



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

TESIS DE GRADO
MAGISTER EN ECONOMIA

Dosque Concha, Pedro Enrique

Julio, 2017



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

**Migración interna en Chile a nivel regional en la Encuesta CASEN 2015:
Efecto de los ingresos, las comodidades y las características individuales**

Pedro Enrique Dosque Concha

Comisión

Emilio Depetris-Chauvin

José Díaz

Jeanne Lafortune

Rolf Lüders

José Tessada

Gert Wagner

Santiago, Julio de 2017

Migración interna en Chile a nivel regional en la Encuesta CASEN 2015: Efecto de los ingresos, las *comodidades* y las características individuales

Pedro Dosque¹

Resumen

La literatura sobre migración interna entre regiones en Chile se basa principalmente en el análisis de estadística descriptiva de datos censales, mientras que los pocos trabajos que realizan análisis con técnicas econométricas para la decisión de migración sólo incluyen variables a nivel agregado, dejando de lado los determinantes individuales, y en general sólo con características de las economías regionales. Sin embargo, los modelos demográficos apuntan a tres tipos de factores relevantes en la migración interna: las características de las regiones, las características del proceso migratorio en sí y las características personales de los migrantes. Por esto, en este trabajo realizamos estimaciones con microdatos de migración de la encuesta CASEN 2015 y variables de características económicas, infraestructura y comodidades a nivel regional que nos permiten incluir estos tres tipos de factores en una estimación econométrica. Además, entre estas características incluimos ingresos estimados para cada individuo para cada región del país considerando sesgos de selección en migración y en trabajo. Los resultados muestran que mayores ingresos potenciales tienen un efecto de atracción de migración interna, los costos de migración un efecto negativo; el ser mujer, tener educación superior, tener menor edad y estar sin pareja tienen en general un efecto positivo en la migración hacia la región metropolitana respecto al resto de las regiones. Mientras que las variables relacionadas con el tamaño de la región en términos demográficos, de infraestructura y diversidad económica tienen efectos negativos en la migración, lo que podría indicar una tendencia hacia la migración de ciudades de tamaño medio que se están desarrollando económicamente y en infraestructura y comodidades.

¹ Tesis para optar al grado de Magíster en Economía, Pontificia Universidad Católica de Chile, Instituto de Economía. Trabajo realizado en el Seminario de Tesis de Magíster EH Clio Lab, Instituto de Economía UC. Agradezco especialmente los comentarios, sugerencias y ayuda continua de los profesores José Tessada y Gert Wagner en el desarrollo de esta investigación. Agradezco también los comentarios y sugerencias de los profesores Emilio Depetris-Chauvin, José Díaz, Jeanne Lafortune y Rolf Lüders. Cualquier error u omisión es de mi completa responsabilidad. Dudas, comentarios y sugerencias a pedro.dosque@gmail.com

I. Introducción

La migración interna es definida desde la demografía como el movimiento de población que cruza un límite geográfico o administrativo definitorio dentro de un mismo país en un periodo determinado de tiempo, involucrando un cambio en el lugar de residencia habitual (INE, 2007). En Chile los determinantes de este tipo de migración han sido poco estudiados en economía a pesar de que si consideramos, por ejemplo, la migración entre regiones del país, hubo alrededor de 880.000 migrantes entre éstas durante los 5 años anteriores al Censo de 1992 y cerca de 783.000 en los 5 años anteriores del Censo de 2002 -esto es, en torno a 7% y 5% de la población chilena respectiva para 1992 y 2002-. Si extendemos el periodo de estudio, estas cifras son aún mayores. En la Encuesta CASEN del 2015, un 22,5% de la población mayor de edad no vivía en la misma región que vivía al momento de nacer. Por tanto, la migración interna es un fenómeno demográfico de una escala importante.

Por otro lado, la migración interna también es relevante dentro de la teoría económica porque la migración es a nivel individual una estrategia de maximización del retorno al capital humano en donde se busca el lugar donde haya más retornos a las habilidades que uno posee, y porque es a nivel agregado un mecanismo para corregir los diferenciales de ingresos que ocurren entre regiones debido a diferencias en las productividades marginales del trabajo en cada región *-ceteris paribus-*. Es decir, la migración interna es una expresión del mercado del trabajo tanto en la optimización de la utilidad individual, como también dentro de la asignación óptima de recursos de un país.

También es relevante en el ámbito socioeconómico ya que, por un lado, las decisiones de migración dependen de -y afectan a- variables como los ingresos esperados, el desempleo, el nivel de pobreza, entre otras, en el lugar al que se migra y en el lugar de residencia original. Además la migración es causada por otros factores que influyen la decisión de residencia y trabajo, como la infraestructura y las comodidades regionales -por ejemplo, la existencia y calidad de hospitales, colegios, caminos, entre otros-, ya que afectan la calidad de vida del individuo y por tanto la utilidad esperada de éste en cada región.

Respecto al análisis de la migración interna regional de Chile, los pocos estudios empíricos sobre este tema en el país se han enfocado en el análisis de determinantes como los ingresos, el desempleo y los costos de migración en la decisión de migración interna, dejando de lado las comodidades (*amenities* en la literatura en inglés) y la infraestructura de las regiones como posibles determinantes de la migración y del “atractivo” de una región. Estos últimos podrían ser factores que influyen en la migración, ya que, por ejemplo, una persona podría aceptar hacer el mismo trabajo que actualmente realiza en una región *g* a un menor salario en otra región *j* si es que ésta última tiene mejores comodidades e infraestructura que la primera. En el caso de Chile éstas podrían tener impacto en la migración considerando por ejemplo que Soto y Torche (2004) encuentran que las “variables económicas” – específicamente, salarios y desempleo - no pueden explicar gran parte de la migración entre

regiones de Chile en los periodos 1987-1992 y 1997-2002, como sí lo hacían para los datos censales de años anteriores.

Por esto, lo que realizamos en este trabajo es estimar los efectos en la migración interna tanto de factores de las economías regionales -salarios, desempleo, nivel de pobreza-, como de otros posibles factores como la provisión de educación y salud, la infraestructura, la seguridad, la contaminación y la conectividad, para establecer así si existen determinantes más allá de los ocupados por la literatura tradicional en la migración regional dentro de Chile.

Una explicación de la falta de interés en el análisis de migración interna en Chile nombrado al inicio de esta introducción podría ser la ausencia de datos desagregados para su estudio, ya que hace menos de una década atrás no existían bases de datos para el país con datos que tuvieran a la vez información de migración interna individual, así como información económica y de características personales de los individuos.² Debido a esto, en Chile la literatura sobre migración interna ha realizado sus estimaciones exclusivamente con datos de flujos de migración agregados a nivel regional y no con decisiones individuales de migración.

Pero en las estimaciones que sólo consideran datos agregados no se puede considerar las características individuales que influyen en la decisión de migración. Por ejemplo, los datos del Censo del 2002 muestran que los inmigrantes a la Región Metropolitana tienen mayor educación que los que emigran de la misma región y que los que se quedan en esa región (no migrantes), lo cual podría deberse a que existen diferencias en los retornos a la educación entre regiones. Estas diferencias no se pueden apreciar con datos agregados en que se toma el ingreso promedio regional como variable explicativa en vez de los ingresos individuales. Por otro lado, también faltaría comprobar si los resultados a nivel agregado se mantienen al considerar el comportamiento individual en la migración, lo cual es importante ya que los coeficientes obtenidos con datos agregados no representan resultados de ecuaciones estructurales provenientes de teorías de comportamiento individual a menos que la población sea homogénea -, que no lo es en realidad-, ni tampoco podría explicar la migración a lugares poco atractivos económicamente, ya que si todas las personas son iguales, todas preferirían ir a la misma región (Robinson y Tomes (1982)).

Por lo que, a diferencia de la literatura anterior (por ejemplo, Aroca (2004) o Soto y Torche (2004)), en este estudio realizamos estimaciones con una muestra de corte transversal con datos a nivel de persona. Para esto ocupamos la Encuesta CASEN 2015, la cual incluye datos tanto de migración interna como de ingresos.³ Estos datos individuales además nos permiten

² Por ejemplo, Soto y Torche (2004) afirman que solo hay disponibles datos agregados, por lo que deben realizar estimaciones con un individuo “promedio” por región.

³ La Encuesta CASEN comenzó a incluir datos de migración desde el año 2011 en adelante (Encuestas de 2011, 2013 y 2015), por lo que no estaba disponible en el tiempo que se realizaron los estudios anteriores de migración interna para Chile. También cabe destacar que el periodo entre las tres encuestas es muy corto como para realizar un estudio comparativo entre las encuestas sobre los efectos de comodidades e infraestructura en la migración.

estimar retornos a la educación por región considerando el hecho que existe autoselección (*sorting*) entre regiones -es decir, que la localización de trabajadores en una región no es aleatoria ya que estos tomaron una decisión previa de migración, como en el ejemplo del párrafo anterior en que a la Región Metropolitana migra más gente con mayor educación, por lo que una estimación considerando solamente los ingresos de los trabajadores presentes en una región tendría un sesgo de autoselección por migración (además del sesgo de autoselección por la decisión de trabajar o no trabajar), -, y así obtener una estimación de los retornos a la educación insesgados. Esto también nos permite encontrar el efecto insesgado de los ingresos en la decisión de migración.

Más concretamente, lo que hacemos empíricamente en este estudio es, primero, estimar por región los ingresos esperados si el individuo viviera allí para cada individuo de la muestra con una ecuación de Mincer modificada para considerar los sesgos de autoselección en la decisión de migrar y en la decisión de trabajar. En esta parte nos basamos en el método de corrección de sesgos de selección muestral en modelos de elección discreta⁴ múltiple desarrollado por Dahl (2002), el cual ha sido utilizado en los últimos años para realizar estimaciones de migración interna corregida por el sesgo de selección muestral por migración. Lo que hacemos en este método es agregar a la ecuación de Mincer una función de corrección que depende de la probabilidad *ex-ante* que tenía el individuo de migrar hacia la región que migró y en la probabilidad *ex-ante* de mantenerse en su región originaria. Estas probabilidades *ex ante* las calculamos como probabilidades frecuentistas en grupos (*celdas*) de individuos de características similares,⁵ los cuales suponemos que tienen beneficios y costos similares de migrar.

Luego, en un segundo paso, estimamos mediante un modelo discreto múltiple Conditional Logit los efectos de los ingresos esperados estimados en el primer paso, de los costos de migración y de características individuales y de las regiones -comodidades, infraestructura, tamaño de la población y características económicas- en la decisión individual de migrar entre regiones de Chile entre i) el año de nacimiento y el año 2015 y ii) entre los años 2010 y 2015. En este tipo de modelo la variable explicada es una variable binaria que toma el valor 1 si el individuo migró hacia la región y 0 si no, y cada individuo “tendrá” N observaciones, siendo N el número de regiones, con una observación con valor 1 -la región en la que vive en 2015- y (N-1) observaciones con valor 0 -las regiones en las que decidió no vivir-. Por su parte, las variables explicativas en este tipo de modelo pueden ser variables relacionadas a

⁴ Por modelo de elección discreta nos referimos a modelos en que la variable explicada es una elección entre dos o más alternativas. En el primer caso hablamos de un modelo discreto binario y en el segundo de un modelo discreto múltiple. En el caso de la migración podemos tener, por ejemplo, modelos en que se decida i) sí migrar o ii) no migrar a una región; o modelos en que se decida migrar entre varias regiones (alternativas).

⁵ Es decir, calculamos la proporción de individuos dentro de cada celda que elige mantenerse en su región y las proporciones de individuos que deciden migrar a cada otra región.

características de las regiones (*específicas a la alternativa*) como a características de las personas (*específicas al caso*).

De esta manera esperamos captar los tres aspectos que afectan a la migración interna según el estudio demográfico de Schiappacasse *et al.* (2001): i) los factores regionales “coyunturales” (ciclos de la economía, por ejemplo) y “estructurales”, es decir, diferencias en características de largo plazo entre las regiones (como tipos de industrias, infraestructura, clima, entre otras); ii) las características intrínsecas del proceso migratorio (volumen y distancia de migración, es decir los costos de migración), y iii) las características propias de la población migrante, como sexo, edad o nivel educativo, las cuales afectarían la selectividad de la migración. De aquí el título de este trabajo: buscamos determinar el efecto de los ingresos y costos de migración como en la literatura pasada, pero también de características de infraestructura y comodidades de las regiones y de las características individuales en la migración interna de Chile.

Lo que buscamos en este trabajo de investigación, mediante un modelo de optimización que considera las utilidades posibles en cada región en la decisión de migración, es responder cuáles son los determinantes que han llevado a la migración dentro del país en los últimos años, si influyen otras características regionales además de las económicas en la atracción de migrantes, qué regiones han tenido mayor atracción de migrantes en los últimos años y cómo afectan las características personales en este tipo de migración en Chile. Por otro lado, esta pregunta sobre los determinantes de la migración interna también es relevante en el sentido que es un insumo para señalar variables que alteren el atractivo relativo entre regiones en un país muy centralizado demográfica y económicamente como lo es Chile.

Los resultados encontrados en el trabajo señalan que si no se consideran las características personales de los individuos, un aumento en los ingresos potenciales en una región no tiene un efecto estadísticamente significativo en la migración entre los años 2010 y 2015, pero cuando incluimos este tipo de características, un aumento en los ingresos potenciales sí atrae una mayor migración interna hacia la región. Para el periodo más largo entre el año de nacimiento y el año 2015, al incluir las características personales, también aumenta el coeficiente asociado a los ingresos potenciales en la región. Por otra parte, los resultados señalan que el efecto de los ingresos esperados en la migración es sobreestimado si no se considera la corrección por sesgo de selección en migración y en trabajo.

Respecto a las variables relacionadas a los costos de migración tienen un efecto negativo, por lo que los lugares más alejados a la región de origen son menos atractivos para migrar. Mientras que respecto a las características personales, el ser mujer, tener educación superior, tener mayor edad y estar sin pareja hacen más probable la migración hacia la Región Metropolitana respecto al resto de las otras regiones. Pero existen particularidades hacia las regiones extremas del país, por ejemplo, el ser más joven y ser hombre aumenta la probabilidad de migrar a estas regiones. Por otro lado, las variables relacionadas con el

tamaño de la región en términos demográficos, de infraestructura y diversidad económica tienen efectos negativos en la migración, lo que podría indicar una tendencia hacia la migración a ciudades de tamaño medio que se están desarrollando económicamente y en infraestructura y *comodidades*. Por último, nuestras variables de *comodidades* muestran un efecto positivo en la decisión de migrar hacia las regiones con mayores *comodidades*.

El trabajo se organiza de la siguiente manera. En la sección II presentamos una revisión de literatura de migración en Chile y de los trabajos que han ocupado la metodología que ocupamos aquí. En la sección III exponemos el modelo teórico y la metodología empírica, mientras que en la sección IV mostramos los datos utilizados. En la sección V presentamos la implementación y los resultados del estudio. Y finalmente en la sección VI concluimos.

II. Revisión de literatura

II.1 Literatura sobre migración interna en Chile

Respecto al tema de este estudio, éste está principalmente relacionado con la literatura de determinantes de la migración interna en Chile, la cual se caracteriza por utilizar datos de migración agregados y enfatizar las características económicas de las regiones en sus estimaciones. Los primeros trabajos sobre Chile en esta área fueron realizados en las décadas de 1970 y 1980. Por ejemplo, Coeymans (1983) realiza una estimación de los determinantes de la migración de orígenes rurales a destinos urbanos dentro y entre provincias⁶ con datos agregados del Censo de 1970 y los estima ocupando modelos de regresión lineal – mínimos cuadrados ordinarios considerando heterocedasticidad-, en vez de un modelo de elección discreta.

Más recientemente, a inicios del presente siglo, podemos encontrar tres publicaciones más sobre esta problemática que realizan estimaciones econométricas, mientras que el resto de los análisis se basan en la revisión de estadística descriptiva obtenida de los Censos. Estos trabajos ocupan modelos de elección discreta múltiple para estimar la migración agregada entre regiones. Sin embargo, continuaron ocupando datos a nivel agregado y como variables explicativas solamente características de las economías regionales (desempleo, PIB, nivel de pobreza, etc.) y de tamaño de la población.

Los primeros de estos tres papers son Aroca (2004), que realiza su estudio con datos de los Censos de 1992 y 2002 y ocupando la misma metodología que Aroca *et al.* (2001), los cuales utilizan datos de los Censos de 1982 y 1992. En estos dos trabajos plantean un modelo de corte transversal sobre los factores determinantes de la probabilidad de migrar de una región *i* a una región *j* con datos de flujos de migración agregados a nivel de región. Sustentan su

⁶ En ese periodo las provincias eran las divisiones administrativas más grandes de Chile. No son las provincias actuales.

modelo con razones y teoría de migración a nivel individual, sin embargo, realizan sus estimaciones con migrantes “promedios” a nivel de región y no decisiones de migración individuales. Su estimación final es una regresión MCO de la razón entre la probabilidad de migrar a j y la probabilidad de quedarse en su región sobre variables económicas regionales, distancia y tamaño de las regiones. Estas probabilidades que ocupan es el porcentaje de personas de la región de origen que migran a la región j , sin diferenciar la probabilidad de migrar según las características de las personas (como hacemos en este trabajo).

El tercero de los estudios es Soto y Torche (2004) y también se basa en la metodología de Aroca *et al.* (2001). Su principal aporte es que, aunque ocupan menos variables explicativas que su predecesor, agregan una variable de subsidios habitacionales -es decir, una variable de políticas públicas- para analizar su impacto en la migración entre regiones, ya que notan que desde el Censo de 1992 la migración se desliga del PIB *per capita* regional, a la vez que no pueden explicar los flujos migratorios considerando solamente las características de las economías regionales. En sus resultados encuentran que estos subsidios tienen un impacto negativo tanto en 1992 como en 2002, pero mucho mayor en el segundo periodo. Sin embargo, la cantidad de subsidios totales son similares en ambos periodos (tanto a nivel nacional como a nivel regional).

