



Revista Científica "Visión de Futuro"  
ISSN: 1669-7634  
ISSN: 1668-8708  
[revistacientifica@fce.unam.edu.ar](mailto:revistacientifica@fce.unam.edu.ar)  
Universidad Nacional de Misiones  
Argentina

# PREDICCIÓN DE LA PROBABILIDAD DE INSERCIÓN OCUPACIONAL DE LOS DESOCUPADOS EN ARGENTINA (2003-2019)

**Staudt, Agustín; Heredia, Juan Luis**

PREDICCIÓN DE LA PROBABILIDAD DE INSERCIÓN OCUPACIONAL DE LOS DESOCUPADOS EN ARGENTINA (2003-2019)

Revista Científica "Visión de Futuro", vol. 25, núm. 2, 2021

Universidad Nacional de Misiones, Argentina

Disponible en: <https://www.redalyc.org/articulo.oa?id=357966632001>

DOI: <https://doi.org/10.36995/j.visiondefuturo.2021.25.02R.001.es>

© 2021 Revista Científica "Visión de Futuro" - Facultad de Ciencias Económicas - UNaM



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial 3.0 Internacional.



## PREDICCIÓN DE LA PROBABILIDAD DE INSERCIÓN OCUPACIONAL DE LOS DESOCUPADOS EN ARGENTINA (2003-2019)

Prediction of transition probability from unemployment to employment in Argentina (2003-2019)

*Agustín Staudt*  
Ministerio de Desarrollo Productivo de la Nación,  
Argentina  
Agusstaudt@gmail.com

DOI: [https://doi.org/10.36995/  
j.visiondefuturo.2021.25.02R.001.es](https://doi.org/10.36995/j.visiondefuturo.2021.25.02R.001.es)  
Redalyc: [https://www.redalyc.org/articulo.oa?  
id=357966632001](https://www.redalyc.org/articulo.oa?id=357966632001)

*Juan Luis Heredia*  
MisES Consulting, Argentina  
Juanluish012@gmail.com

Recepción: 19 Septiembre 2020  
Aprobación: 11 Enero 2021

### RESUMEN:

Pese a su creciente participación en el mercado laboral, las mujeres que deciden salir a buscar un puesto de trabajo enfrentan mayores dificultades para alcanzarlo. La participación de las mujeres en la fuerza laboral es considerablemente menor, inclusive, de ingresar al mercado laboral la posibilidad de encontrar efectivamente un trabajo es también menor a la chance que tienen los varones de hacerlo (CIPPEC, 2019). Poder predecir la probabilidad de inserción ocupacional de varones y mujeres, e indagar sobre los factores que influyen sobre dicha probabilidad, resulta fundamental en pos de entender las brechas de género en el mercado laboral, contribuyendo a mejorar el diseño e implementación de políticas públicas con perspectiva de género, y en última instancia lograr una mayor igualdad de oportunidades. En este marco, el presente trabajo buscará predecir la probabilidad de transición del desempleo al empleo de desocupados en Argentina para los años 2003 a 2019, utilizando la Encuesta Permanente de Hogares, a partir de técnicas tradicionales de predicción y de Machine Learning, con el objetivo de encontrar el modelo más robusto que logre el menor error de predicción.

**PALABRAS CLAVE:** Género, Empleo, Desigualdad, Machine Learning.

### ABSTRACT:

Despite their growing participation in the labor market, women who decide to go out and look for a job face greater difficulties in obtaining it. The participation of women in the labor force is considerably lower, even if entering the labor market the possibility of actually finding a job is also less than the chance that men have of doing so (CIPPEC, 2019). Being able to predict the probability of occupational insertion of men and women, and inquire about the factors that influence this probability, is essential in order to understand gender gaps in the labor market, helping to improve the design and implementation of public policies with a gender perspective, with the final goal to achieve equality of opportunities. In this framework, the present work will seek to predict the probability of transition from unemployment to the employment in Argentina from 2003 to 2019, using the Permanent Household Survey, based on traditional prediction techniques and Machine Learning, with the objective to find the most robust model that achieves the highest level of accuracy.

**KEYWORDS:** Gender, Employment, Inequality, Machine Learning.

## INTRODUCCIÓN

Durante los últimos 60 años se ha presenciado una inserción masiva de las mujeres al mercado laboral. Dicha inserción parte de una profunda transformación ocurrida en el siglo pasado en distintos ámbitos como ser educación, familia y empleo. Por un lado, la toma de decisiones de las mujeres pasó de ser estática, con horizontes limitados o intermitentes, a contemplar decisiones dinámicas, con horizontes a largo plazo. Por otra parte, el enfoque del rol laboral tradicional en el hogar pasa a un segundo plano y comienzan a asociar al empleo como una cuestión de identidad y valor social. Dichos factores implicaron un cambio de enfoque, el concepto de "trabajo" abrió paso al de "carrera" (Goldin, 2006).

