

Documentos de trabajo

Fundamentos de la nueva metodología de
calibración de los factores de expansión
de la Encuesta Nacional de Empleo.

Autores:

Miguel Alvarado

Matías Pizarro

N°12, Diciembre de 2019



INSTITUTO NACIONAL DE ESTADÍSTICAS

Morandé 801, Santiago de Chile

Teléfono: 562 3246 1000

Correo: ine@ine.cl

Facebook: @ChileINE

Twitter: @INE_Chile

Miguel Alvarado[†]

Matías Pizarro

Los autores desean expresar un reconocimiento especial a Diego Inostroza, ex-funcionario de la institución, quien contribuyó en el desarrollo inicial de este trabajo, como parte del equipo de trabajo del Proyecto Marco Muestral de Viviendas del Instituto Nacional de Estadísticas. Naturalmente, cualquier error, dentro de este documento, es de absoluta y exclusiva responsabilidad de ambos autores.

[†] CONICYT - PCHA/Magister para Funcionarios del Sector Público/2019 - 79190218.

Los Documentos de Trabajo del Instituto Nacional de Estadísticas están dirigidos a investigadores, académicos, estudiantes y público especializado en materias económicas, y tienen como objetivo proporcionar un análisis exhaustivo sobre aspectos conceptuales, analíticos y metodológicos claves de los productos estadísticos que elabora la institución y, de esta forma, contribuir al intercambio de ideas entre los distintos componentes del Sistema Estadístico Nacional.

Las interpretaciones y opiniones que se expresan en los Documentos de Trabajo pertenecen en forma exclusiva a los autores y colaboradores y no reflejan necesariamente el punto de vista oficial del INE ni de la institución a la que pertenecen los colaboradores de los documentos.

El uso de un lenguaje que no discrimine ni marque diferencias entre hombres y mujeres ha sido una preocupación en la elaboración de este documento. Sin embargo, y con el fin de evitar la sobrecarga gráfica que supondría utilizar en castellano “o/a” para marcar la existencia de ambos sexos, se ha optado por utilizar -en la mayor parte de los casos- el masculino genérico, en el entendido de que todas las menciones en tal género representan siempre a hombres y mujeres, abarcando claramente ambos sexos.

Fundamentos de la nueva metodología de calibración de los factores de expansión de la Encuesta Nacional de Empleo.

Resumen

Este documento forma parte de la documentación que acompaña y da sustento técnico a la nueva calibración de los factores de expansión en la Encuesta Nacional de Empleo (ENE), la cual se publicará en extenso en febrero de 2020, con especificidad respecto de los criterios de calibración, señalando y describiendo el método estadístico utilizado, así como su implicancia en la serie de la encuesta. En particular, este documento tiene como propósito, describir exhaustivamente los fundamentos estadísticos y económicos bajo los cuales se estructura la nueva metodología de calibración de los factores de expansión en la Encuesta Nacional de Empleo, con especial foco en la reducción de dos importantes sesgos que actualmente tienen las estadísticas del mercado del trabajo provenientes de la ENE: subrepresentación de la *fuerza de trabajo primaria y población extranjera*.

Abstrac

This document is part of the documentation that accompanies and gives technical support to the new calibration of the expansion factors in the National Employment Survey (ENE), which will be published extensively in February 2020, with specificity regarding the criteria calibration, indicating and describing the statistical method used, as well as its implication in the survey series. In particular, this document is intended to describe in detail the statistical and economic foundations under which the new methodology of calibration of the expansion factors in the National Employment Survey is structured, with special focus on reducing two important biases that currently have labor market statistics from the ENE: underrepresentation of the *primary labor force and foreign population*

Palabras clave: *calibración, fuerza de trabajo primaria, migración internacional, muestra, marco muestral, sesgo.*

1. Antecedentes

Las encuestas de hogares son una de las principales fuentes de estadísticas sociales y demográficas en muchos países (UN-Statistics Division, 1982). Estas encuestas tienen el objetivo de entregar información de la población y/o acerca de subpoblaciones de interés especial (Gutiérrez, 2016), a partir de información de una pequeña parte de la población; es decir, a partir de una *muestra*. La selección de una muestra, más precisamente, la selección de los objetos de la población que pertenecerán a la muestra, requiere de un dispositivo que permita identificar, seleccionar y ubicar a todos y cada uno de los objetos de la población, a este dispositivo se denomina *marco de muestreo* (Gutiérrez, 2016) o, también, *marco muestral*. De este modo, el marco muestral se constituye en uno de los insumos más importantes para lograr alcanzar los distintos objetivos de las encuestas de hogares.

Los marcos muestrales son de dos tipos: marco de listas o marco de áreas¹. Un marco muestral de áreas, en general, se conforma a partir de la aglomeración de viviendas contenidas en áreas, tales como los sectores censales, manzanas, zonas e incluso barrios completos. Al interior de estos conjuntos de viviendas, comúnmente denominados como *conglomerados*², se encuentran las unidades de observación; estos son, los hogares y/o las personas en su interior, de las cuales se busca obtener información.

Un marco muestral se dice *perfecto* si, es *exhaustivo* (todos los elementos de la población están en el marco muestral), es *exacto* (todos los elementos de la población se incluyen una sola vez en el marco muestral) y, además, mantiene las anteriores características a través del tiempo, es decir, está *actualizado*. De este modo, un marco muestral actualizado es un marco muestral perfecto. Lamentablemente, los marcos muestrales actualizados o perfectos, suelen ser una excepción más que la regla general. No obstante, se debe procurar minimizar – tanto como se pueda – situaciones donde se hace uso de marcos muestrales *imperfectos*, es decir, marcos muestrales *desactualizados*, puesto que, como se describirá en parte de este documento, cuando una encuesta de hogares utiliza un marco muestral desactualizado, esto constituye una fuente frecuente de error no muestral³ y, en particular, de cobertura insuficiente de importantes subgrupos (i.e. subpoblaciones) de la población (UN-Statistics Division, 1982), pudiendo generar indeseables efectos sobre los resultados de la encuesta, esto es, estimaciones sesgadas⁴.

Un marco muestral desactualizado, entonces, no incluye la totalidad de los elementos de la población; y en el caso particular de los marcos muestrales de áreas, su desactualización suele reflejarse en:

¹Un marco de listas corresponde a listados que permiten identificar y ubicar a los objetos que participan de la selección; en tanto, un marco de áreas corresponde a un conjunto de mapas, fotografías o imágenes geográficas que permiten identificar y delimitar el área en la cual están contenidos los objetos susceptibles de selección. Se enfatiza únicamente en esta última clase de marcos muestrales, pues es la clase a la cual corresponden los marcos muestrales que utiliza la Encuesta Nacional de Empleo.

²Los conglomerados tienen la característica de estar conformados por elementos homogéneos dentro de sí y heterogéneos entre sí. Además, corresponden a los objetos a ser seleccionados en una muestra, permitiendo el acceso a las unidades de observación (Gutiérrez, 2016).

³El error no muestral se refiere al sesgo que se produce en la estimación, producto de errores en el diseño y la ejecución de la encuesta (UN-Statistics Division, 1982).

⁴La falta de cobertura en ciertas subpoblaciones no es causa exclusiva de la imperfección de los marcos muestrales, puede deberse también a eventos como la *no respuesta*.

- I. *La omisión de áreas*; esto es, existe en la población, conglomerados de viviendas, en cuyo interior se encuentran hogares y personas, que no están en el marco muestral⁵, y/o
- II. *La infrarrepresentación al interior de algunas áreas*; esto es, existen en el marco muestral, conglomerados con menos viviendas de las que efectivamente existen en la población⁶.

De este modo, el natural crecimiento en la cantidad de viviendas, de no mediar protocolos que permitan realizar actualizaciones continuas sobre un marco muestral, conducirá irremediablemente hacia un marco muestral desactualizado. En este sentido, es razonable pensar que, la desactualización del marco muestral es un evento que ocurre con mayor probabilidad en aquellas áreas de gran crecimiento habitacional y, de manera menos probable, en áreas más estables; por tanto, la falta de cobertura se concentrará en aquellas subpoblaciones que habitan aquellas áreas con importante expansión habitacional⁷.

Como se ha mencionado, un marco muestral desactualizado excluye conglomerados de viviendas y a las personas que residen en su interior, generando omisión o cobertura insuficiente en aquellas subpoblaciones que residen en aquellas áreas donde se concentra la desactualización del marco muestral. Si la omisión de elementos de la población corresponde a solo una pequeña proporción de la población de estudio, por lo general, no causará un sesgo importante en las estimaciones de la encuesta (UN-Statistics Division, 1982). Sin embargo, si esta exclusión de personas queda concentrada en subpoblaciones cuyas características sociodemográficas son muy distintas de las características de las personas que no son excluidas del marco muestral, entonces, se producirán sesgos en las estimaciones de la encuesta que utiliza tal marco muestral desactualizado⁸ (UN-Statistics Division, 1982).

Durante 2018, previo a la implementación de los vigentes marcos muestrales que utiliza la Encuesta Nacional de Empleo (ENE), los cuales fueron recientemente actualizados con información de PreCenso 2016⁹, se llevó adelante una prueba piloto paralela a la muestra oficial de la ENE¹⁰. La prueba piloto, cuya muestra fue seleccionada desde el actualizado marco muestral de manzanas (MMM2016), tuvo el propósito de verificar si las características sociodemográficas de la población que sistemáticamente venía siendo excluida de las selecciones muestrales de la ENE, son diferentes de las características de la población que efectivamente venía siendo estudiada en la ENE; y, de encontrar diferencias, identificar los efectos que se tendrían sobre los principales indicadores de la ENE, producto de la utilización de muestras

⁵Debido al natural crecimiento horizontal en las zonas pobladas, un marco muestral desactualizado omite todas las nuevas áreas (por ejemplo, las nuevas manzanas) que fueron creadas con posterioridad a la creación del marco muestral; y, por tanto, también omite a las viviendas y a las personas residentes en su interior.

⁶Debido al natural crecimiento vertical en las zonas densamente pobladas, un marco muestral desactualizado infrarrepresenta las nuevas construcciones (por ejemplo, los nuevos edificios de departamentos) que fueron construidas y/o terminadas con posterioridad a la creación del marco muestral; y, por tanto, también omite a las viviendas y a las personas residentes de las nuevas construcciones.

⁷Regresaremos sobre este punto cuando expliquemos por qué la desactualización del marco muestral generó una cobertura insuficiente de la población extranjera.

⁸En la cuarta sección se discute con mayor detalle sobre los efectos que tiene en las estimaciones de una encuesta que utiliza un marco muestral desactualizado.

⁹Para mayor referencia sobre el proceso y la metodología de actualización de los marcos muestrales de manzanas y secciones se puede consultar INE (2019).

¹⁰La prueba piloto se llevó a cabo en el área urbana (CD) de cuatro regiones del país: Tarapacá, O'Higgins, Magallanes y Metropolitana, correspondiente al trimestre móvil mayo, junio y julio de 2018. Para mayor detalle sobre el diseño de la prueba piloto ver Marín y Touron (2018); para detalle sobre los principales resultados de la prueba piloto ver Alvarado y Touron (2018) y (2019).

seleccionadas desde el MMM2016¹¹.

Dentro de los principales resultados de la prueba piloto, Alvarado y Touron (2018) encuentran evidencia respecto a que, la población en edad de trabajar que reside en las manzanas excluidas de las selecciones muestrales de la ENE, presentan una mayor participación en *tramos de edades jóvenes*, en específico, personas entre los 25 y 54 años, la población denominada como, *fuerza de trabajo primaria*¹²(FTP); al igual que una mayor presencia de *personas extranjeras*, respecto de la participación que reflejan estas dos subpoblaciones dentro de la muestra oficial de la ENE, es decir, dentro de la población objeto de estudio de la ENE, esto es, la población en edad de trabajar, la desactualización del marco muestral generó que se omitiera a parte de la población que, por un lado, forma parte de la población objeto de estudio de la encuesta y, por otro, presenta características distintas de la población que si viene formando parte de las muestras oficiales de la ENE y – como mostraremos – tienen una particular importancia dentro del fenómeno que busca estudiar la encuesta. Evidentemente, la omisión de parte de estas subpoblaciones, no solo generó una insuficiente representación de personas extranjeras dentro de las muestras que venía seleccionando la ENE, sino que, además ha sobrerrepresentado a las personas en tramos de edades mayores. De manera más precisa, en términos muestrales y para el dominio geográfico comparable, la prueba piloto recolectó una proporción mayor (3.6 pp.) de personas en la fuerza de trabajo primaria (53.2%), respecto de lo reportado con la muestra oficial de la ENE (49.6%); en tanto, la prueba piloto recolectó una menor proporción (3.3 pp.) de personas de 55 y más años (28.3%), mientras que la ENE recogió una proporción mayor para ese mismo grupo (31.6%). Sobre la población extranjera, la prueba piloto recabó una proporción mayor (2.6 pp.) de personas extranjeras (6.5%) mientras que la ENE reportó una proporción significativamente menor (3.9%)¹³.

Estos resultados, en particular el que concierne a la población extranjera, como señalan Alvarado y Touron (2018), tienen una clara explicación. Ellos muestran que un 83.5% de las personas extranjeras dentro de la muestra de la prueba piloto, provienen de manzanas con nula probabilidad de selección dentro de la muestra de la ENE. En particular, destacan que un 50.7% de las personas extranjeras de la prueba piloto, se encontraron en manzanas que son asociadas principalmente a nuevas construcciones o edificaciones de altura. Este último resultado es coherente con los resultados reportados en el documento “*Características de la inmigración internacional en Chile, Censo 2017*”; dentro del cual señalan que un 40,4% de los hogares de población inmigrantes residen en viviendas del tipo *Departamento en edificio*.