También cabe volver recalcar que estos tres papers, y la literatura general sobre migración interna en Chile, se centra en indicadores de las economías regionales como señala Rodríguez y González (2006). Ellos destacan que ha existido un desinterés en la distribución de la población en Chile por el supuesto que a mediano o largo plazo los flujos migratorios seguirán a las regiones dinámicas económicamente hablando, dejando de lado otras razones de atracción de migración. Sin embargo, la literatura demográfica señala que existen más razones de migración, como ya señalamos en la introducción (Schiappacasse *et al.* (2001)). En nuestro estudio pretendemos acercarnos a la literatura demográfica incluyendo variables que capten los diferentes aspectos considerados en ésta.

II.1 Literatura sobre estimación de determinantes de migración con microdatos

Metodológicamente este trabajo está relacionado con la literatura de estimación de determinantes de la migración que ocupa datos individuales, permitiendo así incorporar la heterogeneidad de los individuos y considerar los sesgos de estimación en los ingresos debido a la migración. Inicialmente esta literatura consideró modelos de elección discreta binaria, como en Robinson y Tomes (1982). Ellos ocupan un modelo Probit para la estimación de salarios en nueve regiones de Canadá, en donde dividen a la población de cada región entre los que nacieron allí y los que inmigraron hacia esa región, sin diferenciar por su origen, controlando por sesgo de selección en la migración para obtener los salarios potenciales para los migrantes si no se hubieran movido y para los no migrantes si se hubieran movido de región.

Pero el problema del sesgo de selección muestral por la migración se vuelve más difícil de corregir si se consideran varias regiones de destino, a menos que se establezcan supuestos muy restrictivos al proceso de selección. Dahl (2002) ofrece una solución a este problema aprovechando el desarrollo en técnicas semiparamétricas para realizar una variación del modelo de Lee (1983), el cual asume que la opción que importa en una decisión múltiple es la primera mejor opción -en nuestro caso, la región hacia donde efectivamente migró el individuo-, la cual podemos observar, y ésta es ocupada para corregir el sesgo de selección en modelos multinomiales. A esta idea de Lee, Dahl incorpora las investigaciones de estimaciones semiparamétricas de *single-index models* desarrolladas entre las décadas de 1980 y 1990, permitiéndole que la función corrección del sesgo de selección muestral tome la forma de una función desconocida que depende sólo de la probabilidad de seleccionar la primera mejor opción. Este método simplifica la estimación -al basarse en la idea de Lee (1983) para reducir la dimensionalidad del problema con la utilización de “la primera mejor opción”⁷- a la vez que ocupa supuestos menos restrictivos -por la utilización de métodos semiparamétricos que permite evadir la necesidad de especificar la distribución conjunta de los errores de las ecuaciones de ingresos y de selección para obtener la función corrección-.

Esta literatura también permite explicar la migración interna de un país a partir de diferencias individuales en el ingreso esperado entre regiones debido a las características de cada persona, como educación, experiencia, sexo, etc.; en vez de diferenciales de ingresos esperados promedios entre regiones como en Aroca *et al.* (2001), Aroca (2004) y Soto y Torche (2004). Agregar características individuales es importante, ya que permite entender por qué existe migración desde y hacia una región en vez de ser solamente unilateral, *ceteris paribus*, a diferencia de un modelo con datos agregados en donde todas las personas de una región “enfrentan” un mismo ingreso esperado promedio y por tanto migran en la misma dirección. En la sección siguiente explicamos más profundamente la metodología de Dahl (2002).

En lo que respecta a las aplicaciones empíricas, la misma publicación de Dahl (2002) aplica su nuevo método en la estimación de determinantes de la migración interna entre los estados de Estados Unidos, donde corrobora la existencia de un sesgo de selección en el salario causado por la migración en el país. Además, encuentra que la diferencia en *amenities* entre estados tiene un efecto en la decisión de migración. Otros trabajos posteriores también utilizan este mismo método para la estimación de migración interna. Este es el caso de Bayer *et al.* (2009) que realizan una estimación de los efectos de la calidad del aire en la localización de los hogares entre diferentes áreas metropolitanas de Estados Unidos; y de De Vroyer *et al.* (2009) que estudian la influencia de los retornos a la educación y los diferenciales de salario en la migración entre 7 capitales de la WAEMU⁸ en 2001 y 2002. También Bertoli *et al.*

⁷ En vez de utilizar N ecuaciones de selección al tener N regiones, sólo utilizamos una variable aleatoria, el *maximum order statistic*, el cual está asociado a “la primera mejor opción” (Ver Dahl (2002) sección 3.1.1).

⁸ WAEMU es la West Africa Economic and Monetary Union.

(2012) utilizan esta metodología para corregir por sesgo de selección en la decisión de migración internacional de ecuatorianos entre 1999 y 2006, pero además corrigen por sesgo de selección en la decisión de trabajar para incluir a las mujeres en la estimación a diferencia de Dahl (2002) que no lo hace, sin embargo, solo ocupan un lugar de origen y no múltiples en su modelo. Por último, Xing y Zhang (2017) utilizan este método para obtener salarios estimados en su estudio sobre las preferencias de ciudades en China, en donde consideran tanto determinantes económicos como comodidades e infraestructura en una muestra de corte transversal de 2005.

Basamos la metodología de este trabajo principalmente en estas publicaciones mencionadas. Nuestra estimación de los ingresos esperados para un individuo en cada región se basa en Dahl (2002), pero incluimos a las mujeres en la muestra ocupando la adaptación que hace Bertoli *et al.* (2012) de este método. Mientras que para la estimación de los determinantes de la migración mediante un modelo Conditional Logit, nos apoyamos principalmente en Xing y Zhang (2017).

III. Modelo

Nuestro modelo empírico de migración interna consta de dos ecuaciones relacionadas, una de elección discreta múltiple de migración entre regiones del país y otra de ingresos ajustada por los sesgos de selección por migración y trabajo por cada región. A continuación presentamos de donde provienen tales ecuaciones.

III.1 El modelo de migración

Para la modelación de la migración interna entre regiones ocupamos un modelo de corte transversal con decisión discreta múltiple. Este modelo se fundamenta en un modelo estructural de migración que sigue las formulaciones de Bayer *et al.* (2009) y Xing y Zhang (2017).

Consideremos el siguiente modelo teórico en que cada individuo i tiene la opción de elegir migrar a N regiones posibles, y por tanto escoge la región que maximiza su utilidad, sujeto a su restricción presupuestaria. Por lo que cada individuo enfrenta el siguiente problema de maximización:

$$\max_{\{C, H\}} U_{ij} = C_{ij}^{\alpha_c} H_{ij}^{\alpha_h} e^{\beta_s \ln S_j + \sum_{k=1}^K \beta_k \ln X_{jk} + M_{ij} + \varepsilon_j + \eta_{ij}} \quad (1)$$

$$s. a. C_{ij} + p_j H_{ij} = I_{ij}$$

Donde C_{ij} es el consumo de un bien compuesto transable en la región j . Asumimos que su precio es el mismo en todo el país y lo normalizamos a 1. Por su parte, H_{ij} es el consumo de un bien compuesto no transable en la región j , y su precio es p_j ; mientras que I_{ij} es el ingreso del individuo i en j . En la función utilidad, S_j es el tamaño poblacional de la región j , el cual

refleja comodidades y beneficios de economías de aglomeración (de localización y urbanas) no observables. X_{jk} es un vector de tamaño k de características de cada región que afectan la utilidad de vivir en j . M_{ij} , por su parte, es el costo para i de migrar a j . Éste costo no lo consideramos únicamente como un costo de moverse de una región a otra que se paga una sola vez, sino que contiene también los costos de migración que se mantienen en el tiempo, como, por ejemplo, costos psicológicos por alejarse de la familia o de perder redes formadas en el lugar de origen. ε_j captura características no observadas de la región j y, por último, η_{ij} es el componente idiosincrático de la utilidad del individuo i .

Luego de resolver el problema del individuo, sustituir las demandas de los bienes transables y no transables en la función utilidad (1), simplificar los términos constantes⁹ y aplicar logaritmo natural, obtenemos la utilidad indirecta expresada de la siguiente manera:

$$V_{ij} = \ln I_{ij} + M_{ij} - \alpha_h \ln p_j + \beta_s \ln S_j + \sum_{k=1}^K \beta_k \ln X_{jk} + \varepsilon_j + \eta_{ij} \quad (2)$$

La ecuación (2) es la ecuación final de modelo teórico, sin embargo, hay algunas variables que no podemos observar directamente. Por lo que debemos realizar ciertos arreglos a la ecuación.

Primero, no podemos observar el precio del bien compuesto no transable.¹⁰ En este caso, supongamos que se podría estimar p_j a partir de características de la región j :

$$\ln p_j = \gamma_s \ln S_j + \sum_{k=1}^K \gamma_k \ln X_{jk} + \vartheta_j \quad (3)$$

Reemplazando (3) en (2) y sustituyendo $\delta_s = \beta_s - \alpha_h \gamma_s$, $\delta_k = \beta_k - \alpha_h \gamma_k$ y $\varphi_j = (\varepsilon_j - \alpha_h \vartheta_j)$ obtenemos

$$V_{ij} = \alpha \ln I_{ij} + M_{ij} + \delta_s \ln S_j + \sum_{k=1}^K \delta_k \ln X_{jk} + \varphi_j + \eta_{ij} \quad (4)$$

Segundo, no podemos observar el ingreso del individuo i en todas las regiones j , sino que sólo en la región en que vive y trabaja el individuo, por lo que debemos estimar el ingreso esperado para el individuo para el resto de las regiones. Por tanto, el ingreso será la suma del ingreso estimado (esperado) más un término de error:

$$\ln I_{ij} = \ln \hat{I}_{ij} + \mu_{ij}$$

\hat{I}_{ij} es estimado ocupando una ecuación del ingreso sobre las características de los individuos que habitan en una región para cada región. En la siguiente subsección explicamos más al respecto.

⁹ Simplificamos: $\left(\frac{\alpha_c}{\alpha_c + \alpha_h}\right)^{\alpha_c} \left(\frac{\alpha_h}{\alpha_c + \alpha_h}\right)^{\alpha_h} = A$ y $\alpha_c + \alpha_h = \alpha$

¹⁰ Una opción sería aproximar el precio de los bienes transables a través, por ejemplo, del precio del arriendo de viviendas ajustado por su calidad con los datos de la Encuesta CASEN.

Por su parte, aproximamos los costos de migración por la distancia geográfica -en logaritmo natural- entre las capitales regionales de las regiones de origen y llegada, ya que suponemos que a mayor distancia del lugar de origen, más difícil el contacto con familiares y conocidos de tal lugar, rompiéndose las redes sociales creadas. Además, que es más probable que el nuevo ambiente sea distinto al de origen. También agregamos una variable dummy “contigua” D_c que es igual a 1 si las regiones son contiguas y 0 si no, de esta manera esperamos considerar ciudades que están cerca de otra región en comparación a la capital regional. También agregamos una variable dummy “no contigua” D_{nc} que toma el valor 1 si la región a la que migra no es contigua.¹¹ Por tanto $M_{ij} = \delta_d D_{gj} + \delta_c D_c + \delta_{nc} D_{nc}$, siendo g la región de origen.

Por tanto, considerando $\epsilon_{ij} = \eta_{ij} + \mu_{ij}$, el modelo estimable es:

$$V_{ij} = \alpha \ln \hat{l}_{ij} + \delta_d D_{gj} + \delta_c D_c + \delta_{nc} D_{nc} + \delta_s \ln S_j + \sum_{k=1}^K \delta_k \ln X_{jk} + \varphi_j + \epsilon_{ij} \quad (5)$$

Esta ecuación la estimamos utilizando un modelo de elección discreta múltiple Conditional Logit. En este tipo de modelo la variable dependiente es una variable binaria que toma el valor 1 si el individuo migró hacia la región y 0 si no. A diferencia de otros modelos multinomiales, el Conditional Logit nos permite tener tanto variables que son específicas al individuo -por ejemplo, sexo o edad-, llamadas *case-specific*; como variables específicas a la opción elegida (región en nuestro caso) -por ejemplo, ingreso potencial en la región, médicos por cada 10.000 habitantes en la región, etc.-, llamadas *alternative-specific*.

Implícitamente en la estimación está el hecho que al elegir una región sobre otras, existe una mayor utilidad de migrar o quedarse allí que en las demás regiones. En un esquema simple, un individuo preferiría migrar a j sobre quedarse en su región g si:

$$Ingreso_g < Ingreso_j - \text{costos de migración}_{gj} + \text{Diferencial de localización}_{jg}$$

Es decir, un individuo tiene mayor utilidad en j que en su región original g si recibe mayores ingresos en j neto de los costos de migración (que son continuos en el tiempo), y considerando la utilidad que proporciona las diferencias regionales de infraestructura, comodidades y otras características específicas de la región, que en esta ecuación llamamos “Diferencial de localización” entre j y g . Nótese que esta estos son flujos que se realizan cada periodo, es decir, en cada periodo el individuo recibe la misma mayor utilidad por estar en j por sobre g . Esta es una simplificación que realizamos debido a que no poseemos datos sobre el ciclo de vida de los individuos.

Antes de realizar la estimación, sin embargo, hay que considerar que al estimar esta ecuación podrían surgir problemas de sesgo de variable omitida y sesgo de simultaneidad. El primer

¹¹ Si no incluyéramos esta segunda variable dummy, entonces la región de origen quedaría dentro del mismo grupo que las no contiguas en la regresión, ya que tiene un valor de 0 en D_c .

sesgo se produce porque el tamaño de la población S_j posiblemente se encuentra correlacionado con características de la región que afectan la utilidad del individuo, incluyendo las características no observables de la región en ε_j . Para solucionar este problema, ocupamos un rezago temporal del tamaño de la población, ya que ésta tiene una alta persistencia en el tiempo y no debería estar relacionado con las características actuales de la región si ocupamos un rezago de suficientes años atrás. Por su parte, el sesgo de simultaneidad es provocado por el hecho que nuestra variable dependiente, la migración, afectaría el tamaño de la población y, posiblemente, las características de la región (infraestructura y comodidades en nuestro caso), como sería el caso, por ejemplo, que un “shock” demográfico a una región obligue al gobierno a crear mayor infraestructura en tal región. Para solucionar este problema, también utilizamos variables rezagadas en el tiempo en caso de ser necesario, como mostramos en la sección siguiente al presentar los datos.

Otro enfoque para enfrentar estos problemas de sesgo es realizar la estimación con un “efecto fijo” por región. En este caso ocupamos un efecto fijo de región θ_j que incluye todas las características observables y no observables de la ecuación (5) que sólo dependen de j :

$$\theta_j = \delta_s \ln S_j + \sum_{k=1}^K \delta_k \ln X_{jk} + \varepsilon_j$$

Por lo que la ecuación a estimar es

$$V_{ij} = \alpha \ln \hat{I}_{ij} + \delta_d D_{gj} + \delta_c D_c + \delta_{nc} D_{nc} + \theta_j + \varepsilon_{ij} \quad (6)$$

De este modo eludimos el problema juntando todas las variables que podrían causar sesgo en θ_j . Aunque esta regresión (6) no nos permite observar el impacto de las variables regionales en la decisión de migración, si nos permite obtener estimaciones de los efectos de los ingresos y costos de migración sin sesgo para comparar con nuestra estimación de (5).

III.2 La ecuación de ingresos

Para obtener \hat{I}_{ij} estimamos una ecuación de Mincer modificada para considerar los sesgos de selección en el ingreso por migración y por trabajo. Si estimáramos los ingresos en una región considerando sólo los individuos que viven y trabajan en tal región, obtendríamos una estimación sesgada, ya que no estaríamos considerando el hecho que nuestra muestra es “seleccionada”. Esto se debe a que los individuos antes de trabajar en una región tienen que decidir primero vivir en tal región y luego decidir si trabajar o no.¹² Por lo que no observamos los ingresos de las personas que no decidieron vivir en tal región (maximizan su utilidad viviendo en otra región), y de los que viven en tal región, pero decidieron no trabajar

¹² Estamos suponiendo que todas las personas viven en la misma región donde trabajan.

(maximizan su utilidad no trabajando), por lo que no observamos que ingreso tendrían si trabajaran.

Esto es relevante en nuestro caso, ya que -en un modelo de decisión discreta múltiple- una persona al enfrentarse a la decisión de migrar tiene que considerar su ingreso “potencial” en cada región, pero este no se desprende simplemente del de las personas que ya están trabajando en esa región, porque estos últimos son los que ya tomaron la decisión de migrar y trabajar a tal región ya que así maximizaban su utilidad. Por lo que no considera, por ejemplo, a alguien que hubiera conseguido un ingreso bajo al migrar hacia esa región y que por esa razón no migró.

