI y TBN.	24
1.9.	Retornos estimados de los distintos niveles educacionales, obtenidos mediante una regresiones separadas para TBN y ESI.	26
1.10.	Ránking de las 100 palabras más frecuentes en títulos de trabajo del sitio web.	29
2.1.	Estimaciones de los 4 primeros momentos de distribuciones empíricas de efectos fijos de empresas.	35
2.2.	Estimaciones de los 4 primeros momentos de distribuciones empíricas de efectos fijos de cargos.	40
4.1.	Retornos estimados de los distintos niveles educacionales y otras estimaciones obtenidas con los datos de vacantes TBN.	50
4.2.	Descomposición de Gelbach para los retornos de los distintos niveles educacionales.	54
5.1.	Ofertas del sitio web para empleadores que quieran publicar avisos.	65
5.2.	Resultados de tests de igualdad entre los retornos a la educación estimados con datos de vacantes TBN y flujo ponderado ESI.	65
5.3.	Distribución por sectores industriales de vacantes TBN para empresas con al menos uno y dos avisos.	66
5.4.	Distribución regional de vacantes TBN para empresas con al menos uno y dos avisos.	66
5.5.	Distribución por tamaños de empresas (medidos en números de trabajadores) de vacantes TBN para empresas con al menos uno y dos avisos.	67
5.6.	Estimaciones obtenidas al regresionar los efectos fijos de las firmas en función de sus respectivas industrias y tamaños (en número de trabajadores).	68

Índice de Ilustraciones

1.1.	Formulario estándar que rellenan los empleadores en el sitio web.	9
1.2.	Distribución por nivel educación de avisos y vacantes TBN y flujo (ponderado) ESI.	14
1.3.	Distribución de (log) salarios reales por hora para datos de TBN y ESI.	16
2.1.	Distribución de efectos fijos de empresas para cada nivel educacional.	33
2.2.	Distribución de efectos fijos de empresas para cada nivel educacional (sin básica).	34
2.3.	Distribución de efectos fijos de cargos para cada nivel educacional.	38
2.4.	Distribución de efectos fijos de cargos para cada nivel educacional (sin básica).	39
4.1.	Descomposición los retornos estimados de cada nivel educacional.	55
5.1.	Retornos estimados para cada industria sobre los efectos fijos de las empresas.	63
5.2.	Retornos estimados para cada categoría de tamaños de firmas (medidos en números de trabajadores) sobre los efectos fijos de las empresas.	64

Introducción

En las últimas décadas se han desarrollado numerosos estudios teóricos y empíricos cuyo propósito es comprender el funcionamiento del mercado laboral. Muchos de estos estudios se enfocan en entender cómo se concreta el equilibrio de oferta y demanda de mano de obra en cada sub-mercado. Una de las típicas variables de interés para los investigadores son los salarios que se les pagan a quienes obtienen un determinado empleo. El modelo de Mincer (1974) [1] permite estimar tanto los retornos de la educación como de la experiencia potencial sobre dichos salarios y es considerado una “piedra angular” para la literatura que abarca estos temas desde un enfoque empírico.

La importancia del estudio de Mincer (1974) [1] se explica en gran medida por ser un punto de partida para modelos más sofisticados que se han estimado en distintos contextos temporales y geográficos, permitiendo: (1) estimar los rendimientos y la calidad de la escolaridad, (2) medir el impacto de la experiencia laboral, (3) medir y entender la dinámica de las brechas salariales entre hombres y mujeres, (4) desarrollar estudios económicos de la educación en países en desarrollo y (5) comprender cómo el mercado laboral premia atributos como la escolaridad y la experiencia, además de las inversiones en capital humano por parte de trabajadores (Heckman, Lochner & Todd, 2003) [2].

El presente estudio se concentra en los retornos de la educación, siendo la experiencia una variable de control. Otra diferencia que tiene este estudio en relación al de Mincer y la mayoría de los que se enuncian a continuación es que la base de datos utilizada describe la demanda de trabajo, representada por vacantes publicadas por empresas en el portal web de búsqueda de empleo www.trabajando.com, y no a datos de personas ya contratadas, recopilados mediante encuestas. Esto implica que las estimaciones que aquí se presentan se realizaron en base a ofertas y requerimientos ingresados por los empleadores. Entre otras variables, las empresas señalan a través de sus avisos qué niveles de educación y experiencia se requiere para un determinado cupo de trabajo, redactan una descripción del cargo e ingresan un salario estimado que estarían dispuestas a pagar. Cabe destacar que este estudio tiene una limitación dada por la misma naturaleza de los datos: los empleadores no exigen años de educación sino un nivel educacional. Los niveles educacionales son: básico, medio científico-humanista, medio técnico-profesional, superior técnico-profesional, universitari y postgrado.

Griliches (1977) [3] analiza algunos problemas estadísticos propios de estimar una relación entre el logaritmo de los ingresos, la escolaridad y otras variables. Este estudio se basa en datos de *National Longitudinal Surveys* (Estados Unidos) y su enfoque está en analizar el rol y la relevancia de las habilidades de los trabajadores, que suelen ser omitidas, en este tipo

de estudios econométricos. En este documento se dan a conocer posibles soluciones ante los sesgos de este factor no observable sobre los estimadores de los retornos a la educación y se señala que no son siempre necesariamente positivos.

En línea con el estudio de Griliches (1977), Angrist & Krueger (1991) [4] reconocen la dificultad de estimar los retornos a la educación sobre los salarios por presuntos problemas de endogeneidad causados por omisión de variables relevantes –asociadas a cualidades o habilidades propias de cada individuo– no observables en los datos del Censo de EEUU ni en los de encuestas de hogares (en general). Para hacerse cargo de esta problemática proponen un diseño cuasi experimental, aprovechando la particular legislación de Estados Unidos. En este país hay asistencia escolar mínima variable, que puede ser hasta los 15, 16 o 17 años, según la ley de cada estado. Como el año escolar se desarrolla entre septiembre y mayo, los estudiantes nacidos poco antes de septiembre no están obligados a completar un curso si cumplen la edad mínima de egreso de su estado. Por otro lado, los nacidos a principios de año están obligados a estudiar por más tiempo. En consecuencia, los trimestres de nacimiento inciden en los años de educación adquiridos, pero no afectarían directamente a los salarios. Los autores justifican que el trimestre de nacimiento es una fuente exógena de variación de los años de educación, que –al ser usado como variable instrumental– permitiría subsanar el problema de endogeneidad antes descrito. Siempre que se cumpla el supuesto de validez del instrumento que sería el trimestre de nacimiento sería posible construir un estimador sin sesgo del retorno de la escolaridad.

Bound, Jaeger & Baker (1995) [5] hacen un llamado de atención respecto a dos problemas asociados con el uso de variables instrumentales. En primer lugar, el uso de instrumentos débiles puede generar inconsistencias en las estimaciones. En segundo lugar, señalan que las estimaciones obtenidas con muestras limitadas mediante variables instrumentales se ven sesgadas en la misma dirección que las obtenidas mediante mínimos cuadrados ordinarios. A modo ilustrativo analiza el estudio de Angrist & Krueger (1991) del cual se indica que la correlación entre la educación alcanzada y el trimestre de nacimiento es baja (este último sería un instrumento débil). Además se señala que hay evidencia médica de que el trimestre de nacimiento sí está correlacionado con factores que a la larga inciden en los salarios, como enfermedades mentales o infecciones. Aunque esta última correlación sea pequeña, al tratarse de un instrumento débil, el sesgo del estimador de variables instrumentales podría llegar a ser alto.

Sin perder de vista las preocupaciones planteadas por los autores mencionados anteriormente y ante la dificultad que supone encontrar variables instrumentales adecuadas, se propone evitar estos problemas de endogeneidad en estimaciones mediante una estrategia diferente. Al estimar los retornos salariales de alcanzar un determinado nivel educacional, una ventaja de usar datos de avisos publicados por empleadores en vez de encuestas es que las características no observables (habilidades o cualidades) de los trabajadores considerados en la muestra pueden influir en sus salarios realizados. Los datos de avisos de empleo publicados por empresas en www.trabajando.com, por su naturaleza, permiten evadir este problema. La razón de ello es que las habilidades no observables del trabajador a contratar no son conocidas a priori por el empleador. Por lo tanto, los datos asociados a demanda de trabajo en el sitio web permiten realizar estimaciones no sesgadas por características individuales de trabajadores a contratar.

Abowd, Kramarz & Margolis (1999) [6] realizaron estimaciones mincerianas incorporando numerosos efectos fijos de trabajadores y firmas usando registros administrativos con más de un millón de trabajadores franceses y más de 500 mil firmas. Algunas de las conclusiones de aquel estudio son: (1) que los efectos fijos (principalmente los no relacionados a variables observables como la educación) de trabajadores son la fuente más importante de variación salarial, (2) que los efectos fijos de las firmas son importantes aunque menos que los de los trabajadores y (3) que los efectos personales explican cerca del 92% de las diferencias salariales interindustriales. En el presente estudio también se incluyen efectos fijos, con el fin de incorporar las variaciones sistemáticas de salarios ofrecidos, existentes entre firmas y cargos. Dado que se estudia la demanda de trabajo, no es posible ni viene al caso incorporar efectos fijos de trabajadores. Sin embargo, sí se controla por características de los empleadores y requerimientos de educación, experiencia, sectores industriales y cargos, que definen el perfil del trabajador esperado. El contar con información de demanda previo a la formación de empleos constituye una cierta ventaja puesto que los sesgos de estimación reportados en literatura subsecuente (por ejemplo, Andrews et al, 2007)

Card (2001) [7] hace una revisión de una serie de investigaciones recientes para la época que intentaron medir el efecto causal de la educación en los ingresos del mercado laboral utilizando características de los establecimientos educacionales como determinantes exógenos de la escolaridad. En este estudio se muestra un modelo teórico simple que resalta la ventaja de decidir un nivel de educación óptimo.

Heckman, Lochner & Todd (2003) [2] reconocen a la función de Mincer como una “piedra angular” de la literatura en la economía empírica y hacen una revisión sobre ella, aplicada a datos de los últimos 50 años hasta entonces. Analizan los fundamentos teóricos del modelo Mincer y examinan su respaldo empírico mediante datos de los censos y las encuestas vigentes. Según los autores, mientras los datos de los años '40 y '50 respaldan al modelo de Mincer, los de las décadas más recientes eran inconsistentes con éste. Estos autores, además de realizar una crítica, examinan los supuestos bajo los cuales los resultados obtenidos pueden interpretarse efectivamente como retornos a la educación. Dado que la base de datos usada en este estudio describe a la demanda de trabajo y que el marco teórico usado por Heckman, Lochner & Todd (2003) [2] se basa en decisiones que toman los individuos que componen la oferta, es difícil usar o adaptar el mismo marco teórico a la realidad que se pretende describir en este estudio.

Considerando que cada postulante elige cuánto educarse en función de los costos y ganancias que ofrece el mercado y que la acumulación de capital humano se traduce en salarios (que presumiblemente estarían dados por precios de mercado, lo cual puede ser comparable con el modelo de Mincer, en el que los precios son exógenos) una manera de explicar las ecuaciones que se estiman en el presente estudio es basándose en modelos de precios hedónicos. Estos precios hedónicos representarían precisamente a los salarios que ofrecen los empleadores en www.trabajando.com. Es posible pensar que los salarios ofrecidos a un postulante que cumple ciertas características es la suma de las valoraciones que se tienen por dichos atributos. Los modelos estimados además incluirían variaciones dadas por características relativas a los puestos de trabajo asociados a las vacantes y a las empresas que publican dichas vacantes. A pesar de que en el presente estudio muchas veces se habla de regresiones “mincerianas” –dada la similitud funcional de las ecuaciones estimadas– se aclara que estas estimaciones

no se basan teóricamente en el modelo de Mincer o alguna de sus variantes por los motivos antes expuestos.

De manera similar a Heckman, Lochner & Todd (2003) [2], Lemieux (2006) [8] intenta responder una serie de preguntas relacionadas a la literatura empírica que surge a partir de la función de Mincer. Menciona, en base a la literatura existente y de nuevas estimaciones, que la ecuación de Mincer sigue siendo un punto de referencia preciso para modelar el comportamiento de los salarios siempre que se incluyan: (1) potencias de hasta orden 4 para la experiencia potencial en lugar de hasta orden 2, (2) un término cuadrático en años de escolaridad para captar la creciente convexidad en la relación entre escolaridad y salarios y (3) efectos de cohorte para capturar el crecimiento del retorno de la educación después del año '50. En el presente estudio se toman en consideración estos tres puntos: se incorporan potencias de hasta orden 4 para la experiencia y se añaden efectos fijos por año, mientras que no viene al caso agregar términos cuadráticos para la educación dado que en este caso se incorpora como una variable categórica y no continua.

Wirz (2008) [9] evalúa si es que los efectos de “derrame” (“*spillovers*”) de educación –es decir, influencia dada por diferencias educacionales entre trabajadores de una firma– sesgan los retornos estimados para la educación. Este fenómeno es estudiado considerando efectos de “derrame” dentro grupos ocupacionales de las empresas, lo cual otorga mayor precisión que si se analizaran, por ejemplo, a nivel de firma o de región (como es el caso de otros estudios). Usando datos de panel de una encuesta suiza de fuerza laboral realizada a empleadores, se encuentra evidencia de que los efectos de derrame afectan significativamente a los salarios individuales. Por lo tanto, los retornos de la educación estarían sesgados hacia arriba por dichos efectos. Martins & Jin (2010) [10] concluyen que los trabajadores menos educados se benefician de la educación de sus colegas. Los autores introducen un modelo de aprendizaje –en el que los trabajadores más educados pueden transferir habilidades generales a los menos educados– y usan datos de panel de empleadores y empleados de Portugal. Lamentablemente, ninguna de las bases de datos en las que se basa este estudio permite incorporar estos efectos de “derrame” en las estimaciones. Las consecuencias de no incluir de ello se discuten brevemente en la sección 6.

Gelbach (2016) [11] desarrolló una variación del modelo de Mincer que permitía estimar la brecha racial de los salarios, ocupando los datos de *National Longitudinal Survey of Youth*. Señala que si bien agregar covariables a un modelo de regresión puede brindar mayor robustez a las estimaciones obtenidas, esto puede ser problemático dada la correlación entre los regresores. Gelbach nota que en estudios previos se ha utilizado la inclusión progresiva de covariables para analizar cambios en estimadores de interés. El problema que identifica es que el orden en el que se agregan las covariables puede incidir en los cambios registrados sobre las estimaciones de los parámetros de interés que se atribuyen a cada una de ellas. Esto lo motivó a utilizar una descomposición, creada por el mismo autor (Gelbach, 2014 [12]), que permite desagregar el sesgo por omisión de variables relevantes y atribuir partes de dicho sesgo a variables intencionalmente omitidas. En el estudio que aquí se presenta, la descomposición de Gelbach toma un rol protagónico por permitir develar los mecanismos a través de los que se manifiesta el retorno a la educación. Es por ello que se incorporaron en este informe una subsección que detalla el contexto que motivó dicha descomposición, así como aspectos teóricos y prácticos de ella.

Marinescu & Wolthoff (2016) [13] usan los títulos de trabajo de www.careerbuilder.com para mejorar el poder predictivo de sus estimaciones. Muestran que los títulos de trabajo explican más del 90% de la varianza de los salarios (de vacantes de que presentan esta información). La base de datos que utilizan cuenta con más de 60.000 ofertas de empleo. En este estudio se truncan los títulos de trabajo a las cuatro primeras palabras. Luego se incorporan en las estimaciones variables binarias que representan a los títulos de trabajo truncados o, en su defecto, las palabras presentes en ellos. De manera similar, este trabajo incorpora variables binarias que indican si una determinada palabra se encuentra entre las cuatro más significativas (sustantivos y adjetivos) de la descripción de un cargo. Estos regresores permiten controlar los eventuales sesgos, que podrían generarse en las estimaciones en caso de ser omitidos, por existir correlaciones entre dichas variables y el hecho de tener un cierto nivel educacional.

Cardoso, Guimarães, Portugal & Reis (2018) [14] analizan los retornos de la educación en el mercado laboral teniendo en cuenta variables como las capacidades de los trabajadores, su título o certificación profesional y características propias de sus compañeros de trabajo y del empleador. Uno de los hallazgos de este estudio es que la educación otorga acceso a firmas y títulos laborales mejor pagados. Este resultado y la metodología expuesta en este documento motivó en gran medida el desarrollo del presente estudio. En ambos casos se utiliza la descomposición de Gelbach para desglosar el retorno de la educación, entendiendo que este se puede manifestar sobre los salarios de manera directa (mediante el beneficio intrínseco esperado por los empleadores que genera la educación sobre la productividad de un trabajador) e indirecta (mediante el acceso a determinadas industrias, firmas, tipos de trabajo, o cargos).

Banfi & Villena-Roldán (2018) [15] utilizan datos de www.trabajando.com y concluyen que efectivamente los anuncios de trabajo cuyos salarios ofrecidos son altos atraen a más postulantes, como es el caso de los modelos de búsqueda dirigida. Banfi, Choi & Villena-Roldán (2019) [16] usan los mismos datos para exponer datos relevantes sobre la búsqueda de trabajo por parte de las personas. Se muestra cómo la postulación a trabajos está influenciada por datos demográficos (como sexo, edad y estado civil), la coherencia entre expectativas salariales de los postulantes y ofertas salariales de los empleadores, la concordancia entre características del trabajador y requerimientos del empleador (como educación, experiencia, ubicación y ocupación), la influencia de variables temporales (como el tiempo que llevan desempleados quienes lo están o la permanencia en su trabajo actual de quienes no lo están) y las condiciones del ciclo económico. El trabajo de Banfi & Villena-Roldán (2018) [15] contempla un procedimiento parecido al de Marinescu & Wolthoff (2016) [13]: se extrajeron y se normalizaron las cuatro palabras más significativas de los títulos de trabajo, para luego incorporar esta información en las estimaciones. En el presente trabajo se toma esta información ya procesada, para usarla en estimaciones y análisis.

Metodológicamente, estos últimos estudios motivan al presente debido a la base de datos utilizada, los criterios para seleccionar la muestra de estudio y la creación de variables que almacenan información sobre los títulos de trabajo. Estas variables han sido de gran ayuda para identificar aquel componente de los retornos a la educación que se debe a la posibilidad de desempeñar determinados cargos dentro de una misma empresa. Aparentemente, tanto a nivel nacional como internacional, es inusual la disponibilidad de información completa de

salarios ofrecidos en este tipo de plataformas de búsqueda de empleo. Con mucha frecuencia no se visualiza el salario ofrecido, y los avisos que lo tienen normalmente son empleos de menor calificación. Además, los datos de muchos sitios no tienen requerimientos codificados de manera estandarizada en cuanto a educación, experiencia y otras variables. En muchos casos, sería posible recuperar esta información a partir de los textos que describen a los avisos, con herramientas de minería de texto. Sin embargo, habrían pérdidas de precisión que podrían variar dependiendo de ciertos requerimientos, como los niveles educacionales. Por lo tanto, sistemas de recolección de información automática en la web (“*scraping*”) tendrían problemas para recuperar estos datos (Banfi y Villena-Roldán, 2019) [16].

Utilizando información extraída mediante minería de texto, distribuciones empíricas, estimaciones de Mincer y descomposiciones de Gelbach, se estudian los retornos asociados a alcanzar distintos niveles educacionales en el contexto del mercado laboral chileno. Específicamente, el análisis se acota al mercado en línea que se desarrolla hasta el día de hoy en la plataforma www.trabajando.com. Adicionalmente, se comparan datos descriptivos y estimaciones que se pueden obtener a partir la Encuesta Suplementaria de Ingresos entre los años 2010 y 2016 para evaluar la representatividad de la muestra de los datos en línea en relación con los de la población chilena.

En el capítulo 1 se describen las bases de datos utilizadas para este estudio. En la sección 1.1 describe la base de datos de demanda por fuerza de trabajo de www.trabajando.com (de ahora en adelante “TBN”). Las estimaciones que se obtienen a partir de esta base son las que se asocian a la motivación principal de este estudio. La sección 1.2 describe la base de datos de la Encuesta Suplementaria de Ingresos (de ahora en adelante “ESI”). El objetivo de utilizar ambas bases de datos radica en poder comparar datos descriptivos y estimaciones obtenidas con cada una, considerando que pueden haber diferencias sistemáticas entre el mercado laboral online y global en Chile. Dicha comparación y los análisis descriptivos de las variables de interés de este estudio se presentan en la sección 1.3. En ella se muestra que algunas variables (como salarios y niveles educacionales) de TBN presentan distribuciones similares a sus contrapartes de la ESI. Sin embargo, hay diferencias considerables en términos de regiones e industrias. De todos modos, se puede utilizar la información de TBN para un grupo importante de demandantes de trabajo en el mercado laboral chileno.

En el capítulo 2 se muestra un análisis exploratorio realizado a partir de las distribuciones de efectos fijos de empresas y palabras en los títulos de trabajo (sobre los logaritmos de los salarios reales por hora) para cada nivel educacional. En la sección 2.1 se compara el comportamiento de los empleadores de www.trabajando.com que contratan, en distintas medidas, a trabajadores de diversos niveles educacionales. En particular, se busca entender cómo difieren las políticas salariales de las empresas que publican vacantes para los distintos niveles educacionales requeridos. En la sección 2.2 se desarrolla un análisis similar para las palabras presentes en los títulos de trabajo. El objetivo es entender cómo difieren las distribuciones remuneraciones de cargos asociados a los distintos niveles educacionales.

A partir de distribuciones estimadas y momentos de distribuciones empíricas de efectos fijos de empresas sobre salarios, mostrados en la sección 2.1, se concluye que existen notables diferencias entre las políticas salariales de las empresas que suelen requerir distintos niveles educacionales. En primer lugar, las empresas más generosas en términos de salarios son

aquéllas que requieren mayores niveles educativos. En segundo lugar, las políticas salariales de las empresas que requieren niveles educacionales más bajos para sus vacantes tienden a parecerse más que las de las firmas que buscan trabajadores más especializados. Finalmente, se observa que las distribuciones de efectos fijos de empresas para niveles educacionales menos avanzados pueden presentar asimetrías negativas, mientras que las de niveles más avanzados presentan asimetrías positivas. Por otro lado, los resultados de la sección 2.2 permiten obtener conclusiones similares para las palabras en los títulos de trabajo: los cargos mejor remunerados son aquellos que requieren mayor capacitación y los cargos que requieren menos educación para sus vacantes tienden a tener remuneraciones más parecidas. En esta ocasión, todas las distribuciones de efectos fijos tienen asimetrías positivas, las que suelen ser menos notorias para los niveles de educación más avanzados. Las comparaciones e interpretaciones se detallan en cada subsección.

En el capítulo 3 se expone la metodología de este trabajo, la cual ocupa estimaciones de Mincer y descomposiciones de Gelbach. En la sección 3.1 se introduce la notación utilizada para los modelos de este estudio. En la sección 3.2 se detallan los modelos de regresión que se utilizarán para estimar los retornos porcentuales de la educación sobre los salarios. En la sección 3.3 se da a conocer la descomposición de Gelbach (por ser un método relativamente novedoso), explicando su contexto de origen y aspectos teóricos y prácticos que son importantes para entender los resultados de este estudio. Tanto en esta sección como en el resto de este documento, se entiende el retorno de la educación como la valoración por el nivel educativo que expresan los empleadores a través de las diferencias de salarios ofrecidos en sus avisos de empleo, controlando o ajustando las comparaciones por otros factores que podrían influir en la determinación de estos salarios, como la experiencia requerida; efectos fijos de empresas; efectos fijos de palabras presentes en las descripciones de los cargos; entre otros.

El capítulo 4 contiene los resultados de las estimaciones llevadas a cabo e interpretaciones pertinentes. En la sección 4.1 se muestran las estimaciones mincerianas obtenidas con y sin incorporar efectos fijos de empresas y cargos. Al incorporar estos efectos fijos, los retornos estimados para los distintos niveles educacionales son considerablemente menores. Esto da un indicio de que parte importante de dichos retornos se pueden explicar por el acceso que proporcionan los distintos niveles educacionales a empresas y cargos mejor remunerados. La sección 4.2 tiene por objetivo estudiar los medios a través de los cuales se manifiesta el retorno a la educación, lo cual se hace analizando las estimaciones relativas a la descomposición de Gelbach. La educación podría tener un retorno intrínseco, que se explicaría por las expectativas que tienen los empleadores de los trabajadores a contratar en función del nivel educacional que han alcanzado. Por otro lado, la educación puede ser entendida como una “llave” que brinda acceso a los trabajadores a cargos y firmas de mayores remuneraciones. Estos accesos podrían ser explicaciones alternativas del valor que tiene la educación en el mercado laboral.

El capítulo 5 se desarrolla una discusión global de este estudio, mencionando consideraciones y limitaciones. Finalmente se dan a conocer las conclusiones generales del presente estudio, los anexos que sirven de apoyo para entender muchas de las descripciones y análisis que se desarrollan en las secciones anteriores y se enumeran las referencias citadas a lo largo del documento.