Alvarado y Touron (2018) y (2019), señalan que la cobertura insuficiente de estas dos subpoblaciones, debido al particular vínculo que tienen estas personas con el mercado del trabajo, tiene efectos no despreciables sobre los principales indicadores que continuamente reporta la ENE. En este sentido, las diferencias encontradas entre la población que venía siendo

¹¹Diversos estudios muestran que las características sociodemográficas y la composición del hogar son determinantes que explican las decisiones de las personas sobre su vínculo con el mercado del trabajo.

¹²La fuerza de trabajo primaria (Prime-Age Workforce) se refiere a las personas que tienen un trabajo (ocupados) o están disponibles para trabajar y están buscando un trabajo (desempleados), comprendidos entre los 25 y 54 años. Estas personas, en general, han concluido su etapa de educación formal y aún no están cerca de la edad de jubilación y, como grupo, tienen la tasa de participación – y ocupación – más alta dentro de la fuerza de trabajo, y corresponden a aproximadamente 2/3 de la fuerza de trabajo (CBO – US, 2018).

¹³Para un mayor detalle sobre estos resultados se puede consultar Alvarado y Touron (2018) y (2019).

excluida de las selecciones muestrales de la ENE – debido a la evidente desactualización de su marco muestral – y la población que venía siendo estudiada, pueden explicar – al menos en parte – los públicos cuestionamientos realizados sobre los principales indicadores reportados por la ENE; en particular, sobre aquellos referidos a que la encuesta estaría subestimando los totales de la fuerza de trabajo, la ocupación y, en particular, la “ocupación Formal”. Evidentemente, aun cuando en los últimos años Chile ha experimentado un significativo influjo de migración internacional (Aldunate, et. al. 2019), la encuesta reporta una estimación significativamente menor para una subpoblación que viene principalmente a trabajar, entonces, es natural pensar que los resultados reportados por la ENE, no corresponden a un adecuado reflejo de la coyuntura del mercado laboral de Chile. Dicho esto, es válido preguntarse *¿por qué la subcobertura de personas en la fuerza de trabajo primaria y de personas extranjeras puede tener efectos sobre fenómenos (entre muchos otros) como la participación, la ocupación y la inactividad?*

Por un lado, las personas en la fuerza de trabajo primaria, es decir, aquellas entre los 25 y 54 años, se caracterizan por presentar altas Tasas de participación y ocupación (CBO – US, 2018), es decir, son personas que participan en el mercado del trabajo y, en particular, lo hacen con bastante éxito; por tanto, tienen una alta probabilidad de estar ocupados. De este modo, su omisión reduce la posibilidad de encontrar personas que tienen una alta probabilidad de estar en la fuerza de trabajo y ser ocupados, lo cual redundaría en menores niveles para los stocks de fuerza de trabajo y Ocupados, así como menores tasas de participación y ocupación, indicadores que son reportados por la ENE. Evidentemente, que las personas en la fuerza de trabajo primaria presenten altas tasas de participación, supone que tienen bajas tasas de inactividad, por cuanto, su omisión aumenta la posibilidad de encontrar personas que tienen una alta probabilidad de estar dentro de la inactividad, lo cual redundaría en niveles elevados para el stock de inactivos que coyunturalmente reporta la ENE. Por su parte, el último Censo de Población y Vivienda de abril de 2017 mostró que, las personas extranjeras entre los 25 y 54 años, correspondían al 72.3 % de los extranjeros en edad de trabajar; de igual modo, la reciente estimación de personas extranjeras a diciembre de 2018 estimó que, las personas extranjeras entre los 25 y 54 años, corresponden al 73.3 % de los extranjeros en edad de trabajar¹⁴; de este modo, es evidente que la población extranjera en edad de trabajar se concentra principalmente dentro de la fuerza de trabajo primaria, la cual – como se mencionó – se caracteriza por presentar altas tasas de participación y ocupación y una baja tasa de inactividad, por cuanto, la omisión de la población extranjera, también afecta a los resultados reportados por la ENE, del mismo modo, que los afecta la omisión de parte de la población dentro de la fuerza de trabajo primaria.

En este mismo sentido, Bravo y Urzúa (2018), utilizando datos de la ENE, muestran que para el trimestre marzo-mayo de 2018, la población extranjera tiene comparativamente una concentración mayor en población en edad de trabajar joven (personas entre 15 y 59 años) correspondiente al 81,8 % (59,6 % para los nacionales) y menor concentración en la población en edad de no trabajar (personas menores de 15 años) correspondiente al 12.1 % (19,7 % para los nacionales) y – menos aún – en población de tercera edad (personas de 60 años y más) correspondiente al 6.2 % (20,7 % para los nacionales), y que estas diferencias de composición en la población, se traducen en que la población extranjera tiene una mayor participación laboral,

¹⁴Para mayor detalle se puede visitar <https://www.ine.cl/estadisticas/demograficas-y-vitales> y consultar el tabulado “Estimación de Personas Extranjeras Residentes en Chile”, así como el documento INE (2018b).

respecto de los nacionales¹⁵. De igual modo, Aldunate et.al. (2019) señala que, la migración hacia Chile ha sido por razones económicas, puesto que los extranjeros son más jóvenes y tienen tasas de participación más altas que los chilenos; en específico, utilizando datos del Censo de 2017, reportan que los extranjeros tienen una tasa de participación de 80.2%, en tanto que los nacionales exhiben una tasa de 61.2%¹⁶. De este modo, dado que la fuerza de trabajo primaria y la población extranjera en Chile son subpoblaciones que exhiben altas tasas de participación y ocupación y bajas tasas de inactividad, su omisión genera sesgos en los resultados que reporta la ENE. En específico, la cobertura insuficiente dentro de estas subpoblaciones, al reducir la posibilidad de encontrar personas que tienen altas tasas de participación y ocupación y bajas tasas de inactividad, evidentemente, redundaría en que la ENE reporte menores niveles para los stocks de fuerza de trabajo y ocupación y mayores niveles para los Inactivos, así como menores tasas de participación y ocupación y mayores tasas de inactividad.

Claramente, la subcobertura de parte de las personas dentro de la fuerza de trabajo primaria y de las personas extranjeras – debido a la desactualización del marco muestral – ha generado sesgos en los resultados coyunturalmente reportados por la ENE; en particular, esto podría explicar, por ejemplo, la subestimación que se genera en las cifras de los *ocupados cotizantes*; sin embargo, esto solo explica una parte del problema. La parte que falta dentro de toda esta historia – que es el foco de este documento – corresponde a la forma en la cual las personas que componen la muestra oficial de la ENE y, por tanto, su información, deben ser ponderadas o calibradas. La ponderación o calibración de los factores de expansión, según señala Gutiérrez (2018), se hace buscando que las estimaciones que se derivan de encuestas de hogares, logren: i) ser consistentes con valores poblacionales conocidos (proyecciones de población, registros administrativos, etc.), ii) las estimaciones sean más precisas y, iii) disminuir sesgos, debido a la ausencia de respuesta o por subcobertura, debido a la imperfección de los marcos muestrales. Sin duda, este último punto es muy importante, para uno de los (muchos) propósitos de la ENE: corregir los sesgos que generan los marcos muestrales desactualizados.

Sobre esto, Alvarado y Touron (2018), relevan la necesidad de contar con un nuevo paradigma sobre la forma en la cual se calibran los factores de expansión. Ellos señalan que, *“Un marco muestral actualizado reduce la subcobertura de subpoblaciones que, por sus características, tienen directas implicancias sobre las estimaciones de los principales indicadores del mercado laboral. Sin embargo, una muestra que refleja de manera más adecuada la realidad nacional, para maximizar sus bondades, debe estar acompañada por una adecuada metodología de calibración”*. En este mismo sentido, Alvarado y Touron (2019), señalan que *“... una metodología de calibración de los factores de expansión que considere a la edad como una variable necesaria para recoger de manera eficiente el correcto vínculo y dinámica que se genera al interior del mercado del trabajo”*. De modo general, Alvarado y Touron (2018), mencionan que, *“... es importante avanzar en una nueva metodología de calibración de los factores de*

¹⁵Bravo y Urzúa (2018), además, presentan una revisión sobre los determinantes de la inmigración, realizan una caracterización de los migrantes, concentrándose en su capital humano y la dinámica que han mostrado algunas nacionalidades dentro del mercado del trabajo chileno; además de aproximar posibles efectos sobre los ingresos del trabajo.

¹⁶Aldunate et.al. (2019), además, presentan una revisión de la literatura sobre los posibles impactos de la inmigración en el mercado del trabajo y, a la luz de la evidencia recopilada por la literatura económica sobre migración, realizan una muy detallada caracterización sociodemográfica de los inmigrantes, puesto que esta puede sugerir los posibles efectos de la inmigración en la economía nacional.

expansión, incorporando variables que estén fuertemente correlacionadas con la situación de las personas dentro y fuera del mercado laboral, como parte de las variables utilizadas para asignar un peso apropiado a las personas de la muestra de la ENE". Estas ideas aluden a dos importantes conceptos: Estrategia de muestreo representativo y al Principio de Representatividad¹⁷, conceptos sobre los cuales se fundamenta la nueva metodología de calibración de los factores de expansión de la Encuesta Nacional de Empleo.

En la siguiente sección, se presenta el soporte teórico de los resultados que se presentan en este documento, además se introducen dos conceptos que serán muy importantes a lo largo de esta discusión: *El principio de Representatividad* y el de *Estrategia de Muestreo Representativa*. En la tercera sección, se revisa la actual metodología de ponderación o calibración de los factores de expansión de la ENE. En la cuarta sección, se presentan los fundamentos de la nueva metodología de calibración de los factores de expansión de la Encuesta Nacional de Empleo, caracterizando el problema, sus causas y efectos. Luego, se describe el por qué se deben considerar los conceptos de *El principio de Representatividad* y el de *Estrategia de Muestreo Representativa* dentro de esta discusión. Además, se muestra cómo al implementar, ambos conceptos y estimadores del tipo que permiten incorporar información auxiliar (exógena) de variables que están correlacionadas con los fenómenos que investigan las encuestas de hogares, por un lado, corrige los efectos nocivos de eventos indeseados que se generan debido a la desactualización de los marcos muestrales y, por otro lado, genera ganancias en la precisión de las estimaciones. Para este propósito, se presentan simulaciones que muestran los claros beneficios que tiene una adecuada calibración de los factores de expansión. La quinta sección contiene las conclusiones. Finalmente, el documento cuenta con una sección de anexos donde se presenta resultados adicionales de los ejercicios realizados.

¹⁷Ver Gutiérrez (2016), capítulo 2.

2. Muestreo probabilístico y el Principio de Representatividad.

En esta sección se desarrolla el necesario soporte teórico de los resultados que se presentan en las siguientes secciones.

2.1. Muestreo probabilístico¹⁸

Como se señala al inicio del documento, las encuestas de hogares buscan entregar información acerca de la población, a partir de información contenida en una *muestra* seleccionada. Las muestras son de dos tipos: *no probabilísticas* y *probabilísticas*; sin embargo, solo estas últimas poseen propiedades estadísticas deseables, por cuanto, se hace ineludible un desarrollo formal de algunos de los conceptos. Con tal propósito, basados en Gutiérrez (2016) y Lumley (2010), se presenta un breve marco teórico del muestreo probabilístico.

2.1.1. Población finita

Una *población finita*, denota por U , es un conjunto de N elementos $\{e_1, \dots, e_k, \dots, e_N\}$, los cuales pueden ser identificados sin ambigüedad por medio de su correspondiente subíndice k ; de este modo, la población puede ser descrita por:

$$U = \{1, \dots, k, \dots, N\} \quad (1)$$

El tamaño de la población se denota por N , el cual no siempre es conocido y en algunas ocasiones el objetivo de la investigación es poder estimarlo (Gutiérrez, 2016).

2.1.2. Muestra aleatoria

Una *muestra aleatoria* es un subconjunto de la población que ha sido extraído mediante un mecanismo estadístico de selección; entonces, toda muestra es función de los elementos que la conforman; en este sentido, la variable aleatoria S denota la muestra aleatoria, mientras que una realización de S , esto es, una muestra seleccionada (realizada) queda denotada por s . De este modo, una muestra seleccionada s , es un subconjunto de elementos de U , los cuales, al igual que en la población, son identificados sin ambigüedad y puede ser descrita por:

$$s = \{1, \dots, k, \dots, n(S)\} \quad (2)$$

El número de componentes de s , denotado por $n(S)$, se denomina tamaño de muestra o, simplemente, tamaño muestral, cantidad que no siempre corresponde a un tamaño fijo; en algunos casos, $n(S)$ es una cantidad aleatoria (Gutiérrez, 2016).

2.1.3. Soporte

Un *Soporte*, denotado por Q , es el conjunto de todas las posibles muestras que pueden ser seleccionadas, dado un método de selección, y se representa por $Q = \{s_1, \dots, s_q, \dots, s_Q\}$.

¹⁸Esta sección, al ser un breve resumen, combina y complementa textos, además, homologa la notación matemática desarrollada en Gutiérrez (2016) y en Lumley (2010); no obstante, debido a la inmejorable claridad de algunos de sus textos, algunos de estos son tomados de manera textual.

2.1.4. Diseño de muestreo¹⁹

Un *diseño de muestreo*, denotado por $p(\cdot)$, es una distribución de probabilidad multivariante definida sobre un soporte Q ; es decir, $p(\cdot)$ es una función que mapea elementos de Q hacia el subconjunto $(0, 1]$, esto es:

$$\begin{aligned} p(\cdot) : Q &\longrightarrow (0, 1] \\ s &\longrightarrow Pr(S = s) = p(s) \end{aligned} \quad (3)$$

tal que:

- I. $p(s) > 0, \quad \forall s \in Q$
- II. $\sum_{s \in Q} p(s) = 1$

En otras palabras, si S es una muestra aleatoria que toma el valor s con probabilidad $p(s)$, se puede escribir esto como:

$$Pr(S = s) = p(s), \quad \forall s \in Q \quad (4)$$

2.1.5. Muestra probabilística

Una *muestra probabilística* es, tal que, para su proceso de selección:

- I. Es posible construir (o al menos definir teóricamente) un soporte Q de todas las muestras posibles de obtener, dado un método de selección.
- II. Las probabilidades de selección que el proceso aleatorio le otorga a cada posible muestra $s \in Q$ son conocidas previo a la selección de la muestra.