Para considerar el sesgo de selección en la migración ocupamos el método semiparamétrico de corrección de sesgo de selección desarrollado por Dahl (2002). En la práctica, en este método primero se divide la población de cada región en celdas mutuamente excluyentes definidas por características observables de los individuos. Se supone que individuos con características similares tendrán costos y beneficios de migrar similares, por lo que a partir de ellas podemos obtener las probabilidades -frecuentitas- de migrar a cada región para cada individuo. Más específicamente, para cada celda se calcula la proporción de individuos que eligen mantenerse en la misma región de origen g (*stayers*) y esa fracción es la probabilidad que un individuo i de esa celda no migre, que llamaremos p_{igg} . Mientras que la fracción de individuos en una celda que migran desde una región g a otra región j es la probabilidad de que un individuo i de esa celda migre a j , p_{igj} . Esta probabilidad es, de hecho, la probabilidad de su mejor opción de migración, ya que en los datos observamos hacia donde decidió migrar i ; y, siguiendo los supuestos de Dahl (2002), esta es la única información que necesitamos para corregir por el sesgo de selección en migración (*index sufficiency assumption*). Aunque también podemos relajar los supuestos, permitiendo la inclusión de otras “buenas” opciones de migración en la función corrección del sesgo, como por ejemplo, la probabilidad de retención -es decir, de no migrar- del individuo que migró (la cual también es observable).¹³ De esta manera obtenemos una estimación simple, y libre de distribución, de las probabilidades de migración.

¹³ En este caso estamos flexibilizando el supuesto ya que no definimos a la primera mejor opción como la única que contiene información relevante sobre como los diferenciales entre cada par de medias poblacionales de la utilidad que se obtiene en cada región posible de migración influyen en la distribución conjunta del término error de la ecuación de ingresos y del *maximum order statistic* de la ecuación de selección región donde vivir; y por tanto la única opción de la maximización de utilidad de la elección de región que importa para estimar el sesgo de selección muestral. Al agregar la probabilidad de retención estamos diciendo que la probabilidad de mantenerse en la región que vive también importa en el sesgo de selección muestral. También quisiéramos hacer notar que con respecto al método de Lee (1983), este es un supuesto menos restrictivo, ya que Lee (1983) asume que la distribución conjunta del término error de la ecuación de ingresos y del *maximum order statistic* de la ecuación de selección región donde vivir no dependen de los diferenciales entre cada par de medias poblacionales de la utilidad que se obtiene en cada región.

Los otros métodos para realizar una corrección de selección en la muestra en casos de varias opciones son los basados en modelo Logit condicional. Hay dos inconvenientes con este tipo de modelos. El primero es que requieren que se cumpla la propiedad de Independencia de alternativas irrelevantes, que puede ser un supuesto muy restrictivo. El otro inconveniente es que la corrección de sesgo de selección puede ser muy sensible a las desviaciones de la verdadera distribución conjunta de los términos de error de la ecuación de ingresos y de las ecuaciones de selección, lo cual genera un sesgo por asumir incorrectamente, por ejemplo, distribuciones conjuntas normales. Por su parte en Dahl (2002) evade la necesidad de especificar la distribución conjunta de los términos de error de estas ecuaciones ocupando el método semiparamétrico descrito más arriba.

En el método que ocupamos la variación en las probabilidades de migración son claves para la correcta identificación de los parámetros en la ecuación de ingresos, ya que lo que necesitamos es que individuos de diferentes características u orígenes tengan similares probabilidades de migración a una región. Por esto también existe un *trade-off* entre cantidad de celdas y precisión de estimación, ya que a más celdas, más parecidos los individuos dentro de la celda y más cantidad de probabilidades tendremos, pero a la vez habrá menos individuos en cada celda para calcular las probabilidades de migración de manera precisa.

En este trabajo aplicamos la variación que hace Bertoli et al (2012) para corregir simultáneamente por el sesgo causado por la decisión de migración y por la decisión de trabajo. Ellos en vez de calcular sólo la probabilidad de migrar de un individuo, calculan la probabilidad de migrar y trabajar en el lugar de llegada. Esto implica solamente hace una división más de cada celda entre las personas que trabajan y no trabajan.

Entonces, la regresión de ingresos a la Mincer modificada por sesgo de selección muestral para cada región j actual (es decir, N regresiones), dado que el individuo proviene de una región g ,¹⁴ con $j=g$ si el individuo no se mueve de región, es:

$$\ln y_{ij} = \alpha_j + educ_i \beta_j + exp_i \beta_j + exp_i^2 \beta_j + N'_i \delta_j + \sum_{j=1}^{12} \{M_{igj} \times \lambda_{g,j}^*(p_{igj}, p_{igg})\} + \omega_{ij} \quad (7)$$

Donde $\ln y_{ij}$ es el logaritmo natural de los ingresos del individuo en j , α_j una constante, $educ_i$ el mayor nivel de educación alcanzado por i , exp_i la experiencia potencial de i ,¹⁵ y N'_i es un vector de otros determinantes de los ingresos; mientras que M_{igj} es una variable dummy igual a 1 si i migró desde g a j , y 0 en caso contrario. El término $\lambda_{g,j}^*(\cdot)$ es la función

¹⁴ Nótese que para calcular las probabilidades se divide la población en sus regiones de origen, pero para la estimación de ingresos se divide a la población según su región de residencia.

¹⁵ Ocupamos experiencia potencial como proxy de la variable experiencia. La experiencia potencial como la edad menos los años de educación menos 5 años.

corrección del sesgo de selección muestral, con $p_{igj} = \Pr(M_{igj} = 1 | celda)$, la probabilidad de migrar de g a j para el individuo i dada la celda a la que pertenece.

Con el método expuesto hemos reducido las funciones corrección a ser N funciones dependientes de a los más dos probabilidades cada una. Sin embargo, sigue siendo un número muy grande de parámetros a estimar para la función corrección. Por lo cual ocupamos el mismo supuesto simplificador que utiliza Dahl (2002): $\lambda_{g,j}^*(\cdot) = \lambda_j^*(\cdot) \forall g \neq j$. Es decir, dada dos personas que eligieron migrar a j , sin importar su región de origen, sus sesgos de selección tienen la misma distribución conjunta, excepto para los no migrantes. De este modo tenemos solamente 2 funciones corrección, en vez de N .

Por último, para estimar la función corrección ocupamos una aproximación polinómica de segundo grado que considera la probabilidad de migrar a la primera mejor opción y trabajar -es decir, la probabilidad de migrar de g a j y trabajar- y la probabilidad de “retención” -o de no migrar y trabajar en el lugar de origen-, además de la interacción entre éstas dos probabilidades. Notar que, si el individuo no migró, entonces $p_{igj} = 0$ y sólo consideramos la primera mejor opción p_{ijj} . Por lo que:

$$\lambda_j^*(p_{igj}, p_{igg}) = p_{igj}\beta_1 + p_{igj}^2\beta_2 + p_{igg}\beta_3 + p_{igg}^2\beta_4 + p_{igj} * p_{igg}\beta_5 \text{ si } g \neq j \quad (8)$$

$$\lambda_j^*(p_{igj}, p_{igg}) = p_{ijj}\beta_1 + p_{ijj}^2\beta_2 \text{ si } g = j \quad (9)$$

En la implementación del método, en la sección V, mostramos como ocupamos esta corrección para nuestro caso.

IV. Datos

IV.1 Fuentes datos individuales

Los datos utilizados en este estudio provienen de distintas fuentes. Para la estimación de la ecuación de ingresos requerimos de datos a nivel individual de características personales de los individuos, de ingresos y datos de migración. Estos datos los podemos encontrar en la Encuesta de Caracterización Socioeconómica Nacional del año 2015 (CASEN 2015),¹⁶ la cual es una muestra de corte transversal a nivel nacional. A partir de ella obtuvimos las variables sexo, edad, estado civil, mayor nivel educativo alcanzado y años de educación, diversas fuentes de ingresos, situación laboral, residencia en zona urbana/rural, y región de residencia actual, en el año 2010 y en el momento de nacimiento.

¹⁶ Las encuestas fueron realizadas entre noviembre de 2015 y enero de 2016. Se podría haber utilizado también las Encuestas de los años 2011 o 2013, pero preferimos ocupar la más cercana en tiempo. No se pueden ocupar las tres Encuestas a la vez, ya que no podemos asegurar que haya observaciones que se repitan entre encuestas. Tampoco se pueden ocupar como un pseudo panel, ya que los periodos de migración son de 5 años, así que se traslapan entre sí las encuestas.

Estos datos nos permiten construir tanto las celdas según características individuales -que se muestra más adelante-, como estimar las probabilidades de migración y la ecuación de Mincer modificada para corregir por sesgo de selección. Para construir la variable explicada de la ecuación de ingresos, Ingreso en logaritmo natural, utilizamos la variable ingreso autónomo del mes anterior a la encuesta que provee la base de datos, de la cual restamos los ingresos que no son generados por trabajo de los mismos individuos.¹⁷

Para la variable máximo nivel de educación alcanzado dividimos a la muestra en “No terminó educación escolar”, “Media humanista”, “Media técnica”, “Superior profesional” y “Superior técnica”, dependiendo del máximo nivel alcanzado.¹⁸ Colocamos a las personas con estudios de postgrado en el grupo Superior profesional.

Es importante notar que la muestra de la CASEN 2015 no contiene observaciones de todas las comunas del país, ya que no es realizada en áreas de difícil acceso. Pero sí contiene observaciones de todas las regiones del país. Es una muestra representativa a nivel país y regional que contiene datos de 139 comunas del país que en conjunto concentran el 80% o más de la población de cada región.¹⁹ Para este estudio acotamos las observaciones a los individuos de 23 a 60 años, para considerar al rango de edad en que la población es más económicamente activa y que posiblemente terminó de estudiar, y para que 5 años atrás fueran mayores de edad; también acotamos la muestra a las personas que trabajan en jornada completa o extendida o más de 30 horas a la semana. Además, debido a que la I Región de Tarapacá y la XV región de Arica y Parinacota tienen muy pocas observaciones, las unimos para crear una sola región de migración. Lo mismo realizamos con la XI Región de Aysén del General Carlos Ibáñez del Campo y la XII Región de Magallanes y de la Antártica Chilena. Mientras que también unimos las X Región de los Lagos y XIV Región de Los Ríos ya que estaban unidas hasta el año 2007; quedando en total N=12 “regiones”.

Respecto a la decisión de migración, tenemos tres variables que indican la región donde residía el individuo en i) su año de nacimiento, ii) en el año 2010 (5 años antes de la encuesta) y iii) en el año 2015. A partir de estas variables formamos doce variables dicotómicas -una para cada “región”- con valor 1 si migró hacia esa región entre su nacimiento y su residencia en el 2015, y 0 si no; y del mismo modo formamos doce variables dicotómicas para el caso de migración entre 2010 y 2015.

¹⁷ Estos ingresos que restamos son: todos los ingresos por pensión de vejez, invalidez, de alimentos, orfandad, etc.; donaciones, remesas, intereses de depósitos y dividendos de acciones, ingresos adeudados de meses anteriores, indemnizaciones por despido y seguros de desempleo. Además, quitamos los ingresos ocasionales.

¹⁸ No dividimos la muestra entre quienes terminaron la educación básica y quienes no, por la reducida cantidad de personas en la muestra que no terminó la educación básica en algunas regiones.

No consideramos dentro de la muestra a las personas con educación diferencial.

¹⁹ En el año 2015 existían 346 comunas en Chile.

IV.2 Estadística descriptiva de datos individuales

En total las observaciones ocupadas para el caso de migración desde el momento de nacer hasta el 2015 son 61.154 individuos migrantes o no migrantes, jefes de hogar o sus cónyuges, entre 23 y 60 años que trabajan a tiempo completo o extendido²⁰. En esta muestra un 22.2% de los individuos son migrantes. Con los mismos criterios, para el caso de la migración entre 2010 y 2015 son 62.626 individuos, de los cuales un 4.4% son migrantes.

En la tabla 1 presentamos la estadística descriptiva de la migración interna en Chile desde el año de nacimiento hasta 2015 y desde el año 2010 hasta 2015. De ésta podemos notar que las regiones extremas se han mantenido en el tiempo como foco de atracción de población interna, mientras que regiones como la II y la VIII se han mantenido como expulsoras de población. La Región Metropolitana se ha convertido también en una expulsora de población en el periodo 2010-2015, lo que viene a confirmar la pérdida de atractivo de la región que se viene produciendo desde el Censo de 1992.

Tabla 1: Estadística descriptiva de migración interna (porcentaje respecto a la población de la región)

	Desde nacimiento			Desde 2010		
	Inmigrante interno	Emigrante interno	Migración neta	Inmigrante interno	Emigrante interno	Migración neta
Chile	22.2%	22.2%		4.4%	4.4%	
XV-I	44.2%	17.7%	26.5%	6.9%	4.5%	2.4%
II	30.5%	36.2%	-5.7%	4.8%	9.6%	-4.8%
III	31.5%	13.5%	18.0%	5.3%	1.8%	3.5%
IV	25.0%	30.9%	-5.9%	6.5%	4.0%	2.5%
V	21.7%	19.3%	2.4%	4.0%	4.7%	-0.7%
VI	23.1%	14.4%	8.7%	5.1%	2.1%	3.0%
VII	15.1%	25.9%	-10.8%	4.9%	3.6%	1.3%
VIII	11.3%	23.5%	-12.2%	3.2%	3.5%	-0.3%
IX	18.0%	27.2%	-9.2%	5.1%	3.8%	1.3%
X-XIV	13.7%	19.7%	-6.0%	4.4%	2.9%	1.5%
XI-XII	35.5%	14.4%	21.1%	7.8%	3.5%	4.3%
RM	25.4%	23.4%	2.0%	2.7%	6.7%	-4.0%

Fuente: Cálculos a partir de Encuesta CASEN 2015

En la tabla 2 mostramos los flujos de migración interna desde el nacimiento del individuo hasta el año 2015, mientras que la tabla 3 presenta los flujos de migración interna desde el 2010 hasta el 2015. Ambas tablas señalan que la migración se produce principalmente entre regiones contiguas o hacia la Región Metropolitana. Y aunque esta última tenga una

²⁰ Según el Artículo 27 del Código de trabajo, la jornada laboral extendida es un tipo de jornada aplicada a trabajos en que se necesita una presencia continua del personal, como restaurantes u hoteles, por ejemplo. En estos casos la jornada puede ser de hasta 12 horas diarias con descansos de por lo menos una hora, y con un máximo de 60 horas semanales.

migración interna negativa, sigue siendo la región que más migración atrae desde el resto del país.

Tabla 2: Flujos de migración interna entre regiones de Chile desde nacimiento a 2015 (número de personas)

Origen	Residencia 2015												total
	XV-I	II	III	IV	V	RM	VI	VII	VIII	IX	X-XIV	XI-XII	
XV-I	1,428	47	59	41	26	86	10	3	17	7	6	5	1,735
II	153	1,033	135	95	47	104	9	4	17	5	10	6	1,618
III	61	44	2,327	143	30	45	15	10	3	3	5	5	2,691
IV	137	113	347	2,039	113	133	22	10	12	3	15	7	2,951
V	148	40	128	74	4,774	450	75	22	49	18	66	68	5,912
RM	292	95	175	193	604	10,415	521	256	293	316	285	147	13,592
VI	31	7	23	23	89	383	4,276	62	37	20	25	20	4,996
VII	60	23	52	21	78	547	202	3,398	92	31	57	26	4,587
VIII	115	51	75	43	166	866	179	134	6,779	215	142	97	8,862
IX	74	14	38	22	65	490	166	56	225	3,760	201	55	5,166
X-XIV	50	17	35	21	74	398	71	40	107	183	5,798	424	7,218
XI-XII	8	2	5	4	33	38	15	9	15	24	109	1,564	1,826
Total	2,557	1,486	3,399	2,719	6,099	13,955	5,561	4,004	7,646	4,585	6,719	2,424	61,154

Fuente: Encuesta CASEN 2015

Tabla 3: Flujos de migración interna entre regiones de Chile desde 2010 a 2015 (número de personas)

2010	Residencia 2015												total
	XV-I	II	III	IV	V	RM	VI	VII	VIII	IX	X-XIV	XI-XII	
XV-I	2,601	17	15	20	10	31	6	4	6	4	7	1	2,722
II	31	1,479	30	34	11	21	4	4	5	8	7	2	1,636
III	3	5	3,262	26	8	1	3	3	4	5	3	0	3,323
IV	13	8	27	2,567	14	20	4	4	3	2	8	4	2,674
V	38	5	21	21	5,951	98	26	7	13	10	21	30	6,241
RM	57	16	46	47	143	14,305	173	115	141	110	111	62	15,326
VI	6	1	3	4	11	29	5,305	17	9	10	13	8	5,416
VII	8	2	14	4	7	51	20	3,844	18	4	13	4	3,989
VIII	19	11	8	10	12	76	25	25	7,462	42	25	16	7,731
IX	5	0	9	3	11	28	11	6	31	4,384	54	16	4,558
X-XIV	11	9	7	7	6	43	9	7	12	34	6,464	51	6,660
XI-XII	1	1	3	3	17	2	6	4	5	7	32	2,269	2,350
Total	2,793	1,554	3,445	2,746	6,201	14,705	5,592	4,040	7,709	4,620	6,758	2,463	62,626

Fuente: Encuesta CASEN 2015

En la tabla 4 presentamos la estadística descriptiva de los individuos, tanto migrantes como no migrantes, para ambos periodos: desde el nacimiento hasta 2015, y entre 2010 y 2015. Para ambos periodos de migración los migrantes internos son levemente más hombres que mujeres, y son más educados que los no migrantes ya que tienen más años de escolaridad y

tienen una mayor proporción de personas con educación superior y tienen mayores ingresos, lo que puede sugerir la existencia de un sesgo en los ingresos por región debido a la migración. La mayor movilidad de las personas con educación superior les permitiría buscar puestos de trabajo en otras regiones con mejores salarios. Los puestos en cada región para las personas con educación superior serían más variables que para las personas sin educación superior y se seleccionarían hacia donde pueden ocupar mejor sus ventajas comparativas obteniendo mejores salarios.