Capítulo 1

Datos

1.1. Trabajando.com (TBN)

1.1.1. Presentación

La página web www.trabajando.com tiene por objetivo ser una plataforma de búsqueda de empleo por parte de personas naturales y de publicación de avisos de trabajo por parte de empresas. Este sitio tiene presencia en Argentina, Brasil, Colombia, Chile, España, México, Perú, Portugal, Puerto Rico, Uruguay y Venezuela. La página es capaz de proveer al menos tres bases de datos útiles para estudios en economía laboral: la primera almacena los datos personales de los postulantes e información sobre sus postulaciones, la segunda contiene datos de los empleadores y la tercera posee información sobre los anuncios de trabajo. La fusión o cruce de estas últimas dos es la que se utiliza en el presente estudio y a la cual denominamos “TBN”.

Los empleadores pueden acceder a los servicios del sitio web pagando por una o más ofertas que les permiten publicar avisos que duran 60 días en la plataforma. Estas ofertas corresponden a paquetes de 1, 3, 5 y 10 avisos de trabajo, cuyos costos y beneficios respectivos se detallan en la tabla 5.1. Una vez que pagan por el servicio pueden completar el formulario en la figura 1.1.

La base de datos de avisos publicados por empresas tiene información de 233.254 de ellos, los cuales fueron subidos a la página entre el 1 de enero de 2008 y el 13 de octubre de 2016. Es importante recalcar que un aviso puede estar ligado a múltiples vacantes, siendo estas últimas las unidades de análisis de los datos descriptivos y estimaciones que se presentarán de aquí en adelante, por lo que se replicó cada fila tantas veces como vacantes tenga el aviso al cual representa. Tras ello la base expandida quedó con 1.129.324 observaciones de vacantes de empleo. Debido a que el enfoque del estudio es la demanda de trabajo, las variables consideradas son aquellas que las firmas ingresan a modo de requerimientos u ofertas para los posibles postulantes.

Figura 1.1: Formulario estándar que rellenan los empleadores en el sitio web.

Mis Avisos Publicar Aviso Ver Postulaciones Buscar Currículum Bodega Currículum Universidad de Chile

Cargar datos desde plantilla de oferta de empleo o de publicaciones anteriores >>

Descripción de la oferta de empleo

Nombre Empresa a Figurar *

Cargo/Puesto *

N° de vacantes *

Tipo de cargo *

Área *

Actividad de la Empresa*

Descripción de la Oferta de empleo *

Disponibilidad para Trabajar / jornada laboral *

Duración del Contrato *

Salario Líquido Mensual aprox. * Pesos Chilenos Mostrar salario en Oferta de empleo

Comentarios del Salario (comisiones/incentivos)

Nota: Actualizado al 5 de mayo de 2015. Los asteriscos rojos indican que el campo requerido es obligatorio.

1.1.2. Muestra del estudio

Una de las variables de interés de este trabajo es el salario que se ofrece en cada aviso. Los empleadores ingresan una estimación de salario mensual que se pagaría además del tipo de jornada laboral que se requiere. La tabla 1.1 muestra la distribución de vacantes según la disponibilidad de jornada requerida. Ante posibles ambigüedades en las definiciones de las jornadas, en este estudio se consideraron solo aquellas vacantes de trabajo que han sido definidas por los empleadores como de jornada completa, las cuales son 783.456 vacantes (casi el 70% de la muestra). De acuerdo a la Dirección del Trabajo de Chile¹, una “jornada completa” corresponde a 45 horas de trabajo semanales como máximo. Sin embargo se tiene evidencia de que muchos empleos comúnmente entendidos como de jornada completa varían considerablemente respecto a esta cifra.

Es de interés de este trabajo realizar las estimaciones a partir de salarios por hora y no mensuales, con el fin de obtener resultados comparables con estudios previos. En consecuencia, sería ideal tener una noción lo más precisa posible sobre las horas de trabajo que contemplaría cada vacante. Resultaría poco preciso asumir que todos los trabajadores a contratar destinarían 45 de horas semanales a sus labores. Por ello, se le imputó a cada vacante de trabajo una predicción de la cantidad de horas laborales por semana. Esta predicción se llevó a cabo a partir de la Encuesta Suplementaria de Ingresos, mediante un modelo que explica el logaritmo natural de las horas que usualmente trabajan los encuestados en función de los niveles educacionales alcanzados y variables dicotómicas de industria y tramo de número de trabajadores en cada empresa.

Para que la muestra sea comparable con los datos de la Encuesta Suplementaria de Ingresos (ESI), de las 783.456 vacantes de jornada completa, se tuvieron en cuenta aquellas 679.555 publicadas entre los años 2010 y 2016. Adicionalmente se eliminaron 185 vacantes restantes en las que se pedía más de 40 años de experiencia laboral. Considerando que el objetivo de este trabajo es estimar los retornos de alcanzar distintos niveles de educación formal sobre

¹<https://www.dt.gob.cl/portal/1628/w3-propertyvalue-22094.html>

Tabla 1.1: Distribución de jornadas requeridas por vacantes del sitio web.

Disponibilidad requerida en el aviso	Frecuencia	Porcentaje
Comisionista	22.545	2,00 %
Jornada Completa	783.456	69,37 %
Media jornada	47.116	4,17 %
Part time	99.035	8,77 %
Por turnos	161.270	14,28 %
Práctica profesional	13.209	1,17 %
Reemplazo	2.417	0,21 %
Sin información	276	0,02 %
Total	1.129.324	100,00 %

Nota: Información de la base de datos original de www.trabajando.com, no de la muestra de estudio. La jornadas descritas son las requeridas por los empleadores en los avisos de empleo que publican.

los salarios, se eliminaron de la muestra 90.919 vacantes restantes que no reportaban salarios o cuyos salarios reportados no eran creíbles².

Extendiendo la lógica propuesta por Banfi y Villena-Roldán (2018) [15] para filtrar vacantes, se descartaron 66.874 vacantes que ofrecen salarios nominales por hora estimados inferiores al salario mínimo legal por hora que trabajaría alguien que desempeña sus labores por el máximo legal de 45 horas semanales. Todo ello considerando la variación temporal del salario mínimo legal en Chile, la cual se puede observar en detalle en la figura 1.2. De este modo, habrían quedado 521.577 observaciones. Además se descartaron 2.045 observaciones con salarios reales (en base al año 2018) superiores a 5 millones de pesos chilenos, por estar esta cifra ampliamente sobre el ingreso por trabajo del percentil 99 según la encuesta CASEN 2017³. Una vez hechos todos estos filtros, se obtuvo una muestra de 519.532 vacantes que serán objeto de este estudio.

1.2. Encuesta Suplementaria de Ingresos (ESI)

1.2.1. Presentación de la base de datos

La Encuesta Suplementaria de Ingresos es un módulo complementario que se aplica dentro de la Encuesta Nacional de Empleo (ENE), cuyo objetivo principal es conocer en detalle la realidad de los ingresos que perciben los hogares chilenos a través del trabajo –de aquellos integrantes de hogares calificados como “ocupados” según la ENE– y otras fuentes. Esta encuesta va dirigida a personas de 15 años o más, que residan habitualmente en alguna de las cerca de 12.000 viviendas seleccionadas aleatoriamente cada mes por el Instituto Nacional

²Típicamente por tratarse de patrones numéricos como “1234567”, “1010101” o “999999”. Cabe destacar que www.trabajando.com brinda la posibilidad a los empleadores de no visibilizar el salario que ofrecerían, lo que explica que muchos empleadores escriban números sin sentido en este campo cuando eligen no mostrar esta información. No es posible afirmar que los empleadores no tengan incentivos a ingresar montos verídicos cuando eligen ocultar los salarios ofrecidos, pues –sea visible o no– el salario ofrecido es una variable que puede ser utilizada como filtro en los motores de búsqueda usados por los postulantes.

³http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/casen_2017.php

Tabla 1.2: Salario mínimo bruto legal en Chile a partir del 1 de julio de 2009.

	Monto nominal (CLP)		Monto real (CLP)		Monto real (USD)		
	Desde	Mensual	Por hora	Mensual	Por hora	Mensual	Por hora
01/julio/2009		165.000	844	212.570	1.087	351	1,79
01/julio/2010		172.000	880	219.135	1.121	361	1,85
01/julio/2011		182.000	931	224.176	1.146	370	1,89
01/julio/2012		193.000	987	231.587	1.184	382	1,95
01/agosto/2013		210.000	1.074	246.662	1.261	407	2,08
01/julio/2014		225.000	1.151	252.922	1.293	417	2,13
01/julio/2015		241.000	1.233	259.442	1.327	428	2,19
01/enero/2016		250.000	1.279	263.808	1.349	435	2,22
01/julio/2016		257.500	1.317	266.013	1.360	439	2,24

Nota: El salario real en CLP está calculado en base al 31 de enero de 2018. Los montos en dólares se obtuvieron mediante el tipo de cambio de esta misma fecha (1 USD = 606,4 CLP). Los montos por hora se calcularon considerando el máximo legal de 45 horas semanales trabajadas.

de Estadísticas (INE). Este muestreo permite capturar información de aproximadamente 35.000 personas mensualmente. La ESI se lleva a cabo durante 3 meses cada año (entre octubre y diciembre), por lo que anualmente se recopila información de entre 100.000 y 110.000 personas. Se tiene documentación de su realización desde el año 2001, sin embargo fue actualizada el año 2010 con el fin de permitir una mayor precisión de los datos. Tanto la información mencionada en este párrafo como las bases de datos utilizadas en este estudio se pueden encontrar en www.ine.cl/estadisticas/ingresos-y-gastos/esi.

1.2.2. Muestra del estudio

Debido a la mayor precisión de los datos a partir del año 2010, la muestra fue restringida al período comprendido entre los años 2010 y 2016. En estas siete versiones se ha capturado información de un total de 865.131 encuestados⁴. Sin embargo, no todas estas personas desarrollan trabajos remunerados. Cuando se restringe la muestra únicamente a la fuerza laboral, el número de observaciones se reduce a 402.550. De esta fuerza laboral sólo 343.896 personas desempeñan trabajos remunerados (es decir, son trabajadores dependientes asalariados). Un filtro específico que se aplicó a esta base de datos fue el de considerar a las personas que han sido contratadas a lo más 12 meses antes de ser encuestadas. El motivo de esta selección se basa en poder hacer una comparación más fiable con los avisos de trabajo publicados en www.trabajando.com, considerando que pueden haber diferencias sistemáticas de salarios entre las personas recientemente contratadas (o por contratar) y aquellas que llevan más de un año trabajando en la misma firma. Una vez aplicado este filtro la muestra se reduce a 117.330 trabajadores remunerados contratados recientemente. De ellos, 117.220 individuos cuentan con información de salarios y nivel educacional.

Consistentemente con la selección muestral llevada a cabo con los datos de TBN, se descartaron aquellos individuos con salarios nominales por hora inferiores al salario mínimo legal por hora que trabajaría alguien que desempeña sus labores por el máximo legal de 45 horas

⁴Esta cifra no considera que una o más personas hayan podido ser encuestadas en más de una versión.

semanales y a quienes reciben salarios reales (en base al año 2018) superiores a 5 millones de pesos chilenos. Además se eliminaron aquellas observaciones de personas que dicen trabajar más de 60 horas a la semana. Una vez hechos todos estos filtros, se obtuvo una muestra de 61.484 encuestados cuyos datos serán utilizados para efectuar los reportes y comparaciones que aquí se presentan.

Una ventaja que tiene esta base de datos por sobre la de TBN es que los encuestados declaran tanto sus ingresos como el número de horas que trabajan a la semana. Esto permite calcular el salario por hora que percibe usualmente cada trabajador sin necesidad de hacer estimaciones ni de filtrar observaciones por proporcionar información ambigua en cuanto a su jornada de trabajo. Tampoco fue necesario eliminar de la base de datos a individuos por proporcionar información no creíble en relación a sus salarios, como sí lo hacen los empleadores que publican vacantes de empleo en www.trabajando.com.

1.3. Comparación TBN-ESI y variables de interés

1.3.1. Resumen

En este apartado se presenta una evaluación sobre la representatividad que tienen los datos de TBN respecto a la realidad de la economía laboral chilena. En primer lugar se comparan los datos considerados en este estudio entre las bases de TBN y ESI entre los años 2010 y 2016. En segundo lugar se estudia la comparabilidad del flujo de creación de trabajo (o demanda por trabajo) en términos de nivel educacional, salarios y otras variables de interés para este estudio. Los datos de TBN presentan distribuciones relativamente similares a aquellas estimadas a nivel poblacional con los datos de ESI. Sin embargo, la comparabilidad a nivel regional y sectorial es bastante más acotada. Es importante notar que la Región Metropolitana y los sectores comercial, de servicios e inmobiliario concentran desmedidamente la demanda de trabajo en TBN (esto último en desmedro de otros como construcción, educación, y agricultura y silvicultura). Es posible utilizar la información de TBN para un grupo importante de demandantes de trabajo en el mercado laboral chileno. Sin embargo, la muestra no es perfectamente representativa en varios aspectos. Por ello sería factible realizar estimaciones a partir de los datos del sitio web con representatividad nacional aproximada, aunque con limitaciones.

Cabe destacar que muchos de los reportes que aquí se muestran se clasifican en cuatro categorías: avisos TBN, vacantes TBN, flujo ESI y flujo ponderado ESI. Como ya se mencionó anteriormente, en www.trabajando.com los potenciales empleadores pueden publicar varias vacantes de trabajo en un mismo aviso. Dado que se pretende representar la demanda de trabajo es más riguroso basar los análisis en las vacantes en lugar de los avisos. Asimismo, los datos de la ESI pueden ser analizados considerando a todos los encuestados –en este caso trabajadores remunerados contratados en los últimos 12 meses– por igual (flujo ESI) o teniendo en cuenta su grado de representatividad en la población chilena mediante la incorporación del factor de expansión propio de la encuesta⁵ (flujo ponderado ESI). Por la construcción misma de la encuesta, resulta razonable considerar para el análisis las observaciones ponde-

⁵https://www.ine.cl/docs/default-source/ingresos-y-gastos/esi/ingreso-de-hogares-y-personas/antecedentes-metodol%C3%B3gicos/documento_metodologico_esi_2017.pdf?sfvrsn=2dbd4fd2_10

radas por el factor de expansión para obtener estimaciones representativas del flujo laboral de la economía chilena.

1.3.2. Niveles educacionales

Los niveles educacionales cumplen un rol fundamental en este estudio, pues el interés de éste radica en entender los retornos comparativos de alcanzar un determinado nivel educacional por sobre uno elegido arbitrariamente como nivel base. Este último sería la educación básica (o primaria), la cual está asociada a los menores salarios. A continuación se comparan los requerimientos educacionales ingresados por los empleadores en www.trabajando.com con los niveles educacionales logrados en el flujo de nuevos empleos ESI. Banfi, Choi y Villena-Roldán (2019) [15] muestran empíricamente y cuantifican la tendencia de quienes buscan cargos de trabajo por ajustarse a los requerimientos de las vacantes durante su proceso de búsqueda. Extrapolando este resultado, es esperable que los niveles educacionales de trabajadores pertenecientes al flujo laboral de la ESI se parezcan a los requerimientos educacionales de las vacantes de trabajo que se están creando en Chile.

En www.trabajando.com existen 6 categorías de requerimientos educacionales: educación básica, media científico-humanista (CH), media técnico profesional (TP), superior técnico-profesional (TP), universitaria de pregrado y universitaria de postgrado. En contraparte, la ESI tiene más información pues en ella se detalla si la persona completó cierto nivel de estudios y, también, el año que está cursando cada encuestado en su respectivo nivel actual. En virtud de lo anterior, se han agregado los datos de la ESI para hacerlos homologables con las categorías de TBN. Este proceso de homologación es natural y parcialmente arbitrario, pues el criterio utilizado es que los encuestados de ESI satisfagan un nivel mínimo educacional que se pueda relacionar a los requerimientos estipulados por los empleadores en el sitio web. En caso de utilizarse un criterio diferente la clasificación de los datos de ESI podrían variar. El resultado de esta clasificación se puede observar en la tabla 1.3, mientras que la distribución por nivel educacional se muestra en la figura 1.2. Cabe destacar que se han agrupado los niveles universitarios de pregrado y postgrado, pues la cantidad de observaciones correspondientes a este último nivel son relativamente escasas en ambas bases de datos.

La figura 1.2 muestra que existen similitudes importantes entre las distribuciones de vacantes TBN y flujo ponderado ESI. Existen diferencias considerables cuando se contrastan estos gráficos con los de avisos TBN. En la muestra de avisos se observa que los que requieren educación básica representan el 2,0 %, mientras que si se observa la muestra de vacantes esta cifra asciende al 20,6 %. Esto quiere decir que los trabajos que requieren enseñanza básica como nivel mínimo de educación formal están concentrados en pocos avisos que tienen numerosas vacantes cada uno. Opuesto es el caso de la educación universitaria la cual concentra el 32,6 % de los avisos, pero sólo el 12,5 % de las vacantes. La idea general es que los avisos que requieren menor calificación tienden a tener más vacantes que los que piden un mayor nivel educacional. Por otro lado, los datos de la ESI no reportan diferencias significativas en esta distribución al momento de incorporar el factor de expansión de esta encuesta.

Cabe destacar que dos quintos de las vacantes del sitio web requieren al menos educación media científico-humanista. En segundo lugar, se tiene la educación básica con más de un quinto de las vacantes. Los dos quintos restantes de las vacantes se distribuyen entre educación

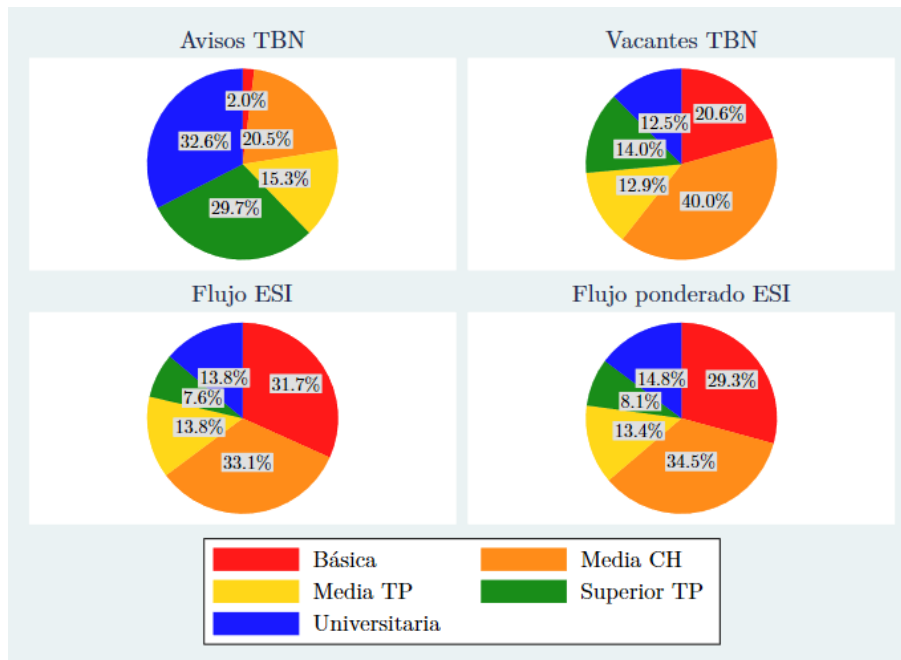
Tabla 1.3: Distribución educacional de avisos y vacantes TBN y flujo (ponderado) ESI.

Educ. req.	Avisos TBN		Vacantes TBN		Educ. alc.	Flujo ESI				Flujo ponderado ESI			
	Frec.	Porc.	Frec.	Porc.		Frec.	Porc.	Frec.	Porc.	Frec.	Porc.	Frec.	Porc.
Básica	2.361	2,0%	107.040	20,6%	Sin básica	5.700	9,3%	19.512	31,7%	5.107	8,3%	17.994	29,3%
					Básica	5.928	9,6%	5.177	8,4%	5.177	8,4%		
					Media inc.	7.884	12,8%	7.710	12,5%	7.710	12,5%		
Media CH	24.536	20,5%	207.852	40,0%	Media CH	12.940	21,0%	24.312	39,5%	12.907	21,0%	24.392	39,7%
					Sup. TP inc.	8.504	13,8%	8.239	13,4%	8.239	13,4%		
					Univ. inc.	2.868	4,7%	3.247	5,3%	3.247	5,3%		
Media TP	18.305	15,3%	67.262	12,9%	Media TP	4.657	7,6%	4.657	7,6%	4.970	8,1%	4.970	8,1%
Superior TP	35.552	29,7%	72.647	14,0%	Sup. TP	4.528	7,4%	4.528	7,4%	5.044	8,2%	5.044	8,2%
Universitaria	38.977	32,6%	64.731	12,5%	Univ.	7.804	12,7%	8.475	13,8%	8.219	13,4%	9.084	14,8%
					Post. inc.	165	0,3%	206	0,3%	206	0,3%		
					Post.	506	0,8%	658	1,1%	658	1,1%		
Total	119.731	100,0%	519.532	100,0%	Total	61.484	100,0%	61.484	100,0%	61.484	100,0%	61.484	100,0%

Nota: En el caso de TBN los niveles educacionales descritos en cada caso son los requeridos por los empleadores, mientras que en ESI son los alcanzados por los encuestados. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta. Los números de encuestados que se reportan para el flujo ponderado ESI están dados por $n_e = (\sum_{i=1}^N FE_i d_{i,e} / \sum_{i=1}^N FE_i) N$, donde $N = 61,484$ es el número total de observaciones en la muestra de la ESI y $d_{i,e}$ es una variable binaria que toma el valor 1 si el encuestado i reporta un nivel educacional e y 0 en caso contrario. Se han agrupado los niveles universitarios de pregrado y postgrado.

media TP, superior TP y universitaria.

Figura 1.2: Distribución por nivel educación de avisos y vacantes TBN y flujo (ponderado) ESI.



Nota: En el caso de TBN los niveles educacionales descritos en cada caso son los requeridos por los empleadores, mientras que en ESI son los alcanzados por los encuestados. Se muestran las distribuciones para avisos y vacantes TBN y flujo (ponderado) ESI. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta. Se han agrupado los niveles universitarios de pregrado y postgrado.

1.3.3. Salarios

La Encuesta Suplementaria de Ingresos mide ingresos corrientes (nominales, provenientes principalmente del mercado del trabajo, las rentas de la propiedad y las transferencias corrientes) que perciben las personas y los hogares en el período de referencia de la encuesta⁶. En este estudio han sido considerados los ingresos netos por ocupación (excluyen los descuentos legales e impuestos) que reciben usualmente aquellas personas que trabajan de manera remunerada y que han sido contratadas en los últimos 12 meses. Esta variable parece ser coherente en su definición con el salario líquido mensual aproximado que ingresan los empleadores de `www.trabajando.com` en el formulario de la figura 1.1.

Tanto para los datos de ESI como para los de TBN, se convirtieron los salarios nominales a reales, utilizando para los cálculos las series del Índice de Precios al Consumidor (IPC), publicadas por el Instituto Nacional de Estadísticas (INE)⁷, utilizando como base el IPC de año 2018. Posteriormente, se estimaron los salarios reales por hora para cada base de datos. Finalmente se le aplicó logaritmo a estos salarios reales por hora, siguiendo la estructura tradicional de la ecuación de Mincer, que permite estimar retornos porcentuales asociados a la educación.

En la figura 1.3 se muestran distribuciones estimadas, mediante estimaciones kernel no-paramétricas (función de densidad de probabilidades), para los logaritmos de los salarios reales por hora estimados para cada uno de los niveles educativos, utilizando los datos de TBN y ESI. La figura muestra que las distribuciones de salarios de TBN son bastante parecidas a las del flujo de ESI. Este parecido se acentúa cuando se ponderan correctamente las observaciones mediante la cantidad de vacantes en el caso de TBN y el factor de expansión en el caso de ESI (a excepción de la educación básica). En el caso del nivel universitario, parece ser que las vacantes de TBN reportan salarios más altos que las de ESI.

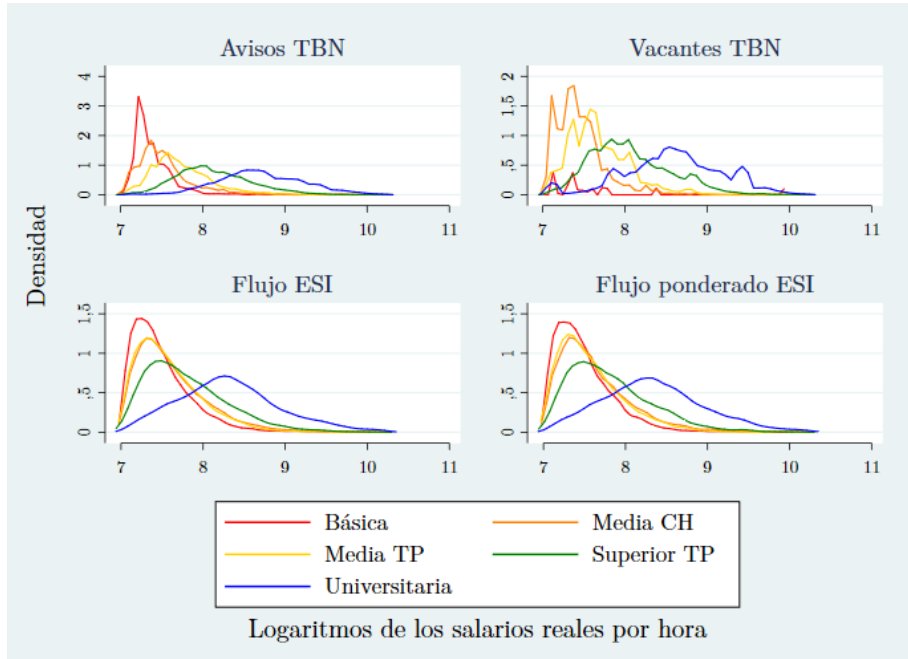
Se puede observar que los datos de TBN proporcionan curvas con más “saltos” que los de ESI, lo cual es más notorio en los niveles educacionales más bajos. En primer lugar, puede ser importante recordar que a los empleadores se les solicitan salarios estimados (no exactos). Es muy probable que muchos de ellos trunquen o redondeen las cifras, lo cual generaría distribuciones más “ruidosas”. Dada la escala logarítmica del eje horizontal, es natural que este “ruido” sea menos notorio cuando se observan salarios mayores. En segundo lugar, se destaca que los avisos de trabajo asociados a niveles educacionales más bajos (que a la vez reportan salarios menores) suelen estar asociados a mayores cantidades de vacantes. Una pequeña cantidad de avisos que ofrecen un mismo salario y que tienen muchas vacantes puede elevar considerablemente la densidad de probabilidad estimada para ese salario, generando un “salto” en la distribución.