2.1.6. Probabilidades de inclusión²⁰

Bajo un diseño de muestreo $p(\cdot)$ y para $k \neq l$, dos elementos cualesquiera de la población. La probabilidad de inclusión del elemento k en la muestra $s \in Q$, denota por π_k , esta dada por:

$$\pi_k = Pr(k \in S) = \sum_{s \ni k} p(s) \quad (5)$$

La probabilidad de inclusión π_k se denomina *probabilidad de inclusión de primer orden*. De igual modo, la probabilidad de inclusión de los elementos k y l en la muestra $s \in Q$, denota por π_{kl} , está dada por:

$$\pi_{kl} = Pr(k \in S \wedge l \in S) = \sum_{s \ni k \wedge l} p(s) \quad (6)$$

La probabilidad de inclusión π_{kl} se denomina *probabilidad de inclusión de segundo orden*.

¹⁹Un detalle más preciso de esta definición y la diferencia que existe con la selección de una muestra, se encuentra en Gutiérrez (2016).

²⁰Para mayor detalle del cálculo de estas probabilidades se puede revisar Gutiérrez (2016).

Adicionalmente, Lumley (2010), señala que, para realizar inferencias basadas en el diseño, se requieren las siguientes importantes propiedades:

- I. Todo individuo en la población debe tener una probabilidad no nula de ser seleccionado en la muestra, es decir, $\forall k : \pi_k > 0$.
- II. La probabilidad π_k debe ser conocida para todo individuo que es seleccionado en la muestra.
- III. Cada par de individuos en la población deben tener una probabilidad no nula de ser ambos seleccionados en la muestra, es decir, $\forall k, l : \pi_{kl} > 0$.
- IV. La probabilidad π_{kl} debe ser conocida para cada par de individuos que son seleccionados en la muestra.

La importancia de estas cuatro propiedades radica en que, las dos primeras propiedades son necesarias para obtener estimaciones validas acerca de la población, mientras que las dos últimas son necesarias para calcular la precisión de las estimaciones (Lumley, 2010).

Lo que resta de esta sección se basa principalmente en Gutiérrez (2011) y (2016), donde se presentan conceptos importantes y se describe un *estimador de muestreo* de particular importancia en las encuestas de hogares.

2.1.7. Característica de interés y parámetro de interés

Las encuestas de hogares tienen como propósito entregar información acerca de una *característica de interés*, denotada por y , la cual se encuentra asociada a cada unidad de la población y corresponde a un valor no aleatorio, y_k , en cada correspondiente elemento k -ésimo de la población. Dicho esto, un *parámetro T* , corresponde a una función de interés que toma por argumentos las características de interés de la población, esto es:

$$T = f(y_1, \dots, y_k, \dots, y_N) \quad (7)$$

Uno de los parámetros de interés más importantes corresponde al *total poblacional* de la característica de interés y , denotado por t_y y definido como:

$$t_y = \sum_{k \in U} y_k \quad (8)$$

Dentro de las encuestas de hogares, el total de personas en la fuerza de trabajo o el total de personas ocupadas, corresponden a un ejemplo de este tipo de parámetros de interés. Otros parámetros de interés son la media poblacional, la varianza poblacional, la mediana poblacional, los percentiles poblacionales, la razón entre dos totales poblacionales, entre otros. Obviamente, todos estos parámetros son desconocidos y, justamente, el objetivo de las encuestas de hogares, que siguen un muestreo probabilístico, es estimar estos parámetros poblacionales a partir de una muestra seleccionada (Gutiérrez, 2016).

2.1.8. Estadística, estimador y estimación

Una *estadística* es una función G de la muestra aleatoria S y solo depende de los elementos pertenecientes a S . Cuando una estadística se usa para estimar un parámetro se dice *estimador* y las realizaciones del estimador en una muestra seleccionada s se dicen *estimaciones* (Gutiérrez,

2016). Una de las propiedades más comúnmente utilizadas de un estimador \hat{T} de un parámetro de interés T es el sesgo, denotado por $B(\hat{T})$, y definido por:

$$B(\hat{T}) = E(\hat{T}) - T \quad (9)$$

Si el sesgo de un estimador es nulo se dice que el estimado es *insesgado* y cuando esto ocurre el *error cuadrático medio* se convierte en la *varianza* del estimador²¹.

Definición 1. Siendo \hat{T} un estimador del parámetro T y $p(\cdot)$ un diseño de muestreo definido sobre Q , se define la dupla $(p(\cdot), \hat{T})$ como una *estrategia de muestreo*.

A continuación, se presenta uno de los estimadores más utilizados en muestreo probabilístico.

2.2. El Estimador de Horvitz–Thompson²²

Para una población U , se quiere estimar el total poblacional de la característica de interés y , denotado por t_y y definido en (8). Entonces, se define el *Estimador de Horvitz–Thompson* para t_y , denotado por $\hat{t}_{y,\pi}$, se define como:

$$\hat{t}_{y,\pi} = \sum_S \frac{1}{\pi_k} y_k = \sum_S d_k y_k \quad (10)$$

Donde π_k corresponde a la *probabilidad de inclusión* del k -ésimo elemento en la muestra, y d_k se suele denominar como *factor de expansión* y corresponde al inverso de la probabilidad de inclusión.

2.2.1. El Principio de Representatividad

La motivación principal detrás del *Estimador de Horvitz–Thompson*, como Brewer (2002) lo indica, descansa en el *Principio de Representatividad*, el cual afirma que *cada elemento incluido en una muestra se representa a sí mismo y a un grupo de elementos que no pertenecen a la muestra seleccionada, cuyas características son cercanas a las del elemento incluido en la muestra.*" (Gutiérrez, 2016).

El *Principio de Representatividad* subraya una importante consideración respecto a qué condición hace que, un elemento dentro de la muestra seleccionada, sea un *adecuado representante* de otros elementos que no están en la muestra seleccionada: *debe representar a elementos cuyas características sean similares a las suyas*. En este sentido, por ejemplo, supongamos una población de elementos cuya única característica sea el sexo de las personas. Según el *Principio de Representatividad*, si el elemento de la muestra seleccionada corresponde a una *mujer*, entonces, dicha *mujer*, *representa adecuadamente a otras mujeres que no están en la muestra, pero no representa adecuadamente a hombres que no están en la muestra*, puesto que no comparten las mismas (en este caso, la única) características: ser *mujer*. De igual modo, por el *Principio de Representatividad*, resulta evidente que, por ejemplo, si un elemento de la

²¹El error cuadrático medio de un estimador \hat{T} de un parámetro de interés T , denotado por $ECM(\hat{T})$ se define como: $ECM(\hat{T}) = E[\hat{T} - T]^2 = Var(\hat{T}) + B(\hat{T})^2$.

²²Este punto sigue a Gutiérrez (2016), donde se puede encontrar un desarrollo preciso y completo de este y otros estimadores de muestreo.

muestra seleccionada corresponde a un *hombre de 25 años*, entonces, este *representa adecuadamente a otros hombres de una edad similar, que están fuera de la muestra seleccionada* pero, por ejemplo, *no representa adecuadamente a hombres mayores de 80 años, que no están en la muestra*; puesto que, aun cuando comparten una de las características (ser hombres), la diferencia entre las edades, los lleva a ser diferentes, en particular, si las características de las personas tiene una alta correlación con el fenómeno que se busca estudiar. No obstante, otro elemento dentro de la muestra seleccionada, por ejemplo, un hombre de 75 años, *representa (más) adecuadamente a los hombres mayores de 80 años, que están fuera de la muestra seleccionada*, de lo que, sin duda, mal representaría aquel hombre de 25 años.

Regresando sobre el Estimador de Horvitz-Thompson, este presenta un resultado muy importante para lo que sigue en este documento.

Teorema 1²³ *Si todas las probabilidades de inclusión de primer orden son mayores que cero ($\pi_k > 0, \forall k$), el Estimador de Horvitz–Thompson es insesgado para el total poblacional de la característica de interés y . Por tanto, se tiene que:*

$$E(\hat{t}_{y,\pi}) = t_y \quad (11)$$

Por otro lado, si el diseño de muestreo $p(\cdot)$ es tal que las probabilidades de inclusión de primer orden π_k conservan una buena correlación positiva con la característica de interés y , esto es ($\pi_k \propto y_k$), el Estimador de Horvitz–Thompson se reduce a una constante y , por tanto, su varianza es nula (Gutiérrez, 2016).

Sin embargo, como bien señala Gutiérrez (2016), *"...en encuestas multi-propósito, donde se quiere estimar parámetros para varias características de interés entre las cuales no hay una buena correlación, al utilizar el Estimador de Horvitz–Thompson es difícil evadir la débil, e incluso negativa, correlación que existe entre las características de interés y las probabilidades de inclusión. Sin embargo, al incluir información auxiliar en la construcción del estimador se puede paliar este hecho."*

Sobre esto último, se presenta una definición muy importante dentro de la teoría del muestreo: *Estrategia de muestreo representativa* y, para esto, consideremos la clase de estimadores $\hat{t}_S(\mathbf{x})$ a los cuales se les incorpora información auxiliar (exógena), denotada por \mathbf{x} .

Definición 2. Una estrategia de muestreo se dice *representativa* con respecto de la información auxiliar \mathbf{x} , sí y solo sí

$$\hat{t}_S(\mathbf{x}) = t_x \quad (12)$$

Es decir, si el estimador aplicado a las variables auxiliares reproduce exactamente el total poblacional de las mismas.

Luego, siguiendo a Gutiérrez (2011) y (2016), si la característica de interés y guarda una estrecha relación lineal con la información auxiliar \mathbf{x} , entonces se puede sostener que la siguiente

²³La prueba puede encontrarse en la página 39 de Gutiérrez (2016).

igualdad se cumple:

$$t_x \approx t_y \quad (13)$$

y, una consecuencia inmediata de esta propiedad, es que:

$$\hat{t}_S(y) \approx t_y \quad (14)$$

Luego, cualquiera sea el diseño de muestreo utilizado, si se cuenta con la información auxiliar t_x , esta se puede utilizar para construir estimadores que mejoran en precisión. De este modo, se tiene el siguiente resultado:

Resultado 1. Si el estimador $\hat{t}_S(\cdot)$ induce una estrategia representativa sobre la información auxiliar x , tal que se satisface (12). Entonces, $\hat{t}_S(x)$ estimará t_x con varianza nula²⁴. Esto es:

$$Var(\hat{t}_S(x)) = Var(t_x) = 0 \quad (15)$$

Desarrollado el soporte teórico necesario, en las siguientes secciones, se hará referencia sobre estos términos desarrollados, en particular, sobre la definición de *estrategia de muestreo representativa* y sobre el *Principio de Representatividad*, pues cobra particular importancia en encuestas de hogares que investigan fenómenos sociales, los cuales están fuertemente vinculados con las características socioeconómicas de las personas. De este modo, aun cuando por la temática propia de este documento, solo se enfatiza sobre el vínculo de las características sociodemográficas de las personas con su situación dentro del mercado del trabajo; se debe tener presente que otros fenómenos sociales como la pobreza, la victimización, entre muchos otros, están estrechamente vinculados a las características socioeconómicas de las personas.

²⁴La prueba puede encontrarse en la página 298 de Gutiérrez (2016).

3. Encuesta Nacional de Empleo: la actual metodología de ponderación de los factores de expansión

La Encuesta Nacional de Empleo tiene un diseño probabilístico, estratificado y bi-etápico, donde en la primera etapa se seleccionan unidades primarias de muestreo (UPM), que en el marco vigente corresponden a manzanas y secciones, y en la segunda etapa, viviendas. Por tanto, el factor de expansión original o teórico da cuenta de estas dos etapas en la selección. Luego, con posterioridad al proceso de recolección de datos en terreno, aun cuando se extremen los esfuerzos operativos por conseguir el contacto y respuesta a la encuesta por parte de los hogares, es habitual la constatación de que no todas las unidades seleccionadas fueron susceptibles de logro y, por tanto, algunas de las unidades seleccionadas no tienen asociada una respuesta. Como medida de mitigación del sesgo que esto podría generar, se realiza un ajuste por no respuesta a los factores de expansión re-ponderando el peso de las unidades para las que no se obtuvo respuesta en aquellas para las que sí se obtuvo, de manera de compensar la pérdida de unidades. Realizado lo anterior, se cuenta con un factor de expansión a nivel de viviendas, con la particularidad que cuyos totales y distribución fue ajustada por la ausencia de respuesta a nivel de UPM. Por tanto, para que la encuesta sea útil en realizar inferencia a nivel de personas, es necesario pasar de ponderadores de nivel de viviendas a individuos, es decir, se hace necesario calibrar.

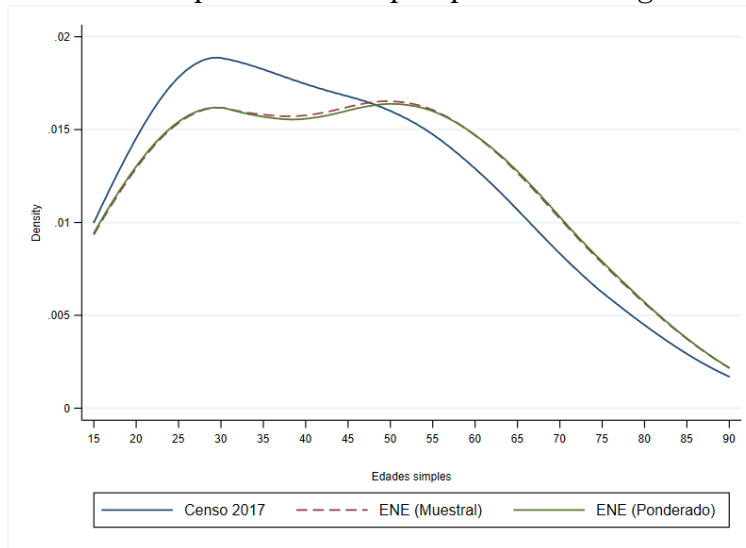
El factor de expansión se calibra con el objetivo de que las estimaciones de población reproduzcan los totales poblacionales según las proyecciones de población para el mes central del trimestre de recolección de datos, considerando dos tramos etarios. La calibración en la ENE se realiza utilizando el método de post-estratificación, siendo los post-estratos de la encuesta, los grupos originados por el cruce de las variables: estrato geográfico, sexo y tramo de edad. Para cada uno de estos grupos se cuenta con información auxiliar proveniente de las proyecciones de población construidas a partir del Censo 2002.