En la tabla 5 mostramos las mismas estadísticas descriptivas, pero para una cohorte de entre 30 y 34 años de edad. Dentro de la cohorte se puede notar más claramente las diferencias vista en la tabla 4 en ingresos, años de escolaridad y nivel de educación entre migrantes y no migrantes sin que estén afectados por la diferencia entre generaciones en el acceso a la educación. Mientras que en la tabla 6 podemos notar que la diferencia en ingresos entre migrantes y no migrantes de un mismo nivel educativo -educación superior profesional o técnico en este caso- se mantiene a favor de los primeros.

Tabla 4: Estadística descriptiva migrantes y no migrantes

Variable	Desde nacimiento a 2015				Desde 2010 a 2015			
	Migrantes		No migrantes		Migrantes		No migrantes	
	Media	Desv. Est.	Media	Desv. Est.	Media	Desv. Est.	Media	Desv. Est.
Hombres	0.597	0.004	0.582	0.002	0.619	0.009	0.583	0.002
Edad	43.433	0.087	42.732	0.468	37.158	0.183	43.092	0.042
Menores de 40 años	0.392	0.004	0.416	0.002	0.659	0.009	0.408	0.002
Casado	0.470	0.004	0.468	0.002	0.375	0.009	0.471	0.002
Años de escolaridad	12.369	0.032	11.460	0.167	13.356	0.067	11.592	0.015
Nivel de educación								
No terminó colegio	0.180	0.003	0.239	0.002	0.104	0.006	0.230	0.002
Media Humanista	0.375	0.004	0.402	0.002	0.343	0.009	0.400	0.002
Media Técnica	0.132	0.003	0.149	0.002	0.140	0.007	0.144	0.001
Superior Técnica	0.100	0.003	0.082	0.001	0.118	0.006	0.085	0.001
Superior Profesional	0.212	0.004	0.128	0.001	0.295	0.009	0.140	0.001
Experiencia potencial	26.065	0.100	26.273	0.059	18.802	0.209	26.500	0.048
Ingresos*	683969	7324.207	513707	3203.146	748265	15244.46	544899	3045.57
Ln (ingresos)*	13.055	0.007	12.81	0.00	13.152	0.017	12.850	0.003
Observaciones	13563		47591		2733		59893	

Fuente: Encuesta CASEN 2015

Notas: Las muestras están acotadas a personas de entre 23 y 60 años, jefes de hogar o pareja, que trabajan a tiempo completo o extendido. *: Según ingresos definidos más arriba en el trabajo a partir de los ingresos autónomos de la CASEN 2015.

Tabla 5: Estadística descriptiva migrantes y no migrantes cohorte 30-34 años

Variable	Desde nacimiento a 2015			
	Migrantes		No migrantes	
	Media	Desv. Est.	Media	Desv. Est.
Hombres	0.581	0.012	0.537	0.007
Edad	31.984	0.035	32.036	0.019
Casado	0.380	0.012	0.318	0.006
Años de escolaridad	13.748	0.012	12.782	0.041
Nivel de educación				
No terminó colegio	0.068	0.006	0.105	0.004
Media Humanista	0.361	0.012	0.416	0.006
Media Técnica	0.147	0.009	0.190	0.005
Superior Técnica	0.099	0.007	0.090	0.003
Superior Profesional	0.322	0.116	0.197	0.005
Experiencia potencial	13.237	0.089	14.253	0.045
Ingresos*	758029	27790.8	548435	8371.3
Ln (ingresos)*	13.185	0.021	12.899	0.01
Observaciones	1611		5810	

Fuente: Encuesta CASEN 2015

Notas: Las muestras están acotadas a personas de entre 30 y 34 años, jefes de hogar o pareja, que trabajan a tiempo completo o extendido. *: Según ingresos definidos más arriba en el trabajo.

Tabla 6: Estadística descriptiva migrantes y no migrantes con educación superior

Variable	Desde nacimiento a 2015			
	Migrantes		No migrantes	
	Media	Desv. Est.	Media	Desv. Est.
Hombres	0.537	0.008	0.462	0.005
Edad	41.585	0.153	40.706	0.982
Casado	0.503	0.008	0.469	0.005
Años de escolaridad	16.417	0.023	16.192	0.148
Nivel de educación				
Superior Técnica	0.321	0.007	0.391	0.005
Superior Profesional	0.679	0.007	0.609	0.005
Experiencia potencial	20.168	0.155	19.513	0.992
Ingresos*	1119063	17817.5	997848	12098.9
Ln (ingresos)*	13.649	0.014	13.452	0.009
Observaciones	4242		10018	

Fuente: Cálculos a partir de Encuesta CASEN 2015

Notas: Las muestras están acotadas a personas educación superior, de entre 23 y 60 años, jefes de hogar o pareja, que trabajan a tiempo completo o extendido. *: Según ingresos definidos más arriba en el trabajo.

IV.3 Datos regionales

Por su parte, para la estimación del modelo de decisión de migración Conditional Logit ocupamos diversas fuentes para obtener las características de la región consideradas en este estudio.²¹ En la tabla 7 presentamos la variable dependiente y todas las variables regionales utilizadas. Éstas se dividen en cuatro tipos: Condiciones económicas, Infraestructura y provisión de bienes públicos, Comodidades y Tamaño de la población. Cuando realizamos la estimación ocupando la migración entre 2010 y 2015, ocupamos las variables a inicios del periodo para evitar el sesgo de simultaneidad y porque refleja las condiciones de la región que los migrantes observaban al momento de migrar.²²

Para el caso de la variable tamaño de la población, como ésta podría estar correlacionada con características no observadas de la región, tenemos que considerar un instrumento para evitar problemas de sesgo de variable omitida. Un instrumento relevante sería la misma variable tamaño de la población para un año anterior a la muestra, ya que la población regional es en gran medida persistente. Pero, a la vez, debe ser un año suficientemente lejano para que no esté correlacionado con las características actuales de la región. Por supuesto, esta condición de exogeneidad no se puede testear directamente. Por esto hemos decidido tomar la población del año 1992, ya que en este año el PIB per capita nacional en dólares era menos de la mitad del PIB per capita nacional en el año 2010 (Díaz *et al.* (2016)) y se ubica a inicios del periodo de los gobiernos de la Concertación (1990-2010).²³

La variable ingresos es una medida de ingresos potenciales a partir de las características actuales del individuo, es una expectativa de ingreso en la región, pero no en el sentido de esperanza matemática; por lo que no considera en sí la posibilidad de desempleo y las oportunidades futuras. Una persona puede ir a otra región a buscar un trabajo según sus características, pero no ser contratado finalmente en ese trabajo. Por eso agregamos variables de condiciones económicas de la región que consideramos que están relacionadas con las expectativas de ingresos futuros. El desempleo para las posibilidades de encontrar ingresos, la pobreza sobre el nivel de ingresos en los trabajos que se consiguen en la región (la diversidad, y probabilidad de obtener un ingreso mínimo) tanto para uno como para sus hijos en el futuro; el crecimiento económico de la región para estimar las posibilidades futuras de la economía (la fortaleza de la economía y el cambio de los ingresos a través del tiempo en la región); y la diversidad económica para la amplitud de posibilidades en el mundo laboral, tanto para uno como para el resto de la familia. Pero también estas variables señalan cuestiones respecto al entorno socioeconómico: un lugar en de mucha pobreza y desempleo

²¹ Las variables de distancia entre regiones y la dummy de fronteras entre regiones las obtenemos de un mapa geográfico común y corriente.

²² En el caso de la estimación ocupando la migración desde el momento hasta 2015 habría que realizar supuestos muy restrictivos para realizar una estimación con variables explicativas de un año específico, ya que el periodo de migración es muy amplio.

²³ Hay correlación superior al 99% entre las poblaciones regionales de los censos de 1992 y 2012.

y con poca diversidad de trabajos y crecimiento económico puede no ser un lugar atractivo donde vivir y criar a los hijos.

Entre las variables de infraestructura tenemos las variables educación, operacionalizada mediante una variable resultado de la educación escolar y las oportunidades de entrar a la educación superior: promedio regional del puntaje PSU. Una variable salud, medida como la disposición de camas hospitalarias en la región. Y una variable de “congestión” de la región -o “saturación” de la región dada la infraestructura actual-, operacionalizada con la variable de caminos pavimentados dividido por la población en miles. En esta variable, mientras menor la cifra, mayor la congestión regional ya que una misma cantidad de personas tiene que compartir menos caminos pavimentados.²⁴ Por su parte, entre las medidas de *comodidades* incluimos la (in)seguridad, con la variable tasa de denuncias de delitos de mayor connotación social. La tasa de personas con conexión a internet que refleja la infraestructura tecnológica y la conectividad (más allá de las distancias físicas) en la región. Y como medida de contaminación ocupamos las emisiones de material particulado en el aire de tamaño menor a 2.5 micrómetros, las cuales tienen en general un origen antropogénico y son muy perjudiciales para la salud por su pequeño tamaño que les permite entrar a través de la respiración. Sin embargo, puede que no reflejar correctamente la contaminación por la diferente geografía (valles, planicies, ciudades costeras, etc.) de cada región.

Tabla 7. Descripción de variables utilizadas en modelo de migración

Variable explicada		
Variable	Descripción	Base de datos
Migración de región g a región j	Es una variable discreta que toma el valor 1 si el individuo migró desde la región g a la región j, y 0 si no.	CASEN 2015
Variables explicativas primera regresión (conditional logit)		
Variable	Descripción	Base de datos
Ln (Ingreso) estimado	Es el ingreso esperado en la región j para el individuo i. Se obtiene de la estimación de una ecuación de Mincer modificada para corregir por sesgo de selección y trabajo.	CASEN 2015
Distancia física entre región g y j	Distancia que tiene que recorrer el migrante para ir de la capital regional de g a la de j. A pesar de ser física, es también un variable proxy de la pérdida de contactos y conocimiento.	
Contigua	Es una variable dummy que toma el valor 1 si las regiones son contiguas, 0 en caso contrario.	
No contiguas	Es una variable dummy que toma el valor 1 si las regiones no son contiguas, 0 en caso contrario.	

²⁴ Consideramos la población en vez de la densidad poblacional, ya que con esta última no existe una relación clara al incluir tres medidas (población, área regional y km. de caminos) en una sola. Tampoco incluimos la variable caminos pavimentado por área de la región, ya que es una medida de “urbanidad” que creemos que ya captamos con las otras variables.

Variables sobre características de regiones		
Variable	Descripción	Base de datos
1) Condiciones económicas		
Desempleo	Promedio anual de nivel de desempleo trimestral a nivel regional.	Nueva Encuesta Nacional de Desempleo
Nivel de pobreza	Proporción de personas bajo la línea de pobreza a nivel regional.	CASEN
Crecimiento región	Crecimiento económico regional anual	INACER
Diversidad de la economía regional ²⁵	Índice de diversidad a partir de desviación estándar de los sectores productivos regionales según ingresos en el SII.	IDERE
2) Infraestructura y provisión de bienes públicos		
Educación	Puntaje PSU promedio por región.	Ministerio de Educación
Salud	Camás hospitalarias por cada 1.000 habitantes de la región	Compendio estadístico (INE)
Caminos: "Congestión"	Kilómetros de caminos pavimentados dividido por población a nivel regional	Dirección de Vialidad (MOP)
Variable	Descripción	Base de datos
3) Comodidades		
Seguridad	Tasa de denuncias de delitos de mayor connotación social cada 100.000 habitantes en la región como medida de violencia e inseguridad.	Informe Anual de Carabineros
Contaminación	Emisión de material particulado fino MP2.5 proveniente de fuentes fijas por región	Compendio estadístico (INE)
Conectividad e infraestructura tecnológica	Tasa de personas con conexión a banda ancha (móvil o fija) respecto al total regional	IDERE
4) Población ²⁶		
Población de 1992	Población regional en el año 1992 como instrumento de la población de 2010-2015.	Censo 1992

Por último, con la variable tamaño de la población en logaritmo natural buscamos captar el bienestar proveniente de las aglomeraciones humana no captadas por el resto de las variables, como economías de escala, economías de localización (cercanía de diversos servicios y entre empresas) y otras economías de urbanización no presentes en las otras variables.

En la tabla 8 presentamos la estadística descriptiva de estas variables para las doce regiones ocupadas en este trabajo con datos del año 2010, el inicio del periodo migratorio 2010-2015.

²⁵ También se probó un Índice de diversidad de Shannon de la industria manufacturera regional considerando la CIU3 o CIU4 de segundo nivel, pero los resultados son similares.

²⁶ No se considera la densidad poblacional porque en Chile está altamente correlacionada con el tamaño de la población (correlación=0.996). Por ejemplo, comparar la Región Metropolitana o la Región de Valparaíso que tienen un gran tamaño demográfico y poco territorio con las regiones extremas del país que tienen a la vez poca población en un territorio muy extenso.

Tabla 8. Estadística descriptiva de variables regionales

Variable	Media	Desv. Est.	Mínimo	Máximo
Tasa de desempleo (2010)	7.61	1.23	5.2	9.6
Nivel de pobreza (2009)	15.76	5.38	8.66	28.3
Crecimiento regional (2009-2011)	3.73	2.88	-1.62	8.48
Índice de diversidad sectores económicos (2010)	0.30	0.10	0.145	0.446
Tasa de camas hospitalarias (2010)	0.35	0.11	0.14	0.521
Puntaje PSU promedio regional (2010)	477.88	8.75	463.7	495.5
Población/Caminos (2010) ("Congestión")	1.95	1.09	0.21	3.58
Tasa de denuncias de delitos (2010)	2714.39	516.45	2112.3	3702.4
Tasa hogares con banda ancha internet (2010)	0.20	0.10	0.048	0.328
Emisiones de MP2.5 (2010)	1459.92	1681.55	76	6160
ln (población) (1992)	1115367	1382211	223699	5257937

Datos calculados usando las variables para las 12 "regiones" establecidas en este trabajo.

V. Estimación y resultados

V.1 Estimación ingresos corregidos por selección

En esta sección seguimos la metodología desarrollada por Dahl (2002). Primero separamos la población por región de origen (ya sea de nacimiento o del 2010). Luego, para cada una de estas regiones, dividimos a la población en celdas mutuamente excluyentes según las siguientes características: hombre/mujer, menores de 40 años/mayores de 40 años, Casados/no casados, con educación superior terminada/sin educación superior. Esta división crea 16 celdas por "región", 192 celdas en total.²⁷ Elegimos estas características porque modifican los costos de migración o el beneficio de migrar. Por ejemplo, dividimos por edad ya que a mayor edad, menor es el retorno por "invertir" en migrar, mientras que el estar casados modificaría la facilidad de movimiento de un individuo, ya que tiene que considerar la situación de su pareja. Por su parte, el sexo y el nivel educacional modifica los salarios recibidos y los sectores en que la población trabaja, lo cual es relevante porque cada región se especializa en diferentes sectores económicos.

Para cada individuo de la región i buscamos la celda a la que pertenece y calculamos su probabilidad estimada de migrar a cada región j , la cual es la proporción de personas en la celda que migraron a j . Para la muestra de migración interna entre el año de nacimiento y el año 2015 el tamaño medio de las celdas es 319, con un máximo de 2057 (en la VIII región: hombre, mayor de 40 años, casado y sin educación superior) y un mínimo de 22 (en la II región: mujeres, menores de 40 años, casadas y con educación superior). Mientras que para la muestra de migración interna entre 2010 y 2015, el tamaño medio de las celdas es 326, con un máximo de 2282 (En la Región Metropolitana: Hombres, de más de 40 años, casados y

²⁷ No es necesario que las celdas sean las mismas para cada región. Podríamos aumentar las celdas para las regiones más pobladas y así obtener grupos más homogéneos para estas regiones.

sin educación superior) y un mínimo de 22 (nuevamente en la II región: mujeres, jóvenes, casadas y con educación superior).

Como la estimación de los ingresos que realizamos en (7) es en dos etapas -estimación de probabilidad de migración y trabajo y luego la estimación de los ingresos-, hay que considerar en la desviación estándar de los estimadores que las probabilidades se estiman con un error en la primera etapa. Para corregir esto utilizamos *bootstrapped standard errors* con 500 repeticiones.

En la tabla A1 del anexo se encuentran los resultados obtenidos para la estimación de las doce ecuaciones -una para cada “región”-con función corrección considerando migración entre el año de nacimiento y el año 2015. Mientras que en la tabla A2, también en el anexo, presentamos los resultados con la función corrección considerando migración entre 2010 y 2015.²⁸ En nuestras estimaciones ocupamos como otros determinantes de los ingresos N'_i ; el sexo, zona urbana/rural y la rama económica a la que pertenece el individuo. La columna 1 de cada región es la estimación sin corrección, mientras que la segunda con corrección.

Antes de interpretar si la función corrección tiene algún efecto en las estimaciones, debemos determinar si la corrección es estadísticamente significativa. Esto lo realizamos con un test de Wald de los 7 términos de la función corrección que se muestran en (8) y (9) para cada región. El test señala que para ninguna región la función corrección es no significativa.

Respecto a las ecuaciones, recordemos que la variable explicada es el logaritmo natural del ingreso mensual del mes anterior a la entrevista, mientras que las variables explicativas el máximo nivel educacional alcanzado, la experiencia potencial, el sexo y si vive en una zona urbana o no. El ajuste de bondad promedio para las 12 regiones es de 0.29 para la ecuación de Mincer, que no es malo siendo que ocupamos una muestra de corte transversal. Mientras que es de 0.296 para la ecuación de Mincer con función de corrección (o de Dahl) para migración entre el nacimiento y el año 2015, y también para la migración entre 2010 y 2015. Por lo que la función corrección no aumenta significativamente el ajuste de la regresión.