Para analizar con mayor precisión las similitudes y diferencias de las distribuciones de salarios reales por hora se ha creado la tabla 1.4. En ella se observan las estimaciones de los primeros cuatro momentos (esperanza, varianza, asimetría y curtosis), además de las medianas y desviaciones estándar de sus distribuciones. A partir de esta tabla, se pueden identificar tres grupos de niveles educacionales, los que se describen a continuación.

⁶https://www.ine.cl/docs/default-source/ingresos-y-gastos/esi/ingreso-de-hogares-y-personas/resultados/2017/sintesis_esi_2017_nacional.pdf

⁷<https://www.ine.cl/estad%C3%ADsticas/precios/ipc>

Figura 1.3: Distribución de (log) salarios reales por hora para datos de TBN y ESI.



Nota: Funciones de densidad de probabilidades, obtenidas mediante estimaciones kernel no-paramétricas, de los salarios reales por hora según el nivel educacional. En el caso de TBN los niveles educacionales y salarios son los requeridos y ofrecidos por los empleadores, mientras que en ESI son los reportados por los encuestados. Se muestran las distribuciones para avisos y vacantes TBN y flujo (ponderado) ESI. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta. El eje vertical del gráfico de avisos TBN tiene una escala distinta al resto. Se han agrupado los niveles universitarios de pregrado y postgrado.

Educación básica y media científico humanista

En general los datos de la ESI reportan mayores salarios en medias y medianas. Mientras que los datos de flujo ponderado de la ESI reportan un salario real por hora esperado de 1.966 CLP para la enseñanza básica, las vacantes TBN apenas llegan a 1.467 CLP. Es decir, los salarios declarados en la encuesta son un 34 % superiores a los requeridos por los empleadores del sitio web. Al momento de observar las medianas, esta diferencia porcentual se reduce a un 19 %. Lo que se observa en el caso de la educación media científico humanista es el mismo fenómeno pero con diferencias considerablemente menores entre ambas bases datos, siendo la brecha porcentual en medias de casi un 27 % y en medianas de casi un 14 %.

En ambos niveles educacionales, los datos de TBN tienen una asimetría más positiva que los de ESI. Esto quiere decir que los salarios ofrecidos para las vacantes del sitio web están más concentrados a la izquierda de su media y más dispersos a su derecha que los de la encuesta. Se infiere que los empleadores de www.trabajando.com en general ofrecen salarios relativamente bajos a los postulantes con niveles educaciones menos especializados, lo cual se compensa (en media y mediana) con una baja proporción de empleos cuyos salarios son relativamente altos. Gráficamente, la distribución de TBN tiene una cola derecha más larga que la de ESI. Cabe destacar que los salarios de vacantes TBN están más concentrados en torno a su media y son menos dispersos en general que los que reciben los encuestados de la ESI. Esto se evidencia numéricamente por mayores curtosis y menores varianzas.

Educación media técnico profesional

Existe un gran parecido en la mayoría de los momentos y estadísticos entre las distribuciones reportadas por ambas bases de datos para este nivel educacional. La media de los salarios reales por hora de vacantes TBN apenas supera en un 9 % a la del flujo ponderado de ESI, mientras que la brecha de medianas es del 13 % aproximadamente. En cuanto a la varianza, asimetría y curtosis parecen no haber diferencias significativas.

Educación superior técnico profesional y universitaria

Curiosamente todos los niveles correspondientes a educación superior reportan los mismos patrones de comparabilidad de distribuciones de salarios entre ambas fuentes de datos, y dichos patrones son opuestos a los de educación básica y media científico humanista. En general los datos de TBN reportan mayores salarios en medias y medianas. Mientras que los datos de vacantes de TBN reportan salarios reales por hora esperados de 3.624 CLP para educación superior técnico profesional, el flujo ponderado ESI apenas llega a 2.877 CLP. Es decir, las vacantes del sitio web tienen cerca de un 26 % más de salario esperado. Al momento de observar las medianas, esta diferencia porcentual aumenta a un 33 % aproximadamente. En los casos de la educación universitaria la brecha porcentual de medias es cercana al 26 % respectivamente, mientras que la de medianas es de aproximadamente un 35 %.

En estos ambos niveles educacionales, los datos de ESI tienen una asimetría más positiva que los de TBN (cuya asimetría sigue siendo positiva). Si bien los gráficos de ambas bases de datos presentan colas derechas más largas que las izquierdas, este fenómeno es menos notorio para los datos de TBN según estas cifras. Esto quiere decir que, si bien las ofertas de salarios de www.trabajando.com son asimétricas (con muchas ofertas bajo la media y pocas considerablemente sobre la media), esta asimetría es más pronunciada en la economía chilena (representada por la muestra de la Encuesta Suplementaria de Ingresos).

Cabe destacar que los salarios que ofrecen los empleadores que publican en el sitio web están menos concentrados en torno a su media y son más dispersos en general que los que reciben los encuestados de la ESI. Esto se evidencia numéricamente por mayores varianzas y menores curtosis.

1.3.4. Regiones

En el sitio web, a los empleadores se les solicita ingresar la región en la cual se desempeñaría el trabajo correspondiente a cada aviso. Lamentablemente, hay 491.348 vacantes sin información en esta variable (más del 94 % de la muestra)⁸. Por este motivo, se utilizó como variable geográfica la región en la que se ubica la casa matriz de la empresa, la cual es reportada en muchos más casos. El problema de esto es que no necesariamente la ubicación de la casa matriz de una empresa coincide con el lugar donde se desempeñaría un determinado trabajo. El caso más ilustrativo de este inconveniente es el de la Región Metropolitana. En ella se ubican las casas matrices de firmas que han publicado 424.532 vacantes (casi el 82 % de

⁸Existen varias hipótesis no excluyentes entre sí sobre por qué ocurre este fenómeno: pueden haber cadenas nacionales que hagan convocatorias transversales para todo el país, trabajos que pueden realizarse sin tener que acudir físicamente a un lugar, puede ser que en algunos casos la región se pueda deducir a partir del título del trabajo o que los empleadores hayan cometido errores al momento de ingresar los datos.

Tabla 1.4: Estadísticos de distribuciones empíricas de salarios reales por hora, por nivel educacional, de avisos y vacantes TBN y flujo (ponderado) ESI.

Nivel educ.	Momento estimado	TBN		ESI	
		Avisos	Vacantes	Flujo	Flujo pond.
Básica	Esperanza	1.708,2	1.466,9	1.959,2	1.965,8
	Varianza	897.803,8	223.838,4	1.419.899,7	1.420.330,6
	Asimetría	10,7	20,4	6,5	6,3
	Curtosis	159,8	606,1	76,6	68,0
	Mediana	1.521,2	1.382,6	1.637,0	1.648,8
	Desviación est.	947,5	473,1	1.191,6	1.191,8
Media CH	Esperanza	2.036,1	1.880,3	2.331,1	2.378,7
	Varianza	1.722.598,0	1.443.454,3	2.804.421,8	3.350.361,0
	Asimetría	8,4	9,1	5,2	5,3
	Curtosis	110,6	118,5	45,3	45,2
	Mediana	1.742,2	1.630,2	1.848,3	1.856,4
	Desviación est.	1.312,5	1.201,4	1.674,6	1.830,4
Media TP	Esperanza	2.543,3	2.393,8	2.192,5	2.195,7
	Varianza	2.458.527,3	1.963.250,8	1.985.296,3	2.244.744,0
	Asimetría	6,5	6,1	6,4	6,5
	Curtosis	71,9	69,7	74,4	71,1
	Mediana	2.166,1	2.088,9	1.813,8	1.804,8
	Desviación est.	1.568,0	1.401,2	1.409,0	1.498,2
Superior TP	Esperanza	3.809,5	3.624,8	2.834,5	2.876,5
	Varianza	4.733.881,1	4.915.584,2	4.014.899,3	4.156.041,3
	Asimetría	3,0	3,0	3,9	3,8
	Curtosis	20,4	19,4	27,7	26,1
	Mediana	3.190,2	2.992,9	2.219,1	2.256,4
	Desviación est.	2.175,7	2.217,1	2.003,7	2.038,6
Universitaria	Esperanza	6.929,9	6.464,1	5.030,2	5.141,6
	Varianza	15.560.113,9	15.556.211,4	14.507.992,8	14.824.150,6
	Asimetría	1,6	1,4	2,5	2,4
	Curtosis	6,5	5,7	11,0	10,9
	Mediana	5.826,5	5.429,7	3.928,7	4.015,8
	Desviación est.	3.944,6	3.944,1	3.808,9	3.850,2

Nota: Información correspondiente a datos de avisos y vacantes TBN y flujo (ponderado) ESI considerados en la muestra de este estudio. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta.

la muestra), mientras que solo 25.507 de ellas con certeza corresponden a avisos de trabajos a desempeñarse en dicha región. Algunos casos particulares pueden ser los de aquellas que tienen centralizadas sus operaciones de contratación en la casa matriz –típicamente ubicada en Santiago– o los de aquellas que externalizan la contratación de personal a agencias dedicadas a esta labor –es posible que muchas de estas también operen en Santiago aunque busquen llenar cargos de trabajo fuera de la Región Metropolitana–.

Una ventaja de los datos de ESI por sobre los de TBN es que tiene nivel nacional tiene información con buena cobertura regional. En la tabla 1.5 se muestra la distribución regional

de vacantes de trabajo publicadas por empleadores en www.trabajando.com (TBN, a la izquierda) y de encuestados en la Encuesta Suplementaria de Ingresos contratados en el último año (ESI, a la derecha). Como se puede apreciar en ella, la distribución regional de las residencias de los encuestados en la ESI están mucho más distribuidas en las distintas regiones que las de TBN y no hay observaciones sin información en este campo. Cerca del 44 % de las observaciones ponderadas de la ESI corresponden a trabajadores de la Región Metropolitana, lo cual es poco más de la mitad que las vacantes de TBN.

Tabla 1.5: Distribución regional de avisos y vacantes TBN y flujo (ponderado) ESI.

Región	Avisos TBN		Vacantes TBN		Flujo ESI		Flujo ponderado ESI	
	Frec.	Porc.	Frec.	Porc.	Frec.	Porc.	Frec.	Porc.
De Tarapacá	142	0,1 %	337	0,1 %	2.284	3,7 %	1.296,6	2,1 %
De Antofagasta	394	0,3 %	966	0,2 %	2.551	4,1 %	2.495,3	4,1 %
De Atacama	81	0,1 %	304	0,1 %	2.082	3,4 %	1.125,3	1,8 %
De Coquimbo	81	0,1 %	182	0,0 %	3.614	5,9 %	2.620,8	4,3 %
De Valparaíso	1.646	1,4 %	2.817	0,5 %	8.617	14,0 %	6.128,3	10,0 %
Del Libertador Gral. Bernardo O'Higgins	374	0,3 %	834	0,2 %	4.014	6,5 %	3.426,0	5,6 %
Del Maule	437	0,4 %	634	0,1 %	3.585	5,8 %	3.188,5	5,2 %
Del Biobío	820	0,7 %	2.530	0,5 %	7.583	12,3 %	6.237,6	10,1 %
De La Araucanía	232	0,2 %	400	0,1 %	2.449	4,0 %	2.710,2	4,4 %
De Los Lagos	184	0,2 %	301	0,1 %	3.051	5,0 %	2.777,5	4,5 %
De Aysén del Gral. Carlos Ibáñez del Campo	18	0,0 %	31	0,0 %	1.617	2,6 %	472,2	0,8 %
De Magallanes y de la Antártica Chilena	69	0,1 %	131	0,0 %	825	1,3 %	639,1	1,0 %
Metropolitana de Santiago	79.288	66,2 %	424.532	81,7 %	16.102	26,2 %	26.769,9	43,5 %
De Los Ríos	56	0,0 %	156	0,0 %	1.432	2,3 %	1.092,2	1,8 %
De Arica y Parinacota	7	0,0 %	9	0,0 %	1.678	2,7 %	504,3	0,8 %
Extranjera	1.794	1,5 %	5.578	1,1 %	0	0,0 %	0	0,0 %
Sin información	34.108	28,5 %	79.790	15,4 %	0	0,0 %	0	0,0 %
Total	119.731	100,0 %	519.532	100,0 %	61.484	100,0 %	61.484	100,0 %

Nota: Información correspondiente a datos de avisos y vacantes TBN y flujo (ponderado) ESI considerados en la muestra de este estudio. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta.

1.3.5. Experiencia

En el caso de la Encuesta Suplementaria de Ingresos es posible estimar la denominada “experiencia potencial” (Mincer, 1974 [1]), la cual se mide en años⁹, como variable de control en las estimaciones que se presentan más adelante. Por otro lado, www.trabajando.com solicita a los empleadores ingresar la experiencia requerida en años en sus avisos de trabajo. Ambos conceptos de experiencia son distintos, por lo que no hace sentido compararlos. Sin embargo, es útil incorporar ambos tipos de experiencias a las estimaciones que se llevan a cabo con cada base de datos, para controlar los efectos que tiene el capital humano que se obtiene en el mundo laboral sobre los salarios.

1.3.6. Industrias

La distribución completa de vacantes entre las distintas industrias para ambas bases de datos se puede apreciar con mayor precisión en la tabla 1.6. En cuanto a las vacantes del

⁹Esta variable se ha obtenido de la siguiente forma: Experiencia potencial = Edad – (Escaridad + 6). El número 6 corresponde a la edad que tiene típicamente la gente antes de ingresar a la educación básica o primaria.

sitio web, las industrias más comunes son: “otros servicios”¹⁰ (25,8%), comercio (25,6%), inmobiliaria (15,6%), transporte y comunicaciones (10,3%) y manufactura (8,1%). Por otro lado, al analizar los datos ponderados de la ESI, se observa que las industrias más comunes en la muestra de encuestados son: comercio (17,1%), construcción (13,3%), manufactura (10,7%), educación (8,4%) e inmobiliaria (7,1%).

Los sectores de servicios, comercial e inmobiliario concentran gran parte de la demanda de trabajo del sitio web (esto último en desmedro de otros como construcción, educación y agricultura y silvicultura). Si bien se pueden usar la información de TBN para un grupo importante de demandantes de trabajo en el mercado laboral chileno, la muestra de este estudio no es perfectamente representativa en cuanto a las distintas industrias de la economía chilena.

Tabla 1.6: Distribución por sectores industriales de avisos y vacantes TBN y flujo (ponderado) ESI.

Industria	Avisos TBN		Vacantes TBN		Flujo ESI		Flujo ponderado ESI	
	Frec.	Porc.	Frec.	Porc.	Frec.	Porc.	Frec.	Porc.
Agricultura y silvicultura	2277	1,9%	3413	0,7%	5608	9,1%	4284,6	7,0%
Pesca	287	0,2%	331	0,1%	467	0,8%	290,6	0,5%
Minería	2120	1,8%	4127	0,8%	4090	6,7%	2829,7	4,6%
Manufactura	17385	14,5%	42002	8,1%	6221	10,1%	6659,4	10,8%
Electricidad, agua y gas	2784	2,3%	8456	1,6%	742	1,2%	639,9	1,0%
Construcción	4295	3,6%	7549	1,5%	7629	12,4%	8120,1	13,2%
Comercio	26754	22,3%	132914	25,6%	10007	16,3%	10636,6	17,3%
Restaurantes y hotelería	1692	1,4%	4675	0,9%	2843	4,6%	2792,7	4,5%
Transporte y comunicaciones	7864	6,6%	53485	10,3%	3892	6,3%	4138,7	6,7%
Financiero	4196	3,5%	13205	2,5%	897	1,5%	1128,3	1,8%
Inmobiliario	23748	19,8%	80947	15,6%	3630	5,9%	4348,9	7,1%
Administración pública	1025	0,9%	3792	0,7%	2945	4,8%	2528,6	4,1%
Educación	6794	5,7%	16425	3,2%	5134	8,4%	5083,5	8,3%
Social y salud	5848	4,9%	14371	2,8%	2505	4,1%	2664,8	4,3%
Otros servicios	12662	10,6%	133840	25,8%	1541	2,5%	1781,2	2,9%
Servicios Domésticos	0	0,0%	0	0,0%	3312	5,4%	3533,8	5,7%
Otra	0	0,0%	0	0,0%	21	0,0%	22,6	0,0%
Total	119.731	100,0%	519.532	100,0%	61.484	100,0%	61.484	100,0%

Nota: Información correspondiente a datos de avisos y vacantes TBN y flujo (ponderado) ESI considerados en la muestra de este estudio. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta.

1.3.7. Tamaño de las empresas

En la tabla 1.7 se muestra la distribución de avisos y vacantes correspondientes a la muestra de www.trabajando.com y de la Encuesta Suplementaria de Ingresos (con y sin ponderar con respecto al factor de expansión de la misma) según el tamaño de las empresas en las que se desempeñan. Es importante recalcar que existen diferencias entre las categorías de tamaños que los empleadores ingresan en el sitio web y las que los encuestados responden mediante la ESI. En virtud de lo anterior fue necesario crear una clasificación que intenta conciliar las categorías de ambas bases de datos.

¹⁰Excluye a servicios domésticos, financieros, sociales y de salud.

Dicha clasificación homologada fue utilizada para las estimaciones que se llevaron a cabo de manera conjunta entre ambas bases de datos. Sin embargo, en aquellas ocasiones en las que se hicieron estimaciones separadamente se ocupó la clasificación de la respectiva base de datos por contener información más precisa. Un detalle que vale la pena destacar es que existe cierto margen de error en la homologación realizada por el hecho de incorporar en la categoría de entre 51 y 150 trabajadores a aquellos encuestados que declararon que la firma en la cual trabaja tiene entre 50 y 199 trabajadores. Sin embargo se espera que esto no altere significativamente los estimadores de interés de este estudio.

Tabla 1.7: Distribución por tamaños de empresas de avisos y vacantes TBN y flujo (ponderado) ESI.

Tam. hom.	Tam. TBN	Avisos TBN		Vacantes TBN		Tam. ESI	Flujo ESI		Flujo pond. ESI	
		Frec.	Porc.	Frec.	Porc.		Frec.	Porc.	Frec.	Porc.
1-10	1-10	29.757	24,9 %	172.969	33,3 %	<5	10.250	16,7 %	9.903,6	16,1 %
						05-10	4.594	7,5 %	4.501,2	7,3 %
11-49	11-50	34.920	29,2 %	120.110	23,1 %	11-49	11.619	18,9 %	11.735,1	19,1 %
51-150	51-150	12.788	10,7 %	47.532	9,1 %	50-199	9.963	16,2 %	10.346,2	16,8 %
>151	151-300	9.181	7,7 %	28.155	5,4 %	≥200	22.639	36,8 %	22.486,4	36,6 %
	301-500	9.860	8,2 %	65.216	12,6 %					
	501-1000	5.470	4,6 %	36.755	7,1 %					
	1001-5000	10.429	8,7 %	27.637	5,3 %					
	5000	803	0,7 %	4.163	0,8 %					
Sin info.	Sin info.	6.523	5,4 %	16.995	3,3 %	Sin info.	2.419	3,9 %	2.511,5	4,1 %
Total	Total	119.731	100,0 %	519.532	100,0 %	Total	61.484	100,0 %	61.484,0	100,0 %

Nota: En el caso de TBN los tamaños son rangos de número de trabajadores que declaran tener los empleadores, mientras que en ESI son rangos declarados por los encuestados. Estos rangos son distintos entre ambas bases por lo que surgió la necesidad de crear una clasificación homologada con cierto margen de error. Información correspondiente a datos de avisos y vacantes TBN y flujo (ponderado) ESI considerados en la muestra de este estudio. Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta.

1.3.8. Retornos de la educación sobre los salarios

En esta subsección se muestran estimaciones para los retornos salariales de alcanzar distintos niveles educativos tanto para los datos de vacantes TBN, como de flujo ponderado ESI. En primer lugar, se muestran los resultados de una regresión lineal hecha a partir de una base de datos fusionada, que contiene datos de ambas fuentes. A partir de estas estimaciones, para cada nivel educacional, se pone a prueba la hipótesis nula de que el retorno sobre los salarios es igual en ambas bases de datos. Los resultados arrojados muestran evidencia contundente de diferencias considerables entre los retornos obtenidos con ambas bases de datos. En particular, se observa que los retornos estimados con los datos de TBN suelen ser mayores. En segundo lugar, se comparan resultados de regresiones lineales llevadas a cabo por separado para cada base de datos. Los resultados son muy semejantes a los de la regresión descrita anteriormente.

Cabe destacar que es natural que los retornos estimados con ambas bases presenten diferencias considerables, teniendo en cuenta una serie de factores. Una primera causa de esta diferencia radica en que los datos que proporciona cada base son distintos, los de TBN corresponden a variables requeridas u ofrecidas por los empleadores y los de ESI a información efectiva que declaran los encuestados. Una segunda causa sería un fenómeno que ya se ha discutido anteriormente: los problemas de endogeneidad que presentan las estimaciones mincerianas cuando se omiten las habilidades no observables. Este problema está presente en las estimaciones de la ESI, por lo que los estimadores obtenidos con esta base serían sesgados. Una tercera causa es que los datos de TBN presentan una sobrerrepresentación de la Región Metropolitana y algunos sectores industriales, como el comercial, el de servicios y el inmobiliario (en desmedro de otros como construcción, educación y agricultura y silvicultura).

En la tabla 1.8 se reportan los retornos a los niveles educativos de acuerdo a las estimaciones realizadas a partir de cada conjunto de datos. El principal interés de estimar modelos a través de una base de datos unificada radica en la posibilidad de probar si es que existen diferencias estadísticamente significativas entre los coeficientes estimados para ESI y TBN. Para poder estimar los coeficientes asociados a cada base de datos se utilizaron interacciones con una variable binaria que tomaba el valor 1 en caso de que la observación proviene de la base de TBN y 0 si proviene de ESI.

Para los 6 modelos estimados y los 4 niveles educativos (omitiendo la educación básica, que se designó arbitrariamente como categoría base) se probó la hipótesis nula de que la brecha de retorno de dichos niveles en relación a la educación básica son iguales para ambas bases de datos. Es decir, se desarrollaron 24 tests con una única restricción lineal cada uno. Los resultados de estos tests se muestran en la tabla 5.2. En todos ellos se rechaza la hipótesis nula con un nivel de significancia del 1 %. Esto permite concluir que hay diferencias estadísticamente significativas para los retornos asociados a todos los niveles educativos entre ambas bases de datos. En la tabla 1.8 se observa que los retornos porcentuales sobre los salarios de la educación media científico humanista (por sobre la educación básica, designada arbitrariamente como categoría base), estimados mediante el modelo (6) con datos de TBN y ESI, son respectivamente 11,1 % y 14,2 % para la educación media científico humanista; 19,5 % y 8,3 % para la media técnico profesional; 48,4 % y 28,6 % para la superior técnico profesional y, finalmente, 97,1 % y 84,1 % para la universitaria .

Tabla 1.8: Retornos estimados de los distintos niveles educacionales, obtenidos mediante una regresión que combina datos de ESI y TBN.

Regresores	(1)	(2)	(3)	(4)	(5)	(6)
Media CH (TBN)	0.190***	0.0889***	0.0823***	0.106***	0.101***	0.111***
Media CH (ESI)	0.140***	0.170***	0.141***	0.151***	0.149***	0.142***
Media TP (TBN)	0.416***	0.221***	0.183***	0.196***	0.185***	0.195***
Media TP (ESI)	0.0892***	0.117***	0.0837***	0.0904***	0.0857***	0.0825***
Superior TP (TBN)	0.790***	0.531***	0.479***	0.494***	0.475***	0.484***
Superior TP (ESI)	0.320***	0.341***	0.298***	0.301***	0.296***	0.286***
Universitaria (TBN)	1.325***	1.010***	0.954***	0.979***	0.958***	0.971***
Universitaria (ESI)	0.845***	0.885***	0.847***	0.855***	0.848***	0.841***
Experiencia		✓	✓	✓	✓	✓
Industria			✓	✓	✓	✓
Tamaño				✓	✓	✓
Región					✓	✓
Año						✓
Observaciones	581,016	580,998	580,998	578,579	578,579	578,579
R^2	0.281	0.290	0.322	0.328	0.333	0.344

Nota: La variable dependiente de los modelos es el logaritmo de los salarios reales por hora. Los datos de TBN corresponden a la información ingresada por los empleadores a nivel de vacantes (no avisos), mientras que los de ESI son información declarada por encuestados que representan al flujo ponderado (se consideran los factores de expansión de la encuesta). Los estimadores de los retornos asociados a los distintos niveles educativos así como los de las variables de control (experiencias y sus potencias de hasta orden 4, industrias, regiones y años) son obtenidos separadamente para cada base de datos mediante el uso de interacciones con una variable binaria que indica a qué base de datos corresponde cada observación. Se muestran las estimaciones de retornos para cada nivel educacional, excepto la educación básica (que fue designada arbitrariamente como categoría base).