Específicamente, la variable tramo de edad es desagregada en dos categorías: menores de 15 años y 15 años o más, es decir, No PET y PET. La variable "estrato" corresponde a las delimitaciones geográficas definidas como estratos de muestreo para la ENE, considerándose 160 a nivel nacional. La cantidad de post-estratos para la calibración se origina a partir de las interacciones entre las variables estrato (160 grupos), tramo de edad (2 grupos) y sexo (2 grupos), obteniéndose en total 640 post-estratos. De esta manera, la actual forma de calibrar los factores de expansión a nivel de personas en la ENE pone énfasis en la desagregación territorial, 160 estratos territoriales, por sobre las características de los individuos en el resguardo de ciertos totales de interés (sólo se distingue por sexo y entre dos grupos de edad). En consecuencia, si la muestra efectiva no tiene características sociodemográficas similares a las de la población real, la actual metodología de calibración no puede corregir este "problema", ya que, los factores de expansión ponderan por igual, independiente de su edad, condición de nacionalidad, u otra variable correlacionada con la situación de las personas en el mercado laboral, a todos los individuos mayores (15 y más años) del mismo sexo y dentro de la misma unidad de muestreo (manzana-sección) en su respectivo estrato geográfico. Por tanto, no hacer una distinción más precisa en la ponderación, mediante la incorporación en la construcción de los factores de expansión de una mayor cantidad de variables y/o categorías relacionada al fenómeno de estudio,

lleva a que algunas subpoblaciones, terminen siendo sub o sobre ponderadas, con evidentes efectos sobre las estimaciones de la encuesta (Alvarado y Touron, 2019), particularmente afectando a dos grupos de extrema sensibilidad para la medición de las estadísticas laborales: *fuerza de trabajo primaria (FTP)* y *población extranjera residente en el país*.

Cuando comparamos la ENE con Censo en el mes y año correspondiente para ello, como se constata en el Gráfico 3.1, la ocurrencia de personas “jóvenes” (<50 años) en la encuesta es menor a lo que se encuentra en la población y la ocurrencia de personas “mayores” (50 y más años) es mayor a la de la población. Por otro lado, dentro de los hechos estilizados del mercado del trabajo, está ampliamente documentado que la edad determina la situación de las personas dentro y fuera del mercado del trabajo, siendo ésta una variable que está fuertemente correlacionada con las principales variables de interés de la encuesta, por tanto, es importante incorporarla como parte de la información auxiliar para calibrar los factores de expansión en la ENE.

Gráfico 3.1: Densidad por edades simples para la PET según ENE y Censo



Fuente: Elaboración propia en base a datos del Censo 2017 y ENE

Otra subpoblación que tiene una subrepresentación importante en la encuesta es la población extranjera residente en Chile. En la encuesta, para el trimestre marzo-mayo 2017 se estimó en 1,4 % del total nacional, mientras que en el Censo de abril de 2017 los extranjeros alcanzaron el 4,3 % de la población residente en Chile. Si se restringe la población a comparar exclusivamente a la FTP, esta diferencia se acrecienta aún más, llegando a consignar 4,3pp. El documento “Estimación de personas extranjeras residentes en Chile al 31 de diciembre de 2018” (Instituto Nacional de Estadísticas y Departamento de Extranjería y Migración, 2019) desarrollado en colaboración con Policía de Investigaciones de Chile, Ministerio de Relaciones Exteriores y Servicio de Registro Civil e Identificación, señala que este importante grupo de la población, se estimó en 1.251.225 personas en la fecha señalada. Esto implica, que los extranjeros a diciembre de 2018 representan el 6,7 % de la población total, al tomar como referencia las proyecciones de población derivadas del Censo 2017. Como ilustra la tabla 3.1, este grupo, según las fuentes antes mencionadas, los extranjeros dentro de la población total llegan a representar el 10,1 % de la FTP, sin embargo, en la ENE, para el mismo periodo comparable, sólo representan el 3,4 % de

la FTP estimada en la encuesta.

Esta brecha, entre lo que reporta la encuesta y lo que nos señalan *benchmark* de comparación como el Censo de 2017 y la Estimación de población extranjera residente en Chile, es una constatación de la pérdida de confiabilidad en estimaciones de relevancia para la medición y monitoreo del mercado laboral chileno como consecuencia de la elevada inmigración de los últimos años.

Tabla 3.1: Participación de la población extranjera: Población, PET y FTP

Extranjeros (%)	Censo Abr. 2017	ENE MAM-2017	Diferencia (p.p.)	Est. Extr. Dic. 2018	ENE NDE-2018	Diferencia (p.p.)
Población	4.3	1.4	-2.8	6.7	2.1	-4.6
PET	4.8	1.6	-3.2	7.5	2.2	-5.2
FTP	6.6	2.3	-4.3	10.1	3.4	-6.7

Fuente: Elaboración propia en base datos del Censo 2017, Estimación de personas extranjeras residentes en Chile y ENE.

De manera genérica, respecto de los fenómenos migratorios, (Maldonado Valera, Martínez Pizarro, Martínez, 2018) destacan que el mercado laboral, *es en lo inmediato una de las dimensiones principales que se ven afectadas por los flujos migratorios*. De esta manera, existe un desafío doble para las mediciones generadas a partir de encuestas por muestreo; la necesidad de disponer de estadísticas del mercado del trabajo confiables aun cuando los marcos de muestreo y los diseños muestrales no hayan sido diseñados para adecuarse a posibles shocks migratorios, y, una buena medición a nivel de totales, en lo posible a nivel de caracterización de la población migrante y sus implicancias en el mercado del trabajo.

Disponer de marcos muestrales desactualizados contribuye a tener estimaciones sesgadas, no obstante, si a lo anterior se añade una metodología de calibración que no corrige estas discrepancias, entonces el problema está lejos de corregirse, o al menos mitigarse. Por tanto, es evidente la necesidad de reformular y contar con una nueva metodología de calibración en la ENE, que aproveche el necesario cambio en el cálculo de los ponderadores producto de la incorporación de las nuevas proyecciones de población derivadas del Censo 2017, relevando las características sociodemográficas pertinentes de resguardar para una calibración en una encuesta de fuerza laboral.

4. Los fundamentos de la nueva metodología de calibración de los factores de expansión de la Encuesta Nacional de Empleo²⁵

En los últimos años se ha puesto en entredicho los principales resultados que coyunturalmente reporta la Encuesta Nacional de Empleo. El principal argumento señala que, la desactualización del marco muestral que utiliza la encuesta, imposibilita incorporar el reciente, pero importante flujo migratorio que ha experimentado Chile y, por tanto, los resultados de la encuesta tampoco incorporan los efectos que dicho fenómeno ha tenido sobre el mercado laboral, llevando a que las cifras entregadas por la encuesta presenten sesgos en alguna dirección, generando la subestimación de algunos indicadores y la sobrestimación de otros.

Lo antes descrito es correcto; sin embargo, el análisis es incompleto puesto que también existe subcobertura dentro de la subpoblación denominada *fuera de trabajo primaria* que, al igual que la población extranjera, tiene un vínculo particular con el mercado del trabajo, con efectos tanto o más importantes que aquellos de la población extranjera. Más importante aún, el análisis del párrafo anterior, ha llevado a que la discusión del problema, y por tanto su solución, pase a ser un tema de responsabilidad exclusiva de la desactualización del marco muestral, siendo que existen adecuadas técnicas estadísticas que, entre sus muchas otras ventajas, permiten disminuir los sesgos que se generan por la falta de cobertura en grupos de la población, producto de la imperfección de los marcos muestrales y/o debido a la no-respuesta. Estas técnicas, que también tiene la ENE, corresponden al ajuste en la ponderación o calibración de los factores de expansión; sin embargo, como se ha ya señalado, la actual metodología de calibración de la ENE no cumple este propósito.

En lo que sigue de esta sección, primero, se describen los problemas que ha generado la desactualización del marco muestral, problemas que no son resueltos por su actual metodología de calibración de los factores de expansión de la ENE. Descrito el problema, en un segundo punto, se reproduce el problema sobre una población que es replicada a partir de la información del Censo de 2017. Esto permite describir los efectos que la desactualización del marco tiene sobre, primero, la estructura sociodemográfica de la población desde la cual se investiga el fenómeno y, segundo, los efectos que se producen sobre las estimaciones de indicadores del mercado laboral, particularmente, cuando los sesgos en la estructura demográfica de la población desde la cual se estudia el fenómeno, no se corrige y/o cuando los ajustes que buscan corregir el problema no consideran el *Principio de Representatividad* y el concepto de *Estrategia de muestreo representativa*, lo cual es, en esencia, el actual problema de la ENE. Finalmente, en un tercer punto, se muestra cómo, considerar el *Principio de Representatividad* y el concepto de *Estrategia de muestreo representativa*, junto con un *estimador de calibración*, permiten resolver los problemas antes descritos. Luego, al ser una reproducción del problema por el cual atraviesa la encuesta, esto mismo puede aplicarse para resolver los pasados y actuales problemas de la ENE. Más aun, la nueva metodología de calibración prevé cualquier futuro problema de desactualización del marco muestral y/o eventos de no-respuesta, garantizando que las estimaciones oficiales de la ENE, contengan al verdadero

²⁵En este documento, pero más precisamente en esta sección, se presentan los fundamentos estadísticos y económicos que dan soporte a esta nueva metodología. Sin embargo, el detalle técnico del método de calibración, denominado *raking*, por su importante componente matemático, se presentará en un próximo documento.

parámetro poblacional y lo hagan con una mayor precisión.

4.1. La Encuesta Nacional de Empleo: Los efectos de sus marcos muestrales desactualizados y su actual método de calibración

Por un lado, es evidente que la desactualización del marco muestral que utiliza la ENE generó una insuficiente cobertura en parte de la población extranjera residente en Chile. No obstante, se debe mencionar que, a partir de la submuestra del mes de noviembre de 2018, la ENE ha comenzado a seleccionar muestras desde marcos muestrales actualizados con información del Pre Censo de 2016, el cual, evidentemente, corresponde a un marco muestral mucho menos desactualizado que el que se venía utilizando hasta octubre de 2018²⁶. Sin embargo, los resultados de la ENE no reflejan una reducción significativa en la subcobertura de la subpoblación de las personas extranjeras que reporta la ENE, por tanto, tampoco se puede esperar una reducción en los sesgos de las principales estimaciones de la encuesta. Entonces, el problema no es exclusivo de la desactualización o no de un marco muestral (Ver Tabla 4.1).

Tabla 4.1: Encuesta Nacional de Empleo: Población extranjera (muestral y ponderado)

Estimación (%)	2017	2018	2019		
	MAM	NDE	JJA	JAS	ASO
Muestral	1.7	2.4	3.0	3.1	3.1
Ponderado	1.4	2.1	2.8	2.8	2.8

Fuente: Elaboración propia en base a datos de la Encuesta Nacional de Empleo.

Por otro lado, los resultados descritos en Alvarado y Touron (2018) y (2019), dan cuenta que, además de la población extranjera, existe subcobertura dentro de otro grupo muy importante para el fenómeno laboral, este es, la *fuerza de trabajo primaria*. Esta subpoblación, no solo por su importancia relativa dentro de la población en edad de trabajar, sino por el particular vínculo que tienen con el mercado del trabajo, es tan o más importante al momento de explicar los sesgos que presenta la ENE. Sin embargo, del mismo modo, utilizar un marco muestral más actualizado, tampoco resuelve el problema de cobertura insuficiente en dicha subpoblación y, por tanto, sus efectos sobre los resultados que reporta la encuesta. Nuevamente, los problemas de la ENE, no son de exclusiva responsabilidad de no haber contado con marcos muestrales actualizados (Ver Tabla 4.2).

Tabla 4.2: Encuesta Nacional de Empleo: Fuerza de trabajo primaria (muestral y ponderado)

Estimación (%)	2017	2018	2019		
	MAM	NDE	JJA	JAS	ASO
Muestral	48.5	48.0	48.0	47.9	47.9
Ponderado	48.0	47.1	47.1	47.1	47.1

Fuente: Elaboración propia en base a datos de la Encuesta Nacional de Empleo.

Por otro lado, los resultados reportados en las Tabla 4.1 y 4.2, ponen en evidente manifiesto lo expresado en la sección anterior. La actual metodología de ponderación o calibración de los

²⁶Previo a la submuestra del noviembre de 2018, la ENE utilizaba, dos marcos muestrales: el marco de manzanas MMM2008, construido con información del Censo 2002 y actualizado en 2008 y el marco de secciones MMS2002, construido a partir del Censo 2002 y sin ninguna posterior actualización. A la fecha, las muestras de la ENE corresponden a viviendas que fueron seleccionadas desde los marcos de manzanas MMM2016 y de secciones MMS2016, ambos actualizados con información de Pre Censo 2016.

factores de expansión de la Encuesta Nacional de Empleo, lejos de corregir los sesgos que se generan producto de la imperfección de sus marcos muestrales y/o la falta de respuesta, su actual metodología de calibración agrava las subcoberturas, por tanto, también agrava los efectos nocivos que se producen sobre los principales indicadores que coyunturalmente reporta. En el caso de la población extranjera, los resultados muestrales si bien señalan un incremento relativo de personas extranjeras dentro de la muestra, estas observaciones, al ser ponderadas por el factor de expansión, en todos los trimestre móviles de la Tabla 4.1, esta subpoblación pierde importancia relativa dentro de la población, estimando una proporción significativamente menor de la verdadera participación de la población extranjera dentro de la población total²⁷. Un similar escenario se aprecia en el caso de las personas en la fuerza de trabajo primaria. De este modo, los evidentes problemas que se generan debido a la desactualización de los marcos muestrales que utiliza la ENE, no solo no son corregidos por su actual método de ponderación o calibración de sus factores de expansión, al contrario, los problemas se ven agravados.