Los resultados de las estimaciones por región muestran los resultados esperados. Un mayor nivel de educación aumenta los ingresos, al igual que mayor experiencia potencial -de manera decreciente en el tiempo- y el ser hombre. Mientras que vivir en zonas urbanas aumenta los ingresos, excepto para los casos de la I, III, V, VI, X-XIV, y Región Metropolitana en donde no tiene efecto significativo, y la XI-XII región, donde tiene un efecto negativo. Esta variedad de resultados se puede deber a la alta tasa de urbanidad que tiene el país (alrededor de 90% en el año 2015), lo cual hace que ya no sea tan relevante la diferencia entre vivir en una zona rural, que posiblemente se encuentre cerca de un pueblo o ciudad usando un automóvil, o en una zona urbana.

²⁸ No mostramos las funciones corrección ya que no tienen interpretación económica.

En las regresiones con corrección por sesgo los resultados son similares excepto para la variable sexo. El retorno a la educación media es prácticamente la misma para todas las regiones en ambos casos. En el caso de la educación superior los retornos cambian hasta en un 9% hacia arriba o hacia abajo con un promedio de 2.1% y 3.2% en la tabla A1 y A2 respectivamente. En términos absolutos estas diferencias varían entre -0.04 y 0.05.²⁹ Por lo que al igual que Bertoli et al. (2013), el efecto de los sesgos por decisión de migración y de trabajo serían muy bajos en los retornos a la educación. En el caso de la variable sexo, al añadir la función corrección con sesgo por decisión de trabajo el efecto positivo en los ingresos de ser hombre disminuye, y de manera considerable en algunas regiones.³⁰

Si comparamos los retornos a la educación entre regiones -considerando las ramas económicas en la regresión- podemos notar que la educación superior profesional genera mayores retornos en la Región metropolitana y, un poco más atrás, la III región, la VIII región y las regiones XI y XII. Mientras que se obtienen menos retornos por un título profesional en la II región, IV región y en las regiones X y XIV. En el caso de los títulos técnicos, se obtiene un mayor retorno, VIII región, IX región, las regiones XI y XII y un poco menos en la Región metropolitana y en la III región. Mientras que se obtiene bajo retorno en la I y II región del norte, la V región y en la X y XIV región. Por último, la educación media humanística y media técnica generan mayores ingresos en la zona sur del país que en el norte y centro-V, RM y VI regiones- de Chile, a excepción de la III región.

Si comparamos estos retornos a la educación regionales con la tabla 1 de migración interna, podemos notar que efectivamente la III, XI y XII regiones -que tienen mayores retornos- han atraído una alta cantidad de migrantes relativo a las otras regiones, mientras que la región II, IV, X y XIV -que tienen menos retornos- han expulsado migrantes. Sin embargo, las regiones VIII y IX también tienen bajos resultados en migración, a pesar de tener mayores retornos a la educación superior. Incluso si consideramos la migración desde el año de nacimiento tienen una migración neta de -9,2% y -6% respectivamente. Por lo que es necesario revisar otros determinantes de la migración además de los ingresos.

Para la predicción de ingresos preferimos utilizar la estimación con corrección de sesgo ya que la función corrección no es estadísticamente significativa, por lo que es relevante incluirla en la regresión. A partir de estos resultados podemos predecir para cada individuo i un ingreso “potencial” en cada una de las 12 “regiones” del país, las cuales se hacen sin considerar la función corrección, aunque estén incluidas en la regresión.

²⁹ También realizamos las estimaciones considerando solo a los hombres para sólo corregir por sesgo por la decisión de migrar y no por la decisión de trabajar, ya que los hombres en Chile suelen trabajar siempre. En estas estimaciones los parámetros son levemente más positivos, pero más volátiles. Esto se puede deber a que la muestra se reduce a dos tercios de la original, provocando que el método de Dahl sea más impreciso.

³⁰ Realizando la función corrección solo para el sesgo por decisión de migración, esta caída del parámetro asociado al sexo no se produce, por lo que la podemos asociar al sesgo por decisión de trabajo.

V.2 Modelo de elección de migración

V.2.1 Estimación con “efectos fijos” por región

Primero, para la estimación del modelo de decisión discreta múltiple de migración, estimamos (6) mediante Conditional Logit usando como variables explicativas los ingresos predichos que obtuvimos en la subsección anterior, además de la distancia de migración y si migró a una región que es contigua o no, además de sólo una constante por región que captura el “efecto fijo” de la región o “Disposición a pagar por vivir” en tal región, la cual capta todas las características de la región.

En los modelos de decisión discreta múltiple solo podemos estimar libremente (m-1) probabilidades, siendo m la cantidad de regiones. Por lo que debemos normalizar asignando una región como categoría base. En nuestro caso elegiremos para este propósito a la Región Metropolitana, ya que es la región que más migrantes recibe.³¹

En la tabla 9 se presentan los resultados de esta regresión para la migración entre el año de nacimiento y el año 2015, además del efecto marginal en la media y el efecto marginal promedio.

Podemos notar en la parte superior de la tabla que las variables referentes a la distancia de migración tienen el signo esperado ya que a mayor distancia disminuye la probabilidad de migrar a la región elegida, y si la región no es contigua, la probabilidad disminuye aún más que si es contigua.³² Por lo que alejarse de su lugar de origen efectivamente reduce la migración, lo que podríamos asociar a un costo por la pérdida de contactos y redes sociales en lugar de residencia original y diferencias entre el lugar de origen y llegada que van aumentando en la medida que se aleja geográficamente más. Además, son todas las variables estadísticamente significativas. Por su parte el coeficiente de ingresos potenciales, estimados considerando la corrección por sesgo de selección en migración y trabajo, confirma que estos son relevantes en la decisión de migración, como postula la teoría económica vista más arriba.

En la siguiente parte de la tabla presentamos los efectos fijos por región, los cuales reflejarían la disposición a pagar por vivir en tal región neto de ingresos potenciales y costos de migración, o la “calidad de vida” en términos amplios. Los resultados son relativos a la Región Metropolitana -región base- que toma valor 0. Como podemos ver, la región

³¹ La elección de la categoría base afecta la interpretación de los resultados ya que los parámetros estimados de las variables individuales (*case-specific*) serán interpretados como parámetros de un modelo binario entre la región j y la categoría base (Región Metropolitana en nuestro caso). Para la primera estimación que realizamos no hay variables *case-specific*, por lo que no afecta tales resultados. En el caso de los resultados mostrados en la tabla 11, sí afecta.

³² Ser región contigua también disminuye la probabilidad de migración respecto a quedarse en la región de origen, es decir, no migrar.

metropolitana junto a la V región y las regiones más al norte tienen el mayor “atractivo” en el largo periodo entre el nacimiento (año 1992 o antes) y 2015, mientras que las regiones extremas también tienen altos efectos fijos luego de ser considerados los costos de migración para llegar a ellas. Las regiones del centro-sur (VII a IX región) son las menos atractivas para migrar, lo cual podría provenir tanto de razones de actividad económica como infraestructura y comodidades, ya que según los datos regionales, estas regiones aparecen con altas cifras relativas de pobreza, desempleo y bajos niveles relativos de actividad económica, a la vez que sus resultados relativos en infraestructura y comodidades se encuentran en general en el promedio del país y algunas por debajo.

Tabla 9: Resultados de regresión modelo de migración interna entre el año de nacimiento y el año 2015

Variable	Coefficiente	Error est.
ln (ingreso) estimado	0.41***	0.075
ln (distancia)	-0.543***	0.018
Región contigua	-0.617***	0.076
Región no contigua	-0.748***	0.102
Efectos fijos por región	Sí	
Wald Chi2 (4) p-value	0.0000	
Observaciones	733.848	
Casos	61.154	

Región	Efecto fijo	MEM	AME
I-XV	0.07	0.020	0.0075
Región Metropolitana	0	0.082	0.0282
V	-0.07	0.042	0.0177
XI-XII	-0.37*	0.017	0.0069
III	-0.46*	0.024	0.0096
VI	-0.68*	0.043	0.0186
X-XIV	-0.68*	0.026	0.0129
II	-0.73*	0.008	0.0047
IV	-0.86*	0.016	0.0085
VIII	-1.07*	0.032	0.0172
IX	-1.09*	0.022	0.0124
VII	-1.14*	0.024	0.0131

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones. MEM: Efectos Marginales en la Media; AME: Efectos Marginales Promedio, ambos de un aumento de 1% en el ingreso potencial de un individuo en la región j en la probabilidad de que éste migre hacia dicha región.

A continuación, en la misma tabla, presentamos los Efectos Marginales en la Media (MEM) y los Efectos Marginales Promedio (AME) de un aumento de 1% en el ingreso potencial de un individuo en la región j en la probabilidad de que éste migre hacia dicha región. Por ejemplo, en el caso de la I y XV región un aumento de 1% en los ingresos que se pueden obtener allí, aumenta la probabilidad de migrar hacia tal región en un 0.02 (dos puntos porcentuales) para el individuo con características promedio de la muestra, mientras que en un 0.0075 (0.75 puntos porcentuales) en promedio para la muestra.

Los resultados para todas las regiones son similares y rondan entre 1.5 puntos porcentuales y 4.5 puntos porcentuales a excepción de la II región de Antofagasta y la Región Metropolitana en el caso de los Efectos Marginales en la Media. Todos estos efectos marginales son estadísticamente significativos al 1%.³³ Por su parte, el efecto marginal promedio de la muestra de un aumento de 1% en el ingreso es menor para todas las regiones y varía entre 0.47 puntos porcentuales para la II región y 2.82 puntos porcentuales para la Región metropolitana. El efecto marginal promedio es más similar entre regiones que el efecto marginal en la media. Sin embargo, no existe una relación exacta en la diferencia entre AME y MEM.³⁴

En el anexo 3 presentamos los resultados para la migración entre los años 2010 y 2015. Los resultados respecto a los costos de migración muestran efectos más negativos, señalando que en este periodo más corto la migración es menor (ya que los coeficientes de contigua y no contigua son al menos 5 veces más grandes), pero si la hay, la distancia afecta menos, ya que el coeficiente de \ln (distancia) es -0.368 , menor a -0.543 de la regresión en la tabla 9. Esto es coincidente con los datos de migración que presentamos en las tablas 2 y 3, donde la migración entre 2010 y 2015 es más “dispersa” hacia todo el país que en los datos entre nacimiento y 2015. Es decir, durante los últimos años los costos producidos por alejarse del lugar de origen han disminuido su impacto en la decisión de migración, lo cual puede deberse a la mayor conectividad del país, tanto espacial como en comunicaciones.

Por su parte, el coeficiente del ingreso potencial tiene un efecto positivo no estadísticamente significativo. Por lo que al no considerar las características personales -y sólo ocupando un “efecto fijo” por región en vez de las características regionales en sí- los resultados señalarían que las diferencias en ingresos entre regiones no explicarían la migración. Un resultado similar al de Soto y Torche para los datos del Censo de 2002 (migración entre 1997 y 2002).

³³ Por espacio no presentamos los efectos marginales cruzados (cuánto disminuye la probabilidad de migrar a las otras regiones por un aumento de 1% en los ingresos potenciales de la región j para el individuo i . Sin embargo, estos efectos son negativos y significativos al 1% de significancia. Tampoco mostramos los efectos marginales de las variables de costo de migración, pero se pueden entregar de ser requeridos.

³⁴ Estos resultados no son comparables a los de Aroca (2004) y Soto y Torche (2004), sin embargo, para dar unas cifras de referencias, ambos trabajos encuentran que un aumento del diferencial de ingresos entre dos regiones de 1% provoca un aumento de alrededor de 2% en la migración a la región de mayor ingreso relativo.

Respecto a los coeficientes de efecto fijo, la Región Metropolitana cae considerablemente respecto al resto, solo siendo mayor que la IV y II región, el resto mantiene más o menos el mismo orden. Sin embargo, lo más importante es que la mayoría de las regiones tienen un “efecto fijo” que no es estadísticamente distinto de cero. Esto podría señalarnos que en el periodo corto entre 2010 y 2015 entre estas regiones (que se ubican entre la IV y la VII región, además de las regiones más extremas del norte), las características de tales regiones no son relevantes en la decisión de migración.

V.2.2 Estimación de efecto de las características de las regiones

En esta subsección presentamos los resultados de la estimación de la ecuación (5). Es decir, incluyendo las características de las regiones en vez de “efectos fijos” por región. De esta manera esperamos explicar parte de la “disposición a pagar por vivir” en cada región a partir de las características mismas de cada región.

A diferencia de la subsección anterior, en la tabla 10 presentamos los resultados de la regresión para el periodo de migración 2010 a 2015. Esto ya que el periodo de migración entre el año de nacimiento y 2015 es muy amplio para explicar el “efecto fijo” regional con características tomadas de un año particular.³⁵

En la primera parte de la tabla 10 podemos observar las mismas características ya ocupadas en la sección anterior (tabla en anexo 3), las cuales no se modifican ya que varían según el individuo, mientras que el resto de variables varía según región.

Respecto a las características regionales ocupadas la mayoría tiene el signo esperado. Respecto a las variables de condiciones económicas de la región, la tasa de desempleo y el nivel de pobreza de la región reducen el atractivo de la región, ya que reflejan una mayor precariedad laboral en la zona además que podrían reflejar un entorno con menor calidad de vida. El crecimiento regional de los años iniciales del periodo de migración tiene un signo positivo, pero no estadísticamente significativo, lo cual puede relacionarse porque es una medida transitoria de la economía regional. Por último, una mayor diversidad de sectores económicos disminuye la migración hacia la región, lo cual refleja la emigración interna neta que poseen las regiones con grandes centros comerciales e industriales diversos del país como la Región de Valparaíso, la Región Metropolitana y la del Biobío.³⁶ La gente podría estar

³⁵ Para mostrar resultados con la migración entre el año de nacimiento y 2015 presentamos resultados en el anexo 4 ocupando las mismas variables que para el periodo 2010 a 2015 pero tomadas en el año 2015 -o años cercanos si no se encuentran los datos-. Hay que considerar que estas variables no reflejan fielmente las condiciones al momento de migrar y solo las podemos ocupar bajo el supuesto muy restrictivo de que los individuos sabían cómo serían las regiones en el futuro, es decir, las variables reflejarían expectativas de condiciones económicas y de vida más que las condiciones que se veían al momento de migrar.

³⁶ Ocupamos medidas de diversidad de sectores económicos a partir de la diversidad según producción por sector como empleos por sector. En ambos casos el resultado es un coeficiente negativo.

migrando hacia ciudades medias que se están desarrollando en las últimas décadas (Schiappacasse et al (2001) y Rodríguez y González (2006)).

Tabla 10: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 ocupando características regionales

Variable	Coefficiente	Error est.
ln (ingreso) estimado	0.035	0.144
ln (distancia)	-0.368***	0.039
Región contigua	-3.304***	0.165
Región no contigua	-3.568***	0.227
Tasa de desempleo (2010)	-1.229***	0.104
Nivel de pobreza (2009)	-0.046*	0.026
Crecimiento regional (2009-2011)	0.154	0.108
Índice de diversidad sectores económicos (2010)	-13.551***	2.000
Tasa de camas hospitalarias (2010)	0.908	1.533
Puntaje PSU promedio regional (2010)	0.034**	0.016
Caminos/Población (2010) ("Congestión")	0.908***	0.371
Tasa de denuncias de delitos (2010)	0.003***	0.001
Tasa hogares con banda ancha internet (2010)	-9.061***	2.363
Emisiones de MP2.5 (2010)	0.000***	0.000
ln (población) (1992)	-3.58***	0.382
Efectos fijos por región	No	
Wald Chi2 (4) p-value	0.0000	
Observaciones	751.512	
Casos	62.626	

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

Siguiendo el análisis con las características de infraestructura y comodidades de las regiones, podemos notar que la tasa de camas hospitalarias tiene el signo positivo esperado, pero no es significativo. Esto podría deberse a que las condiciones mínimas de infraestructura de salud son similares en el país,³⁷ por lo que habría que considerar medidas como la disponibilidad de especialistas -como oncólogos, geriatras, etc.- en cada región para medir diferencias más relevantes en el sector de la salud.

El promedio regional de puntaje PSU, como medida de resultado de la educación y la capacidad de entrar a la educación superior, tiene un efecto atractor en la migración. Por lo

³⁷ Otras medidas como la tasa de médicos por cada 10.000 habitantes tampoco tienen resultados estadísticamente significativos,

que una buena educación en la región podría atraer más familias migrantes o personas que están pensando formar una familia en la región de destino. Este resultado es interesante en la medida que regiones como la II, la VIII y la Región Metropolitana tienen buenos promedios de puntaje PSU, tienen sin embargo emigración neta, por lo que tienen otras características que anulan la atracción que ejerce la educación.

Por su parte, nuestra variable de “congestión”, caminos pavimentados divididos en población, exhibe un coeficiente positivo, lo que significa que a mayor “congestión”, menor atracción migratoria. Esto apoya la tesis de que existe una emigración desde las grandes ciudades muy congestionadas hacia ciudades de menor tamaño en desarrollo de Schiappacasse et al (2001) y Rodríguez y González (2006).

Nuestra medida de violencia e inseguridad, la tasa de denuncias de delitos de mayor connotación social, presenta un signo positivo y significativo, lo cual es contrario a lo esperado. Mirando los datos regionales directamente, no se ve una relación clara entre esta variable y la migración. La tasa de hogares con internet, como medida de conectividad e infraestructura tecnológica, tiene un efecto negativo en la migración, lo cual también podría verse como contraintuitivo, pero tiene sentido si se observa desde la perspectiva que las regiones con grandes ciudades -como ya hemos señalado- poseen una emigración neta. Por su parte, nuestra medida de contaminación, emisiones de MP2.5, tiene un efecto positivo en la inmigración interna regional, pero con una cifra poco relevante. Podría señalar la complicación de medir la contaminación en espacio tan amplios como regiones.