Aunque los datos del sitio web fuesen representativos de la economía chilena, estos retornos no tendrían por qué parecerse. La razón de lo anterior es que las estimaciones se obtienen en base a información de diferente naturaleza. Como se discutió en la introducción, los salarios reportados mediante datos de encuestas y censos dependen de habilidades y características no observables de los trabajadores, mientras que los datos del sitio web permiten cuantificar las valoraciones que tienen los empleadores por ciertos atributos requeridos por ellos mismos. En consecuencia, es de suponerse que las estimaciones obtenidas con los datos de TBN estén menos sesgadas que las de ESI, pues estas últimas se verían afectadas por omisión de variables relevantes en la determinación de salarios.

A pesar de las diferencias estadísticamente significativas entre los coeficientes de interés para cada fuente de datos, existen claras similitudes entre dichos retornos y es posible identificar ciertos patrones entre ellos. Observando el modelo (6) de la tabla 1.8 es posible notar que el retorno del nivel medio científico humanista estimado con datos de la ESI supera al estimado con datos de TBN por apenas un 3,1%. En los otros niveles educacionales son mayores los retornos estimados con datos de TBN y las diferencias oscilan entre un 11,3% (media técnico profesional) y un 19,8% (superior técnico profesional).

Cabe destacar que, al momento de enumerar los niveles educacionales según sus respectivos

retornos, se observan órdenes similares para ambas fuentes de datos, consistentes entre todos los modelos reportados en la tabla 1.8 e intuitivos (los niveles educacionales más avanzados suelen reportar retornos más altos). Como era de esperarse, todos los retornos estimados son positivos por lo que el nivel base (educación básica o primaria) es el que efectivamente evidencia menores salarios. La única diferencia entre los resultados de TBN y ESI radica en que en el sitio web tener educación media técnico profesional resulta ser mejor retribuido que tener educación media científico humanista, mientras que según los datos de la ESI esta realidad se invierte.

Adicionalmente, se desarrollaron regresiones lineales por separado para cada base de datos. La finalidad de hacer esto es comparar nuevamente los retornos de la educación considerando información no homologada entre ambas bases que podrían aportar a una estimación más precisa de los retornos de la educación. Ejemplos claros son las categorías de tamaño de la empresa (que, como se explicó anteriormente, difieren entre ambas empresas) y las regiones (cuya información parece ser más precisa en la base de datos de ESI, pues las de TBN corresponden a las ubicaciones de las casas matrices de las empresas). Una limitación de este procedimiento es no poder realizar computacionalmente los tests antes mencionados. Los resultados de estas regresiones separadas se muestran en la tabla 1.9.

Tabla 1.9: Retornos estimados de los distintos niveles educacionales, obtenidos mediante una regresiones separadas para TBN y ESI.

	TBN	ESI
Media CH	0.110***	0.144***
Media TP	0.184***	0.0855***
Superior TP	0.480***	0.287***
Universitaria	0.959***	0.842***
Observaciones	502,537	59,047
R^2	0.649	0.329

Nota: La variable dependiente de los modelos es el logaritmo de los salarios reales por hora. Los datos de TBN corresponden a la información ingresada por los empleadores a nivel de vacantes (no avisos), mientras que los de ESI son información declarada por encuestados que representan al flujo ponderado (SE consideran los factores de expansión de la encuesta). Los estimadores son obtenidos separadamente para cada base de datos mediante el uso de interacciones con una variable binaria que indica a qué base de datos corresponde cada observación. Se muestran las estimaciones de retornos para cada nivel educacional, excepto la educación básica (que fue designada arbitrariamente como categoría base).

A pesar de que en esta nueva regresión en general los retornos de los distintos niveles educacionales se parecen más entre las distintas bases de datos, los resultados y las brechas siguen siendo muy similares a las de los modelos presentados en la tabla 1.8. Por lo tanto, las conclusiones que se pueden obtener de la tabla 1.9 no difieren considerablemente de las antes expuestas.

Como se mencionó al comienzo de esta subsección, algunas posibles explicaciones de las diferencias entre los retornos estimados con ambas bases de datos son: (1) las diferencias entre las naturalezas de ambas bases de datos (TBN contiene información de la demanda de trabajo, mientras que ESI contiene datos efectivos que declaran los encuestados); (2) los eventuales sesgos generados por las habilidades no observables de los trabajadores encuestados en la ESI y (3) el hecho de que los datos de TBN tienen una sobrerrepresentación de la Región Metropolitana y algunos sectores industriales, como se explicó anteriormente.

1.3.9. Empresas (TBN)

En la muestra del estudio hay un total de 9.717 firmas. Cada una de ellas cuenta con un identificador único que permite distinguirla. A partir de estos identificadores se crearon 9.717 variables binarias que permiten estimar efectos fijos de empresas. A partir de estos efectos fijos se construyeron gráficos de sus distribuciones por nivel educacional. Dichos gráficos permiten comparar los retornos a nivel de firmas para los distintos niveles educacionales, lo cual ayuda a deducir cuáles de ellos están ligados a empresas que suelen remunerar más o menos. Al controlar por estas variables es posible aislar el retorno a la educación que se manifiesta a través del acceso a empresas que suelen pagar más a sus trabajadores. Mediante la descomposición planteada por Gelbach (2016) [11], estas variables permitirán cuantificar aquel componente del retorno a la educación en los salarios que se manifiesta a través del acceso a firmas que suelen presentar mayores remuneraciones.

Las firmas presentes en esta muestra son variadas en cuanto a sus tamaños, las industrias a las que se dedican, las regiones en las que se ubican y los tipos de cargos que buscan llenar. A su vez, esto genera variabilidad en los requerimientos y ofertas de salarios que ingresan para cada vacante de trabajo que publican.

Las empresas presentes que han publicado solo un aviso de trabajo se excluyen, ya que no es posible estimar sus efectos fijos. Las empresas que han publicado al menos dos avisos son 5.042 y las vacantes asociadas a ellas son 502.447 (de un total de 574.362).

Una pregunta que naturalmente puede surgir es si las muestras de empresas con al menos un y al menos dos avisos publicados varían considerablemente en ciertas características que pudiesen influir en las estimaciones. Algunas variables que podrían ser relevantes son las industrias a las que se dedican, las regiones en las que operan las casas matrices y sus tamaños (medidos en número de trabajadores). En las tablas 5.3, 5.4 y 5.5 es posible observar que no existen diferencias considerables en las distribuciones porcentuales de las empresas en torno a estas variables, por lo que la selección de empresas descrita anteriormente pareciera no ser un problema mayor.

1.3.10. Palabras presentes en las descripciones de los cargos (TBN)

Al momento de publicar los avisos de empleo, los empleadores tienen la posibilidad de escribir un título del cargo correspondiente. Aprovechando el trabajo realizado con herramientas simples de minería de texto por parte de Banfi y Villena-Roldán (2019) [16] se obtuvieron las cuatro primeras palabras más significativas (sustantivos o adjetivos) en los títulos de cada aviso, con el fin de desarrollar un procedimiento similar al de Marinescu y Wolthoff (2015) [13]. De un total de 447 palabras distintas, se creó el mismo número de variables binarias, las cuales indicaban si una determinada palabra estaba o no presente entre las cuatro primeras significativas de cada título. Dado que estos títulos frecuentemente identifican cargos específicos, en lo restante de este documento, se referirá a estos como “cargos”.

Todas las palabras referentes a cargos fueron normalizadas a minúsculas y se eliminaron las tildes y los caracteres no alfabéticos. Como muchas de estas palabras no son neutrales en cuanto al género por tratarse del idioma español, estas han sido neutralizadas de manera que no se perciba diferencia entre aquellas que apelan al género femenino y masculino. Un inconveniente de este procedimiento es la posible pérdida de información, ya que el empleador podría definir un género deseado para el o la postulante. En la tabla 1.10 se muestra un listado de las 100 palabras más frecuentes en las descripciones de las vacantes publicadas dentro de la muestra considerada para este estudio, una vez hechos los arreglos descritos.

Al igual que con las variables binarias correspondientes a las distintas empresas, se han estimado los efectos fijos correspondientes a la presencia de cada palabra en la descripción de los cargos de los avisos de trabajo publicados en la plataforma de empleo. La importancia de estas variables es que permiten aislar el retorno intrínseco de recibir un cierto grado de educación de aquel que se manifiesta a través del acceso a cargos de mayor remuneración. Mediante la descomposición de Gelbach es posible atribuir una parte del retorno a la educación a la posibilidad de acceso a cargos específicos.

Tabla 1.10: Ránking de las 100 palabras más frecuentes en títulos de trabajo del sitio web.

N°	Palabra	Vacantes	N°	Palabra	Vacantes
1	asistente	10.734	51	rrhh	883
2	ejecutiv	10.382	52	remuneraciones	881
3	vendedor	8.002	53	time	880
4	analista	7.269	54	humanos	861
5	venta	6.491	55	profesional	848
6	administrativ	5.473	56	tienda	828
7	tecnico	5.224	57	diseñador	813
8	contable	4.256	58	asesor	787
9	atencion	4.247	59	programa	747
10	promotor	3.871	60	psicolog	738
11	jefe	3.852	61	informatico	732
12	ing	3.818	62	repcionista	730
13	comercial	3.769	63	administracion	705
14	secretari	3.618	64	iquique	688
15	cliente	3.573	65	proyecto	683
16	encargad	3.308	66	contabilidad	673
17	contador	2.606	67	reemplazo	659
18	auxiliar	2.478	68	vendedores	659
19	seguridad	2.324	69	cocina	653
20	supervisor	2.250	70	chofer	649
21	practica	2.249	71	gestion	634
22	centro	2.185	72	personal	627
23	full	2.185	73	manager	624
24	bodega	1.972	74	mantencion	621
25	guardia	1.963	75	cuenta	617
26	terreno	1.600	76	ordenador	597
27	operari	1.536	77	seguros	595
28	ayudante	1.468	78	calidad	581
29	junior	1.391	79	santiago	577
30	area	1.294	80	electrico	576
31	control	1.286	81	antofagasta	571
32	operador	1.274	82	agente	569
33	aseo	1.266	83	desarrollador	565
34	servicio	1.251	84	consultor	547
35	soporte	1.248	85	reponedor	542
36	cajero	1.196	86	gerencia	532
37	obra	1.168	87	senior	522
38	call	1.145	88	jornada	520
39	produccion	1.108	89	alumn	512
40	repcion	1.068	90	ejecutivo	501
41	pre	1.065	91	prog	497
42	center	1.039	92	maestro	473
43	administrativa	1.021	93	web	453
44	coordinador	1.014	94	marketing	448
45	general	988	95	linea	447
46	farmaceutico	962	96	prevencionista	446
47	administrador	942	97	trainee	445
48	auditor	937	98	cajer	444
49	mecanico	936	99	sistema	444
50	curso	903	100	dibujante	442

Nota: Estas palabras son únicamente sustantivos y adjetivos homogeneizados (se neutralizó el género de muchas palabras, se cambiaron mayúsculas por minúsculas y se eliminaron tildes). No se consideraron artículos, conectores y otras palabras que no aportan mayormente a la caracterización de los cargos. Este ránking se basa únicamente en la muestra utilizada en este estudio.

Capítulo 2

Distribuciones de efectos fijos: análisis exploratorio

2.1. Políticas salariales de las empresas

El objetivo de esta sección es dar a entender cómo difieren las políticas salariales de las empresas que publican vacantes para los distintos niveles educacionales requeridos. Naturalmente, algunas preguntas que pueden surgir son: ¿existen diferencias sistemáticas entre las políticas salariales de empresas que suelen requerir determinados niveles educacionales para sus vacantes?, ¿qué niveles educacionales son requeridos por empresas con políticas salariales más generosas? o ¿hay niveles educacionales en los que las empresas tiendan a ofrecer salarios de manera más parecida que en otros?

En las figuras 2.1 y 2.2 se muestran gráficamente tres versiones de distribuciones de efectos fijos de empresas sobre los logaritmos de los salarios por nivel educacional. La versión E1 corresponde a la distribución obtenida cuando se regresiona el logaritmo del salario real por hora ofrecido por las empresas en sus vacantes de trabajo únicamente en función de la experiencia (y sus potencias hasta el grado 4) y los efectos fijos de empresa. Progresivamente, las versiones E2 y E3 incluyen efectos fijos por industrias y por palabras presentes en las descripciones de los cargos de trabajo. Se muestran las tres versiones para todos los niveles educacionales. Como se puede observar, la educación básica tiene sus distribuciones de efectos fijos de empresas muy condensadas cerca de sus valores esperados (-0,36, -0,33 y -0,14 respectivamente). Esto genera un problema de escala que dificulta la visualización de las otras curvas, que se muestran de mejor manera en la figura 2.2, que omite el nivel básico.

Las distribuciones de efectos fijos de las figuras 2.1 y 2.2 revelan las diferencias de políticas salariales entre firmas que publican vacantes para un mismo nivel educacional, así como las diferencias sistemáticas de dichas políticas salariales de empresas que existen entre los distintos niveles educacionales. Para analizar el comportamiento de las empresas dentro de un mismo nivel educacional es necesario observar cada una de las curvas correspondientes a cada nivel por separado. Por otro lado, para estudiar las diferencias entre niveles educacionales es pertinente concentrarse en cómo varían las distintas curvas asociadas a los diversos niveles.

Los efectos fijos de las empresas se normalizan de tal modo que la suma de ellos sea cero. Por ende, un efecto fijo positivo representa una política salarial de una empresa que suele pagar más que el promedio de los empleadores de que publican vacantes en el sitio web, controlando por las variables correspondiente a las versiones E1, E2 y E3 ya detalladas.

En la tabla 2.1. en primer lugar, se destaca que las vacantes de empleo destinadas a trabajadores con niveles educacionales más avanzados están relacionadas en general a empresas que ofrecen salarios más altos. Concretamente, se evidencia que las medias de los efectos fijos asociados a la educación básica, media científico humanista, media técnico profesional y superior técnico profesional son negativas, variando entre -0,22 y -0,04 en la versión E3. Avisos que requieren estos niveles se asocian a empresas que suelen pagar menos que el promedio. Por otro lado, las medias de los efectos fijos de la educación universitaria son siempre positivas, por lo que estarían ligadas a empresas que suelen tener políticas salariales sobre el promedio.

En segundo lugar, se observa que la dispersión de los efectos fijos es mayor para los niveles educacionales más avanzados. La tabla 2.1 muestra que las varianzas asociadas a las distribuciones de efectos fijos de empresas para los niveles superior técnico profesional y universitario son al menos 10 veces mayores que las reportadas para la educación básica. Esto denota que las empresas que contratan trabajadores menos calificados tienen políticas salariales parecidas, mientras que las que contratan a trabajadores más educados ofrecen salarios más heterogéneos. Una posible explicación es que los empleadores suponen que el nivel de productividad que puedan manifestar los trabajadores en empleos de baja especialización es relativamente invariante si se compara con aquellos que requieren un mayor nivel educacional. En otras palabras, en los trabajos con mayores requerimientos educacionales (como cargos de gerencia o mayor responsabilidad) hay un mayor margen para tener un desempeño bueno o malo que en cargos con bajas exigencias de formación (como operarios o trabajadores de aseo). Por lo tanto, cuando se trata de trabajos que requieren poca educación, los empleadores considerarían que la mano de obra es relativamente homogénea y no tendrían mayores incentivos para diferenciarse en cuanto a los salarios que ofrecen. En el caso contrario, dichos incentivos sí existen, pues algunas empresas estarían dispuestas a ofrecer salarios más altos por trabajadores considerablemente más productivos.

Cardoso, Guimaraes, Portugal & Reis (2018) [14] reportan gráficos similares para efectos fijos asociados a firmas, trabajadores y títulos de trabajo. Al analizar los efectos fijos de los trabajadores para distintos niveles educacionales (primario, secundario y superior) señalan que “la dispersión de las habilidades de los trabajadores es considerablemente mayor entre los graduados universitarios que entre los otros niveles escolares”. Esta afirmación parece coherente con la conclusión expuesta recientemente. En mercados laborales con mayores requerimientos educacionales la productividad de los trabajadores tiende a variar más, los empleadores así lo creen y manifiestan estas creencias mediante sus políticas salariales, las cuales son relativamente heterogéneas. Por el contrario, en mercados laborales con menores requerimientos educacionales los trabajadores no se diferencian considerablemente en sus productividades. Si los empleadores creen esto, supondrán que la calidad de la fuerza de trabajo (y, por lo tanto, la productividad asociada a ella) no variará mucho entre un trabajador y otro. En consecuencia, los empleadores no tendrían suficientes incentivos a ofrecer salarios considerablemente mayores que los que ofrecen las otras empresas ni deberían poder ofrecer

salarios muy bajos debido a la existencia del sueldo mínimo.

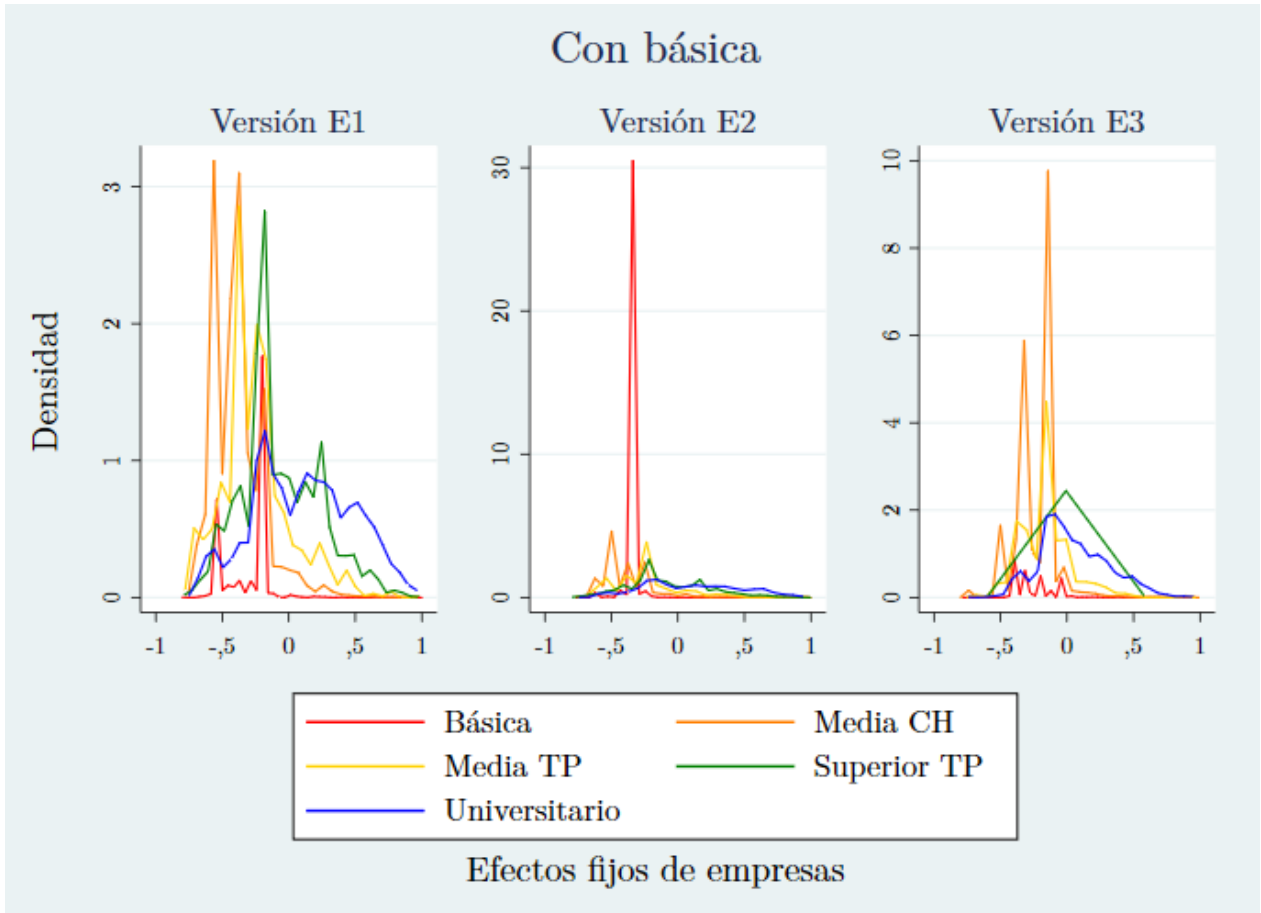
De manera complementaria al estudio de las varianzas, se puede notar que las curtosis tienen un comportamiento análogo que permite complementar el relato del párrafo anterior. Una curtosis grande implica que los efectos fijos se concentran más cerca de la media y las colas de ella. Se observa que en todos los casos las varianzas y las curtosis se mueven en sentidos opuestos. La educación básica presenta la menor varianza (menor dispersión) y mayor curtosis (mayor concentración en la media y los extremos), mientras que en la educación universitaria y de postgrado pasa lo opuesto.

En tercer lugar, cabe destacar que en su mayoría las distribuciones de efectos fijos presentan asimetría positiva. Gráficamente, esto quiere decir que las colas derechas de las distribuciones tienden a ser más largas que las colas izquierdas. Una asimetría positiva implica que la mayoría de las vacantes del nivel educacional observado corresponden a empresas cuya política salarial es menos generosa que el promedio de dicho nivel. Un ejemplo claro de asimetría positiva es la de la educación media científico-humanista (con valores entre 1,63 y 2,17). La distribución asociada a esta categoría es la que presenta una cola derecha más larga, siendo muy probable que una vacante elegida arbitrariamente corresponda a una empresa con políticas salariales menos generosas que las que en promedio requieren nivel universitario.

Tras analizar las distribuciones de los efectos fijos de empresas, puede resultar interesante preguntarse qué es lo que caracteriza a aquellas firmas con políticas salariales más y menos generosas. Algunas características, de las que se tiene información, que podrían ser relevantes y que no han sido consideradas en los modelos mediante los cuales se obtuvieron los gráficos y estadísticos recientemente expuestos son los tamaños de las empresas (medidos en números de empleadores) y las industrias a las cuales se dedican. A continuación se desarrolla un breve apartado en el cual se analizan estos factores a partir de algunas estimaciones. Es posible concluir que las empresas más generosas en términos salariales suelen dedicarse a industrias como la minería y restaurantes y hotelería, además de coincidir con aquellas que tienen menos trabajadores (destacan las que tienen menos de 150) o las que se encuentran en la categoría de entre 1.001 y 5.000 empleados. Por otro lado, las empresas menos generosas se caracterizan por tener una cantidad moderadamente alta de trabajadores (destacan aquellas que tienen entre 301 y 1.000 trabajadores).

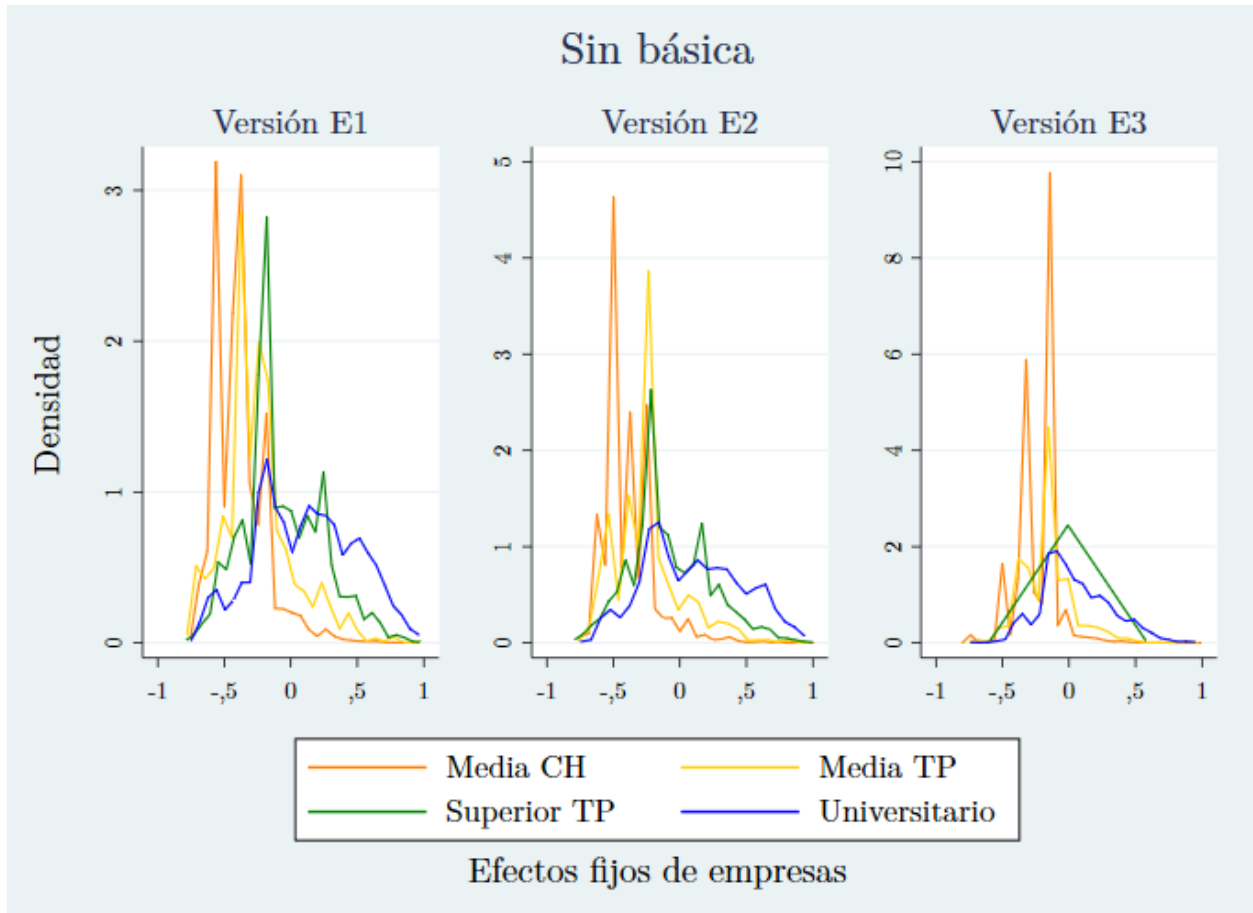
Para poder desarrollar el análisis antes descrito, fue necesario estimar nuevamente los efectos fijos de empresas. La versión E3 permitía obtener dichas estimaciones a través de una regresión que buscaba explicar los logaritmos de los salarios reales por hora en función de la experiencia requerida por los empleadores (y sus potencias de hasta orden 4), las industrias y los títulos de trabajo. En dicha especificación se habían omitido los niveles educacionales, ya que se condicionó en ellos, justamente porque es de interés de este estudio entender las diferencias sistemáticas que hay entre los efectos fijos asociados a cada categoría educacional. En esta ocasión, se incorporan los niveles educacionales al modelo. El motivo de ello es evitar que estas diferencias sistemáticas sesguen las estimaciones de los parámetros de interés. Luego, a cada aviso se le asignó el efecto fijo asociado a la empresa que lo publicó. Finalmente, se regresionaron estos efectos fijos en función de la industria y los tamaños (en número de trabajadores) que describen a las empresas que publicaron cada vacante. Los resultados de estas estimaciones se pueden observar en la tabla 5.6. Al ser las vacantes las unidades de

Figura 2.1: Distribución de efectos fijos de empresas para cada nivel educacional.