En lo que va de este documento, se ha destacado el importante papel que tiene un marco muestral para los objetivos de las encuestas de hogares. Con igual énfasis, se ha mencionado que un marco muestral imperfecto, constituye una fuente frecuente de error no muestral y de cobertura insuficiente en importantes subpoblaciones de la población, lo que inclusive puede generar severos sesgos en los resultados de estas encuestas, cuando la falta de cobertura se concentra en subpoblaciones cuyas características sociodemográficas son distintas de las características del resto de los subgrupos de la población que no tienen problemas de cobertura. Por otro lado, equivocadamente se suele pensar que un marco muestral que presenta cobertura insuficiente en subpoblaciones de características sociodemográficas distintas de las del resto de la población que no tienen problemas de cobertura, deriva en muestras cuya composición, al "*no ser proporcional*" a la composición de la población, genera una aparente "*falta de representatividad*" de dichas muestras. Al respecto, es necesario mencionar que este no es un problema es sí, como señala Gutiérrez (2016), *en algunos casos es fundamental sobrerrepresentar algunas categorías o incluso seleccionar unidades con probabilidades desiguales*. Entonces, si sobre-representar algunas subpoblaciones, y por tanto, sobrerrepresentar a otras dentro de la muestra, en ocasiones suele ser una práctica adecuada del muestreo, resulta obvio preguntarse, *si un marco muestral desactualizado genera también sub y sobrerrepresentación en ciertas subpoblaciones, ¿por qué es esto un problema?*. Al respecto, se debe señalar que, aunque el efecto pueda ser "*parecido*", un marco muestral imperfecto implica algo que, no necesariamente implican las selecciones con probabilidades desiguales.

A continuación, se describe la implicancia directa que tienen los marcos muestrales desactualizados.

4.2. Aproximación de los efectos de un marco muestral imperfecto y el uso insuficiente de información poblacional exógena

Recordemos que el Teorema 1 de la segunda sección señala que, sí *todos* los elementos de la población tienen una probabilidad estrictamente positiva de ser seleccionados ($\pi_k > 0, \forall k$),

²⁷La ENE estimó para el trimestre móvil que tiene a diciembre de 2018 como mes central (NDE - 2018), un total de 393613 personas extranjeras (2.1 %), siendo que la estimación oficial del INE y el Departamento de Extranjería y Migraciones a diciembre de 2018, estimó un total de 1251225 personas extranjeras, lo que corresponde a un 6.7 % de la población total.

entonces, el Estimador de Horvitz–Thompson, $\hat{t}_{y,\pi}$, es *insesgado*; es decir, la esperanza de dicho estimador es igual a t_y , es decir, $E(\hat{t}_{y,\pi}) = t_y$. Sin embargo, esta situación no necesariamente ocurre cuando la selección de la muestra es realizada desde un marco muestral imperfecto, pues, por definición, este omite a elementos de la población; por tanto, existen elementos en la población que (al no estar en el marco muestral) tiene probabilidad nula de ser seleccionados ($\exists k : \pi_k = 0$); entonces, esto implica que, el Estimador de Horvitz–Thompson, $\hat{t}_{y,\pi}$, no necesariamente mantiene la propiedad de *insesgades*; es decir:

$$E(\hat{t}_{y,\pi}) \neq t_y \quad (16)$$

Esto último, por la ecuación (9), implica que:

$$B(\hat{t}_{y,\pi}) \neq 0 \quad (17)$$

Entonces, debido a la imperfección de marco muestral, el Estimador de Horvitz–Thompson, $\hat{t}_{y,\pi}$, estima la característica y con sesgo.

Con el propósito de visualizar los efectos de un marco muestral desactualizado, se presentan simulaciones que permiten, por un lado, describir los efectos sobre la estructura demográfica desde la cual se investiga el fenómeno social de interés y , por otro lado, aproximar las consecuencias que esto supone en indicadores laborales, similares a los que se reportan en la ENE.

4.2.1. Población simulada con datos del Censo de 2017

A partir de datos del Censo de 2017, se simula una población que reproduce de manera precisa la estructura demográfica de las variables sexo, tramos de edad²⁸ y dos nacionalidades (chileno/a - extranjero/a). Para las personas dentro de la población en edad de trabajar se simulan cuatro estados dentro del mercado del trabajo. Los estados se denotan por Y_i , con $i = \{1, 2, 3, 4\}$, los cuales corresponden, respectivamente, a estar en la *fuerza de trabajo* ($i = 1$), estar *ocupado* ($i = 2$), estar *desocupado* ($i = 3$) y ser *inactivo* ($i = 4$).

Para asignar el o los estados que le corresponden a cada persona dentro de la población en edad de trabajar de nuestra población simulada, se utilizan dos distribuciones de probabilidades, las cuales se determinan a partir de la información de las personas del Censo de 2017, según el sexo, $s = \{\text{hombre, mujer}\}$, el tramo de edad, $e = \{15-19, 20-24, \dots, 80-84, 85+\}$, y la nacionalidad, $k = \{\text{chileno/a, extranjero/a}\}$: la probabilidad de estar en la *fuerza de trabajo* (denotado por $p_{1,(s,e,k)}$) y la probabilidad condicional de estar *ocupado* dado que se está en el primer estado (denotado por $p_{2,(s,e,k)}$)²⁹. De este modo, los estados de estar (éxito) en la *fuerza de trabajo*, ($Y_1 = 1$), y estar (éxito) *ocupado*, ($Y_2 = 1$), dado que se está en el primer estado, se asignan por medio de una distribución *Binomial*, para una familia de parámetros $(n_{(s,e,k)}, p_{i,(s,e,k)})$, con

²⁸La población se reproduce considerando 16 grupos de edad: el tramo de 0-14 años, 14 quinquenios al interior del tramo entre 15 y 84 años y el tramo de 85 y más años.

²⁹El cálculo de las probabilidades se realizó a partir de la información de la pregunta 17 del Censo de 2017. Para esto, distinguiendo según el sexo s , el tramo de edad e y la nacionalidad k de las personas, se realizó el conteo de las mismas, considerando que quienes contestaron entre 1 y 4, están en la *fuerza de trabajo* ($i = 1$); en tanto, las que contestan entre 1 y 3, son *ocupados* ($i = 2$). Ambas probabilidades y para cada uno de los grupos descritos, se calculan considerando únicamente a las personas de 15 y más años, que hayan contestado sobre su lugar de nacimiento y que hayan contestado en alguna de las opciones de la pregunta 17.

$i = \{1, 2\}$. De este modo:

$$Y_1 \sim \text{Bin}(n_{(s,e,k)}, p_{1,(s,e,k)}) \quad (18)$$

$$Y_2 \sim \text{Bin}(n_{(s,e,k)}, p_{2,(s,e,k)}) \quad (19)$$

Donde, el parámetro $n_{(s,e,k)}$, corresponde al número de personas de sexo s , en el tramo de edad e y de nacionalidad k , mientras que, $p_{1,(s,e,k)}$ y $p_{2,(s,e,k)}$, corresponden a las probabilidades antes descritas; todo esto, para una secuencia de $n_{(s,e,k)}$ ensayos independientes. Por último, los estados de ser *inactivo*, ($Y_4 = 1$), y estar *desocupado*, ($Y_3 = 1$), corresponden, respectivamente, al complemento (fracaso) de los dos primeros estados.

De este modo, en nuestra población simulada, los parámetros demográficos de interés son la población en edad de trabajar y la población según nacionalidad, puesto que en ellos se reflejarán los efectos de la subcobertura. Estos corresponden a:

Tabla 4.3: Parámetros demográficos de la Población simulada: Totales y Participación

Población en edad de trabajar						Nacionalidad	
15-24	25-34	35-44	45-54	55-64	65+	Chilenos/as	Extranjeros/as
2561067	2706276	2362311	2304165	1859049	1955392	16403918	746465
18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.6

Fuente: Elaboración propia en base a la población simulada con datos del Censo 2017.

Del mismo modo, en nuestra población simulada, los parámetros de interés del mercado laboral son el total personas en edad de trabajar (y_0), en la fuerza de trabajo (y_1), ocupados (y_2), desocupados (y_3) e inactivos (y_4); en tanto, las tasas de participación (r_1), ocupación (r_2), desocupación (r_3) e inactividad (r_4). Estos corresponden a:

Tabla 4.4: Parámetros del mercado laboral de la Población simulada: Totales y Tasas

Totales					Tasas (%)			
y_0	y_1	y_2	y_3	y_4	r_1	r_2	r_3	r_4
13748260	8503965	7905541	598424	5244295	61.9	57.5	7.0	38.1

Fuente: Elaboración propia en base a la población simulada con datos del Censo 2017.

En el anexo 1 de este documento, se presentan los totales y_i y tasas r_i , para cada grupo de personas de sexo s , dentro del tramo de edad e y de nacionalidad k , dentro de nuestra población simulada a partir de la información del Censo de 2017.

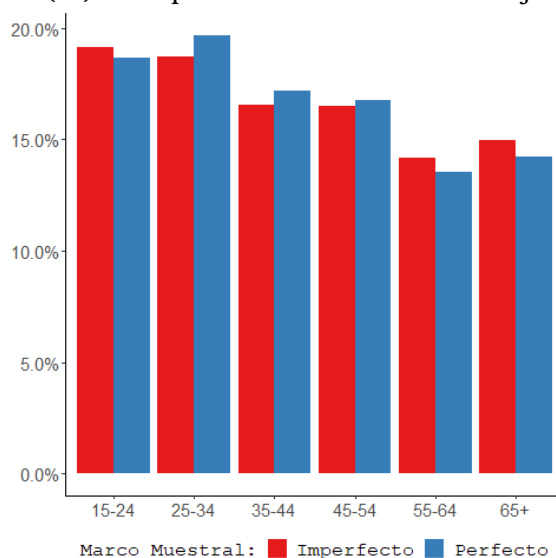
4.2.2. Aproximación de los efectos de un marco muestral imperfecto

Como se ha señalado al inicio de esta sección, un marco muestral imperfecto o desactualizado implica que, existen elementos en la población que no están en el marco muestral, de este modo, tienen nula probabilidad de ser seleccionados, lo que genera que el Estimador de Horvitz–Thompson, no necesariamente mantenga su propiedad de *insesgadez*, al momento de estimar el total de cualquier característica de interés, y . De este modo, con el propósito de visualizar los efectos de un marco muestral, se eliminan (omitir) personas de la población simulada. Para que los efectos puedan ser comparables con los efectos que se producen en la ENE, la omisión de personas se realizada de manera aleatoria, pero restringida a únicamente las dos subpoblaciones sobre las cuales se sabe existe cobertura insuficiente debido a

la desactualización del marco muestral.

Dentro de nuestra población simulada en base al Censo de 2017, se omitió (eliminó) de manera aleatoria al 65 % de las personas extranjeras, lo que corresponde a eliminar 485202 personas de las 746465 personas extranjeras de nuestra población simulada³⁰. Por otro lado, en la subpoblación de personas en la fuerza de trabajo primaria, es decir, aquellas comprendidas entre los 25 y 54 años, se omitió (eliminó) de manera aleatoria al 5 % de estas personas³¹, lo que corresponde a omitir 344503 personas nacionales entre 25 y 54 años³². En síntesis, se omitieron a 829705 personas, lo que equivale a una omisión del 4.8 % de la población total simulada.

Gráfico 4.1: Distribución (%) de la población en edad de trabajar, según tramos de edad



Fuente: Elaboración propia en base a la población simulada (con y sin subcobertura) con datos del Censo 2017

La antes descrita omisión (eliminación) de parte de las personas dentro de la subpoblación de personas extranjeras y personas en la fuerza de trabajo primaria (25-54 años), tuvo el mismo efecto reportado en los resultados de la prueba piloto (ver Gráfico 4 en Alvarado y Touron (2018)), esto es, la omisión (eliminación) de personas en la población replicada con datos del Censo de 2017 (*marco muestral perfecto*, desde donde se calculan por verdaderos parámetros

³⁰Para el trimestre móvil MAM-2017, que es comparable con los resultados del Censo de 2017, la ENE estimó 261042 personas extranjeras, es decir, presentó una falta de cobertura de aproximadamente un 65% del total de personas extranjeras reportadas según el Censo de 2017.

³¹De este proceso de omisión aleatoria, se excluye a las personas extranjeras comprendidas entre los 25 y 54 años y que no fueron eliminadas en el primer proceso de omisión, puesto que se podría generar una omisión mayor de lo que la ENE reporta para tal subpoblación.