Por último, nuestra medida de escala, el tamaño de la población, tiene un efecto negativo en la atracción migratoria interna. Este resultado puede señalar que los efectos de bienestar de la concentración demográfica ya no atraen como lo hicieron en Chile durante finales del siglo XIX y el siglo XX hasta la década de 1970, confirmando nuevamente la tesis de emigración hacia concentraciones urbanas de menor escala.

V.2.3 Estimación incluyendo características de las personas

- Características individuales y migración entre el nacimiento y 2015

Una de las ventajas de poseer microdatos sobre migración es que podemos estimar el efecto de las características individuales en la probabilidad de migrar hacia -o quedarse en- una región j . Para esto, primero utilizamos el mismo modelo (6) con un “efecto fijo” por región, pero añadiendo como variables explicativas las siguientes características personales: sexo, edad, el estar casado o conviviendo, y el tener educación superior (técnica o profesional) o no. Este tipo de variables son del tipo *case-specific*, es decir, no varían con la alternativa - región en nuestro caso- elegida. Por lo que su efecto se interpreta del siguiente modo: si el coeficiente es positivo (negativo), un aumento de la variable en j conlleva a un aumento (disminución) de la probabilidad de elegir la región j por sobre la Región metropolitana, la región base en nuestro modelo.

Del mismo modo que en V.2.1 al ocupar “efectos fijos”, podemos mostrar resultados tanto para la migración desde el nacimiento hasta 2015, como desde 2010 a 2015. En la tabla 11 presentamos los resultados para entre el nacimiento hasta el año 2015 como un “ejercicio intermedio”³⁸ antes de los resultados finales mostrados a continuación, dejando los resultados entre 2010 y 2015 en el anexo 5. Más adelante mostraremos resultados en un modelo más completo para este periodo de migración.

Lo primero que podemos notar en parte superior de la tabla 11 es que los coeficientes relacionados al ingreso estimado y costos de migración son similares a los de la tabla 9, por lo que son una prueba de que nuestros resultados son robustos para la migración entre el año de nacimiento y el año 2015. Por su parte, en la parte inferior de la tabla mostramos los coeficientes de las características individuales para cada una de las 12 “regiones” en que dividimos a Chile para este trabajo. Como podemos observar, el ser hombre aumenta la probabilidad de migrar o mantenerse en cada una de las regiones por sobre migrar a la Región Metropolitana, o alternativamente, el ser mujer aumenta la probabilidad de migrar hacia la Región Metropolitana. Esto concuerda con los resultados de los últimos censos realizados en el país, los cuales muestran una mayor migración femenina hacia esta región, y con el hecho que la Región Metropolitana tiene un mayor sector terciario, donde trabaja una mayor proporción de mujeres. También podemos ver que el efecto es mayor hacia las regiones extremas, lo que indicaría una mayor movilidad espacial de los hombres.

En lo que respecta a la edad, el efecto es negativo y significativo para todas las regiones exceptuando la V Región. A mayor edad, se prefiere migrar entre las regiones de Valparaíso y del Libertador General Bernardo O’Higgins. También vale la pena notar que mientras más joven más probabilidad de migrar -según la estimación puntual- hacia las regiones extremas (XV, I, II, X y XIV) con respecto a la Región Metropolitana.

Las personas con pareja en 2015, en general tuvieron más probabilidad de migrar hacia otras regiones que a la Región metropolitana. Pero en las regiones del norte, entre Antofagasta y Coquimbo, las personas sin pareja en 2015 tuvieron más probabilidad de migrar hacia allá. Aunque en el caso de las regiones II y IV, los resultados no son estadísticamente distintos de cero. Hay que notar que esto podría indicar una causalidad reversa, las personas que migraron hacia las regiones II, III y IV, tuvieron menor probabilidad de encontrar pareja estable.

Por su parte, las personas con educación superior en 2015 tuvieron mayor probabilidad de migrar hacia la Región metropolitana con respecto a cualquier otra región a excepción de la Región de Coquimbo y las regiones más australes del país. Este resultado tiene sentido en la medida que la Región Metropolitana posee más puestos de trabajo para personas con mayores

³⁸ El espacio temporal de migración es muy amplio, por lo que puede haber una causalidad reversa que explique los resultados obtenido. Sin embargo, colocamos los resultados por dos razones, primero, los resultados son similares, aunque no iguales, a los obtenidos posteriormente en el periodo más corto de 2010 a 2015, y, por otro lado, nos permite ver que un nivel de consistencia en los resultados obtenido en la tabla 9.

estudios. Pero también podría tener sentido en un argumento de causalidad reversa en que las personas que migraron a la Región Metropolitana tienen en términos absolutos mayores oportunidades de estudiar en instituciones de educación superior. Sin embargo, hay tres regiones que poseen una mayor tasa de matrícula de educación superior sobre su población y otras dos regiones con tasas similares (IDERE 2016). Además que más adelante mostraremos que este resultado se mantiene para el periodo 2010 a 2015, donde consideramos la migración de sólo mayores de edad.

Tabla 11: Resultados de regresión modelo de migración interna entre el año de nacimiento y el año 2015 con características individuales

Variable	Coeficiente		Error estándar		
ln (ingreso) estimado	0.476***		0.083		
ln (distancia)	-0.554***		0.018		
Región contigua	-0.576***		0.076		
Región no contigua	-0.701***		0.102		
Efectos fijos por región	Sí				
Wald Chi2 (4) p-value	0.0000				
Observaciones	733.848				
Casos	61.154				
Región	Hombre	Edad	Con pareja	Ed. Superior	Efecto fijo
I-XV	0.52***	-0.02***	0.29***	-0.45***	0.86***
II	0.54***	-0.02***	-0.07	-0.62***	0.17
III	0.46***	-0.01***	-0.24***	-0.24***	-0.07
IV	0.15**	-0.02***	-0.03	-0.02	-0.19
V	0.09*	-0.00	0.02	-0.23***	0.11
VI	0.10**	-0.01***	0.16***	-0.60***	-0.15
VII	0.07	-0.02***	0.33***	-0.82***	-0.44***
VIII	0.13**	-0.02***	0.40***	-0.64***	-0.37***
IX	0.21***	-0.02***	0.22***	-0.31***	-0.49***
X-XIV	0.29***	-0.04***	0.34***	-0.27***	0.59***
XI-XII	0.26***	-0.01***	0.18**	-0.16	0.09

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

Por último, el “efecto fijo” por región se mantiene negativo para la mayoría de las regiones y de manera significativa, por lo que habría características regionales que reducen la “disposición a pagar” por migrar hacia otras regiones en comparación a la Región Metropolitana y las regiones de Antofagasta, Atacama y del extremo sur del país. Vale la pena notar también que la Región de la Araucanía -junto a otras regiones cercanas- es la

menos atractiva para migrar o quedarse, lo cual podría estar relacionado tanto a las malas condiciones económicas de la región como a la pobreza o alto desempleo u otras características como la infraestructura o servicios en la zona.³⁹

- Características regionales e individuales y migración entre 2010 y 2015

Con los microdatos sobre migración también podemos agregar a la vez tanto las características personales como las regionales para tener nuestro modelo “completo” en que consideramos los tres aspectos que afectan la migración interna según Schiappacasse et al (2001): las características de la región (economía regional, de infraestructura y comodidades), las características de propias de la población migrante (ingresos potenciales y características personales) y las características intrínsecas del proceso migratorio (costos de migración).

En este caso ocupamos el modelo de la ecuación (5), al que agregamos características individuales. Presentamos los resultados de este modelo en la tabla 12.

En la primera parte de la tabla presentamos los resultados para la variable ingresos potenciales en la región y los costos de migración, tradicionales en la literatura. Las variables de costos de migración son similares a las obtenidas anteriormente. Sin embargo, a diferencia de los resultados obtenidos en la tabla A3, cuando no consideramos características personales en la decisión de migración, los resultados muestran que los ingresos potenciales sí tienen un efecto positivo y significativo en la migración interna. Además de ser un efecto estadísticamente similar al del periodo migratorio más amplio entre nacimiento y 2015. Este resultado muestra que las estimaciones que no consideran las características de las personas, como en Soto y Torche (2004) que encuentran un bajo impacto de los diferenciales de ingresos entre regiones, estarían subestimando el impacto de los ingresos potenciales en la migración.

En lo que respecta a las variables sobre características regionales, éstas tienen el mismo signo que en la estimación en que no consideramos las características individuales, con la excepción de la variable “congestión”, que ahora es negativa y significativa. Es decir, las regiones con más aglomeración urbana, de modo que la infraestructura está siendo sobrepasada, estarían atrayendo migración interna. Por otra parte, ahora el crecimiento económico sí es significativo en la atracción de migración interna. Por lo que las expectativas

³⁹ Los resultados para el periodo 2010 a 2015, en el anexo 5, son similares cualitativamente. Sin embargo, algunas diferencias que podemos notar son que: el ser hombre tiene un coeficiente puntual mayor, al igual que la edad, mientras que el tener educación superior tiene un efecto menor. Respecto al efecto fijo, o disposición a pagar por vivir allí, cabe destacar que en los últimos años el efecto aumenta para las regiones extremas (I, XV; XI y XII) y la III región -una región minera-, mientras que el efecto se hace vuelven positivas para las regiones del centro-sur del país.

de una economía creciente atraen personas a la región. Por otra parte, las variables asociadas a la contaminación y a la educación dejan de ser estadísticamente significativas.

Tabla 12: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 ocupando características individuales y regionales

Variable	Coficiente	Error estándar		
ln (ingreso) estimado	0.406**	0.172		
ln (distancia)	-0.416***	0.041		
Región contigua	-3.133***	0.174		
Región no contigua	-3.364***	0.239		
Tasa de desempleo (2010)	-2.490***	0.470		
Nivel de pobreza (2009)	-0.114**	0.051		
Crecimiento regional (2009-2011)	0.365**	0.172		
Índice de diversidad sectores económicos (2010)	-25.888***	5.900		
Tasa de camas hospitalarias (2010)	5.355**	2.194		
Puntaje PSU promedio regional (2010)	-0.017	0.037		
Caminos/Población (2010) ("Congestión")	-4.663***	1.527		
Tasa de denuncias de delitos (2010)	0.007***	0.002		
Tasa hogares con banda ancha internet (2010)	-18.801***	5.249		
Emissiones de MP2.5 (2010)	0.000	0.000		
ln (población) (1992)	-4.723***	1.653		
Efectos fijos por región		No		
Wald Chi2 (4) p-value		0.0000		
Observaciones		751.512		
Casos		62.626		
Región	Hombre	Edad	Con pareja	Ed. Superior
I-XV	0.55***	-0.06***	0.66***	-0.32
II	0.05	-0.02**	-0.02	-0.71***
III	0.22*	-0.04***	-0.32***	-0.08
IV	0.35***	-0.01	0.18	0.03
V	0.14	0.00	0.25**	-0.56***
VI	0.21**	-0.02***	0.16*	-0.48***
VII	0.10	-0.02***	0.45***	-0.73***
VIII	0.24**	-0.01*	0.43***	-1.20***
IX	0.56***	-0.02***	0.28**	-0.79***
X-XIV	0.44***	-0.03***	0.25**	-0.04
XI-XII	0.25*	-0.06***	0.10	0.42**

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

En la tercera parte de la tabla mostramos los coeficientes relacionados a las características individuales. Como podemos notar, el ser hombre aumenta la probabilidad de migrar o mantenerse en cada una de las regiones por sobre migrar a la Región Metropolitana, o alternativamente, el ser mujer aumenta la probabilidad de migrar hacia la Región Metropolitana. Esto concuerda con los resultados de los últimos censos realizados en el país, los cuales muestran una mayor migración femenina hacia esta región, por lo que sería una tendencia que se mantiene en el tiempo en Chile. También podemos ver que el efecto es mayor hacia las regiones extremas, lo que indicaría una mayor movilidad espacial de los hombres.

En lo que respecta a la edad, hacia las regiones extremas el efecto se hace cada vez más negativo, por lo que mientras más joven una persona tiene más probabilidad de migrar hacia los extremos geográficos del país, con respecto a la Región Metropolitana. Solamente en la IV y V región la edad no afecta la decisión de migración.

En lo que respecta a las variables de tener pareja y haber cursado educación superior, habíamos señalado que en el amplio periodo de entre el nacimiento y 2015 podía producirse una causalidad reversa. En el periodo migratorio de 2010 a 2015 esta posibilidad disminuye notoriamente tanto por la reducción del periodo de tiempo como por el hecho que consideramos personas que migraron siendo mayores de edad. Aun así, los resultados son similares. El tener pareja aumenta en general la probabilidad de migrar hacia otras regiones que a la Región metropolitana, especialmente en las zona central y sur de Chile. Pero en las regiones del norte, en Antofagasta y Atacama, las personas sin pareja en 2015 tuvieron más probabilidad de migrar hacia allá. Por su parte, las personas con educación superior tienen mayor probabilidad de migrar hacia la Región metropolitana con respecto a cualquier otra región a excepción de las regiones de Aysén y de Magallanes. Este resultado tiene sentido en la medida que la Región Metropolitana posee más oportunidades de trabajo y desarrollo para personas con mayores estudios.

Por tanto, en nuestro modelo “completo” para la migración interna de Chile entre los años 2010 y 2015 los resultados esperados en los tres tipos de factores considerados son confirmados en general.

V.2.4 Estimación del modelo sin corrección por selección y ejercicios de robustez

Todos los resultados vistos más arriba en el trabajo han sido realizados con los salarios estimados con corrección por sesgo de selección en migración y trabajo. A continuación, en la tabla 13 presentamos los resultados para el modelo de migración interna con características regionales e individuales, pero sin esta corrección para ver si los resultados difieren. Para su comparación rápida, al lado aparecen los resultados con corrección por sesgo de selección de migración y trabajo de la tabla 12 para la primera parte de la tabla.

Tabla 13: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 con características individuales y regionales, sin corrección por selección en migración y trabajo

Variable	Coef. sin corrección	Error estándar	Coef. con corrección	Error estándar
ln (ingreso) estimado	0.534***	0.173	0.406**	0.172
ln (distancia)	-0.415***	0.035	-0.416***	0.041
Región contigua	-3.137***	0.152	-3.133***	0.174
Región no contigua	-3.370***	0.202	-3.364***	0.239
Tasa de desempleo (2010)	-2.478***	0.422	-2.490***	0.470
Nivel de pobreza (2009)	-0.044	0.034	-0.114**	0.051
Crecimiento regional (2009-2011)	0.052	0.111	0.365**	0.172
Índice de diversidad sectores económicos (2010)	-21.223***	5.023	-25.888***	5.900
Tasa de camas hospitalarias (2010)	1.207	1.319	5.355**	2.194
Puntaje PSU promedio regional (2010)	0.014	0.034	-0.017	0.037
Caminos/Población (2010) ("Congestión")	-5.037***	1.389	-4.663***	1.527
Tasa de denuncias de delitos (2010)	0.004***	0.002	0.007***	0.002
Tasa hogares con banda ancha internet (2010)	-13.253***	3.975	-18.801***	5.249
Emisiones de MP2.5 (2010)	0.001***	0.000	0.000	0.000
ln (población) (1992)	-4.908***	1.540	-4.723***	1.653
Efectos fijos por región	No			
Wald Chi2 (4) p-value	0.0000			
Observaciones	751.512			
Casos	62.626			
Coeficientes de regresión sin corrección por selección en migración y trabajo				
Región	Hombre	Edad	Con pareja	Ed. Superior
I-XV	0.50***	-0.06***	0.66***	-0.29*
II	0.04	-0.02**	-0.02	-0.67***
III	0.21*	-0.04***	-0.32**	-0.06
IV	0.33**	-0.01	0.18	0.05
V	0.16*	0.00	0.25**	-0.52***
VI	0.18*	-0.02***	0.16	-0.44***
VII	0.06	-0.02***	0.45***	-0.71***
VIII	0.20*	-0.01**	0.43***	-1.19***
IX	0.53***	-0.02***	0.28**	-0.78***
X-XIV	0.41***	-0.03***	0.25**	-0.01
XI-XII	0.30**	-0.06***	0.10	0.46***

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados sin y con corrección por sesgo de selección de migración y trabajo. Para la segunda de estas estimaciones se ocupó errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

Lo primero que podemos notar es que al no considerar los sesgos de selección en migración y trabajo dentro de la estimación, el ingreso esperado en la región de destino tiene un coeficiente positivo 32% mayor que cuando corregimos por el sesgo. Es decir, existe un sesgo positivo por selección en migración y trabajo, y al no considerarlo estaríamos sobreestimando el efecto de los salarios en la decisión de migración entre regiones. Sin embargo, los coeficientes no son estadísticamente distintos y se encuentran a menos de un error estándar de diferencia. Respecto a las variables relacionadas a los costos de migración y las variables regionales, estas son en general similares, a excepción que las variables Nivel de pobreza, Crecimiento regional y Tasa de camas hospitalarias son estadísticamente significativas al aplicar la función corrección, mientras que Emisiones de MP2.5 deja de ser estadísticamente significativas. Las variables de características individuales son similares en ambos casos. Por lo que los resultados de ambas regresiones son muy similares y solamente la variable ingreso esperado en la región de destino cambia sustantivamente, pero sin resultados contundentes que sugieran que la corrección por sesgo sea cuantitativamente necesaria para los resultados de migración interna. Pero recordemos que la función corrección es estadísticamente significativa para la estimación de salarios esperados.

A continuación, presentamos tres ejercicios de robustez de los resultados obtenidos para la migración interna entre 2010 y 2015 ocupando primero una submuestra de individuos de 40 o menos; segundo, una submuestra que considera solamente a los hombres; y, tercero, la utilización del método de “análisis de componentes principales” para ocupar definiciones más amplias de las variables relacionadas a provisión y calidad de educación y salud. Por espacio, presentamos las tablas con estos resultados en el anexo 6.