Sin embargo, las brechas de género aún persisten: en América Latina, las mujeres participan menos que los varones en el mercado de trabajo, y cuando lo hacen suelen trabajar en empleos de peor calidad, a tiempo parcial, menor remuneración y suelen estar subrepresentadas en puestos gerenciales y ejecutivos. De hecho, mientras que el 95 por ciento de los varones adultos de 25 a 54 años trabaja o busca empleo activamente, esa proporción cae a 66 por ciento en el caso de las mujeres (CAF 2019, Ministerio de Trabajo 2018). En Argentina, el 62 por ciento de las mujeres entre 16 y 59 años participan en el mercado laboral, lo cual representa una brecha de 19 puntos porcentuales con respecto a la fuerza laboral masculina, que se encuentra en el 81 por ciento según datos de la Encuesta Permanente de Hogares (EPH) del cuarto trimestre de 2018 (CIPPEC, 2019).

Dichas desigualdades laborales surgen principalmente por distorsiones que limitan o sesgan decisiones de formación de capital humano, familia y empleo a lo largo de la vida de las personas, ya sea debido a la presencia de la división tradicional del trabajo y roles en la familia de mujeres y varones<sup>1</sup>, presencia de discriminación laboral, o a variables asociadas a las fluctuaciones en la actividad económica (CAF 2019, CIPPEC 2019).

Si bien en los últimos años la literatura ha puesto el foco en indagar acerca de los determinantes de la participación laboral de mujeres y varones, la brecha en la inserción ocupacional de género es una problemática muy relevante, debido a que una mayor participación laboral podría no traducirse en resultados de empleo efectivo. Según CIPPEC (2019), en el cuarto trimestre de 2018, 11 por ciento de las mujeres de 16 a 59 años se encontraba desempleada, en comparación con 9 por ciento de los varones. Esto a su vez se agrava con la permanencia en el desempleo de las mujeres: utilizando datos de la EPH se observa que, del total de varones que buscan trabajo en el primer trimestre de 2018, el 50 por ciento lo encuentra en ese mismo trimestre del siguiente año, mientras que ese porcentaje para mujeres cae al 32 por ciento.

Por dichas razones, se torna muy relevante el análisis predictivo de transiciones del desempleo, con el objetivo de lograr un entendimiento preciso acerca de la probabilidad de inserción ocupacional de mujeres y varones. Esto permitiría comprender en mayor profundidad qué posibilidades tienen los desocupados y desocupadas<sup>2</sup> de encontrar efectivamente un empleo en un período futuro y, al mismo tiempo, entender qué factores subyacentes podría estar detrás de las diferencias en oportunidades laborales entre mujeres y varones. De esta forma, el presente trabajo contribuye a mejorar el diseño e implementación de políticas públicas con perspectiva de género a través de un estudio predictivo de transiciones laborales del desempleo al empleo, buscando contribuir a cerrar las brechas de género, para logar en última instancia una mayor igualdad entre mujeres y varones<sup>3</sup>.

Si bien en la región ya se han realizado trabajos previos que enfocan el análisis al entendimiento de los determinantes del desempleo, a través de estimaciones de transiciones laborales hacia el empleo, son pocos los que centran el análisis desde una perspectiva de género. Por otro lado, en los últimos años en distintos campos de la economía se comenzaron a intensificar usos de métodos en machine learning (mayormente enfoque de aprendizaje supervisado), los cuales buscan optimizar problemas de predicción, poniendo en segundo plano la cuestión del estimador insensado para permitir un trade-off entre sesgo y varianza del estimador, centrándose, a su vez, en obtener buenas predicciones fuera de la muestra (Varian 2014, James et al. 2013, Kleingerg et al. 2015). Dicho enfoque se las arregla para ajustar formas funcionales complejas y muy flexibles a los datos sin simplemente caer en el sobreajuste, encontrando funciones que, a la hora de predecir, se desenvuelven bien fuera de la muestra (Mullainathan y Spiess, 2017).

Teniendo en cuenta este último enfoque, las investigaciones que buscan predecir transiciones laborales con métodos de aprendizaje supervisado son aún escasas en nuestra región, siendo que estos giran en torno al problema de producir predicciones de a partir de (Mullainathan y Spiess, 2017).

En este marco, el presente trabajo indaga las características de transiciones laborales de género desde la desocupación a la ocupación mediante una estrategia de predicción de probabilidades, para el mercado laboral argentino durante el período 2003-2019, con el objetivo de encontrar el modelo de mayor consistencia y

poder predictivo, comparando así el desempeño que presenta el enfoque tradicional y comúnmente utilizado por la literatura<sup>4</sup>, con respecto al de aprendizaje supervisado.