³²Para el trimestre móvil MAM-2017, la ENE estimó 7100706 personas en la fuerza de trabajo primaria, en tanto, según el Censo de 2017 existían 7372752 personas para tal subpoblación. Es evidente que, en términos absolutos, la ENE exhibe una subestimación ligeramente superior a las 270 mil personas dentro de tal subpoblación, cantidad que corresponde al 3.7% del total de dicha subpoblación, según el Censo de 2017. Sin embargo, para MAM-2017 la ENE estima un total poblacional (18473452) distinto al total reportado, ex-ante la conciliación, en el Censo de 2017 (17150383). Esto se traduce al interior de las subpoblaciones que estima la ENE; de este modo, el total de personas en la fuerza de trabajo primaria que estima la ENE corresponden al 48.0% del total de personas dentro de la estimación que realiza para la población en edad de trabajar, en tanto, para el Censo de 2017, el total de esta subpoblación corresponde al 53.6% del total de personas en edad de trabajar, es decir, se produce una diferencia de 5.7 puntos porcentuales en contra de la ENE; es decir, en términos relativos, la ENE también subestima a la fuerza de trabajo primaria. De este modo, se optó por omitir (eliminar) aleatoriamente a únicamente el 5% de las personas nacionales comprendidas entre los 25 y 54 años

poblacionales), generó una distorsión en la estructura demográfica (Ver Gráfico 4.1), generando un subconjunto de la población original, que denominaremos *población simulada con subcobertura*, desde la cual se investigaran los fenómenos de interés, es decir, estimar los verdaderos parámetros poblacionales, mismos que, muy rara vez, se conocen.

Con el propósito de estimar los verdaderos parámetros poblacionales, la población simulada con subcobertura (que mostramos busca reproducir la omisión de personas que ocurre en la ENE), corresponderá a nuestro *marco muestral imperfecto* desde el cual, se realizarán selecciones muestrales. Los elementos de las muestras son seleccionadas a partir de un muestreo aleatorio simple, conformando muestras de una tamaño igual al 1% del total de personas que están en nuestro *marco muestral imperfecto*. Con el propósito de mostrar que los resultados corresponden a un efecto estable y robusto, se realizan 1000 simulaciones de selecciones muestrales, que son utilizadas para estimar los totales de las características de interés del mercado del trabajo y sus tasas correspondientes³³. También se estiman algunas variables demográficas, con las cuales comenzaremos la descripción de resultados.

Tabla 4.5: Distribución (%) promedio de la población en edad de trabajar y población extranjera

Estimadores y Parámetros	Población en edad de trabajar						Nacionalidad	
	15-24	25-34	35-44	45-54	55-64	65+	Chilenos/as	Extranjeros/as
Horvitz-Thompson	19.1	18.7	16.5	16.5	14.2	14.9	98.40	1.60
Parámetros	18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.4

Elaboración propia en base simulaciones con la población simulada con subcobertura.

La imperfección del marco muestral que utilizamos para seleccionar las 1000 muestras, se refleja en la distorsión que sufrió la estructura demográfica de la población desde donde se buscarán estudiar las distintas características del mercado laboral. En la Tabla 4.5 se aprecia que, las estimaciones a partir del estimador de Horvitz–Thompson, subestima a las personas en la fuerza de trabajo primaria (25-54 años) y, consecuentemente, sobreestima a las personas en el resto de los tramos de edades. De igual modo, este mismo estimador, subestima la participación de la población extranjera dentro del total poblacional (ver Tabla 4.5).

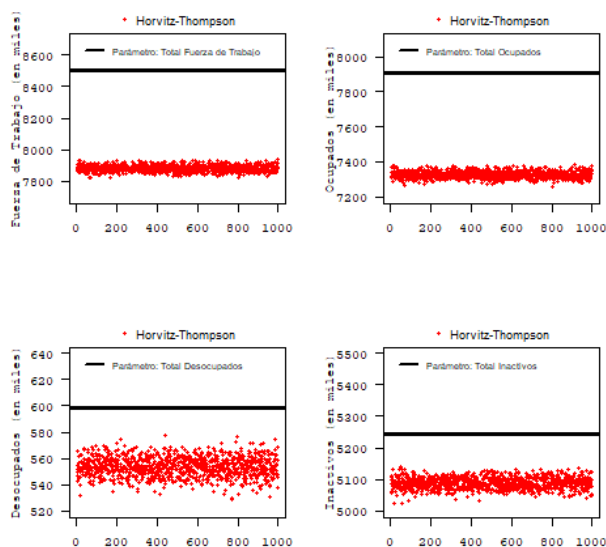
Por otro lado, según queda descrito en el Gráfico 4.2, y consistente con lo señalado en las ecuaciones (16) y (17) de la subsección 4.2, muestra que, cuando existen elementos dentro de la población que tienen nula probabilidad de ser seleccionados ($\exists k : \pi_k = 0$), es decir, existen elementos (personas, en este caso) de la población que no están en el marco muestral (puesto que fueron omitidos de este), las estimaciones que se realizan mediante el Estimador de Horvitz–Thompson (puntos rojos) quedan sesgadas respecto del correspondiente parámetro poblacional verdadero (línea negra). En específico, se aprecia que, en todos los casos se *subestiman* los totales poblacionales para el total de personas en la fuerza de trabajo, el total de ocupados, el total de desocupados y el total de inactivos.

4.2.3. Uso insuficiente de información exógena

Para finalizar esta sección, consideremos la clase de estimadores $\hat{t}_S(\mathbf{x})$ que incorpora información auxiliar (exógena) \mathbf{x} . Como se mencionó en la Sección 2, con esta clase de

³³Los totales de las variables de interés, corresponden al total de personas en la fuerza de trabajo, el total de ocupados, el total de desocupados y el total de inactivos. En tanto, las tasas de participación, ocupación, desocupación e inactividad, se presentan como parte de los anexos finales

Gráfico 4.2: Parámetros y estimaciones de las muestras simuladas



Fuente: Elaboración propia en base a simulaciones con la población simulada con subcobertura

estimadores se busca mejorar la eficiencia de las estimaciones de características de interés y , que estén relacionadas con x . Como señala Gutiérrez (2016), "Estos estimadores, que incorporan información auxiliar, apuntan a la mejora dramática en la eficiencia de las estrategias de estimación de totales poblacionales. Además de esta característica, existen otras que tienen que ver con la consistencia y el insesgamiento. Sin embargo, una característica importante de un estimador construido a partir de información auxiliar está dado la definición de estrategia de muestro representativa."; esto debido a que estos estimadores, por su construcción, reproducen exactamente los totales poblacionales de la variable auxiliar x .

Sin entrar en mayor detalle, suponga³⁴ se cuenta con información auxiliar de la población x . En particular, supongamos se cuenta con información del mismo tipo que la utilizada en la ENE. De este modo, podemos aproximarnos no solo a los efectos que ha generado la desactualización del marco muestral, además, podremos aproximar los efectos derivados de hacer un uso limitado de la información poblacional exógena, tal como lo hace la ENE. Dicho lo anterior, supongamos se cuenta con información sobre el *verdadero total poblacional* de las personas según su *sexo* y según si son *menores de 15 años* o *de 15 y más años (mayores)*. Entonces, tomemos esta información auxiliar, $x = \{\text{Sexo, Menores-Mayores}\}$ e incorpómosla en las simulaciones antes presentados, a partir del estimador $\hat{t}_S(x)$, el cual incorpora la información de x , y utilicémoslo en la estimación de las características de interés como y , a partir de $\hat{t}_S(y)$. Antes de ver los resultados que genera el incorporar información auxiliar de la variable x , precisemos que el propósito de incorporar x es que nuestra estrategia de muestreo sea *representativa* respecto de dicha variable, es decir, que esta reproduzca exactamente el total poblacional de la variable x . Luego, si alguna variable de

³⁴En este caso, no es necesario suponer nada, puesto que conocemos estos valores de la población simulada con datos del Censo de 2017. En el caso de las encuestas de hogares reales, estos valores se pueden obtener de conteos censales, proyecciones de población, registros administrativos, etc.

interés y está correlacionada con la información auxiliar x , la estimación que se deriva de este tipo de estimadores, es por tanto próxima al verdadero valor poblacional de y (ver ecuación 14).

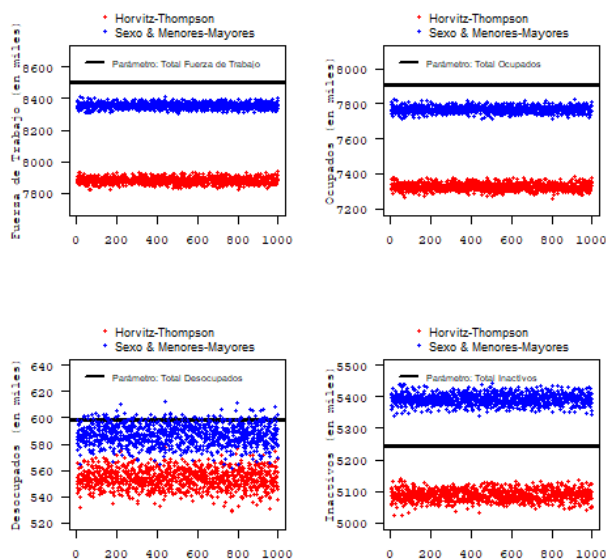
Tabla 4.6: Distribución (%) promedio de la población en edad de trabajar y población extranjera

Estimadores y Parámetros	Población en edad de trabajar						Nacionalidad	
	15-24	25-34	35-44	45-54	55-64	65+	Chilenos/as	Extranjeros/as
Horvitz-Thompson	19.1	18.7	16.5	16.5	14.2	14.9	98.40	1.60
Sexo & Menores-Mayores	19.1	18.7	16.5	16.5	14.2	14.9	98.39	1.61
Parámetros	18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.4

Fuente: Elaboración propia en base simulaciones con la población simulada con subcobertura.

En la Tabla 4.6 se aprecia que las estimaciones que se derivan del estimador $\hat{t}_S(x)$, el cual incorpora la información auxiliar x , sobre el verdadero total poblacional de las personas según su sexo y según si son menores de 15 años o de 15 y más años (mayores), al igual que las estimaciones del estimador de Horvitz–Thompson, aún subestiman a las personas en la fuerza de trabajo primaria (25-54 años) y, consecuentemente, sobreestiman a las personas en el resto de los tramos de edades. De igual modo, este mismo estimador, subestima la participación de la población extranjera dentro del total poblacional. Los resultados, que corresponden al promedio (de ahí que no se vean efectos sino hasta el tercer decimal) de las participaciones que se derivan de las 1000 selecciones muestrales, señalan que este estimador que incorpora información auxiliar x , no logra reducir los sesgos en la estructura demográfica de la población, para la cual buscamos estimar sus parámetros poblacionales de las características del mercado del trabajo, por cuanto, tampoco se podría esperar que los sesgos en las estimaciones se eliminen.

Gráfico 4.3: Parámetros y estimaciones de las muestras simuladas



Fuente: Elaboración propia en base a simulaciones con la población simulada con subcobertura

Evidentemente, las estimaciones que se derivan del estimador $\hat{t}_S(x)$, el cual incorpora información sobre el total de la población por sexo y el total de personas menores de 15 y el total de las personas de 15 y más años, aún presentan sesgos importantes para las cuatro características del mercado laboral que se buscan investigar. Más allá que algunas estimaciones,

para el total de desocupados, estén próximas e incluso puedan coincidir con el verdadero parámetro poblacional, el sesgo persiste. Por otro lado, llama la atención que este mismo estimador, en el caso del total de inactivos, genere una sobrestimación para tal total.

Luego, como el ejercicio buscó estar lo más próximo posible a la situación que atraviesa la ENE, de los resultados del Gráfico 4.3 y considerando los problemas propios que se generan con marcos desactualizados y la actual metodología de calibración, la cual únicamente considera a la variable sexo como una variable que aporta verdadera información para reducir los sesgos en las estimaciones de variables del mercado del trabajo, se puede señalar que, las estimaciones coyunturales que reporta la ENE, para el total de personas en la fuerza de trabajo, el total de ocupados y el total de desocupados, corresponden a estimaciones sesgadas, pues *subestiman* estos totales, debido a que su actual metodología de calibración es insuficiente al momento de corregir la falta de cobertura de personas extranjeras y personas en la fuerza de trabajo primaria; más aún, como la actual metodología de calibración considera un único tramo de edad para las personas en edad de trabajar, los resultados que de ahí se derivan, van en una única dirección: sobre ponderar a las personas que mayor participación relativa tengan dentro de sus muestras y, por tanto, de sus realidades dentro del mercado laboral. En específico, la ENE sobre representa la realidad de las personas de 55 y más años de edad, pero también de aquellas entre los 15 y 24 años; ambos grupos caracterizados por bajas tasas de participación y ocupación y, además, altas tasas de inactividad. Lo anterior, explica no solo la subestimación de la participación y la ocupación, sino, además, explica que las estimaciones coyunturales para el total de personas fuera de la fuerza de trabajo o inactivos, estén *sobreestimadas*, al igual que en las simulaciones.

Estas distorsiones y sesgos que hasta ahora persisten; sin embargo, pueden, y más importante, deben corregirse. En la siguiente subsección, se buscará resolver estos problemas, para lo cual, se consideran los conceptos de *El Principio de Representatividad* y la de una *Estrategia de muestreo representativa*.

4.3. Principio de Representatividad y Estrategia de muestreo representativa

Comencemos viendo los resultados de incorporar información auxiliar \mathbf{x} en un estimador del tipo que permite hacer aquello, como el estimador antes introducido, $\hat{t}_S(\mathbf{x})$, el cual incorpora la información auxiliar \mathbf{x} y utilicémoslo en la estimación de las características de interés como y , a partir de $\hat{t}_S(y)$. A diferencia del caso anterior, \mathbf{x} corresponde a información sobre el *verdadero total poblacional* de las personas según su sexo y según *varios tramos de edad*. En particular, para estos ejercicios se utilizaron 8 tramos de edad, 7 de los cuales corresponden tramos de edad dentro de la población en edad de trabajar³⁵. Entonces, tomemos esta información auxiliar $\mathbf{x} = \{\text{Sexo}, \text{Edades}\}$ e incorporémosla en las simulaciones antes presentados, a partir del estimador $\hat{t}_S(\mathbf{x})$, el cual incorpora la información de \mathbf{x} y utilicémoslo en la estimación de las características de interés como y , a partir de $\hat{t}_S(y)$.