En la tabla A6 mostramos los resultados al ocupar sólo la submuestra de personas menores de 40 años en el modelo de migración. En este caso hemos dividido a la población entre menores y mayores de 31 años para la característica joven. Los resultados señalan que los jóvenes reaccionan más ante cambios en los ingresos potenciales en la región de destino que la población total y de manera estadísticamente significativa. Respecto a los costos de migración, los resultados son cualitativamente similares.

En tanto a las características regionales, éstas tienen el mismo signo en todas las variables. Sin embargo, la población más joven reacciona más a las características económicas de la región, lo que viene a reforzar lo encontrado con la variable de ingresos esperados. También se dirigen a lugares más especializados económicamente y con más conectividad y tecnología (mayor conexión de banda ancha), con mayor puntaje PSU regional, menor congestión y población, además que les preocupa menos la delincuencia y la contaminación del lugar de destino que a la población general. Mientras que respecto a las características individuales, el sexo importa menos en la decisión de migración, excepto para las regiones extremas del país, donde los hombres tienen más probabilidad de ir en comparación a la región base. El tener educación superior también es menos relevante, pero entre los menores de 41 años, las

personas con más educación tienen más probabilidad de migrar hacia la III región y las del sur del país (X a XII región) respecto a la región base. En general, los resultados son cualitativamente similares a los encontrados con la muestra completa de individuos.

En la tabla A7 presentamos los resultados para una muestra que considera solamente a los hombres de la población, para ver si existen diferencias en la migración entre sexos. En esta submuestra los resultados de las variables asociadas a los costos de migración y las características individuales son muy similares a las de la muestra completa, mientras que respecto a las características regionales tienen el mismo signo, pero con coeficiente distinto, por ejemplo, tienen una mayor reacción ante cambios en la situación económica de la región. Sin embargo, existe una gran diferencia respecto a los ingresos esperados en la región de destino. Para los hombres el coeficiente asociado al ingreso esperado es negativo y no es estadísticamente distinto de cero. Este resultado podría estar relacionado al hecho que los hombres suelen trabajar casi siempre, por lo que estarían menos influenciados por los ingresos esperados que las mujeres -que tienen un coeficiente positivo (0.472) y estadísticamente significativo- y se podría intuir que los hombres también tendrían mayor aceptación a cambiar de región de ser necesario para encontrar trabajo. Esto también se ve reforzado por el hecho que los hombres reaccionan más a la tasa de desempleo que las mujeres en su decisión de migrar.⁴⁰ Por lo que en general, exceptuando por la variable ingresos esperados, los resultados son similares a los encontrados con la muestra completa de individuos.

Por último, realizamos una regresión ocupando el método de análisis de componentes principales. Usamos componentes principales para ocupar definiciones más amplias de las variables relacionadas a provisión y calidad de educación y salud sin ocupar muchas variables más.^{41, 42} Presentamos los resultados asociados en la tabla A8.

Para caracterizar la educación a nivel regional ocupamos las variables para el año 2010: Puntaje PSU, Alumnos por profesor en educación básica y media, Matrícula educación parvularia, Matrícula de educación superior profesional, y Matrícula de educación superior técnica.⁴³ El análisis de componentes principales nos deja con tres componentes, el primero correlacionado principalmente con las variables de educación superior: Matrícula de educación superior técnica y Matrícula de educación superior profesional; el segundo con las variables de educación de menores de edad: Matrícula de educación parvularia

⁴⁰ El coeficiente de tasa de desempleo para los hombres es de -15.303 con un error estándar de 7.623, mientras que para las mujeres es de -5.751 con un error estándar de 2.630. Los resultados para la submuestra mujeres no se presentan en este trabajo.

⁴¹ Nótese que ya que tenemos 12 regiones, podemos colocar un máximo de 11 variables regionales sin que se produzca colinealidad exacta entre las variables.

⁴² Para la elección de los componentes principales ocupamos la Regla de Kaiser y Rotación Varimax. Por espacio no presentamos estas estimaciones, pero se pueden entregar a petición.

⁴³ La fuente de estas variables educacionales es MINEDUC.

(negativamente), y Puntaje PSU; y el tercero con Alumnos por profesor en el colegio (y en menor medida Matrícula de educación parvularia (positivamente)). Los resultados muestran que el primer y tercer componentes son estadísticamente significativos y positivos, por lo que las variables relacionadas a las posibilidades de acceso a educación superior en la región y de calidad de educación colegial (medido como alumnos por profesor) están relacionadas positivamente con la decisión de migración.

Por su parte, para caracterizar la calidad e infraestructura de la salud regional ocupamos 4 variables en el año 2010: Disponibilidad de camas hospitalarias, Cantidad de médicos por cada 10.000 habitantes en la región⁴⁴, Tasa de consultas médicas en el sector público asociadas a morbilidad, la Tasa de niños menores de 6 años obesos.⁴⁵ En este caso tenemos dos componentes principales, el primer componente se relaciona con las enfermedades: obesidad y morbilidad, mientras que el segundo componente está correlacionado con la oferta de salud: médicos per capita y disponibilidad camas hospitalarias. Los resultados muestran que el primer componente es negativo y estadísticamente significativo, por lo que las regiones con peores condiciones de salud atraen menos migración. Por otro lado, el componente relacionado con la oferta de salud tiene un efecto positivo en la migración hacia la región, pero con menor coeficiente y significancia estadística. Por lo que las variables de resultado tienen más influencia que las variables de input en la salud en la atracción de migración.

También ocupamos componentes principales para reducir las variables relacionadas a la economía regional en la regresión. En este caso ocupamos cinco variables en el año 2010: Tasa de desempleo, Nivel de pobreza, Tasa de crecimiento económico (2009-2011), un Índice de diversidad de la industria manufacturera según empleo⁴⁶, y un Índice de diversidad de la economía regional según ingresos. En este caso tenemos dos componentes principales, el primero relacionado con todas las variables excepto nuestro Índice de diversidad de la industria manufacturera, y el segundo con este último y la Tasa de desempleo y de crecimiento de la economía regional (negativamente). El segundo componente es negativo y estadísticamente significativo al 1%, por lo que a menor tasa de desempleo, mayor crecimiento y menos diversidad económica, más atractiva la región; lo cual es acorde a nuestra regresión original. Por su parte, el primer componente es positivo y pero no significativo, reduciendo el efecto de la tasa de desempleo y la diversidad económica provenientes del otro componente principal.

Las características individuales y los costos de migración tienen resultados iguales que en la regresión original. Por tanto, los resultados al ocupar otras variables para salud y educación

⁴⁴ Extraído del Compendio estadístico 2010 del INE.

⁴⁵ Las últimas dos variables se obtienen de Vial Cossani (2016).

⁴⁶ Esta variable la formamos a partir de un Índice de Shannon (Entropía) con datos extraídos de ENIA 2010 y a partir de la clasificación de industrias CIUU 3 en segundo nivel.

muestran un mayor efecto de éstas características regionales en la decisión de migración, pero se encuentran dentro del marco de resultados de la regresión original.

VI. Conclusión

El aporte de este estudio radica en tres aspectos. Primero, ocupamos una muestra amplia de decisiones individuales de migración en vez de datos de migración agregados. Esto nos permitió realizar estimaciones empíricas con microfundamentos para distinguir si las estimaciones a nivel agregado que se han realizado en Chile se mantenían al considerar datos individuales o no. Segundo, estimamos a nivel regional los ingresos esperados considerando los sesgos de selección en la migración y en el trabajo, lo cual no puede realizarse con datos agregados que utilizan un trabajador promedio, como en Aroca *et al.* (2001), Aroca (2004) y Soto y Torche (2004). Y tercero, con el modelo econométrico desarrollado hemos podido integrar la literatura demográfica con la literatura económica sobre migración interna a nivel regional en Chile, y así considerar los diversos factores que afectan la decisión de migrar: las características de las regiones, las características del proceso migratorio en sí y las características individuales.

Los resultados encontrados en el trabajo muestran que al considerar todos los aspectos de la migración, mayores ingresos potenciales en una región tienen un efecto que atrae la migración interna, tanto para el periodo entre el año de nacimiento como entre los años 2010 y 2015. Por su parte, los costos de migración, un efecto negativo, por lo que los lugares más alejados de nuestra región de origen son menos atractivos. El ser mujer, tener educación superior, tener menor edad y estar sin pareja tienen en general un efecto positivo en la migración hacia la región metropolitana respecto al resto de las regiones. Aunque existen particularidades hacia las regiones extremas del país. Mientras que las variables relacionadas con el tamaño de la región en términos demográficos, de infraestructura y diversidad económica tienen efectos negativos en la migración, lo que podría indicar una tendencia hacia la migración de ciudades de tamaño medio que se están desarrollando económicamente y en infraestructura y comodidades. Aunque en el modelo final, la variable “congestión” tiene un efecto que cuestiona en parte esta hipótesis. Por último, nuestras variables de *comodidades* muestran un efecto positivo en la decisión de migrar hacia la región con mejores *comodidades*.

Aún existen muchos aspectos que mejorar en la modelación de la migración interna en Chile. Primero que todo, la falta de datos a niveles geográficos más pequeños, como provincias, por ejemplo, complica la estimación precisa de los beneficios de vivir en cierta “localización”, ya que dentro de una región los empleos, y la calidad de vida en general, varían según la localidad en que se vive. Una mejor disposición de datos públicos al respecto sería deseable.

Por otro lado, la Encuesta CASEN empezó a realizar preguntas sobre migración hace poco tiempo, por lo que sólo hay tres encuestas disponibles-2011, 2013 y 2015- que poseen datos al respecto. El corto periodo que abarcan estas encuestas implica que no se puede realizar análisis de datos de panel con éstas, ya que los periodos de migración se traslapan y porque el periodo es muy corto para ver cambios relevantes en las condiciones de las regiones. Sin embargo, en la medida que vayan apareciendo más encuestas, estos problemas pueden ser solucionados, y así sería posible aplicar mejores métodos de estimación.

Por último, se pueden realizar mejoras al modelo aquí utilizado. Un ejemplo sería ocupar precios de las viviendas ajustados por su calidad como proxy del precio de los bienes no transables. Esto permitiría considerar las diferencias en costo de vida entre regiones. También se podría integrar variables de política pública directamente, como la entrega de subsidios habitacionales, para analizar si las políticas de los gobiernos tienen efectos directos en la distribución de la población. Por otro lado, sería interesante incluir en un mismo modelo datos de migración interna como extranjera, ya que en los últimos años la inmigración extranjera a Chile ha aumentado considerablemente. Así podríamos tener un modelo más “completo” del fenómeno demográfico de la distribución poblacional dentro de Chile.

Referencias

- Aroca, P., Hewings, G. J., y Paredes Godoy, J. (2001). "Migración interregional y el mercado laboral en Chile: 1977-82 y 1987-92". *Cuadernos de economía*, 38(115), 321-345.
- Aroca, P. (2004). "Migración interregional en Chile. Modelos y resultados 1987-2002", *Notas de población*, CEPAL, n° 78, 97-154.
- Bayer, P., Keohane, N., y Timmins, C. (2009). "Migration and hedonic valuation: The case of air quality", *Journal of Environmental Economics and Management*, 58(1), 1-14.
- Bertoli, S., Fernández-Huertas, J. y Ortega, F. (2013). "Crossing the border: Self-selection, earnings and individual migration decisions". *Journal of Development Economics*, 101, 75-91.
- Coeymans, J. E. (1983). "Determinantes de la migración rural-urbana en Chile según origen y destino". *Cuadernos de Economía*, 59, 43-64.
- Dahl, G. B. (2002). "Mobility and the return to education: Testing a Roy model with multiple markets". *Econometrica*, 70(6), 2367-2420.
- De Vreyer, P., Gubert, F., y Roubaud, F. (2010). "Migration, Self-selection and Returns to Education in the WAEMU". *Journal of African Economies*, 19(1), 52-87.
- Díaz, J., Lüders. R. y Wagner, G. (2016). *Chile 1810 – 2010. La República en cifras. Historical statistics*. Santiago: Ediciones Universidad Católica de Chile.
- INE (2007). *Migraciones internas regionales. 1992-2002*. Santiago, Chile.
- Lee, L. F. (1983). "Generalized econometric models with selectivity". *Econometrica: Journal of the Econometric Society*, 507-512.
- Robinson, C., y Tomes, N. (1982). "Self-selection and interprovincial migration in Canada". *Canadian Journal of Economics*, 474-502.
- Rodríguez, J., y González, D. (2006). "Redistribución de la población y migración interna en Chile: continuidad y cambio según los últimos cuatro censos nacionales de población y vivienda". *Revista de Geografía Norte Grande*, 35, 7-28.
- Schiappacasse, P., Contreras, M., y Fuensalida, C. (2001). "Migraciones internas hacia la Región Metropolitana de Santiago de Chile: una comparación con planteamientos teóricos". *Investigaciones Geográficas*, 35, 1-25.
- Soto, R., y Torche, A. (2004). "Spatial inequality, migration and economic growth in Chile". *Cuadernos de economía*, 41(124), 401-424.
- Vial Cossani, C. (2016). Índice de Desarrollo Regional IDERE 2016. Recuperado de <http://www.idere.cl/>
- Xing, C., and Zhang, J. (2017). "The preference for larger cities in China: Evidence from rural-urban migrants", *China Economic Review*, 43, 72-90.

Anexos

- Anexo 1

Tabla A1: Ecuación de ingresos (con función corrección por decisión de trabajo y migración desde nacimiento)

Variable explicada: ln (ingreso)

región especificación	[I-XV]		[II]		[III]		[IV]		[V]		[VI]	
	Mincer	Dahl										
Educ. med. Hum.	0.18*** (0.04)	0.18*** (0.04)	0.16** (0.06)	0.16** (0.06)	0.24*** (0.03)	0.25*** (0.03)	0.18*** (0.03)	0.18*** (0.03)	0.16*** (0.03)	0.16*** (0.03)	0.15*** (0.02)	0.14*** (0.02)
Educ. med. Tec.	0.29*** (0.05)	0.29*** (0.05)	0.23*** (0.07)	0.23*** (0.07)	0.31*** (0.04)	0.33*** (0.04)	0.26*** (0.05)	0.26*** (0.05)	0.25*** (0.04)	0.25*** (0.04)	0.25*** (0.04)	0.25*** (0.04)
Educ. sup. Tec.	0.46*** (0.06)	0.46*** (0.07)	0.49*** (0.08)	0.50*** (0.08)	0.58*** (0.05)	0.59*** (0.05)	0.53*** (0.06)	0.50*** (0.06)	0.46*** (0.04)	0.43*** (0.04)	0.50*** (0.05)	0.51*** (0.05)
Educ. sup. Prof.	1.13*** (0.06)	1.14*** (0.05)	0.95*** (0.08)	0.95*** (0.09)	1.17*** (0.05)	1.16*** (0.04)	1.05*** (0.06)	1.02*** (0.06)	1.11*** (0.04)	1.06*** (0.04)	1.12*** (0.05)	1.12*** (0.05)
Exp. potencial	0.01*** (0.00)	0.01*** (0.00)	0.01 (0.01)	0.01* (0.01)	0.02*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.01)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)
Exp. potencial^2	-0.00*** (0.00)	-0.00*** (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00*** (0.00)							
Sexo (hombre)	0.30*** (0.03)	0.22*** (0.05)	0.34*** (0.04)	0.23** (0.09)	0.30*** (0.03)	0.24*** (0.05)	0.33*** (0.03)	0.30*** (0.05)	0.28*** (0.02)	0.33*** (0.04)	0.29*** (0.02)	0.31*** (0.05)
Zona (urbano)	0.02 (0.04)	0.02 (0.04)	0.36*** (0.09)	0.38*** (0.10)	0.01 (0.04)	0.01 (0.04)	0.07** (0.04)	0.07** (0.04)	-0.03 (0.02)	-0.04* (0.02)	-0.01 (0.02)	-0.01 (0.02)
Constante	11.80*** (0.09)	11.46*** (0.47)	12.02*** (0.21)	10.86*** (0.54)	12.24*** (0.07)	12.56*** (0.38)	11.85*** (0.07)	11.78*** (0.50)	12.05*** (0.05)	11.38*** (0.50)	11.96*** (0.05)	12.70*** (0.49)
Observaciones	2,557		1,486		3,399		2,719		6,099		5,561	
R-squared	0.34 0.34		0.27 0.28		0.35 0.35		0.31 0.31		0.27 0.27		0.28 0.28	
región especificación	[VII]		[VIII]		[IX]		[X-XIV]		[XI-XII]		[RM]	
	Mincer	Dahl										
Educ. med. Hum.	0.26*** (0.03)	0.25*** (0.04)	0.24*** (0.02)	0.24*** (0.02)	0.25*** (0.03)	0.25*** (0.03)	0.18*** (0.02)	0.18*** (0.02)	0.28*** (0.04)	0.27*** (0.04)	0.19*** (0.02)	0.20*** (0.02)
Educ. med. Tec.	0.32*** (0.04)	0.31*** (0.04)	0.31*** (0.03)	0.30*** (0.03)	0.34*** (0.04)	0.34*** (0.04)	0.30*** (0.03)	0.29*** (0.03)	0.41*** (0.05)	0.41*** (0.05)	0.23*** (0.02)	0.24*** (0.02)
Educ. sup. Tec.	0.56*** (0.06)	0.56*** (0.06)	0.57*** (0.04)	0.62*** (0.04)	0.61*** (0.05)	0.63*** (0.05)	0.50*** (0.05)	0.50*** (0.05)	0.61*** (0.05)	0.60*** (0.06)	0.57*** (0.03)	0.58*** (0.03)
Educ. sup. Prof.	1.15*** (0.06)	1.15*** (0.07)	1.21*** (0.03)	1.25*** (0.04)	1.14*** (0.05)	1.16*** (0.05)	1.00*** (0.05)	1.00*** (0.06)	1.19*** (0.05)	1.16*** (0.06)	1.34*** (0.03)	1.35*** (0.03)
Exp. potencial	0.02*** (0.01)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)
Exp. potencial^2	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00** (0.00)	-0.00** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Sexo (hombre)	0.36*** (0.03)	0.36*** (0.06)	0.32*** (0.02)	0.26*** (0.03)	0.28*** (0.02)	0.29*** (0.04)	0.29*** (0.02)	0.24*** (0.05)	0.23*** (0.03)	0.24*** (0.05)	0.30*** (0.01)	0.30*** (0.03)
Zona (urbano)	0.12*** (0.03)	0.12*** (0.03)	0.12*** (0.02)	0.12*** (0.02)	0.05** (0.02)	0.05** (0.02)	0.01 (0.02)	0.01 (0.02)	-0.10** (0.05)	-0.10** (0.05)	-0.04 (0.03)	-0.05 (0.03)
Constante	11.69*** (0.08)	12.45*** (0.54)	11.65*** (0.05)	12.80*** (0.35)	11.82*** (0.06)	12.58*** (0.40)	12.04*** (0.06)	12.46*** (0.31)	12.24*** (0.09)	11.98*** (0.43)	12.08*** (0.06)	12.20*** (0.21)
Observaciones	4,004		7,646		4,585		6,719		2,424		13,955	
R-squared	0.22 0.22		0.3 0.3		0.3 0.3		0.18 0.19		0.34 0.34		0.37 0.37	

Standard errors en paréntesis. Bootstrapped standard error con 500 repeticiones incluidos en estimación tipo Dahl (2002).