Nota: Funciones de densidad de probabilidades, obtenidas mediante estimaciones kernel no-paramétricas, de los efectos fijos de empresas, estimados mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real por hora. La versión E1 usa como variables de control la experiencia requerida por los empleadores y sus potencias de hasta grado 4. La versión E2 adicionalmente agrega efectos fijos por industrias. La versión E3 agrega además efectos fijos por palabras presentes en las descripciones de los cargos. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a empresas con políticas salariales más/menos generosas que el promedio.

Figura 2.2: Distribución de efectos fijos de empresas para cada nivel educacional (sin básica).



Nota: Para una mejor visualización de las curvas asociadas a otros niveles educacionales se omite la educación básica. Funciones de densidad de probabilidades, obtenidas mediante estimaciones kernel no-paramétricas, de los efectos fijos de empresas, estimados mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real por hora. La versión E1 usa como variables de control la experiencia requerida por los empleadores y sus potencias de hasta grado 4. La versión E2 adicionalmente agrega efectos fijos por industrias. La versión E3 agrega además efectos fijos por palabras presentes en las descripciones de los cargos. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a empresas con políticas salariales más/menos generosas que el promedio.

Tabla 2.1: Estimaciones de los 4 primeros momentos de distribuciones empíricas de efectos fijos de empresas.

Nivel educacional	Momento estimado	Versión		
		E1	E2	E3
Básica	Esperanza	-0,36	-0,33	-0,14
	Varianza	0,01	0,01	0,00
	Asimetría	2,18	1,86	-0,03
	Curtosis	22,75	41,53	46,91
Media CH	Esperanza	-0,37	-0,36	-0,22
	Varianza	0,04	0,04	0,03
	Asimetría	2,02	2,17	1,63
	Curtosis	12,88	16,12	17,67
Media TP	Esperanza	-0,25	-0,26	-0,14
	Varianza	0,07	0,07	0,04
	Asimetría	0,88	1,02	0,54
	Curtosis	5,09	5,79	9,06
Superior TP	Esperanza	-0,06	-0,08	-0,04
	Varianza	0,10	0,10	0,27
	Asimetría	0,52	0,57	43,18
	Curtosis	3,62	3,70	2.297,72
Universitaria	Esperanza	0,16	0,16	0,11
	Varianza	0,19	0,18	0,14
	Asimetría	0,46	0,49	1,53
	Curtosis	2,99	2,80	6,03

Nota: Estos efectos fijos se obtienen mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real. La versión E1 usa como variables de control la experiencia requerida por los empleadores y su factor cuadrático. La versión E2 adicionalmente agrega efectos fijos por industrias. La versión E3 agrega además efectos fijos por palabras presentes en las descripciones de los cargos. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a empresas con políticas monetarias más/menos generosas que el promedio.

análisis de la regresión, es lógico que las empresas que más vacantes han publicado son las que tienen más peso en las estimaciones.

En la figura 5.1 se puede observar un gráfico que muestra a qué industrias están asociadas las empresas que reportan políticas salariales altas y bajas. En orden descendente, se tiene que –en promedio– las industrias cuyas empresas son relativamente generosas en términos salariales son las de minería; restaurantes y hotelería; pesca; construcción; inmobiliaria y administración pública. Posteriormente se encuentra la industria de agricultura y silvicultura, que ha sido elegida arbitrariamente como categoría base. Manteniendo el orden descendente, pero esta vez observando a las industrias cuyas empresas –en promedio– tienen políticas salariales relativamente bajas es posible enumerar a las de: electricidad, agua y gas; social y salud; financiera; educación; manufactura; comercio; otros servicios; y transporte y comunicaciones.

En la figura 5.2 se muestra un gráfico parecido al anterior, pero enfocado esta vez en los

tamaños de las empresas (medidos en números de trabajadores). Según se observa en este gráfico, en general las empresas con políticas salariales ofrecidas más generosas son las que tienen menos trabajadores. En primer lugar se tienen aquellas con entre 1 y 10 empleados. En segundo lugar se tienen aquellas empresas con entre 11 y 150. Posteriormente, y alterando un poco el patrón, se encuentran las empresas con entre 1.001 y 5.000 trabajadores. Luego se encuentran, en el siguiente orden, las empresas con entre 151 y 300 trabajadores, con más de 5.000, con entre 301 y 500 y con entre 501 y 1.000. A grandes rasgos, se puede decir que las empresas de mayores tamaños remunerar menos, sin embargo esto cambia drásticamente cuando se comparan las empresas que tienen más y menos de 1.000 trabajadores. Es posible que la muestra de *trabajando.com* tenga una composición no representativa de la población de empresas chilenas, lo que podría dar cuenta de este resultado particular.

2.2. Rentabilidades salariales de los cargos

Análogamente al caso de las empresas, es posible estudiar las distribuciones de los efectos fijos de los cargos y plantearse una serie de preguntas: ¿existen diferencias sistemáticas entre las rentabilidades de los cargos de las vacantes que requieren determinados niveles educacionales?, ¿qué niveles educacionales están asociados a cargos más rentables? o ¿hay niveles educacionales en los que los cargos tiendan presenten rentabilidades salariales similares? Nuevamente se intenta responder a preguntas como éstas, intentando hallar las razones de los fenómenos que quedan en evidencia a partir de las distribuciones de efectos fijos.

Nuevamente, se consideraron tres versiones de regresiones para las estimaciones de efectos fijos. En la versión C1 sólo se estiman los logaritmos de los salarios reales por hora en función de las cuatro primeras potencias de la experiencia requerida, mientras que en las versiones C2 y C3 se incorporan progresivamente los efectos fijos de industrias y los de empresas. La figura 2.3 muestra las tres versiones para todos los niveles educacionales. Otra vez, la educación básica tiene sus distribuciones de efectos fijos muy condensadas cerca de sus valores esperados (entre -0,39, -0,37 y -0,30 respectivamente). Esto genera un problema de escala que dificulta la visualización de las otras curvas, que se muestran de mejor manera en la figura 2.4, que omite el nivel básico.

Una vez más, se impuso que la suma de los efectos fijos de las palabras de los cargos sean cero, lo que permite asociar los valores positivos y negativos a palabras que tienden a estar sobre o bajo el promedio de las que están contenidas en las descripciones de los cargos, respectivamente. A continuación se detallan interpretaciones que surgen a partir de las figuras 2.3 y 2.4 y los momentos de las distribuciones empíricas de efectos fijos que se reportan en la tabla 2.2.

Como es de esperarse, los cargos que suelen estar destinados niveles educacionales más avanzados suelen estar mejor remunerados. En efecto, las medias de efectos fijos relacionados a cargos destinados a personas con educación básica, media científico humanista y media técnico profesional son negativas (-0,30, -0,11 y -0,04 respectivamente en la versión C3), mientras que las de educación superior técnico profesional y universitaria son positivas (0,11 y 0,37 respectivamente).

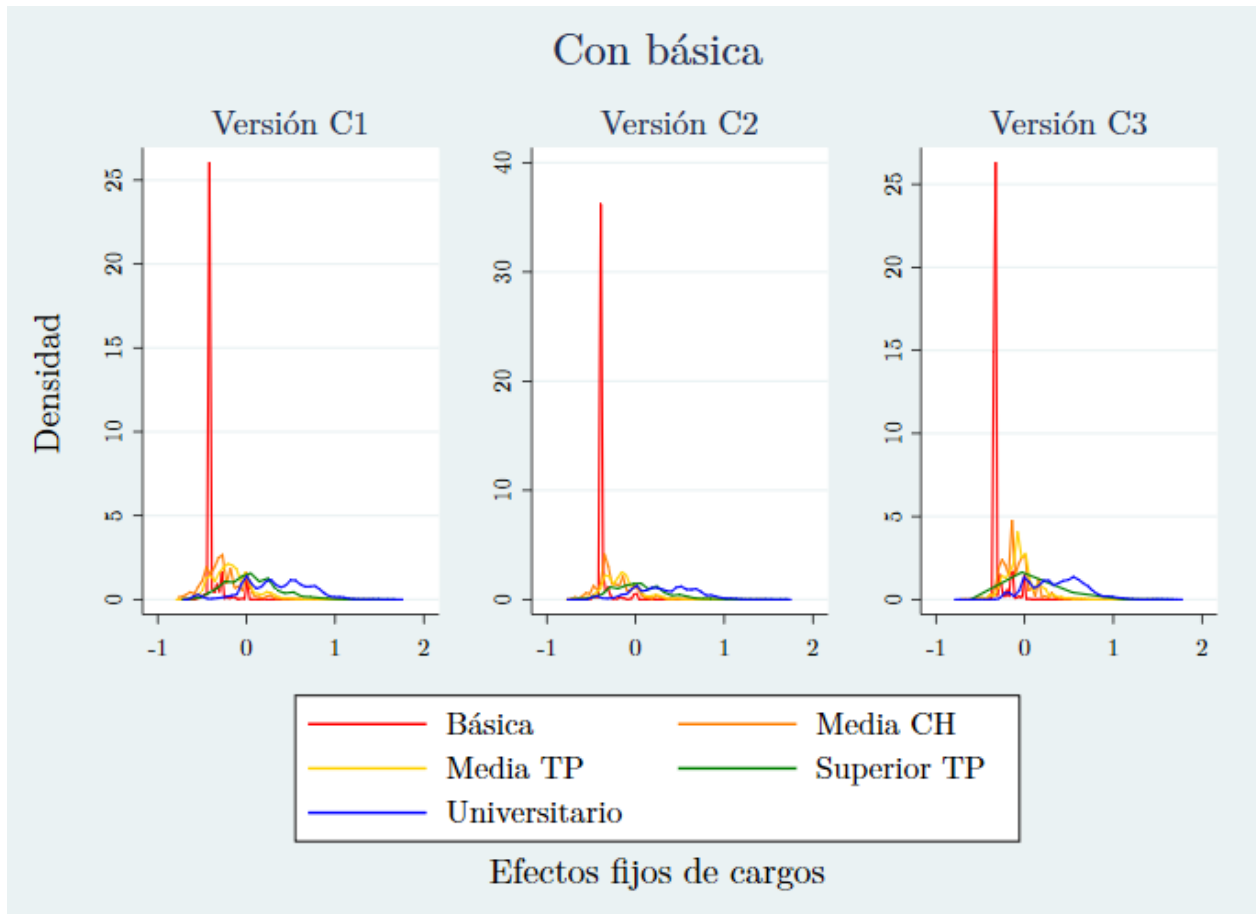
Al igual que como sucede con las empresas, la dispersión de los efectos fijos de palabras

contenidas en las descripciones de los cargos es creciente a medida que se observan niveles educacionales de mayor especialización. La tabla 2.2 muestra que, por ejemplo, la varianza de la distribución de la educación univesitaria es entre cuatro y cinco veces mayor que la de educación media científico-humanista. Se infiere que los cargos que requieren postulantes menos educados tienen remuneraciones relativas más parecidas entre ellas, en comparación a los cargos que suponen una mayor especialización.

Una posible explicación es que los empleadores suponen que el nivel de productividad que puedan manifestar los trabajadores en empleos de baja especialización es relativamente invariante si se compara con aquellos que requieren un mayor nivel educacional. Se infiere que quienes tienen una formación educacional más avanzada –al tener productividades más heterogéneas, como indican Cardoso, Guimaraes, Portugal & Reis (2018) [14]– pueden acceder a cargos ofrecidos por empresas que difieren considerablemente en cuanto a sus remuneraciones, mientras que los distintos cargos destinados a postulantes con educación menos especializadas son más parecidos entre ellos en términos salariales. También se observa que las curtosis, a diferencia del caso de las empresas, no tienen un comportamiento monótono en relación a los niveles educacionales. Por lo tanto, no es posible sacar conclusiones generales respecto a la concentración de los efectos fijos cerca y lejos de la media de cada distribución.

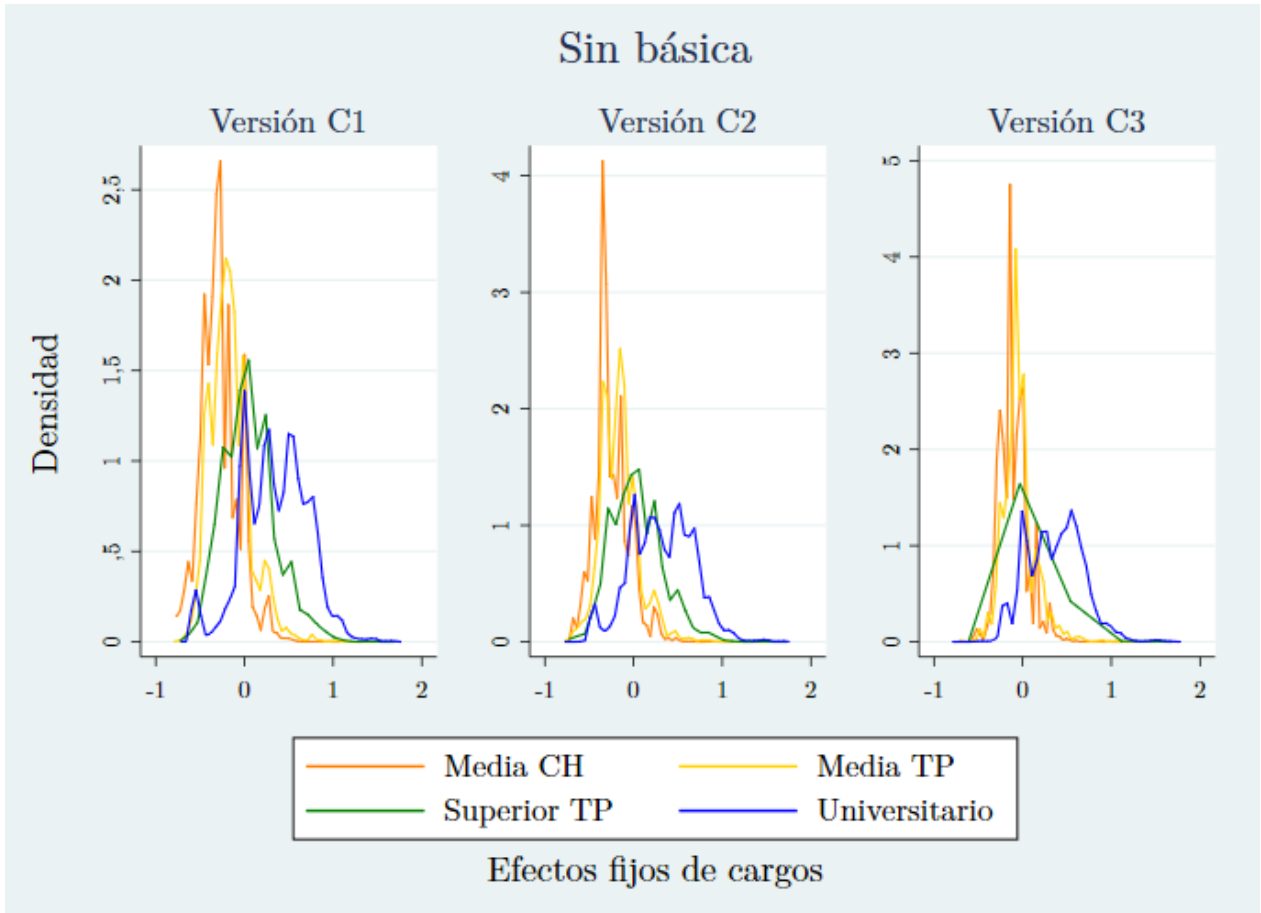
Casi todas las distribuciones observadas en la tabla 2.2 presentan asimetrías positivas. De aquí se concluye que suele ser más probable que, al seleccionar una vacante de manera aleatoria dentro de aquellas que requieren un determinado nivel educacional, su cargo de trabajo sea menos rentable que el promedio de dicho nivel. Las colas derechas de estas distribuciones evidencian que existe una minoría de cargos que son altamente remunerados y que, por lo tanto, elevan las medias estimadas. Generalmente la educación básica es el nivel que reporta las asimetrías más pronunciadas, por lo que es la categoría donde el fenómeno antes descrito es más notorio. En contraparte, niveles educacionales más avanzados suelen presentar distribuciones menos asimétricas. Por lo tanto, el fenómeno de aquellas minorías de cargos bien remunerados que elevan las medias estimadas es menos pronunciado para vacantes destinadas a postulantes con mayor educación formal.

Figura 2.3: Distribución de efectos fijos de cargos para cada nivel educacional.



Nota: Funciones de densidad de probabilidades, obtenidas mediante estimaciones kernel no-paramétricas, de los efectos fijos de palabras en títulos de trabajo, estimados mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real por hora. La versión C1 usa como variables de control la experiencia requerida por los empleadores y su factor cuadrático. La versión C2 adicionalmente agrega efectos fijos por industrias. La versión C3 agrega además efectos fijos por empresas. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a cargos de mayor/menor remuneración que el promedio.

Figura 2.4: Distribución de efectos fijos de cargos para cada nivel educacional (sin básica).



Nota: Para una mejor visualización de las curvas asociadas a otros niveles educacionales se omite la educación básica. Funciones de densidad de probabilidades, obtenidas mediante estimaciones kernel no-paramétricas, de los efectos fijos de palabras en títulos de trabajo, estimados mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real por hora. La versión C1 usa como variables de control la experiencia requerida por los empleadores y su factor cuadrático. La versión C2 adicionalmente agrega efectos fijos por industrias. La versión C3 agrega además efectos fijos por empresas. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a cargos de mayor/menor remuneración que el promedio.

Tabla 2.2: Estimaciones de los 4 primeros momentos de distribuciones empíricas de efectos fijos de cargos.

Nivel educacional	Momento estimado	Versión		
		C1	C2	C3
Básica	Esperanza	-0,39	-0,37	-0,30
	Varianza	0,01	0,01	0,01
	Asimetría	3,37	3,90	2,69
	Curtosis	15,94	20,41	10,80
Media CH	Esperanza	-0,28	-0,25	-0,11
	Varianza	0,04	0,04	0,02
	Asimetría	0,56	0,68	0,73
	Curtosis	3,97	4,05	4,70
Media TP	Esperanza	-0,16	-0,16	-0,04
	Varianza	0,05	0,05	0,03
	Asimetría	0,85	0,79	0,99
	Curtosis	4,68	4,89	6,17
Superior TP	Esperanza	0,05	0,04	0,11
	Varianza	0,09	0,08	0,27
	Asimetría	0,52	0,49	-38,33
	Curtosis	3,55	3,50	1.967,56
Universitaria	Esperanza	0,35	0,33	0,37
	Varianza	0,13	0,12	0,10
	Asimetría	-0,13	0,03	0,23
	Curtosis	3,05	2,90	3,02

Nota: Estos efectos fijos se obtienen mediante distintas regresiones que tienen como variable dependiente el logaritmo del salario real. La versión C1 usa como variables de control la experiencia requerida por los empleadores y su factor cuadrático. La versión C2 adicionalmente agrega efectos fijos por industrias. La versión C3 agrega además efectos fijos por empresas. En todas las regresiones se impuso como restricción que la suma de los efectos fijos estimados sean cero, por lo que los efectos fijos positivos/negativos corresponden a cargos de mayor/menor remuneración que el promedio.

Capítulo 3

Metodología

A continuación se describen aspectos relevantes de las estimaciones de Mincer realizadas, tanto para estimar los retornos a la educación como para desglosar estos mismos más adelante –mediante la descomposición de Gelbach–. Las estimaciones que se presentan a partir de aquí se hacen únicamente en base a los datos de vacantes (no a nivel de avisos) de `www.trabajando.com`. Esto permite estimar de forma más precisa los retornos a la educación, pues dicha base de datos contiene información importante sobre las empresas que publican cada vacante y las palabras presentes en las descripciones de cada cargo que los empleadores han buscado rellenar. Esta información ha sido incorporada en las regresiones mediante efectos fijos, tal como se anticipó en capítulo 1. Este capítulo se desarrolla de la siguiente manera: primero se introduce la notación utilizada en los modelos, luego se describe el modelo general de regresión lineal a estimar (además de sus variaciones) y finalmente se muestran y se discuten los resultados de las regresiones estimadas.

3.1. Notación

La descomposición de Gelbach se basa en omitir intencionalmente algunos regresores con el fin de desmenuzar el sesgo generado por omisión de variables relevantes, atribuyendo aditivamente las componentes de este sesgo a grupos de dichos regresores omitidos. Por ello, es necesario distinguir los regresores de las especificaciones “base” o “incompleta” y la especificación “completa”, que incorpora los regresores omitidos en primera instancia. La matriz de regresores de la especificación base es \mathbf{X}_1 , mientras que la de la especificación completa es $\mathbf{X} \equiv [\mathbf{X}_1 \ \mathbf{X}_2]$. La especificación base incorpora los niveles educativos, la experiencia requerida por los empleadores –y sus potencias de hasta orden 4, teniendo en consideración uno de los resultados de Lemieux (2006) [8]– y los años en los que fueron publicados los distintos avisos. Por su parte, la especificación completa, además, incorpora las variables binarias asociadas a las empresas que publican cada aviso y a las palabras relevantes presentes en las descripciones de los cargos. El vector asociado a la variable explicada (logaritmos de los salarios reales por hora) es notado como \mathbf{Y} .

3.1.1. Niveles educativos

Los niveles educativos se clasifican en 5 categorías: educación básica, media científico humanista, media técnico profesional, superior técnico profesional y universitaria (agrupa pregrado y postgrado). Para evitar multicolinealidad, se ha elegido la educación básica como categoría base y se ha omitido del modelo general a estimar. Se define la matriz de niveles educativos distintos del básico como $\mathbf{E} = [\mathbf{E}_{\text{MCH}} \mathbf{E}_{\text{MTP}} \mathbf{E}_{\text{STP}} \mathbf{E}_{\text{U}}]$, la cual contiene 4 columnas de variables binarias que indican si cada vacante requiere o no cada nivel educativo. Los retornos a estimar de los respectivos niveles educativos son escritos como $\boldsymbol{\beta} = (\beta_{\text{MCH}}, \beta_{\text{MTP}}, \beta_{\text{STP}}, \beta_{\text{U}})$, utilizando el superíndice “*base*” cuando se trate de la especificación base y “*full*” cuando se trate de la especificación completa.

3.1.2. Experiencia

La experiencia requerida por los empleadores es una variable numérica y ha sido incorporada hasta sus potencias de hasta orden 4. El conjunto de estas 4 variables será notado matricialmente como $\mathbf{P} \equiv [\mathbf{P}^1 \mathbf{P}^2 \mathbf{P}^3 \mathbf{P}^4]$. Los coeficientes a estimar asociados a las distintas potencias de la experiencia son notados por ρ_p con $p \in \{1, 2, 3, 4\}$, usando nuevamente el superíndice “*base*” cuando se trate de la especificación base y “*full*” cuando se trate de la especificación completa.

3.1.3. Años

En la muestra considerada para este estudio se han tenido en cuenta los 7 años del período entre 2010 y 2016. Se ha elegido el año 2010 como categoría base y se ha omitido del modelo general a estimar. Estas 6 variables serán notadas como $\mathbf{A} = [\mathbf{A}_{2011} \dots \mathbf{A}_{2016}]$. Los coeficientes a estimar correspondientes a cada año se escriben como $\boldsymbol{\alpha} = (\alpha_{2011}, \dots, \alpha_{2016})$, agregando el superíndice “*base*” cuando se trate de la especificación base y “*full*” cuando se trate de la especificación completa.