En la Tabla 4.7 se aprecia que las estimaciones que se derivan del estimador $\hat{t}_S(\mathbf{x})$, el cual incorpora la información auxiliar \mathbf{x} , sobre el verdadero total poblacional de las personas según su sexo y varios tramos de edad, logra eliminar la subestimación de las personas en la fuerza de

³⁵Los 8 tramos de edad utilizados: 0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+.

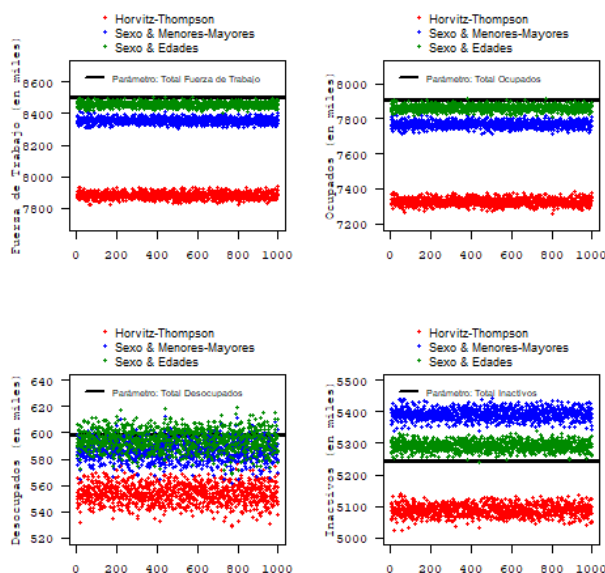
Tabla 4.7: Distribución (%) promedio de la población en edad de trabajar y población extranjera

Estimadores y Parámetros	Población en edad de trabajar						Nacionalidad	
	15-24	25-34	35-44	45-54	55-64	65+	Chilenos/as	Extranjeros/as
Horvitz-Thompson	19.1	18.7	16.5	16.5	14.2	14.9	98.40	1.60
Sexo & Menores-Mayores	19.1	18.7	16.5	16.5	14.2	14.9	98.39	1.61
Sexo & Edades	18.6	19.7	17.2	16.8	13.5	14.2	98.36	1.64
Parámetros	18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.4

Fuente: Elaboración propia en base simulaciones con la población simulada con subcobertura.

trabajo primaria (25-54 años) y, consecuentemente, elimina las distorsiones al interior de las otras subpoblaciones dentro de la población en edad de trabajar. No obstante, este mismo estimador, aún subestima la participación de la población extranjera dentro del total poblacional. Luego, al reducir los sesgos de parte de la estructura demográfica de la población, para la cual buscamos estimar sus parámetros poblacionales de las características del mercado del trabajo, se puede esperar que los sesgos en las estimaciones se minimicen e incluso se eliminen.

Gráfico 4.4: Parámetros y estimaciones de las muestras simuladas



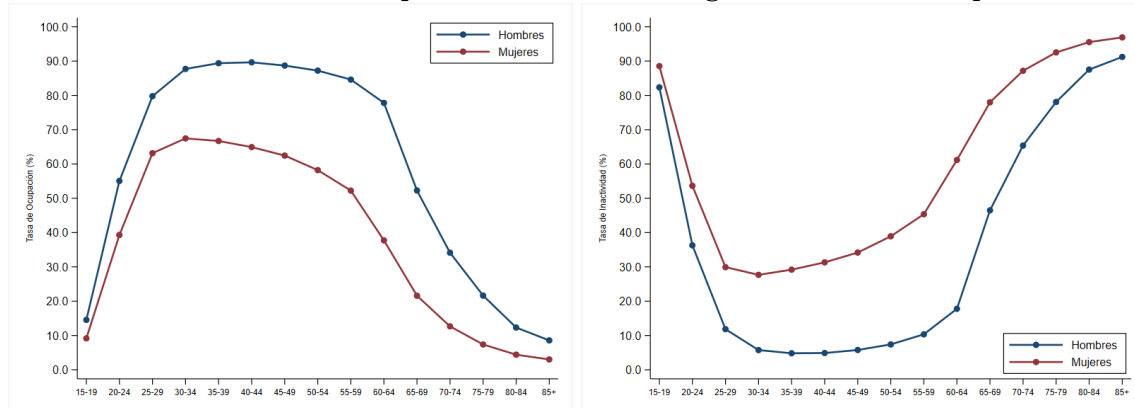
Fuente: Elaboración propia en base a simulaciones con la población simulada con subcobertura

Evidentemente, las estimaciones que se derivan del estimador $\hat{t}_S(\mathbf{x})$, el cual incorpora información sobre el total de la población por sexo y varios tramos de edad, reduce de manera significativa los sesgos para las cuatro características del mercado laboral que se buscan investigar. Incluso, para el total de desocupados, se podría decir que el sesgo está muy próximo a ser nulo.

Entonces, la pregunta natural es *¿Por qué el incorporar información referida a la edad reduce de manera significativa los sesgos?* La respuesta se encuentra en el Gráfico 4.5. En tal gráfico, lo primero que se distingue es que, tanto la tasa de ocupación (probabilidad de estar ocupado), como la tasa de inactividad (probabilidad de no estar en la fuerza de trabajo o estar inactivo), es muy distinta para las mujeres comparando con la de los hombres. Luego, el sexo es una variable

relacionada con la situación de las personas, tanto dentro como fuera del mercado del trabajo; justificando de este modo que, la variable *sexo* debe ser una de las variables que debe formar parte de cualquier set de variables auxiliares x para la calibración. Luego, por ejemplo, como la probabilidad de estar ocupado, depende del sexo, es decir, una mujer no se vincula con el mercado laboral del mismo modo que un hombre lo hace, por el *Principio de Representatividad*, las mujeres de la muestra, al compartir una característica (el sexo) que está fuertemente relacionada con el fenómeno de interés, representan de mejor manera a otras mujeres que no están en la muestra.

Gráfico 4.5: Tasas de ocupación e inactividad, según tramos de edad por sexo



Fuente: Elaboración propia en base a la población simulada con datos del Censo de 2017

Del mismo gráfico, también es evidente que, tanto la tasa de ocupación (probabilidad de estar ocupado), como la tasa de inactividad (probabilidad de no estar en la fuerza de trabajo o estar inactivo) que tiene una persona, incluso controlando por su sexo, difiere según el tramo de edad en el cual se encuentre. Por ejemplo, una persona que está en el tramo de edades de 30 a 34 años, no tiene las mismas probabilidades de estar ocupado, que tiene otra persona en el tramo de edades de 75 a 79 años. En el caso de los hombres, por ejemplo, en el tramo de 30 a 34 años, la tasa de ocupación es próxima al 90 %, mientras que, para un hombre en el tramo de 70 a 74 años, esta misma tasa o probabilidad de estar ocupado, disminuye de manera dramática por bajo el 30 %. Ocurre lo mismo, en el caso de las mujeres, las mujeres entre los 30 a 34 años, tienen una probabilidad de estar ocupadas (65 %), mientras que una mujer que acaba de ingresar a la población en edad de trabajar, tiene una probabilidad de estar ocupada por debajo del 10 %. De este modo, la forma en la cual se vinculan las personas con el mercado del trabajo, está fuertemente correlacionada con la edad, tal como lo señalan las diversas investigaciones referidas a investigar los fundamentos (hechos estilizados) del mercado del trabajo (Ver Contreras et.al. 2010). Entonces, por el *Principio de Representatividad* una persona de la muestra, representará de mejor manera a otras personas con las cuales, además de compartir el mismo sexo, comparta una edad similar. En este sentido, las mujeres de la muestra que tengan entre 30 y 34 años, representaran mejor a las mujeres entre 30 y 34 años que no están en la muestra, puesto que, ellas (las mujeres entre 30 y 34 años) tienen características (el sexo y la edad en la que se encuentran) que las vinculan con el mercado laboral de un modo distinto que a otras mujeres en otros tramos de edad.

De este modo, es posible incorporar información poblacional x que permita corregir los

sesgos que se generan por, justamente, la subcobertura de subpoblaciones que se concentran en tramos de edad específicos. En este sentido, una variable como la edad, la cual además de estar fuertemente correlacionada con los distintos fenómenos de estudio dentro del mercado laboral (participación, ocupación, inactividad, informalidad, etc.), redundaría en reducir los sesgos que se generan en las variables de interés (en este caso, la ocupación) producto de la correlación que existe con la característica demográfica (la edad) sobre la cual se entrega información exógena.

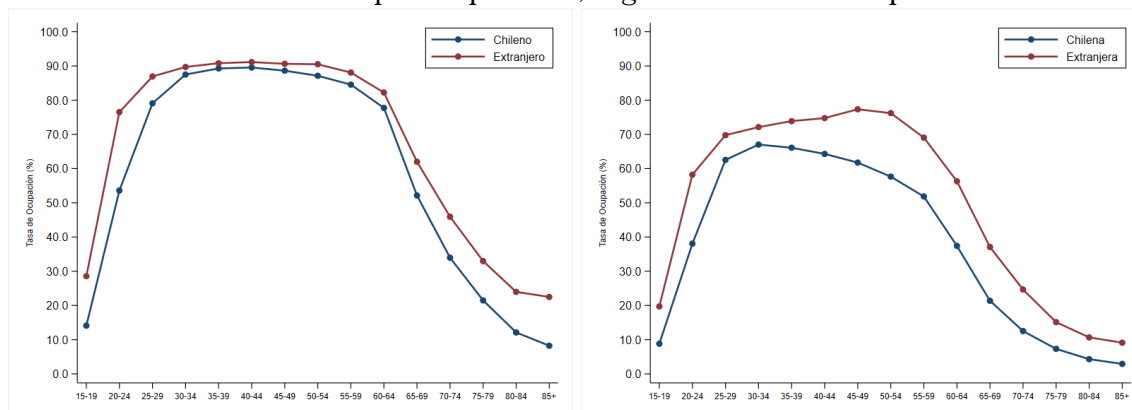
Sin embargo, la realidad (al menos la que se refleja dentro de las simulaciones) es que la edad no es suficiente, para cerrar las brechas demográficas y con esto minimizar los sesgos que se tramiten sobre las características de interés dentro del mercado laboral. Mas aún, se sabe que la población extranjera están concentrada en tramos de edades “jóvenes”³⁶, entonces, resulta natural al menos pensar que al distinguir por varios tramos de edad al momento de calibrar los factores de expansión, esto podría ser suficiente para corregir la sub-cobertura que se genera en ambas subpoblaciones: Fuerza de trabajo primaria y en la población extranjera; y, por tanto, minimizar los sesgos que estas subcoberturas generan sobre las estimaciones.

Los resultados de la Tabla 4.7, claramente señalan que la edad, como parte de la información auxiliar dentro de x , es insuficiente para alcanzar este propósito. Entonces, corresponde nuevamente preguntarnos *¿Por qué el incorporar información referida a la edad no reduce completamente los sesgos y, si estos sesgos solo se minimizan por completo al incorporar información referida a la condición de nacionalidad?*. Se debe volver a razonar como en el argumento de la edad. Del Gráfico 4.5, es evidente que la probabilidad de estar ocupado que tiene un hombre es distinta que la que tiene una mujer, cualquiera sea el tramo de edad que se mire. De este modo, si la probabilidad de estar ocupado es distinta según el sexo, *es posible que la probabilidad de estar ocupado o estar en la inactividad, sea distinta para dos hombres dentro del mismo tramo de edad y que solo los distinga su condición migratoria?*, es decir, *un hombre extranjero y un chileno, en el mismo tramo de edad, tienen las mismas probabilidades de estar ocupados?*. La respuesta se encuentra en el Gráfico 4.6. En ese gráfico, es evidente que la probabilidad de estar ocupado de un chileno en el tramo de 20 a 24 años es cercana a un 55 %, mientras que un extranjero, para el mismo tramo de edades, su probabilidad de estar ocupado es cercana al 80 %. Estas diferencias, ocurren también entre las mujeres de ambas subpoblaciones. En general, los extranjeros como una subpoblación, presenta tasas de ocupación mayores a la de los nacionales, ocurre lo contrario con las tasas de inactividad. Entonces, un similar argumento al planteado para justificar la incorporación de la variable edad, aplica para justificar la variable nacionalidad, en el sentido que, el vínculo con el mercado del trabajo no es el mismo para dos personas de tramos de edades distintas, sino que además el vínculo con el mercado del trabajo no es el mismo, para dos personas dentro del mismo tramo de edad, pero que tienen nacionalidades diferentes. Entonces, por el *Principio de Representatividad* una persona de la muestra que tiene una nacionalidad distinta a la chilena, es decir, una persona extranjera, representará de mejor manera a las otras personas extranjeras que no están en la muestra, de este modo, se hace absolutamente necesario incorporar información auxiliar dentro de x , de forma que esto permita distinguir, de manera adecuado, el vínculo distinto que ambas subpoblaciones tienen con el mercado laboral, pero además, al reproducir de manera precisa la información auxiliar de la variable x , que como vimos recoge variables como el sexo, la edad y la

³⁶Ver Características de la inmigración internacional en Chile, Censo 2017".

condición de ser nacional o extranjero, las cuales están directamente correlacionadas con los fenómenos que se buscan estudiar dentro del mercado del trabajo, tenemos una *Estrategia de muestreo representativa* en x , lo cual permitirá estimar de manera más precisa, características y , con las cuales x guarda directa relación, como los fenómenos del mercado laboral.