Función de corrección polinomial de segundo grado. No mostrados. *** p<0.01, ** p<0.05, * p<0.1

- Anexo 2

Tabla A2: Ecuación de ingresos (con función corrección por decisión de trabajo y migración desde 2010)

Variable explicada: ln (ingreso)

especificación	[I-XV]		[II]		[III]		[IV]		[V]		[VI]	
	Mincer	Dahl										
Educ. med. Hum.	0.18*** (0.04)	0.18*** (0.04)	0.16** (0.06)	0.17*** (0.06)	0.24*** (0.03)	0.24*** (0.03)	0.18*** (0.03)	0.18*** (0.03)	0.16*** (0.03)	0.16*** (0.03)	0.15*** (0.02)	0.14*** (0.02)
Educ. med. Tec.	0.29*** (0.05)	0.28*** (0.05)	0.23*** (0.07)	0.25*** (0.06)	0.31*** (0.04)	0.31*** (0.04)	0.26*** (0.05)	0.25*** (0.05)	0.25*** (0.04)	0.26*** (0.04)	0.25*** (0.04)	0.24*** (0.03)
Educ. sup. Tec.	0.46*** (0.06)	0.46*** (0.06)	0.49*** (0.08)	0.53*** (0.08)	0.58*** (0.05)	0.53*** (0.05)	0.53*** (0.06)	0.50*** (0.07)	0.46*** (0.04)	0.47*** (0.04)	0.50*** (0.05)	0.55*** (0.05)
Educ. sup. Prof.	1.13*** (0.06)	1.16*** (0.06)	0.95*** (0.08)	0.98*** (0.08)	1.17*** (0.05)	1.13*** (0.06)	1.05*** (0.06)	1.01*** (0.07)	1.11*** (0.04)	1.11*** (0.05)	1.12*** (0.05)	1.15*** (0.05)
Exp. potencial	0.01*** (0.00)	0.02*** (0.00)	0.01 (0.01)	0.01 (0.01)	0.02*** (0.00)							
Exp. potencial^2	-0.00*** (0.00)	-0.00*** (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00*** (0.00)							
Sexo (hombre)	0.30*** (0.03)	0.13* (0.08)	0.34*** (0.04)	0.29** (0.12)	0.30*** (0.03)	0.21*** (0.07)	0.33*** (0.03)	0.26*** (0.09)	0.28*** (0.02)	0.30*** (0.05)	0.29*** (0.02)	0.15** (0.08)
Zona (urbano)	0.02 (0.04)	0.03 (0.04)	0.36*** (0.09)	0.36*** (0.10)	0.01 (0.04)	0.01 (0.04)	0.07** (0.04)	0.09** (0.04)	-0.03 (0.02)	-0.03 (0.02)	-0.01 (0.02)	-0.00 (0.02)
Constante	11.80*** (0.09)	12.40*** (0.64)	12.02*** (0.21)	12.82*** (0.69)	12.24*** (0.07)	12.39*** (0.36)	11.85*** (0.07)	11.96*** (0.48)	12.05*** (0.05)	12.38*** (0.52)	11.96*** (0.05)	15.16*** (0.81)
Observaciones	2,793		1,554		3,445		2,746		6,201		5,592	
R-squared	0.34 0.34		0.27 0.28		0.35 0.35		0.31 0.32		0.27 0.27		0.28 0.29	
especificación	[VII]		[VIII]		[IX]		[X-XIV]		[XI-XII]		[RM]	
	Mincer	Dahl										
Educ. med. Hum.	0.26*** (0.03)	0.25*** (0.03)	0.24*** (0.02)	0.24*** (0.02)	0.25*** (0.03)	0.24*** (0.03)	0.18*** (0.02)	0.18*** (0.02)	0.28*** (0.04)	0.27*** (0.04)	0.19*** (0.02)	0.20*** (0.02)
Educ. med. Tec.	0.32*** (0.04)	0.31*** (0.04)	0.31*** (0.03)	0.30*** (0.03)	0.34*** (0.04)	0.33*** (0.04)	0.30*** (0.03)	0.30*** (0.03)	0.41*** (0.05)	0.39*** (0.05)	0.23*** (0.02)	0.24*** (0.02)
Educ. sup. Tec.	0.56*** (0.06)	0.54*** (0.07)	0.57*** (0.04)	0.55*** (0.04)	0.61*** (0.05)	0.59*** (0.05)	0.50*** (0.05)	0.47*** (0.06)	0.61*** (0.05)	0.63*** (0.05)	0.57*** (0.03)	0.55*** (0.03)
Educ. sup. Prof.	1.15*** (0.06)	1.13*** (0.07)	1.21*** (0.03)	1.18*** (0.04)	1.14*** (0.05)	1.14*** (0.05)	1.00*** (0.05)	1.00*** (0.06)	1.19*** (0.05)	1.19*** (0.05)	1.34*** (0.03)	1.34*** (0.03)
Exp. potencial	0.02*** (0.01)	0.02*** (0.01)	0.02*** (0.00)	0.02*** (0.00)	0.01*** (0.00)	0.01** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.00)	0.02*** (0.01)	0.02*** (0.00)	0.02*** (0.00)
Exp. potencial^2	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00** (0.00)	-0.00* (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
Sexo (hombre)	0.36*** (0.03)	0.26*** (0.09)	0.32*** (0.02)	0.17*** (0.04)	0.28*** (0.02)	0.14*** (0.05)	0.29*** (0.02)	0.15** (0.07)	0.23*** (0.03)	0.26*** (0.05)	0.30*** (0.01)	0.25*** (0.03)
Zona (urbano)	0.12*** (0.03)	0.12*** (0.03)	0.12*** (0.02)	0.12*** (0.02)	0.05** (0.02)	0.05** (0.02)	0.01 (0.02)	0.01 (0.02)	-0.10** (0.05)	-0.10** (0.05)	-0.04 (0.03)	-0.04 (0.03)
Constante	11.69*** (0.08)	12.71*** (0.80)	11.65*** (0.05)	12.82*** (0.33)	11.82*** (0.06)	13.14*** (0.38)	12.04*** (0.06)	13.03*** (0.51)	12.24*** (0.09)	12.40*** (0.50)	12.08*** (0.06)	12.47*** (0.26)
Observaciones	4,040		7,709		4,620		6,758		2,463		14,705	
R-squared	0.22 0.21		0.30 0.30		0.30 0.30		0.18 0.18		0.34 0.34		0.37 0.37	

Standard errors en paréntesis. Bootstrapped standard error con 500 repeticiones incluidos en estimación tipo Dahl (2002).

Función de corrección polinomial de segundo grado. No mostrados. *** p<0.01, ** p<0.05, * p<0.1

- Anexo 3: Resultados modelo de migración interna (Conditional Logit) para migración entre 2010 y 2015

Tabla A3: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015

Variable	Coefficiente	Error est.
ln (ingreso) estimado	0.035	0.148
ln (distancia)	-0.368***	0.039
Región contigua	-3.304***	0.165
Región no contigua	-3.568***	0.228
Efectos fijos por región	Sí	
Wald Chi2 (4) p-value	0.0000	
Observaciones	751.512	
Casos	62.626	
Región	Efecto fijo	
X-XIV	0.42***	
VIII	0.22**	
III	0.22***	
XI-XII	0.19**	
VI	0.15	
IX	0.14	
I-XV	0.02	
VII	0.01	
V	0.004	
Región Metropolitana	0	
IV	-0.04	
II	-0.91	

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

- Anexo 4: Resultados de modelo de migración interna (Conditional Logit) con características regionales para la migración entre el periodo de nacimiento y el año 2015

Tabla A4: Resultados de regresión modelo de migración interna entre el año de nacimiento y el año 2015 ocupando características regionales

Variable	Coefficiente	Error est.
ln (ingreso) estimado	0.347***	0.038
ln (distancia)	-0.543***	0.016
Región contigua	-0.621***	0.069
Región no contigua	-0.752***	0.091
Tasa de desempleo (2015)	-0.295***	0.039
Nivel de pobreza (2015)	-0.010	0.016
Crecimiento regional (2012-2015)	0.010	0.396
Índice de diversidad sectores económicos (2014)	-0.391***	0.134
Tasa de camas hospitalarias (2014)	-0.953***	0.260
Puntaje PSU promedio regional (2015)	-0.065***	0.121
Caminos/Población (2015) ("Congestión")	-0.517***	0.120
Tasa de denuncias de delitos (2015)	0.001***	0.000
Tasa hogares con banda ancha internet (2014)	0.986	1.080
ln (población) (1992)	-0.175	0.269
Efectos fijos por región	No	
Wald Chi2 (4) p-value	0.0000	
Observaciones	751.512	
Casos	62.626	

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo.

- Anexo 5: Resultados de modelo de migración interna (Conditional Logit) con características individuales para el periodo de migración 2010 a 2015

Tabla A5: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 con características individuales

Variable	Coeficiente		Error estándar		
ln (ingreso) estimado	0.406**		0.172		
ln (distancia)	-0.416***		0.041		
Región contigua	-3.133***		0.174		
Región no contigua	-3.364***		0.239		
Efectos fijos por región	Sí				
Wald Chi2 (4) p-value	0.0000				
Observaciones	751.512				
Casos	62.626				
Región	Hombre	Edad	Con pareja	Ed. Superior	Efecto fijo
I-XV	0.54***	-0.06***	0.66***	-0.32	2.08***
II	0.06	-0.02**	-0.02	-0.71***	-0.11
III	0.22*	-0.04***	-0.32***	-0.08	1.99***
IV	0.35***	-0.01	0.18	0.03	0.28
V	0.14	0.00	0.25**	-0.56***	-0.16
VI	0.21**	-0.02***	0.16*	-0.48***	0.03
VII	0.10	-0.02***	0.45***	-0.73***	0.60**
VIII	0.24**	-0.01*	0.43***	-1.20***	0.46*
IX	0.56***	-0.02***	0.28**	-0.79***	0.35
X-XIV	0.44***	-0.03***	0.25**	-0.04	1.15***
XI-XII	0.25*	-0.06***	0.10	0.42**	2.46***

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

- Anexo 6: Resultados de ejercicio de robustez

Tabla A6: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 ocupando características individuales y regionales, submuestra menores de 40 años

Variable	Coficiente	Error estándar		
ln (ingreso) estimado	1.285***	0.179		
ln (distancia)	-0.476***	0.045		
Región contigua	-2.372***	0.192		
Región no contigua	-2.504***	0.261		
Tasa de desempleo (2010)	-5.553***	0.931		
Nivel de pobreza (2009)	-0.238***	0.074		
Crecimiento regional (2009-2011)	0.807***	0.257		
Índice de diversidad sectores económicos (2010)	-63.899***	11.585		
Tasa de camas hospitalarias (2010)	-2.667	2.329		
Puntaje PSU promedio regional (2010)	0.325***	0.076		
Caminos/Población (2010) ("Congestión")	-19.089***	3.413		
Tasa de denuncias de delitos (2010)	0.017***	0.004		
Tasa hogares con banda ancha internet (2010)	-36.437***	9.016		
Emisiones de MP2.5 (2010)	0.002***	0.000		
ln (población) (1992)	-20.722***	3.819		
Efectos fijos por región	No			
Wald Chi2 (4) p-value	0.0000			
Observaciones	310.392			
Casos	25.866			
Región	Hombre	Edad	Con pareja	Ed. Superior
I-XV	0.52**	-0.06***	1.15***	0.02
II	-0.24	-0.03	0.33	-0.03
III	0.11	-0.04**	-0.05	0.45***
IV	0.03	0.02	0.21	0.08
V	0.01	0.03**	0.43***	-0.32**
VI	0.13	0.00	0.36***	-0.00
VII	-0.04	0.03**	0.61***	-0.47***
VIII	0.01	0.01	0.66***	-0.95***
IX	0.60***	0.02	0.43***	-0.45**
X-XIV	0.22	-0.00	0.63***	0.53**
XI-XII	0.25	-0.04**	0.47**	1.07***

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

Tabla A7: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 ocupando características individuales y regionales, submuestra hombres

Variable	Coefficiente	Error estándar	
ln (ingreso) estimado	-0.225	0.192	
ln (distancia)	-0.393***	0.049	
Región contigua	-3.208***	0.211	
Región no contigua	-3.437***	0.287	
Tasa de desempleo (2010)	-11.166	6.897	
Nivel de pobreza (2009)	0.117	0.127	
Crecimiento regional (2009-2011)	0.758	0.540	
Índice de diversidad sectores económicos (2010)	-88.219*	49.827	
Tasa de camas hospitalarias (2010)	-31.071	26.960	
Puntaje PSU promedio regional (2010)	0.276	0.200	
Caminos/Población (2010) ("Congestión")	-24.707*	14.969	
Tasa de denuncias de delitos (2010)	0.026	0.016	
Tasa hogares con banda ancha internet (2010)	-45.189**	23.049	
Emisiones de MP2.5 (2010)	0.002**	0.001	
ln (población) (1992)	-19.265*	10.270	
Efectos fijos por región		No	
Wald Chi2 (4) p-value		0.0000	
Observaciones		439.344	
Casos		36612	
Región	Edad	Con pareja	Ed. Superior
I-XV	-0.07***	0.60***	-0.37
II	-0.01	-0.18	-0.97***
III	-0.03***	-0.42***	-0.25
IV	-0.01	0.24	-0.40*
V	0.00	0.28**	-0.66***
VI	-0.02***	0.09	-0.74***
VII	-0.02**	0.57***	-0.83***
VIII	-0.02**	0.50***	-1.46***
IX	-0.03***	0.27	-0.77***
X-XIV	-0.03***	0.23	-0.44*
XI-XII	-0.05***	-0.08	0.28

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.

Tabla A8: Resultados de regresión modelo de migración interna entre el año 2010 y el año 2015 ocupando características individuales y regionales, y componentes principales

Variable	Coficiente	Error estándar		
ln (ingreso) estimado	0.406**	0.182		
ln (distancia)	-0.416***	0.041		
Región contigua	-3.133***	0.174		
Región no contigua	-3.364***	0.238		
Economía regional: Componente principal 1	0.990	0.603		
Economía regional: Componente principal 2	-5.981***	1.598		
Educación: Componente principal 1	4.701***	1.222		
Educación: Componente principal 2	0.017	0.276		
Educación: Componente principal 3	1.015*	0.527		
Salud: Componente principal 1	-1.476***	0.245		
Salud: Componente principal 2	1.228*	0.632		
Tasa de denuncias de delitos (2010)	0.011	0.003		
Tasa hogares con banda ancha internet (2010)	-46.606***	12.821		
Emisiones de MP2.5 (2010)	-0.001**	0.001		
ln (población) (1992)	-0.887**	0.392		
Efectos fijos por región	No			
Wald Chi2 (4) p-value	0.0000			
Observaciones	751.512			
Casos	62.626			
Región	Hombre	Edad	Con pareja	Ed. Superior
I-XV	0.55***	-0.06***	0.66***	-0.32
II	0.05	-0.02**	-0.02	-0.71***
III	0.22*	-0.04***	-0.32***	-0.08
IV	0.35***	-0.01	0.18	0.03
V	0.14	0.00	0.25**	-0.56***
VI	0.21**	-0.02***	0.16*	-0.48***
VII	0.10**	-0.02***	0.45***	-0.73***
VIII	0.24**	-0.01*	0.43***	-1.20***
IX	0.56***	-0.02**	0.28**	-0.79***
X-XIV	0.44***	-0.03***	0.25**	-0.04
XI-XII	0.25*	-0.06***	0.10	0.42**

*, **, ***: Significancia estadística al 10%, 5% y 1% respectivamente.

Notas: Estimaciones calculadas con ingresos estimados considerando corrección por sesgo de selección de migración y trabajo. Errores estándares estimados con Bootstrapped standard errors con 200 replicaciones.