3.1.4. Empresas

Como bien se dijo en el capítulo 1, la muestra de 574.362 vacantes de trabajo considerada para este estudio cuenta con un total de 9.717 empleadores. En las estimaciones que consideran efectos fijos de empresas se han excluido a aquellas firmas que solo han publicado un aviso. De este modo, el número de empresas consideradas en el análisis se reduce a 5.042, mientras que la cantidad de vacantes que permanecen en la muestra es de 502.447. Esto quiere decir que las estimaciones que incorporan dichos efectos fijos representan a casi el 87 % de las vacantes de trabajo en la muestra.

Se ha denotado mediante \mathbf{F} a la matriz que representa a las empresas que publican cada vacante. Esta matriz contiene 5.042 columnas de variables binarias, donde cada una de ellas toma el valor 1 cuando la empresa representada publicó una determinada vacante y 0 en caso contrario. Adicionalmente, se define mediante \mathcal{F} al conjunto de estas 5.042 empresas. Los efectos fijos a estimar son escritos como ϕ . Otra vez, se usa el superíndice “*base*” cuando se trate de la especificación base y “*full*” cuando se trate de la especificación completa.

3.1.5. Cargos

Se ha notado mediante \mathbf{W} a la matriz que representa a las palabras presentes en las descripciones de los cargos. Esta matriz contiene 447 columnas de variables binarias, donde cada una de ellas toma el valor 1 cuando la empresa representada publicó una determinada vacante y 0 en caso contrario. Adicionalmente, se define mediante \mathcal{W} al conjunto de estas 447 palabras. Los efectos fijos a estimar son escritos como $\boldsymbol{\omega}$. Nuevamente, se usa el superíndice “*base*” cuando se trate de la especificación base y “*full*” cuando se trate de la especificación completa.

3.2. Estimaciones de Mincer

A partir de la notación introducida anteriormente, se puede proponer la siguiente “especificación completa” del modelo regresión lineal general a estimar. Se normalizan los efectos fijos de empresas y cargos de modo que sumen cero, por lo que aquellos efectos fijos positivos describen a empresas y cargos que son más rentables que el promedio.

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\beta}_B^{full} + \mathbf{E}\boldsymbol{\beta}^{full} + \mathbf{P}\boldsymbol{\rho}^{full} + \mathbf{A}\boldsymbol{\alpha}^{full} + \mathbf{F}\boldsymbol{\phi}^{full} + \mathbf{W}\boldsymbol{\omega}^{full} + \boldsymbol{\varepsilon}^{full} \quad (3.1)$$

$$\begin{aligned} Y_i &= \beta_B^{full} + \mathbf{E}_i\boldsymbol{\beta}^{full} + \mathbf{P}_i\boldsymbol{\rho}^{full} + \mathbf{A}_i\boldsymbol{\alpha}^{full} + \mathbf{F}_i\boldsymbol{\phi}^{full} + \mathbf{W}_i\boldsymbol{\omega}^{full} + \varepsilon_i^{full} \\ &= \beta_B^{full} + \mathbf{E}_i\boldsymbol{\beta}^{full} + \mathbf{P}_i\boldsymbol{\rho}^{full} + \mathbf{A}_i\boldsymbol{\alpha}^{full} + \sum_{f \in \mathcal{F}} F_{if}\phi_f^{full} + \sum_{w \in \mathcal{W}} W_{iw}\omega_w^{full} + \varepsilon_i^{full} \end{aligned} \quad (3.2)$$

La normalización, matemáticamente, consiste en imponer las condiciones siguientes.

$$\sum_{f \in \mathcal{F}} \phi_f = 0 \quad \wedge \quad \sum_{w \in \mathcal{W}} \omega_w = 0 \quad (3.3)$$

Es posible elegir arbitrariamente una empresa y una palabra presente en descripciones de cargos y notarlas respectivamente como f_1 y w_1 . De este modo es posible reescribir las condiciones de la expresión 3.3 del siguiente modo.

$$\sum_{f \in \mathcal{F} \setminus \{f_1\}} \phi_f = -\phi_{f_1} \quad \wedge \quad \sum_{w \in \mathcal{W} \setminus \{w_1\}} \omega_w = -\omega_{w_1} \quad (3.4)$$

Desarrollando la ecuación 3.2 y utilizando las condiciones de la expresión 3.4, es posible expresar el modelo general con efectos fijos de empresas y cargos normalizados de la siguiente forma.

$$\begin{aligned}
Y_i &= \beta_B^{full} + \mathbf{E}_i \beta^{full} + \mathbf{P}_i \rho^{full} + \mathbf{A}_i \alpha^{full} + \sum_{f \in \mathcal{F} \setminus \{f_1\}} F_{if} \phi_f^{full} + F_{i1} \phi_{f_1}^{full} + \sum_{w \in \mathcal{W} \setminus \{w_1\}} W_{iw} \omega_w^{full} + W_{i1} \omega_{w_1}^{full} + \varepsilon_i^{full} \\
&= \beta_B^{full} + \mathbf{E}_i \beta^{full} + \mathbf{P}_i \rho^{full} + \mathbf{A}_i \alpha^{full} + \left[\sum_{f \in \mathcal{F} \setminus \{f_1\}} (F_{if} - F_{i1}) \phi_f^{full} \right] + \left[\sum_{w \in \mathcal{W} \setminus \{w_1\}} (W_{iw} - W_{i1}) \omega_w^{full} \right] + \varepsilon_i^{full}
\end{aligned}$$

Definiendo $\widetilde{W}_{iw} \equiv W_{iw} - W_{i1}$, $\widetilde{F}_{if} \equiv F_{if} - F_{i1}$, $\widetilde{\phi}^{full}$ y $\widetilde{\omega}^{full}$ (donde estos últimos vectores son análogos a ϕ^{full} y ω^{full} pero sin las componentes f_1 y w_1 respectivamente), la expresión anterior se puede escribir matricialmente del siguiente modo. Las matrices $\widetilde{\mathbf{F}}$ y $\widetilde{\mathbf{W}}$ tienen una columna menos que \mathbf{F} y \mathbf{W} respectivamente, debido a que las columnas asociadas a f_1 y w_1 serían siempre cero, por lo que no sería posible incorporarlas a las estimaciones.

$$\mathbf{Y} = \beta_B^{full} + \mathbf{E} \beta^{full} + \mathbf{P} \rho^{full} + \mathbf{A} \alpha^{full} + \widetilde{\mathbf{F}} \widetilde{\phi}^{full} + \widetilde{\mathbf{W}} \widetilde{\omega}^{full} + \boldsymbol{\varepsilon}^{full} \quad (3.5)$$

Finalmente, se considerará el modelo de la ecuación anterior desviado respecto a su media. Restando \bar{Y} a ambos lados se obtiene el modelo de la ecuación 3.6, cuyas estimaciones se presentan más adelante. Por simplicidad, se ocupan letras minúsculas para representar a cada variable desviada respecto a su media y se define $\xi_i \equiv \varepsilon_i - \bar{\varepsilon}$ (agregando el superíndice “base” cuando se trate de la especificación incompleta y “full” cuando sea la completa).

$$\mathbf{y} = \mathbf{e} \beta^{full} + \mathbf{p} \rho^{full} + \mathbf{a} \alpha^{full} + \widetilde{\mathbf{f}} \widetilde{\phi}^{full} + \widetilde{\mathbf{w}} \widetilde{\omega}^{full} + \boldsymbol{\xi}^{full} \quad (3.6)$$

De manera análoga, el modelo base (que omita efectos fijos de empresas y cargos) a estimar, está dado por la siguiente expresión

$$\mathbf{y} = \mathbf{e} \beta^{base} + \mathbf{p} \rho^{base} + \mathbf{a} \alpha^{base} + \boldsymbol{\xi}^{base} \quad (3.7)$$

Los parámetros de interés de las estimaciones que se han realizado son aquéllos que representan los retornos porcentuales de los niveles educativos β^{base} y β^{full} , por lo que los resultados e interpretaciones se centran en ellos.

Como se señaló anteriormente, para efectos de la descomposición de Gelbach es necesario definir \mathbf{X}_1 y \mathbf{X}_2 . La especificación base tiene como matriz de regresores a \mathbf{X}_1 , mientras que la especificación completa tiene como matriz de regresores a $\mathbf{X} \equiv [\mathbf{X}_1 \ \mathbf{X}_2]$. Es de interés de este estudio cuantificar aquellas componentes del sesgo por omisión de variables relevantes de los retornos de la educación estimados con el modelo base que se deben a la omisión de empresas y cargos respectivamente. Por ello, se declaran las matrices \mathbf{X}_1 y \mathbf{X}_2 de la siguiente forma.

$$\mathbf{X} \equiv \underbrace{[\mathbf{e} \ \mathbf{p} \ \mathbf{a}]}_{\equiv \mathbf{X}_1} \underbrace{[\widetilde{\mathbf{f}} \ \widetilde{\mathbf{w}}]}_{\equiv \mathbf{X}_2} \quad (3.8)$$

3.3. Descomposición de Gelbach

Debido a que este método es relativamente novedoso, parece pertinente describir el contexto en el que se originó, sus aspectos teóricos y prácticos y cómo puede ser utilizada para descomponer los retornos a la educación que se pueden estimar a partir de los datos de www.trabajando.com. Después de estos alcances, se detallan los resultados e interpretaciones obtenidos a partir de la implementación de esta descomposición.

3.3.1. Contexto de origen

Gelbach (2016) [11] desarrolló un modelo que permitía estimar la brecha racial sobre los logaritmos de los salarios de los encuestados en *National Longitudinal Survey of Youth*. Señala que si bien agregar covariables a un modelo de regresión puede brindar mayor robustez a las estimaciones obtenidas, esto puede ser problemático dadas las correlaciones entre los regresores. En términos metodológicos, la descomposición busca responder a las preguntas: ¿cuándo importan las covariables o regresores?, ¿cuáles importan? y ¿qué tanto importan?.

Sean \mathbf{Y} los logaritmos de los salarios de los individuos considerados en la muestra; \mathbf{X}_1 una matriz de covariables que contiene a la variable de interés –en este caso un vector de variables binarias que indican si la persona es o no afroamericano– y otras de control –un vector de variables binarias que indican si la persona es o no hispana–; y \mathbf{X}_2 una matriz de variables cuya omisión podría sesgar las estimaciones de los coeficientes asociados a las variables de interés obtenidos mediante regresión lineal. En el caso de dicho estudio, las variables incorporadas en la matriz \mathbf{X}_2 son el puntaje obtenido en la Prueba de Calificación de las Fuerzas Armadas (*AFQT*, por sus siglas en inglés)¹ y los años de educación de los encuestados.

Si se quiere entender cómo incide la incorporación de cada variable perteneciente a la matriz \mathbf{X}_2 en los coeficientes estimado que constituyen el $\hat{\beta}_1$ asociados a las variables contenidas en \mathbf{X}_1 , se puede partir considerando un modelo simple –sin dichos controles–, agregar estas variables de manera secuencial y registrar los cambios en $\hat{\beta}_1$. El problema de este método es que el orden en el que se agregan las covariables en \mathbf{X}_2 generalmente incide en los cambios registrados en la estimación de dicho coeficiente. Gelbach demuestra que agregar primero los puntajes de *AFQT* y luego los años de educación no reporta los mismos resultados que hacerlo en el orden opuesto. Esto quiere decir que dicha metodología secuencial es deficiente por su arbitrariedad. Este es el inconveniente que busca resolver la descomposición de Gelbach.

3.3.2. Descripción teórica

En primer lugar, se propone considerar los siguientes modelos, denominados como especificación base (ecuación 3.9) y completa (ecuación 3.10).

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{U} \quad (3.9)$$

¹<https://afqttest.com>

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \quad (3.10)$$

Por mínimos cuadrados ordinarios, se tiene que el estimador de $\boldsymbol{\beta}_1$ en la regresión base está dado por la siguiente expresión.

$$\widehat{\boldsymbol{\beta}}_1^{\text{base}} = (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{Y} \quad (3.11)$$

Suponiendo que la especificación completa es acertada, se puede reemplazar la ecuación 3.10 en la ecuación 3.11. Después de un poco de álgebra, se puede identificar el sesgo por variables omitidas, que queda a la derecha del término $\boldsymbol{\beta}_1$.

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_1^{\text{base}} &= (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta}_1 + (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2 \boldsymbol{\beta}_2 + (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \boldsymbol{\varepsilon} \end{aligned} \quad (3.12)$$

Tomando límite en probabilidad, bajo supuestos de exogeneidad de covariables, se observa que el estimador del modelo base no es consistente para el parámetro $\boldsymbol{\beta}_1$. Para efectos de lo que se explica más adelante, es necesario introducir un poco de notación, como se hace a continuación.

$$\begin{aligned} \underbrace{\text{plim}[\widehat{\boldsymbol{\beta}}_1^{\text{base}}]}_{\equiv \boldsymbol{\beta}_1^{\text{base}}} &= \boldsymbol{\beta}_1 + \underbrace{\text{plim}[(\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{X}_1^T \mathbf{X}_2]}_{\equiv \boldsymbol{\Gamma}} \boldsymbol{\beta}_2 \\ &= \boldsymbol{\beta}_1 + \underbrace{\boldsymbol{\Gamma} \boldsymbol{\beta}_2}_{\equiv \boldsymbol{\delta}} \end{aligned} \quad (3.13)$$

En lo que sigue, es necesario desglosar la definición de $\boldsymbol{\Gamma} \in \mathbb{R}^{2 \times 2}$ para comprender la implementación y el sentido de la descomposición de Gelbach. Recordar que $\boldsymbol{\beta}_2$ en este caso está asociado a dos covariables: el puntaje obtenido por los encuestados en la *AFQT* y sus años de educación formal, por lo que $\boldsymbol{\beta}_2 \in \mathbb{R}^{2 \times 1}$. Por conveniencia, se puede considerar la siguiente notación (Gelbach no lo detalla así).

- $\boldsymbol{\beta}_2 = \begin{pmatrix} \beta_2^{\text{AFQT}} \\ \beta_2^{\text{educ}} \end{pmatrix}$
- $\boldsymbol{\Gamma} = \begin{pmatrix} \Gamma_{\text{race}}^{\text{AFQT}} & \Gamma_{\text{race}}^{\text{educ}} \\ \Gamma_{\text{hisp}}^{\text{AFQT}} & \Gamma_{\text{hisp}}^{\text{educ}} \end{pmatrix}$

De este modo, se puede reescribir el desarrollo anterior como sigue.

$$\begin{aligned}
\beta_1^{\text{base}} - \beta_1 &= \begin{pmatrix} \Gamma_{\text{race}}^{AFQT} & \Gamma_{\text{race}}^{\text{educ}} \\ \Gamma_{\text{hisp}}^{AFQT} & \Gamma_{\text{hisp}}^{\text{educ}} \end{pmatrix} \begin{pmatrix} \beta_2^{AFQT} \\ \beta_2^{\text{educ}} \end{pmatrix} \\
&= \begin{pmatrix} \Gamma_{\text{race}}^{AFQT} \\ \Gamma_{\text{hisp}}^{AFQT} \end{pmatrix} \beta_2^{AFQT} + \begin{pmatrix} \Gamma_{\text{race}}^{\text{educ}} \\ \Gamma_{\text{hisp}}^{\text{educ}} \end{pmatrix} \beta_2^{\text{educ}} \\
&= \Gamma^{\text{AFQT}} \beta_2^{AFQT} + \Gamma^{\text{educ}} \beta_2^{\text{educ}}
\end{aligned} \tag{3.14}$$

Observando la primera coordenada y recordando la definición de δ , se puede definir δ_{race} de la siguiente forma.

$$\delta_{\text{race}} \equiv \underbrace{\Gamma_{\text{race}}^{AFQT} \beta_2^{AFQT}}_{\equiv \delta_{\text{race}}^{AFQT}} + \underbrace{\Gamma_{\text{race}}^{\text{educ}} \beta_2^{\text{educ}}}_{\equiv \delta_{\text{race}}^{\text{educ}}} \tag{3.15}$$

Este parámetro representa la diferencia entre el límite en probabilidad del estimador correspondiente a la especificación base y el parámetro de interés $\beta_{1,\text{race}}$. En términos muestrales, esta variable ($\widehat{\delta}_{\text{race}}$) sería el sesgo por omisión de variables relevantes que se genera en el estimador $\widehat{\beta}_{1,\text{race}}^{\text{base}}$ de la brecha racial en los logaritmos de los salarios al no incorporar en el modelo base a las variables de la matriz \mathbf{X}_2 (puntaje en *AFQT* y años de educación). Este sesgo, a su vez, se puede desglosar en componentes atribuibles a cada una de las variables omitidas en la especificación base.

3.3.3. Descripción práctica

Premultiplicando la especificación completa (ecuación 3.10) por $(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$ se obtiene lo siguiente.

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}_1 \widehat{\beta}_1^{\text{full}} + \mathbf{X}_2 \widehat{\beta}_2 + \widehat{\varepsilon} \\
\Rightarrow \underbrace{(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{Y}}_{=\widehat{\beta}_1^{\text{base}}} &= \underbrace{(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_1}_{=\mathbf{I}_{\mathbf{K}_1}} \widehat{\beta}_1^{\text{full}} + (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 \widehat{\beta}_2 + \underbrace{(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \widehat{\varepsilon}}_{=0} \\
\Rightarrow \widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}} &= (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 \widehat{\beta}_2
\end{aligned} \tag{3.16}$$

De la expresión anterior se evidencia que cualquier descomposición del sesgo por variables omitidas $(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 \widehat{\beta}_2$ será descomposición de $\widehat{\delta} \equiv \widehat{\beta}_1^{\text{base}} - \widehat{\beta}_1^{\text{full}}$.

A continuación se introduce más notación.

- \mathbf{X}_{2k} : columna de observaciones de la k -ésima variable en \mathbf{X}_2 .
- $\widehat{\Gamma}_k = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_{2k}$: estimador MCO de regresionar \mathbf{X}_{2k} en función de \mathbf{X}_1 .
- $\widehat{\beta}_{2k}$: coeficiente asociado a la k -ésima variable de \mathbf{X}_2 en el modelo *full*.

La parte del sesgo generado por omisión de variables relevantes atribuible a la k -ésima variable omitida \mathbf{X}_{2k} es $(\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_{2k} \hat{\beta}_{2k}$, la cual es notada por conveniencia como $\hat{\delta}_k$. De aquí se obtiene que el estimador propuesto para el límite en probabilidad del sesgo de $\hat{\beta}_1^{\text{base}}$ se puede descomponer de la siguiente forma.

$$\hat{\delta} = \sum_{k=1}^{K_2} \underbrace{\hat{\Gamma}_k \mathbf{X}_{2k}}_{\equiv \hat{\delta}_k} \quad (3.17)$$

Sea $g \subseteq \{1, \dots, K_2\}$, donde K_2 es el número de variables (columnas) presentes en la matriz \mathbf{X}_2 . Se puede escribir la parte de dicho sesgo atribuible al grupo de variables g de la siguiente forma.

$$\hat{\delta}^g = \sum_{k \in g} \hat{\delta}_k \quad (3.18)$$

Sea G una partición de $\{1, \dots, K_2\}$, lógicamente se cumple la siguiente identidad.

$$\hat{\delta} = \sum_{g \in G} \hat{\delta}^g \quad (3.19)$$

De esta manera es que se pueden descomponer los retornos de los distintos niveles educacionales en grupos de variables, como lo son las binarias correspondientes a las empresas y las palabras presentes en descripciones de cargos.

El algoritmo que propone Gelbach para implementar su descomposición es el siguiente:

1. Estimar por mínimos cuadrados ordinarios la especificación completa, obteniendo $\hat{\beta}_2$.
2. Estimar \mathbf{X}_{2k} en función de \mathbf{X}_1 , obteniendo $\hat{\Gamma}_{2k} \forall k \in \{1, \dots, K_2\}$.
3. Multiplicar $\hat{\beta}_{2k}$ por $\hat{\Gamma}_{2k}$, obteniendo $\hat{\delta}_k \forall k \in \{1, \dots, K_2\}$

Este algoritmo está implementado en el paquete `b1x2` del software Stata, creado por el mismo autor (Gelbach, 2014) [12].

Capítulo 4

Resultados

4.1. Estimaciones de Mincer

Las estimaciones de Mincer para las especificaciones base y completa se muestran en la tabla 4.1. En ella, es posible observar que el efecto de incluir conjuntamente los efectos fijos de empresas y de palabras es negativo sobre los retornos de los distintos niveles educacionales. En el caso de la educación media científico humanista se observa una reducción de un 14,5 % a un 4,4 %; en el de la media técnico profesional se observa una reducción de un 27,4 % a un 10,1 %; en el caso de la educación superior técnico profesional se observa una reducción de un 58,8 % a un 30,1 % y en el caso de la educación universitaria se observa una reducción de un 107,3 % a un 59,4 %.

Esto da un indicio de que alcanzar un mayor nivel educacional está correlacionado con la oportunidad de acceder a empresas y cargos asociados a mayores salarios. En otras palabras, los retornos a la educación que se estiman en el modelo base son mayores, pues absorben el hecho de que educarse más permite acceder a cargos y empresas mejor remuneradas. Naturalmente, una pregunta que surge a partir de este resultado es: ¿qué partes de las diferencias entre los retornos estimados en ambas especificaciones son explicadas por los cargos y por las empresas respectivamente? Lo interesante de esta pregunta es que responderla puede ayudar a entender los medios a través de los cuales se manifiestan los retornos a la educación.

A modo de ejemplo, se podría decir –a partir del modelo base– que la brecha salarial entre tener educación básica y universitaria (manteniendo las otras variables constantes) es de un 107,3 %. Parte de este porcentaje puede ser atribuido al hecho de que las universidades proveen a los trabajadores de herramientas y habilidades que son valoradas por los empleadores, pues hacen que las personas tiendan a ser más productivas. Sin embargo, puede resultar ingenuo pensar que los retornos de la educación se manifiestan únicamente a través de esta brecha de productividad esperada. Evidentemente, alcanzar un cierto nivel educacional resulta ser un requisito para poder alcanzar determinados cargos, los cuales pueden –por supuesto– tener requisitos de productividad que podrían ser generados por la educación formal. Asimismo, puede ser que hayan empresas más exigentes que otras al momento de imponer la educación como requerimiento para satisfacer un mismo cargo. En definitiva, se propone poner a prueba

Tabla 4.1: Retornos estimados de los distintos niveles educacionales y otras estimaciones obtenidas con los datos de vacantes TBN.

Regresores	Base	Completa
Media CH	0,1450***	0,0442***
Media TP	0,2738***	0,1010***
Superior TP	0,5882***	0,3012***
Universitaria	1,0732***	0,5942***
Experiencia	0,1557***	0,0972***
Experiencia ²	-0,0015***	-0,0018***
Experiencia ³	-0,0006***	-0,0003***
Experiencia ⁴	0,0000***	0,0000***
Año 2011	-0,0492***	-0,0034
Año 2012	0,0055	0,0276***
Año 2013	-0,0060	0,0797***
Año 2014	0,0839***	0,1283***
Año 2015	0,0807***	0,1168***
Año 2016	0,0989***	0,1302***
Empresas		✓
Palabras		✓
R^2	0,6278	0,8082
R^2 ajustado	0,6278	0,8060
Número de obs.	502.447	502.447

Nota: La variable dependiente de los modelos es el logaritmo de los salarios reales por hora. Los datos utilizados corresponden a información ingresada por los empleadores (requerimientos y ofertas) a nivel de vacantes (no avisos). Los estimadores fueron obtenidos mínimos cuadrados ordinarios. Se muestran las estimaciones de retornos para cada nivel educacional, excepto la educación básica (que fue designada arbitrariamente como categoría base).

la hipótesis de que el retorno salarial de alcanzar un determinado nivel educacional se puede manifestar a través de al menos tres vías distintas cuando se observa la demanda de trabajo a nivel agregado.

La primera de estas vías es el retorno intrínseco de la educación, que se explica a partir de la valoración de los empleadores por la brecha de habilidades que esperan que haya entre alguien con educación básica y alguien con un nivel de formación más avanzado. A modo ilustrativo, se podría suponer un empleador que publica una vacante de trabajo para un programador. Este podría elegir como requisito que los postulantes hayan alcanzado al menos la educación técnico profesional o bien universitaria. Es posible que alguien que haya estudiado una carrera universitaria relacionada al tema maneje más herramientas computacionales. El empleador podría tener una valoración por esta brecha de habilidades laborales entre ambos niveles educativos. También es posible que quienes hayan terminado la educación universitaria cuenten en mayor medida con aptitudes no directamente ligadas a lo laboral, que sean esperables y valoradas por los empleadores.

Para ejemplificar, se podría especular que generalmente quienes egresan de la educación universitaria serían más cultos que quienes tienen educación técnico profesinal (por tener

que aprobar cursos de formación general). Naturalmente, no es posible cuantificar cómo se desglosan en detalle estas valoraciones esperadas de los empleadores. Sin embargo, se podría cuestionar la relevancia de llevar a cabo este desglose, considerando que su utilidad sería hacer comparaciones con contrafactuales imposibles o poco frecuentes, como sería el caso de un universitario que no haya aprobado cursos de formación general.