Gráfico 4.6: Tasas de ocupación por sexo, según tramos de edad por nacionalidad



Fuente: Elaboración propia en base a la población simulada con datos del Censo de 2017

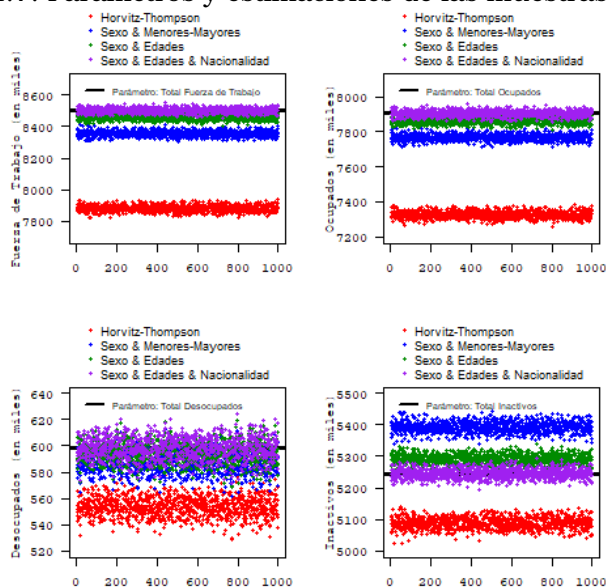
Dicho esto, veamos los notables beneficios de incorporar, además del sexo y la edad, la condición de nacionalidad, como parte del set de variables auxiliares x , que serán utilizadas dentro de la nueva metodología de calibración de los factores de expansión, reconociendo no solo *El Principio de Representatividad*, más importante aún, implementando una *Estrategia de muestreo representativa* en un set de variables auxiliares x (el sexo, la edad y la condición de ser nacional o extranjero, además de otras de sentido geográfico), que están directamente relacionadas con los fenómenos dentro del mercado laboral.

Tabla 4.8: Distribución (%) promedio de la población en edad de trabajar y población extranjera

Estimadores y Parámetros	Población en edad de trabajar						Nacionalidad	
	15-24	25-34	35-44	45-54	55-64	65+	Chilenos/as	Extranjeros/as
Horvitz-Thompson	19.1	18.7	16.5	16.5	14.2	14.9	98.40	1.60
Sexo & Menores-Mayores	19.1	18.7	16.5	16.5	14.2	14.9	98.39	1.61
Sexo & Edades	18.6	19.7	17.2	16.8	13.5	14.2	98.36	1.64
Sexo & Nacionalidad	18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.4
Parámetros	18.6	19.7	17.2	16.8	13.5	14.2	95.6	4.4

Fuente: Elaboración propia en base simulaciones con la población simulada con subcobertura.

Gráfico 4.7: Parámetros y estimaciones de las muestras simuladas



Fuente: Elaboración propia en base a simulaciones con la población simulada con subcobertura

5. Conclusiones

Vincular el propósito de una medición de estadísticas laborales con su calibración, permite mitigar sesgos propios del muestreo y de otras contingencias naturales en encuestas; marcos desactualizados, eventuales problemas en el trabajo de campo, entre otros aspectos. Análogamente, uno de los hechos estilizados del mercado laboral, independiente del desarrollo de este, es que la edad tiene una correlación alta con la situación de las personas dentro y fuera del mercado laboral, específicamente con la inactividad, ocupación y desocupación. En particular, poder resguardar totales a nivel de varios grupos de edad implica minimizar sesgos y mejorar la precisión de las estimaciones en su interior.

Por otra parte, el mercado laboral, es en lo inmediato la principal dimensión que se ve afectada por los flujos migratorios, implicando un desafío doble para las mediciones generadas a partir de encuestas por muestreo; en el caso de la ENE disponer de estadísticas del mercado del trabajo confiables aun cuando los marcos de muestreo y los diseños muestrales no fueron diseñados salvaguardando posibles shocks migratorios y, una buena medición a nivel de totales, en lo posible a nivel de caracterización de la población migrante y sus implicancias en el mercado del trabajo. Mejorar la confiabilidad de la medición del mercado laboral posibilita tomar decisiones de política pública de manera más informada, redundando en una mayor eficacia de estas. El resguardo de totales de interés en la encuesta permite una caracterización más precisa, contribuyendo, en el caso de la migración, no sólo a estimar totales, sino que eventualmente a identificar las fases del proceso, y por tanto a actuar de manera oportuna, ya que, aun cuando en el largo plazo las familias y los países se beneficien del proceso migratorio, en el corto y mediano plazo los costos pueden ser muy altos, principalmente por los mayores obstáculos de los inmigrantes para acceder al trabajo decente (CEPAL y OIT, 2017).

A partir de escenarios que simulan marcos desactualizados, incompletos de manera intencionada en dos importantes grupos para el mercado laboral: fuerza de trabajo primaria y extranjeros residentes, se muestra que calibrar por variables fuertemente correlacionadas con las variables de estudio (situación de vínculo con el mercado laboral: ocupados, desocupados e inactivos) minimiza el sesgo de las estimaciones, aun ante la presencia de marcos muestrales incompletos.

La Nueva Metodología de calibración de los factores de expansión ajustará los totales de personas a grupos de interés en la medición de las estadísticas laborales; al interior de la PET, distinguiendo por sexo y condición de nacionalidad. Por tanto, minimizará sesgos, pero también mejorará la precisión de las estimaciones en subpoblaciones que están correlacionadas con la información exógena a la cual se está calibrando, FTP y desagregaciones en el interior.

En el mediano plazo, con el asentamiento de la población migrante en el país de destino y, eventualmente, con la reintegración del núcleo familiar, incorporándose con mayor propensión niños y mujeres, los desafíos de una adecuada integración familiar al país de destino excede con creces el mercado laboral y pasa a abarcar una multiplicidad de dimensiones, teniendo implicancias en pobreza, salud, educación, cohesión social, entre otros aspectos, de manera que estimar de manera confiable y precisa a este importante grupo de la población se transforma en una necesidad latente en el resto de las estadísticas sociales del país.

6. Referencias

Aldunate R., Contreras G., de la Huerta C., y Tapia M. (2019) Characterization of the Recent Immigration to Chile. *Documentos de Trabajo N° 830, Banco Central de Chile*. Enero, 2019.

Alvarado, M. y Touron, I. (2018). Actualización del Marco Muestral de Manzanas a 2016 en la Encuesta Nacional de Empleo: Resultados prueba piloto. Documentos de Trabajo, Serie Documentos Metodológicos N5, diciembre 2018. Instituto Nacional de Estadísticas, Santiago de Chile.

Alvarado, M. y Touron, I. (2019). Actualización del Marco Muestral de Manzanas a 2016 en la Encuesta Nacional de Empleo: Aproximación del impacto sobre los principales indicadores de la ENE. Documentos de Trabajo, Serie Documentos Metodológicos N9, marzo 2019. Instituto Nacional de Estadísticas, Santiago de Chile.

CEPAL y OIT. (2017). Coyuntura Laboral en América Latina y el Caribe. Santiago: Naciones Unidas.

Contreras, D., Mello, L. d., Puentes, E. (21 de Septiembre de 2010). The determinants of labour force participation and employment in Chile. *Applied Economics*, iFirst, 1-12.

Gutiérrez, A. (2016). Estratедias de muestreo, diseño de encuestas y estimación de parámetros. Ediciones de la U, Bogotá.

Instituto Nacional de Estadísticas y Departamento de Extranjería y Migración. (2019). ESTIMACIÓN DE PERSONAS EXTRANJERAS RESIDENTES EN CHILE AL 31 DE DICIEMBRE DE 2018. SANTIAGO.

Lumley, T. (2010) *Complex Surveys. A guide to analysis using R*. Wiley Series in Survey Methodology.

Maldonado Valera, C., Martínez Pizarro, J., Martínez, R. (2018). Protección social y migración: una mirada desde las vulnerabilidades a lo largo del ciclo de la migración y de la vida de las personas. Santiago: Naciones Unidas.

United Nations Statistics Division (1982), *Designing Household Survey Samples: Practical Guidelines*, ST/ESA/STAT/SER.F/98, United Nations, New York.

7. Anexo 1

Tabla 7.1: Población simulada de hombres nacionales: Estados y Tasas, según tramos de edad

Tramo de edad	Totales					Tasas (%)			
	y_0	y_1	y_2	y_3	y_4	r_1	r_2	r_3	r_4
15-19	598,774	102624	84427	18197	496150	17.1	14.1	17.7	82.9
20-24	639670	398741	342904	55837	240929	62.3	53.6	14.0	37.7
25-29	659521	577755	521520	56235	81766	87.6	79.1	9.7	12.4
30-34	571172	537407	499831	37576	33765	94.1	87.5	7.0	5.9
35-39	536596	510161	479022	31139	26435	95.1	89.3	6.1	4.9
40-44	542266	515318	485613	29705	26948	95.0	89.6	5.8	5.0
45-49	530307	499244	469965	29279	31063	94.1	88.6	5.9	5.9
50-54	544044	503330	474008	29322	40714	92.5	87.1	5.8	7.5
55-59	480486	430467	406263	24204	50019	89.6	84.6	5.6	10.4
60-64	385709	316831	299889	16942	68878	82.1	77.8	5.3	17.9
65-69	293140	156393	152870	3523	136747	53.4	52.1	2.3	46.6
70-74	225061	77545	76459	1086	147516	34.5	34.0	1.4	65.5
75-79	149927	32631	32216	415	117296	21.8	21.5	1.3	78.2
80-84	91014	11174	11031	143	79840	12.3	12.1	1.3	87.7
85+	72844	6148	6004	144	66696	8.4	8.2	2.3	91.6

Fuente: Elaboración propia en base simulaciones con datos del Censo 2017

Tabla 7.2: Población simulada de mujeres nacionales: Estados y Tasas, según tramos de edad

Tramo de edad	Totales					Tasas (%)			
	y_0	y_1	y_2	y_3	y_4	r_1	r_2	r_3	r_4
15-19	574149	63447	50698	12749	510702	11.1	8.8	20.1	88.9
20-24	626516	281960	238520	43440	344556	45.0	38.1	15.4	55.0
25-29	655955	454885	410316	44569	201070	69.3	62.6	9.8	30.7
30-34	580903	416241	389318	26923	164662	71.7	67.0	6.5	28.3
35-39	554520	388320	366486	21834	166200	70.0	66.1	5.6	30.0
40-44	566055	384636	363957	20679	181419	68.0	64.3	5.4	32.0
45-49	562839	366208	347619	18589	196631	65.1	61.8	5.1	34.9
50-54	585890	354233	337790	16443	231657	60.5	57.7	4.6	39.5
55-59	525756	284983	272582	12401	240773	54.2	51.8	4.4	45.8
60-64	430815	165673	161172	4501	265142	38.5	37.4	2.7	61.5
65-69	337733	73474	72189	1285	264259	21.8	21.4	1.7	78.2
70-74	273835	34713	34274	439	239122	12.7	12.5	1.3	87.3
75-79	201266	14863	14707	156	186403	7.4	7.3	1.0	92.6
80-84	138883	6074	6005	69	132809	4.4	4.3	1.1	95.6
85+	144988	4318	4229	89	140670	3.0	2.9	2.1	97.0

Fuente: Elaboración propia en base simulaciones con datos del Censo 2017

Tabla 7.3: Población simulada de hombres extranjeros: Estados y Tasas, según tramos de edad

Tramo de edad	Totales					Tasas (%)			
	y_0	y_1	y_2	y_3	y_4	r_1	r_2	r_3	r_4
15-19	19329	6380	5520	860	12949	33.0	28.6	13.5	67.0
20-24	43415	36533	33223	3310	6882	84.1	76.5	9.1	15.9
25-29	63875	59974	55518	4456	3901	93.9	86.9	7.4	6.1
30-34	59256	56773	53161	3612	2483	95.8	89.7	6.4	4.2
35-39	46406	44714	42140	2574	1692	96.4	90.8	5.8	3.6
40-44	32392	31246	29518	1728	1146	96.5	91.1	5.5	3.5
45-49	21465	20676	19454	1222	789	96.3	90.6	5.9	3.7
50-54	15088	14422	13653	769	666	95.6	90.5	5.3	4.4
55-59	9667	8989	8515	474	678	93.0	88.1	5.3	7.0
60-64	6268	5463	5154	309	805	87.2	82.2	5.7	12.8
65-69	4045	2621	2508	113	1424	64.8	62.0	4.3	35.2
70-74	2974	1409	1366	43	1565	47.4	45.9	3.1	52.6
75-79	1959	663	646	17	1296	33.8	33.0	2.6	66.2
80-84	1527	370	366	4	1157	24.2	24.0	1.1	75.8
85+	1716	401	386	15	1315	23.4	22.5	3.7	76.6

Fuente: Elaboración propia en base simulaciones con datos del Censo 2017

Tabla 7.4: Población simulada de mujeres extranjeras: Totales y Tasas, según tramos de edad

Tramo de edad	Totales					Tasas (%)			
	y_0	y_1	y_2	y_3	y_4	r_1	r_2	r_3	r_4
15-19	18087	4358	3569	789	13729	24.1	19.7	18.1	75.9
20-24	41127	27590	23952	3638	13537	67.1	58.2	13.2	32.9
25-29	60028	46815	41888	4927	13213	78.0	69.8	10.5	22.0
30-34	55566	44023	40081	3942	11543	79.2	72.1	9.0	20.8
35-39	47786	38160	35302	2858	9626	79.9	73.9	7.5	20.1
40-44	36290	29038	27129	1909	7252	80.0	74.8	6.6	20.0
45-49	25947	21310	20068	1242	4637	82.1	77.3	5.8	17.9
50-54	18585	14974	14165	809	3611	80.6	76.2	5.4	19.4
55-59	12443	9100	8592	508	3343	73.1	69.1	5.6	26.9
60-64	7905	4691	4451	240	3214	59.3	56.3	5.1	40.7
65-69	4854	1887	1800	87	2967	38.9	37.1	4.6	61.1
70-74	3179	813	784	29	2366	25.6	24.7	3.6	74.4
75-79	2241	343	339	4	1898	15.3	15.1	1.2	84.7
80-84	1650	186	176	10	1464	11.3	10.7	5.4	88.7
85+	2556	246	233	13	2310	9.6	9.1	5.3	90.4

Fuente: Elaboración propia en base simulaciones con datos del Censo 2017