También es posible que quienes han alcanzado la educación universitaria tengan una mayor capacidad de adaptarse a las condiciones del mercado. Esto se explicaría por el hecho de que sus estudios permiten una formación más general en comparación a quienes egresan de la educación técnico profesional. Retomando el ejemplo del programador, un universitario puede tener más conocimientos sobre matemáticas y computación, los cuales pueden ser transversales a los lenguajes de programación que utilizaría en la práctica y por ello tenga una mayor capacidad de aprender nuevos lenguajes. Es posible que la persona contratada, en primera instancia, no requiera ocupar estos conocimientos. Sin embargo, el empleador sí podría tener una valoración por el potencial de creatividad y adaptación de la persona contratada. Esto podría explicar en cierta medida que las personas con mayor educación formal tengan salarios más altos y obtendrían lo que entendemos como “retorno intrínseco” de la educación. Nuevamente, con los datos que se tienen no es posible aislar esta valoración de los retornos estimados. A pesar de ello, otra vez parece poco relevante hacer comparaciones con contrafactuales imposibles o poco frecuentes, como el de un universitario sin conocimientos propios y fundamentales de su formación.

La segunda y la tercera vía a través de las cuales se pueden manifestar los retornos a la educación estarían dadas por el acceso a vacantes de trabajo relacionadas respectivamente a empresas y cargos que suelen presentar mayores remuneraciones. Al igual que en como explica Spence (1973) [17], los trabajadores pueden elegir si educarse o no en base a incentivos dados por la señalización que estos puedan generar en el mercado laboral. La educación en este contexto teórico, es solamente una señal costosa que sólo los que poseen la habilidad pueden pagar, aunque la educación formal obtenido no tenga ningún impacto real en sus productividades.

Por otro lado, alcanzar un cierto nivel educacional se puede interpretar, por parte de los empleadores, como una señal de que la persona es apta para desempeñar ciertos trabajos en determinadas empresas, sea que la educación formal haya incidido en la generación de éstas o no. De este modo, se entienden los distintos niveles de educación como “llaves” que pueden servir o no para abrir algunas de las “puertas” que serían las vacantes de trabajo. Cuando los empleadores ingresan como requisito un determinado nivel educacional en un aviso de trabajo, deciden directamente qué “llaves” son útiles para acceder a las vacantes de empleo correspondientes. En otras palabras, los empleadores restringen el acceso de vacantes de trabajo con mayores remuneraciones a quienes poseen un menor grado de educación formal. Por lo tanto, la educación no solo tendría un retorno intrínseco dado por el aumento esperado de productividad de los trabajadores, sino que también brinda el acceso a trabajar en empresas y cargos de mayor exigencia y mayores salarios.

A pesar de que la tabla 4.1 evidencia una notable influencia del hecho de agregar o no las variable binarias relacionadas a empresas y palabras presentes en las descripciones de los cargos (en conjunto), no permite distinguir en qué medida cada uno de estos grupos

de variables es responsable de los sesgos o diferencias en las estimaciones de los retornos de los distintos niveles educativos. Un mecanismo posible para responder a esta pregunta sería observar cómo cambian los estimadores de dichos retornos a medida que se agregan secuencialmente estos dos grupos de variables binarias. De este modo, se pueden atribuir las primeras variaciones en los estimadores al primer grupo de variables en ser incorporado a la regresión y las segundas variaciones al segundo grupo de variables en ser agregado. Una posibilidad sería incorporar primero los efectos fijos de empresas y luego los de las palabras presentes en las descripciones de los cargos, mientras que una segunda posibilidad sería hacerlo en el orden opuesto.

El problema del procedimiento descrito en el párrafo anterior es que –como explica Gelbach (2016) [11]– sus resultados son naturalmente ambiguos, pues puede entregar resultados diferentes dependiendo del orden en el que se incorporen los distintos grupos de variables en la regresión. Este es el problema que motiva el uso de su descomposición, por permitir atribuir simultáneamente y sin ambigüedades los sesgos generados por la omisión de ambos grupos de variables (empresas y descripciones de cargos). Es por ello que en la siguiente sección se retoman los análisis expuestos en esta y se profundizan en virtud de los resultados relativos a las estimaciones hechas mediante la descomposición de Gelbach.

4.2. Descomposición de Gelbach

Aplicando la identidad 3.19 al caso de este estudio, se obtiene lo siguiente.

$$\begin{pmatrix} \widehat{\delta}_{MCH} \\ \widehat{\delta}_{MTP} \\ \widehat{\delta}_{STP} \\ \widehat{\delta}_U \end{pmatrix} = \begin{pmatrix} \widehat{\delta}_{MCH}^f \\ \widehat{\delta}_{MTP}^f \\ \widehat{\delta}_{STP}^f \\ \widehat{\delta}_U^f \end{pmatrix} + \begin{pmatrix} \widehat{\delta}_{MCH}^w \\ \widehat{\delta}_{MTP}^w \\ \widehat{\delta}_{STP}^w \\ \widehat{\delta}_U^w \end{pmatrix} \quad (4.1)$$

El término que se encuentra al lado izquierdo de la igualdad representa las variaciones totales en los estimadores de los retornos. Si para un determinado nivel educacional e se tiene que $\widehat{\delta}_e > 0$ esto significa que la incorporación de las empresas y descripciones de cargos en la regresión tiene un efecto negativo sobre el retorno estimado para a dicho nivel ($\beta_e^{full} < \beta_e^{base}$). Esto se traduce en que dicho nivel educacional estaría ligado a empresas y cargos que, en suma, tienen salarios más altos para sus vacantes. Esta misma lógica se puede aplicar para los términos asociados únicamente a las firmas ($\widehat{\delta}_e^f$) y únicamente a las palabras presentes en las descripciones de los cargos ($\widehat{\delta}_e^w$). Si $\widehat{\delta}_e^f > 0$ entonces el nivel educacional e está asociado a vacantes de trabajo publicadas por firmas que tienden a pagar más que el promedio. Del mismo modo, si $\widehat{\delta}_e^w > 0$ entonces el nivel educacional e está asociado a vacantes de trabajo cuyos cargos tienden a pagar más que el promedio.

Como se discutió en la sección anterior, los retornos de los niveles educativos se reduce al agregar ambos tipos de efectos fijos. Gracias a la descomposición de Gelbach (2016) [11] es posible dar a entender qué roles juegan tanto las empresas como las descripciones de los cargos sobre las variaciones en los retornos de los distintos niveles educativos. De esta forma, es posible comprender en qué sentido y magnitud afectan a dichos retornos y mediante

qué vías se manifiestan los impactos de alcanzar distintos niveles educativos.

Antes de describir lo que se observa para cada nivel educativo, puede ser conveniente introducir un par de distinciones que serán útiles para comunicar los análisis de manera más clara y eficiente. En primer lugar, se puede escribir la ecuación 4.1 para un determinado nivel educativo e del siguiente modo.

$$\widehat{\delta}_e = \widehat{\delta}_e^f + \widehat{\delta}_e^w \quad (4.2)$$

Recordando que $\widehat{\delta}$ es la diferencia entre los estimadores obtenidos mediante las especificaciones base y completa, esta identidad se puede escribir como en la ecuación 4.3. A partir de aquí, el término $\widehat{\beta}_e^{base}$ será denominado retorno "bruto" del nivel educativo e sobre la educación básica. El término "bruto" se debe a que no está aislado, sino que absorbe las diferencias de empresas y cargos que tiene dicho nivel educativo respecto a la educación básica. Por lo tanto, cuando se analiza el retorno "bruto" de la educación no es posible identificar netamente la valoración que tienen los empleadores por la brecha esperada de habilidades y conocimientos que pueda tener un trabajador más educado. Esta última valoración será denominada como retorno "neto" o "intrínseco" de la educación y se estima mediante $\widehat{\beta}_e^{full}$. El estimador $\widehat{\beta}_e^{base}$, está sesgado por la omisión de las variables relevantes que contienen información sobre las empresas y los cargos. Por lo tanto, su estimación se confunde con el hecho de que trabajadores más educados puedan acceder a empresas con políticas salariales más generosas o cargos de trabajo que presentan mayores remuneraciones. El sesgo estimado $\widehat{\delta}_e$ es parte de lo que se mide en el retorno bruto estimado $\widehat{\beta}_e^{base}$. Una parte de este sesgo es atribuible a las empresas y otra parte a las descripciones de los cargos.

$$\underbrace{\widehat{\beta}_e^{base}}_{\text{Retorno bruto}} = \underbrace{\widehat{\beta}_e^{full}}_{\text{Retorno intrínseco}} + \underbrace{\widehat{\delta}_e^f}_{\text{Retorno por acceso a firmas}} + \underbrace{\widehat{\delta}_e^w}_{\text{Retorno por acceso a cargos}} \quad (4.3)$$

A continuación se presentan la tabla 4.2 y la figura 4.1 que resumen los resultados de las descomposiciones de los retornos asociados a cada uno de los niveles educativos en las tres componentes descritas anteriormente. Luego se desarrolla un breve análisis sobre la descomposición de Gelbach para cada nivel educativo. Para facilitar la lectura de la tabla 4.2, se definen $\widehat{\delta}_e^{empresas} \equiv \widehat{\delta}_e^f$ y $\widehat{\delta}_e^{cargos} \equiv \widehat{\delta}_e^w$.

4.2.1. Educación media científico humanista

El retorno bruto de la educación media científico humanista estimado es de un 14,5%. Esto quiere decir que, en valor esperado, quienes han alcanzado este nivel educativo ganan considerablemente más que quienes solo han alcanzado la educación primaria. Al momento de incorporar conjuntamente las empresas y cargos a la regresión, se obtiene un retorno intrínseco de un 4,4%, lo cual es bastante menor. Es decir, se experimenta una reducción de un 10,1%.

En la tabla 4.2 se observa que la incorporación de variables de empresas generan una reducción del 3,1% del retorno estimado para este nivel, mientras que las de las palabras presentes en las descripciones de los avisos generan un aumento del 13,2%. Como se observó

Tabla 4.2: Descomposición de Gelbach para los retornos de los distintos niveles educacionales.

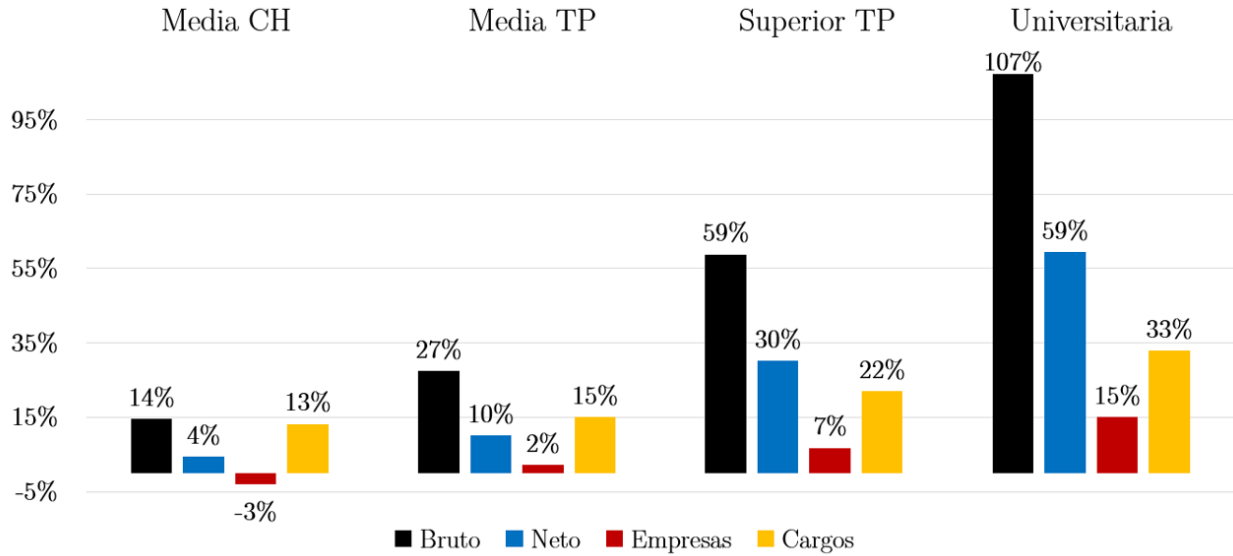
	Estimador	Valor	Desv. est.	Porc.
Media CH	$\widehat{\beta}_{MCH}^{base}$	0,1450	0,0017	100
	$\widehat{\beta}_{MCH}^{full}$	0,0442	0,0022	30
	$\widehat{\delta}_{MCH}$	0,1008	0,0022	70
	$\widehat{\delta}_{MCH}^{empresas}$	-0,0308	0,0013	-21
	$\widehat{\delta}_{MCH}^{cargos}$	0,1316	0,0019	91
	Media TP	$\widehat{\beta}_{MTP}^{base}$	0,2738	0,0022
$\widehat{\beta}_{MTP}^{full}$		0,1010	0,0026	37
$\widehat{\delta}_{MTP}$		0,1727	0,0026	63
$\widehat{\delta}_{MTP}^{empresas}$		0,0218	0,0016	8
$\widehat{\delta}_{MTP}^{cargos}$		0,1510	0,0022	55
Superior TP		$\widehat{\beta}_{STP}^{base}$	0,5882	0,0023
	$\widehat{\beta}_{STP}^{full}$	0,3012	0,0028	51
	$\widehat{\delta}_{STP}$	0,2870	0,0027	49
	$\widehat{\delta}_{STP}^{empresas}$	0,0672	0,0019	11
	$\widehat{\delta}_{STP}^{cargos}$	0,2198	0,0023	37
	Universitaria	$\widehat{\beta}_U^{base}$	1,0732	0,0024
$\widehat{\beta}_U^{full}$		0,5942	0,0030	55
$\widehat{\delta}_U$		0,4790	0,0030	45
$\widehat{\delta}_U^{empresas}$		0,1499	0,0021	14
$\widehat{\delta}_U^{cargos}$		0,3291	0,0026	31

Nota: Esta descomposición explica cómo se descomponen los sesgos por omisión de variables relevantes de los estimadores relacionados al retorno de los distintos niveles educacionales al momento de omitir intencionalmente los efectos fijos de empresas y de palabras presentes en descripciones de cargos. La descomposición se hizo en base a las regresiones de la tabla 4.1.

en la tabla 2.1, las vacantes que requieren educación media científico humanista en general están asociadas a empresas con políticas salariales levemente menos generosas que las empresas promedio. Esto explica que el retorno por acceso a firmas de la educación media científico humanista sea moderadamente negativo. Se puede interpretar que las vacantes que requieren educación media científico humanista en general están asociadas a empresas con políticas salariales poco generosas, pero también a cargos mejor remunerados.

Es posible concluir que los empleadores valoran considerablemente que los postulantes alcancen el nivel medio científico humanista. Esta valoración (14,5 %) se explica, en primer lugar, por el acceso a cargos mejor remunerados (13,2 %) y, en segundo lugar, por las herramientas que intrínsecamente entrega dicha formación (4,4 %). Estos dos factores contrarrestan el hecho de que las empresas que suelen pedir este nivel suelen ofrecer salarios menores, generando un retorno por acceso a empresas negativo (-3,1 %).

Figura 4.1: Descomposición los retornos estimados de cada nivel educacional.



Nota: Los retornos brutos estimados (negro) se descomponen en retornos netos o intrínsecos (azul) y relativos al acceso a firmas (rojo) y cargos (amarillo). Los datos de este gráfico fueron extraídos de la tabla 4.2.

4.2.2. Educación media técnico profesional

El retorno bruto del nivel medio técnico profesional por sobre el básico es de un 27,4% (12,9% mayor que el de la educación media científico humanista). Al incorporar conjuntamente las empresas y cargos a la regresión, se obtiene un retorno intrínseco de un 10,1% (5,7% mayor que el de la educación media científico humanista), el cual es considerablemente menor que el retorno bruto estimado (disminución de un 17,3%). Al analizar la tabla 4.2 se evidencia que tanto la introducción de variables de empresas como de cargos generan, por separado, disminuciones en el retorno estimado para este nivel educativo (pues $\hat{\delta}_{MTP}^{empresas}, \hat{\delta}_{MTP}^{cargos} > 0$). Dicha reducción del 17,3% se desglosa en un 15,1% relativo a cargos y un 2,2% relativo a empresas (estos retornos por accesos son respectivamente 1,9% y 5,3% mayores que los estimados para la educación media científico humanista), por lo que la mayor parte de esta diferencia se debe al acceso a cargos mejor remunerados que el promedio.

En síntesis, el retorno bruto de la educación media técnico profesional (27,4%) se explica principalmente por el acceso a mejores cargos (15,1%), luego por las habilidades propias de la formación (10,1%) y finalmente por el acceso a empresas que tienden a pagar mejor a sus trabajadores (2,2%).

Al comparar la educación media técnico profesional con la media científico humanista se observa que todos los retornos estimados (bruto, neto, por acceso a empresas y por acceso a cargos) son mayores. Por ello se puede concluir que: (1) quienes tienen educación media TP pueden acceder a vacantes con salarios ofrecidos más altos, (2) las herramientas propias que entrega la educación media TP son más valoradas en el mercado (para quienes alcanzan sólo este nivel), (3) quienes han alcanzado educación media TP pueden acceder a cargos mejor remunerados y (4) a empresas con políticas salariales más generosas.

4.2.3. Educación superior técnico profesional

El retorno bruto de la educación superior técnico profesional es de un 58,8 % (31,4 % mayor que el de la educación media TP). Esto se traduce en que este nivel educacional tiene un retorno esperado considerablemente mayor que el de la educación media técnico profesional y la educación media científico humanista. Este 58,8 % se explica principalmente por el retorno intrínseco (30,1 %, 20,0 % mayor que el de la media TP) y por acceso a cargos (22,0 %, 6,9 % mayor que el de la media TP). El acceso a empresas con políticas salariales más generosas sólo explica un 6,7 % (4,5 % mayor que el de la media TP) de dicho retorno bruto.

Al comparar la educación superior técnico profesional con la media técnico profesional se observa que todos los retornos estimados (bruto, neto, por acceso a empresas y por acceso a cargos) son mayores. Por ello se puede concluir que: (1) quienes tienen educación superior TP pueden acceder a vacantes con salarios ofrecidos más altos, (2) las herramientas propias que entrega la educación superior TP son considerablemente valoradas por los empleadores, (3) quienes han alcanzado educación superior TP pueden acceder a cargos bastante mejor remunerados y (4) a empresas con políticas salariales más generosas.

4.2.4. Educación universitaria

Como era de esperarse, este nivel educacional es el que reporta los mayores retornos estimados. El retorno bruto de este nivel es de un 107,3 % respecto al nivel básico. La mayor parte de este retorno se manifiesta como retorno intrínseco (59,4 %), en segundo lugar se explica por el acceso a cargos mejor remunerados (32,9 %) y en tercer lugar por el acceso a empresas con políticas salariales más generosas (15,0 %). Esto da a entender que la educación universitaria es mejor remunerada que otros niveles tanto por las habilidades relativas a la formación y las mejoras en productividad que los empleadores esperan de este nivel respecto a otros, así como por los cargos y empresas a las que permite acceder.

Al comparar la educación universitaria con la superior técnico profesional se observa que todos los retornos estimados (bruto, neto, por acceso a empresas y por acceso a cargos) son mayores. Las diferencias son de un 48,5 % de retorno bruto, 29,3 % de neto, 8,3 % por acceso a empresas y de un 10,9 % por acceso a cargos. Esto permite concluir que existen diferencias considerables en todos los aspectos analizados de los retornos estimados a la educación cuando se comparan ambos niveles asociados a educación superior.

Es posible observar fenómenos parecidos al comparar la educación universitaria con la media científico humanista. Las diferencias son de un 92,8 % de retorno bruto, 55,0 % de neto, 18,1 % por acceso a empresas y de un 19,7 % por acceso a cargos. Esto permite concluir que existen diferencias considerables en todos los aspectos analizados de los retornos estimados a la educación cuando se comparan ambos niveles asociados a formaciones científico-humanistas (no técnicas).

Capítulo 5

Discusión

Como se mencionó al comienzo de este documento, las estimaciones mincerianas han experimentado una evolución constante a lo largo del tiempo debido a la incorporación de algunos factores que no estaban presentes en el modelo original. Muchos de estos aspectos han sido considerados en el presente estudio con el fin de entender de forma precisa y detallada varios aspectos relevantes asociados a los retornos salariales de la educación. En este capítulo se habla sobre algunos problemas econométricos clásicos de esta literatura y cómo es que se han tenido presentes en las estimaciones realizadas. Finalmente, se incluye un breve párrafo acerca las limitaciones que tiene el uso de datos de www.trabajando.com en términos de representatividad del mercado laboral chileno, tema que se toca en profundidad en la sección 1.3.

Parece razonable pensar que pueden existir diferencias sistemáticas entre las habilidades de los trabajadores más y menos educados. Una persona, a medida que se educa, desarrolla habilidades que pueden estar o no asociadas a su formación profesional. Del mismo modo, alguien que tiene mayores habilidades puede tener acceso y/o facilidades para alcanzar niveles de educación más avanzados. Esta correlación entre la educación y variables que típicamente no son observables mediante datos de encuestas genera un problema de endogeneidad (u omisión de variables) que ha sido vastamente estudiado en la literatura, por autores como Griliches (1977) [3], Angrist & Krueger (1991) [4] y Bound, Jaeger & Baker (1995) [5]. Una forma recurrente de solucionar este problema es mediante el uso de variables instrumentales. A pesar de su utilidad, suele ser complicado encontrar instrumentos que cumplan con los supuestos de relevancia y exclusión. Un enfoque alternativo que se ha propuesto en este documento se basa en considerar una base de datos descriptiva de la demanda de trabajo. Por su naturaleza, esta información y las estimaciones que se hagan a partir de ella no se ven afectadas por las habilidades propias de trabajadores específicos, sino únicamente por las expectativas y valoraciones de los empleadores. Cabe destacar que las expectativas que tienen los empleadores sobre las habilidades de los trabajadores podrían estar expresadas parcialmente por variables como las descripciones de los cargos o (en cierta medida) las empresas. Estas variables pueden contener información descriptiva de las habilidades que se espera que tenga un trabajador que ha terminado un determinado nivel educacional.

En la década del 2000, algunos autores como Heckmann, Lochner & Todd (2003) [2] y

Lemieux (2006) [8] reconocieron la importancia de la función de Mincer en la economía laboral e hicieron revisiones de estimaciones relacionadas a ella, que se llevaron a cabo en las últimas décadas hasta entonces. Ambos estudios señalan que existen contextos en los que las estimaciones presentaban inconsistencias. Lemieux (2006) [8] indica que el modelo de Mincer sigue siendo un punto de referencia en la medida de que se incluyan ciertas mejoras, que han sido consideradas. Se agregaron potencias de hasta orden 4 para la experiencia y efectos fijos para los años, con el fin de capturar diferencias por cohortes. El autor también sugiere agregar un término cuadrático para la escolaridad, entendiendo esta última como una variable continua (lo cual no es el caso de este estudio porque los requerimientos de las empresas son categóricos).

Heckmann, Lochner & Todd (2003) [2] examinan los supuestos bajo los cuales los resultados obtenidos pueden interpretarse efectivamente como retornos a la educación. Podría haber resultado relevante examinar estos supuestos en caso de tratarse de un estudio basado teóricamente en el modelo de Mincer (1974) [1]. Como se señaló en la introducción, esto no es rigurosamente así, debido a que los datos observados corresponden a la demanda de trabajo y no a la oferta. En consecuencia, resulta difícil usar o adaptar el mismo marco teórico a la realidad que se ha descrito en este estudio.

Teniendo en cuenta que: (1) los trabajadores a contratar eligen su nivel educacional en base a costos y ganancias que caracterizan al mercado; (2) la acumulación de capital humano reporta remuneraciones salariales propias del mismo y (3) el modelo de Mincer se basa en salarios que también son exógenos, es posible explicar las ecuaciones que se estiman en el presente estudio basándose en modelos de precios hedónicos. Estos precios hedónicos representarían precisamente a los salarios que ofrecen los empleadores en el sitio web, los cuales pueden ser descritos como la suma de las valoraciones que manifiestan estos empleadores por dichos atributos. Los modelos estimados además incluirían factores asociados a los títulos de trabajo y a las empresas que publican las vacantes. Entendiendo que este estudio no se basa teóricamente en el modelo de Mincer (ni de alguna de sus variantes), parece relevante plantear el desafío a futuro de desarrollar un modelo económico que de un mayor sustento a estas “estimaciones mincerianas” vistas desde la perspectiva de la demanda de trabajo.

Otro presunto problema de endogeneidad puede ser causado por la omisión de variables relevantes en la determinación de salarios, que pueden ser observables y que van más allá de la educación y la experiencia. Algunos ejemplos pueden ser aquellas que contienen información sobre trabajadores, firmas, cargos o títulos de trabajo. La incorporación de efectos fijos que controlan por este tipo de variables no es algo nuevo en esta literatura y, por lo general, reporta considerables beneficios en cuanto a estimaciones, interpretaciones de resultados y mejoras en cuanto al poder predictivo de los modelos. Algunos estudios que han utilizado esta estrategia son los de Abowd, Kramarz & Margolis (1999) [6] y Cardoso, Guimarães, Portugal & Reis (2018) [14], con registros administrativos de datos pareados de empresas y trabajadores); y Marinescu & Wolthoff (2016) [13] y Banfi, Choi & Villena-Roldán (2019) [16], con datos de portales de búsqueda de empleo en internet. En este estudio se optó por incorporar variables binarias que indican qué empresas publicaron cada vacante de trabajo y qué palabras están presentes en las descripciones de los cargos asociados a cada vacante. Este último grupo de variables fue tomado del trabajo realizado por Banfi, Choi & Villena-Roldán (2019) [16], en el cual se extrajeron las cuatro palabras más relevantes de cada descripción.

Los resultados expuestos en el capítulo anterior evidencian que la incorporación de estos efectos fijos es crucial para evitar descomponer e interpretar las estimaciones de los retornos de la educación sobre los salarios.

Otro fenómeno que puede tenerse en consideración es el de los efectos de “derrame” (spillover) de educación, que ha sido estudiado por autores como Wirz (2008) [9] y Cardoso, Guimarães, Portugal & Reis (2018). Se podría especular que los empleadores que publican avisos de trabajo en www.trabajando.com –conociendo características de los trabajadores con los que actualmente cuentan– se anticiparían a los posibles efectos de “derrame” de educación que podrían darse en ambos sentidos (tanto desde el trabajador a contratar hacia sus colegas como desde los trabajadores actuales hacia el trabajador a contratar). A pesar de que no se tiene la información necesaria para incorporar dichos efectos, se cuenta con efectos fijos de empresas y palabras presentes en descripciones de cargos. Estas variables en alguna medida contienen información sobre la educación de los colegas que tendría el trabajador a contratar, pues dicha educación tendría relación con los empleadores y los cargos que buscan llenar. Si bien esto podría mitigar los eventuales sesgos generados por los efectos de “derrame”, tampoco se puede descartar de raíz que estos existan, pues incluso dentro de una misma empresa y un mismo cargo pueden haber distintos grupos de trabajadores que difieran en su educación y/o habilidades.

Podría ser relevante analizar en un futuro la presencia de salarios de eficiencia en el contexto del mercado laboral en línea que se desarrolla en www.trabajando.com. Akerlof (1982) [18] desarrolló un modelo mediante el cual busca explicar por qué algunos trabajadores tienden a esforzarse más que lo exigido por sus empleadores, mientras que algunas empresas pagan salarios mayores a los de limpieza de mercado. El autor denomina su modelo como de “intercambio de regalos”, dado el comportamiento de los agentes. Por un lado, el trabajador “regala” aquellos esfuerzos que exceden los estándares mínimos de trabajo. Por otro lado, la empresa “regala” bonificaciones de salarios (de manera que el salario percibido por los empleados sean mayores a los que recibirían si dejaran sus trabajos actuales). El modelo de intercambio de regalos se diferencia de modelos económicos estándar en al menos dos aspectos: (1) en modelos neoclásicos los trabajadores solo cumplen con los estándares mínimos que exige la empresa¹ y (2) no necesariamente se cumplen las condiciones de limpieza de mercado. Esto último sería una posible explicación para la existencia de desempleo involuntario. Autores como Krueger & Summers (1988) [19] han encontrado evidencia empírica que respalda la teoría de salarios de eficiencia. Estos autores rechazan teorías clásicas de determinación de salarios examinando diferencias salariales para trabajadores igualmente calificados en distintas industrias. Además presentan evidencia de que una mayor rotación de trabajadores está ligada a industrias con salarios menores, lo que sugiere que los trabajadores de industrias con mayores salarios reciben sueldos no competitivos.

Teniendo en vista el trabajo de Akerlof (1982) [18] es posible plantear una serie de reflexiones e interrogantes. Cabe destacar que los salarios estudiados son ofertas estimadas que ingresan los empleadores en el sitio web para postulantes que cumplan requisitos mínimos. Mientras varios candidatos pueden postular a una misma vacante, el empleador que la publica es capaz de elegir a su postulante preferido. Lógicamente, los salarios que ingresan los empleadores serían aquellos que pagarían al postulante que finalmente contratarían. Es

¹Siempre que perciban utilidades marginales negativas por trabajar más.

esperable que –en muchos casos– dichas ofertas de salarios incorporen primas por eficiencia (“regalos”) que concederían los empleadores a los trabajadores que presumiblemente sean más productivos (tanto si se comparan con el resto de los postulantes como con los trabajadores que ya ejercen en las respectivas empresas). ¿Es posible que esta hipótesis explique parcialmente las ofertas de salarios ingresadas en el sitio web? ¿Afectaría esto a las estimaciones de los retornos a la educación? ¿En qué medida?

A pesar de todas las consideraciones incluidas en este trabajo, es necesario reconocer las limitaciones que tiene este estudio al momento de querer extrapolar los resultados a la economía chilena en su conjunto. Como se vio en la sección 1.3, los datos del sitio web están muy concentrados en empresas cuyas casas matrices se ubican en la Región Metropolitana. Además hay industrias como las de comercio, servicios e inmobiliaria que están sobrerrepresentadas en la demanda de www.trabajando.com, mientras que sectores como el agrícola y el de construcción carecen de representatividad en el sitio web. Es por ello que se concluye que las estimaciones obtenidas representan en mayor medida a ciertas industrias de la Región Metropolitana. Esto, a su vez, explica que hayan diferencias considerables en los retornos de la educación estimados en la sección 1.3 (además del hecho de que las estimaciones realizadas con datos de la ESI puedan estar sesgadas por las habilidades no observables de los trabajadores).

Conclusión

Spence (1973) [17] manifiesta que el valor de la educación en el mercado laboral puede expresarse a través de lo que él denomina “señalización”. En su modelo, los trabajadores eligen si educarse –asumiendo los costos que esto implica– o no. La decisión que toman los trabajadores es interpretada como una señal de sus habilidades por parte de los empleadores, quienes distintamente valoran el trabajo de empleados educados y no educados. Una reflexión importante que puede surgir a partir de este modelo es el hecho de que el valor de la educación tiene que ver con una cuestión de acceso. Lo anterior motiva la idea de que los retornos “brutos” estimados para la educación se pueden descomponer en las tres componentes antes detalladas: (1) retorno intrínseco, (2) retorno por acceso a firmas y (3) retorno por acceso a cargos. A continuación se discuten aspectos relevantes sobre estos mecanismos a través de los que se manifiesta el retorno de la educación.

Posteriormente, se hace una revisión general de los resultados expuestos en este documento. Algunas preguntas que naturalmente surgen acerca de estos resultados son: ¿son coherentes entre ellos?, ¿son intuitivos?, ¿qué información nueva proporcionan? y ¿qué limitaciones tienen? Adicionalmente se discute acerca de los mecanismos a través de los que la educación se manifiesta en los salarios.

En primer lugar, cabe destacar la relevancia del nivel de educación sobre los salarios. Este resultado es aplicable tanto para los datos de la ESI como los de TBN. En el último caso, se observa que la educación universitaria es el nivel que reporta los retornos más altos una vez que se controla por variables como la experiencia requerida (y sus potencias de hasta grado 4), los años, las empresas y los cargos. Para este nivel, el retorno bruto estimado (sin considerar controles por empresas y cargos) por sobre la educación básica es aproximadamente de un 107,3%, mientras que el neto es cercano al 59,4%. Seguidamente se encuentran la educación superior técnico profesional, media técnico profesional, media científico humanista y básica. Es fácil ver que el mercado laboral en línea premia considerablemente la educación formal, evidenciándose que aquellos niveles de mayor especialización reportan los mayores retornos. Es posible extender esta conclusión para las tres vías analizadas de los retornos a la educación. Esto quiere decir que un nivel educacional más avanzado reporta mayores retornos intrínsecos así como acceso a empresas y cargos de mayores remuneraciones. Todo esto se puede evidenciar en detalle en la tabla 4.2 y la figura 4.1.

En segundo lugar, cabe destacar que los accesos a cargos y a empresas asumen un rol importante de los retornos a la educación. Esto permite concluir que aquellas estimaciones que omitan esta información podrían estar sesgadas, o no desagregan esta estimación en

otras causas como acceso a empresas y cargos de mayores salarios. Los primeros indicios de este fenómeno se manifiestan en las distribuciones de efectos fijos reportadas en el capítulo 2 y en las diferencias de los estimadores correspondientes a las especificaciones básicas y completas para las ecuaciones de Mincer desarrolladas en el capítulo 3. Posteriormente, y a través de la descomposición de Gelbach, es posible develar cómo afectan distintos grupos de variables al retorno de la educación al momento de ser incorporados a las regresiones. Dicha descomposición revela que el retorno intrínseco de la educación y el que se explica por el acceso a cargos mejor remunerados suelen asumir las partes más considerables del retorno bruto de la educación. A pesar de ello, el acceso a firmas mejor remuneradas no deja de ser una componente relevante.

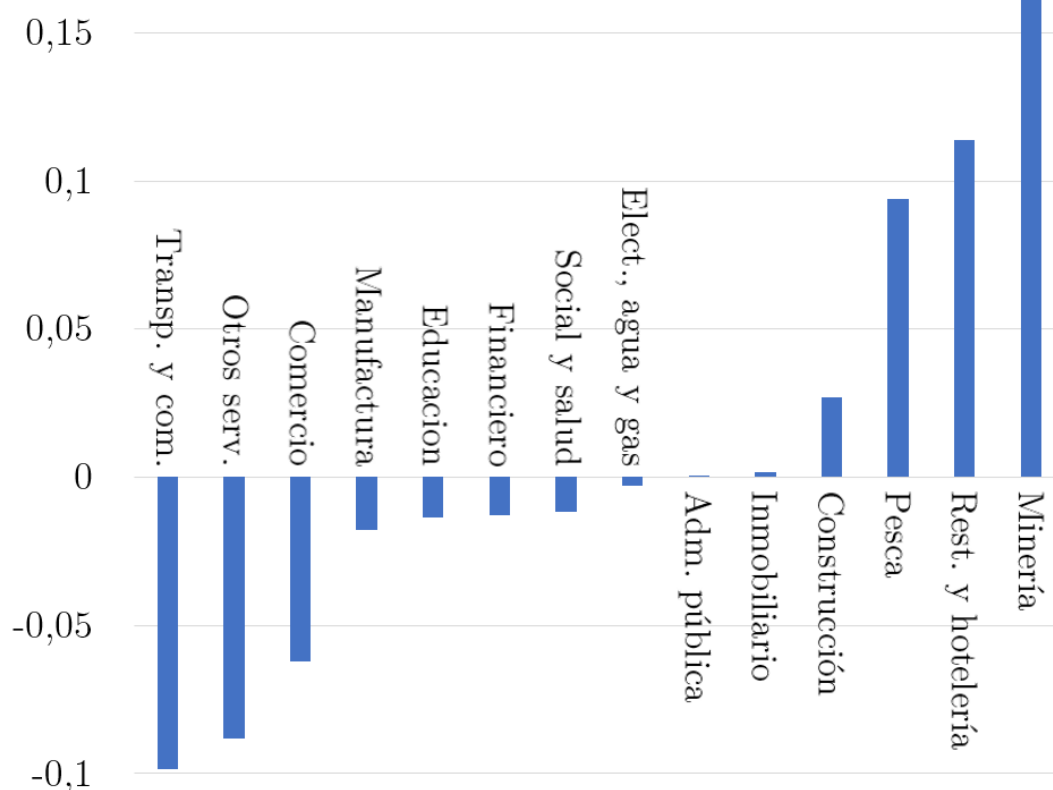
En tercer lugar, se destaca que un mayor nivel educacional es requerido en vacantes publicadas por empresas que poseen políticas salariales más generosas y cuyos cargos ofrecidos tienden a tener remuneraciones relativamente mayores. Se puede observar que existe cierto grado de coherencia entre los resultados asociados a estimaciones kernel no-paramétricas y la descomposición de Gelbach. En general, el acceso a cargos y a empresas de mayores remuneraciones generan retornos más grandes a medida que se observan los niveles educacionales más avanzados. Este resultado parece ser intuitivo si se entiende la educación como una llave de acceso a empresas y cargos ligados a salarios más altos.

En cuarto lugar, se destaca que las políticas salariales de las empresas se parecen más al momento de ofrecer cargos con bajos requerimientos educacionales. Una explicación intuitiva de este fenómeno sería que los empleadores consideran que la fuerza de trabajo es más homogénea para trabajos que requieren niveles educativos bajos. Asumiendo que diferencias notorias de habilidades en este tipo de cargos no tendrían mayores repercusiones en la rentabilidad de sus empresas, los empleadores no tendrían los mismos incentivos a diferenciarse en cuanto a sus ofertas de salarios, como sí es el caso de los niveles educativos superiores. De manera similar, se observa que los efectos fijos de los cargos son más dispersos para niveles educacionales más altos. Se puede decir que una mayor formación educacional es requerida para cargos de trabajo más heterogéneos en términos de rentabilidad salarial (una persona educada puede esperar salarios en un rango relativamente más amplio que una persona menos educada).

Anexos

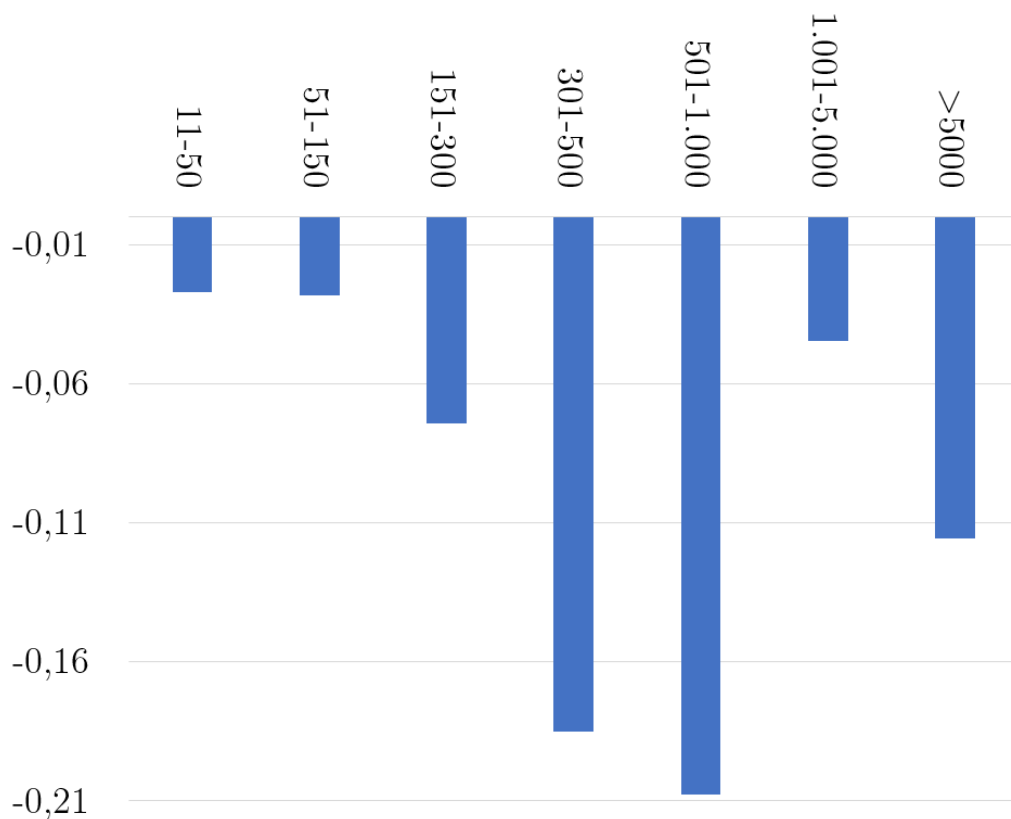
Figuras

Figura 5.1: Retornos estimados para cada industria sobre los efectos fijos de las empresas.



Nota: Este gráfico corresponde coeficientes estimados para las industrias, reportados en la tabla 5.6. En dichas estimaciones, la categoría base es la industria de agricultura y silvicultura. La variable dependiente del modelo correspondiente a este gráfico son los efectos fijos de empresas. Estos fueron estimados al regresionar –a nivel de vacantes– los logaritmos de los salarios reales por hora en función de los niveles educacionales, la experiencia requerida (y sus potencias de hasta orden 4), las industrias, la palabras presentes en las descripciones de los cargos y –lógicamente– las empresas.

Figura 5.2: Retornos estimados para cada categoría de tamaños de firmas (medidos en números de trabajadores) sobre los efectos fijos de las empresas.



Nota: Este gráfico corresponde coeficientes estimados para los tamaños de empresas, reportados en la tabla 5.6. En dichas estimaciones, la categoría base es el tamaño dado el hecho de tener entre 1 y 10 trabajadores. La variable dependiente del modelo correspondiente a este gráfico son los efectos fijos de empresas. Estos fueron estimados al regresionar –a nivel de vacantes– los logaritmos de los salarios reales por hora en función de los niveles educacionales, la experiencia requerida (y sus potencias de hasta orden 4), las industrias, la palabras presentes en las descripciones de los cargos y –lógicamente– las empresas.

Tablas

Tabla 5.1: Ofertas del sitio web para empleadores que quieran publicar avisos.

	1 aviso	Pack 3 avisos	Pack 5 avisos	Pack 10 avisos
Cantidad de avisos	1	3	5	10
Descargas de CV	0	1	2	3
Precio antes de impuestos	\$69.900	\$194.990	\$292.000	\$474.215
Precio después de impuestos	\$83.181	\$232.038	\$347.480	\$564.316
Cuentas de usuario	2	3	4	5

Nota: Actualizado al 27 de julio de 2019. Fuente: <https://gestion.trabajando.cl/companies/planSelection>.

Tabla 5.2: Resultados de tests de igualdad entre los retornos a la educación estimados con datos de vacantes TBN y flujo ponderado ESI.

Nivel educacional	Modelo	(1)	(2)	(3)	(4)	(5)	(6)
Media CH	Estadístico F	49,6	103,3	54,9	31,3	35,2	16,1
	P-valor	0,00	0,00	0,00	0,00	0,00	0,00
Media TP	Estadístico F	1.401,9	114,4	104,4	116,2	103,2	138,5
	P-valor	0,00	0,00	0,00	0,00	0,00	0,00
Superior TP	Estadístico F	1.292,3	195,2	173,4	207,3	180,5	218,5
	P-valor	0,00	0,00	0,00	0,00	0,00	0,00
Universitaria	Estadístico F	1.577,2	92,5	52,6	71,9	57,3	79,9
	P-valor	0,00	0,00	0,00	0,00	0,00	0,00
Grados de libertad		(1,581.006)	(1,580.980)	(1,580.950)	(1,578.525)	(1,578.495)	(1,578.483)

Nota: Los datos de flujo ponderado ESI consideran los factores de expansión de la encuesta. Los tests que aquí se presentan se hicieron en base a las estimaciones de la tabla 1.8.

Tabla 5.3: Distribución por sectores industriales de vacantes TBN para empresas con al menos uno y dos avisos.

Industria	Todas		Al menos dos avisos	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Agricultura y silvicultura	3.413	0,66	3.091	0,62
Pesca	331	0,06	324	0,06
Minería	4.127	0,79	4.012	0,8
Manufactura	42.002	8,08	40.260	8,01
Electricidad, agua y gas	8.456	1,63	8.143	1,62
Construcción	7.549	1,45	6.869	1,37
Comercio	132.914	25,58	128.979	25,67
Restaurantes y hotelería	4.675	0,9	4.186	0,83
Transporte y comunicaciones	53.485	10,29	52.132	10,38
Financiero	13.205	2,54	12.879	2,56
Inmobiliario	80.947	15,58	76.029	15,13
Administración pública	3.792	0,73	3.735	0,74
Educación	16.425	3,16	15.660	3,12
Social y salud	14.371	2,77	13.606	2,71
Otros servicios	133.840	25,76	132.542	26,38
Total	519.532	100	502.447	100

Nota: Información correspondiente a datos de vacantes TBN considerados en la muestra de este estudio.

Tabla 5.4: Distribución regional de vacantes TBN para empresas con al menos uno y dos avisos.

Región	Todas		Al menos dos avisos	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
De Tarapacá	337	0,06	246	0,05
De Antofagasta	966	0,19	730	0,15
De Atacama	304	0,06	199	0,04
De Coquimbo	182	0,04	70	0,01
De Valparaíso	2.817	0,54	2.329	0,46
Del Libertador General Bernardo O'Higgins	834	0,16	687	0,14
Del Maule	634	0,12	520	0,1
Del Biobío	2.530	0,49	2.173	0,43
De La Araucanía	400	0,08	321	0,06
De Los Lagos	301	0,06	197	0,04
De Aysén del Gral. Carlos Ibáñez del Campo	31	0,01	27	0,01
De Magallanes y de la Antártica Chilena	131	0,03	114	0,02
Metropolitana de Santiago	424.532	81,71	413.535	82,3
De Los Ríos	156	0,03	53	0,01
De Arica y Parinacota	9	0	4	0
Extranjera	5.578	1,07	5.435	1,08
Sin información	79.790	15,36	75.807	15,09
Total	519.532	100	502447	100

Nota: Información correspondiente a datos de vacantes TBN considerados en la muestra de este estudio.

Tabla 5.5: Distribución por tamaños de empresas (medidos en números de trabajadores) de vacantes TBN para empresas con al menos uno y dos avisos.

Tamaño	Todas		Al menos dos avisos	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
1-10	172.969	34,42	168.636	34,62
11-50	120.110	23,9	115.638	23,74
51-150	47.532	9,46	45.412	9,32
151-300	28.155	5,6	26.851	5,51
301-500	65.216	12,98	64.560	13,25
501-1000	36.755	7,31	36.295	7,45
1001-5000	27.637	5,5	26.387	5,42
5000	4.163	0,83	3.348	0,69
Total	502.537	100	487.127	100

Nota: Información correspondiente a datos de vacantes TBN considerados en la muestra de este estudio.

Tabla 5.6: Estimaciones obtenidas al regresionar los efectos fijos de las firmas en función de sus respectivas industrias y tamaños (en número de trabajadores).

Regresores		Coefficiente
Industria	Transporte y comunicaciones	-0,099***
	Otros servicios	-0,088***
	Comercio	-0,062***
	Manufactura	-0,018***
	Educación	-0,013***
	Financiero	-0,013***
	Social y salud	-0,012***
	Electricidad, agua y gas	-0,003
	Administración pública	0,000*
	Inmobiliario	0,002
	Construcción	0,027***
	Pesca	0,094***
	Restaurantes y hotelería	0,114***
Minería	0,164***	
Tamaño	11-50	-0,027***
	51-150	-0,028***
	151-300	-0,074***
	301-500	-0,185***
	501-1.000	-0,208***
	1.001-5.000	-0,045***
	5.000	-0,116***
Constante	0,009**	
R^2	0,1806	
Número de obs.	487.127	

Nota: Los datos utilizados corresponden a información auto-descriptiva ingresada por los empleadores en el sitio web. Las unidades de análisis de esta regresión son las vacantes, no las firmas. Esto explica la cantidad de observaciones y que el intercepto estimado mediante la regresión no sea cero a pesar de la normalización realizada, que exige que a nivel de empresas el promedio de los efectos fijos sea cero. Los estimadores fueron obtenidos mínimos cuadrados ordinarios.

Bibliografía

- [1] Jacob A Mincer et al. Schooling, experience, and earnings. *NBER Books*, 1974.
- [2] James Joseph Heckman, Lance J Lochner, and Petra E Todd. Fifty years of mincer earnings regressions. Technical report, IZA Discussion paper series, 2003.
- [3] Zvi Griliches. Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society*, pages 1–22, 1977.
- [4] Joshua D Angrist and Alan B Keueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.
- [5] John Bound, David A Jaeger, and Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450, 1995.
- [6] John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.
- [7] David Card. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160, 2001.
- [8] Thomas Lemieux. The “mincer equation” thirty years after schooling, experience, and earnings. In *Jacob Mincer a pioneer of modern labor economics*, pages 127–145. Springer, 2006.
- [9] Aniela Maria Wirz. Private returns to education versus education spill-over effects. *Empirical Economics*, 34(2):315–342, 2008.
- [10] Pedro S Martins and Jim Y Jin. Firm-level social returns to education. *Journal of Population Economics*, 23(2):539–558, 2010.
- [11] Jonah B Gelbach. When do covariates matter? and which ones, and how much? *Journal of Labor Economics*, 34(2):509–543, 2016.
- [12] Jonah Gelbach. B1X2: Stata module to account for changes when X2 is added to a base model with X1. Statistical Software Components, Boston College Department of Economics, March 2014.

- [13] Ioana Marinescu and Ronald Wolthoff. Opening the black box of the matching function: The power of words. Technical report, National Bureau of Economic Research, 2016.
- [14] Ana Rute Cardoso, Paulo Guimaraes, Pedro Portugal, and Hugo Reis. The returns to schooling unveiled. *IZA Discussion Papers*, 11419, 2001.
- [15] Stefano Banfi and Benjamin Villena-Roldan. Do high-wage jobs attract more applicants? directed search evidence from the online labor market. *Journal of Labor Economics*, 37(3):715–746, 2019.
- [16] Stefano Banfi, Sekyu Choi, and Benjamin Villena-Roldán. Deconstructing job search behavior. *Available at SSRN 3323545*, 2019.
- [17] Spence Michael. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.
- [18] George A Akerlof. Labor contracts as partial gift exchange. *The quarterly journal of economics*, 97(4):543–569, 1982.
- [19] Alan B Krueger and Lawrence H Summers. Efficiency wages and the inter-industry wage structure. *Econometrica: Journal of the Econometric Society*, pages 259–293, 1988.