Por lo tanto, en primer lugar, se utiliza una regresión logística para estimar las probabilidades mencionadas anteriormente. No obstante, con el objetivo de buscar la mejor predicción se realiza un análisis con metodologías de clasificación de Machine Learning. De todas formas, dado que el modelo lineal tiene claras ventajas en términos de inferencia y a menudo es sorprendentemente competitivo en relación con los métodos no lineales, se busca mejorarlo a partir de la contracción (shrinkage) de los coeficientes estimados. Es decir, con el objetivo de mejorar la predicción del modelo disminuyendo su varianza, los coeficientes son contraídos hacia cero en relación a las estimaciones tradicionales como ser mínimos cuadrados ordinarios, logit, probit, entre otros (James et al. 2013). Por consiguiente, se utilizan los métodos de regresión ridge y lasso, que en el primer caso contrae los coeficientes asintóticamente a cero y en el segundo caso los contrae a valores iguales a cero.

El trabajo se organiza de la siguiente forma. En la sección 2 se realiza una revisión de la literatura, la sección 3 describe la fuente de datos como también la metodología utilizada. Por último, en la sección 4 se presentan los principales resultados, tanto del análisis descriptivo como del econométrico.

## DESARROLLO

### Revisión de la literatura

Existe amplia literatura que estudia las transiciones de los individuos a distintos estados ocupacionales, basados en sus características socio-demográficas, institucionales y coyunturales, algunos haciendo foco en determinado grupo de la población o el total de un país determinado. De entre los distintos trabajos, muchos difieren en la metodología de estimación de la matriz de transición, en las formas funcionales o en la amplitud geográfica del estudio, trátese de economías a nivel nacional, regional u otro. En el caso de Russell y O'Connell (2001) analizan la probabilidad de transición de los individuos jóvenes en 9 países de Europa teniendo en cuenta las características individuales, sociodemográficas y coyunturales, para de esta forma lograr aislar las diferencias cruzadas entre países. Lo mismo realiza Fabrizi y Mussida (2009), quienes estiman las transiciones de los individuos en Italia mediante un modelo logit multinomial con un enfoque de cadenas de Markov para el inicio y final de la década bajo estudio. En este trabajo concluyen que hay menos oportunidades laborales para los jóvenes y las mujeres, teniendo los varones mayor probabilidad de estar empleados. Para el caso específico de Argentina, Iturriza, Bedi, y Sparrow (2008) analizan la transición de desempleo a empleo 2002 a 2003, período posterior a la crisis del 2001, para individuos que recibían asistencia social a través del “Plan Jefes de Hogar”. Estiman mediante un análisis logit, la probabilidad de transición con el fin de probar la hipótesis de que los beneficiarios tendrían menor incentivo a buscar trabajo una vez que comenzaban a recibir el beneficio. Estos concluyen en primer lugar, que la transición al empleo suele ser retrasado ya que el costo de estar desempleado es menor. En segundo lugar, la transición de las mujeres al empleo es menor que el de los varones, y consecuentemente, la composición de los participantes de la asistencia social y desempleados tiende a ser feminizada. Por último, Favata (2020) no solo analiza las transiciones, sino que dentro de su análisis incluye el estudio de la duración promedio del desempleo, encontrando que el rango etario, región, jefatura de hogar y el hecho de ser mujer o no, pueden ser factores que influyan la posibilidad de permanecer desempleado.

## Fuente de datos

Para realizar el trabajo se utilizaron los microdatos de la Encuesta Permanente de Hogares (EPH), el mayor relevamiento de información sociolaboral de Argentina, realizado por el Instituto Nacional de Estadísticas y Censos (INDEC). Dicho relevamiento es representativo de la población urbana de los 31 aglomerados más grandes del país y se realiza de forma continua desde el año 2003.

Esta encuesta es llevada a cabo de manera individual, entrevistando a cada persona una vez durante un trimestre. Adicionalmente, la misma persona es seguida a lo largo de tres trimestres más, alternando cada período mediante el sistema de rotación 2-2-2, es decir, se sigue al individuo a lo largo de dos trimestres de forma sucesiva, sale del panel por los dos siguientes, y vuelve a ser encuestado por dos trimestres consecutivos, para luego salir definitivamente del relevamiento. De esta forma, el seguimiento al mismo individuo es de un año y medio (INDEC, 2003) 5.

Un problema que suele encontrarse en encuestas de hogares como la EPH, es la tasa de deserción (attrition) positiva, que disminuye la cantidad de casos encuestados durante todo el período de seguimiento. Sin embargo, para aplicar las estrategias de predicción se requiere que el individuo sea seguido al menos tres trimestres, lo cual ocurre para el 61 por ciento de las personas relevadas por la EPH (Beccaria et al., 2016).

Para estudiar las transiciones laborales del desempleo, con el enfoque que pretende este trabajo, se construyen diferentes cohortes de hogares y personas, utilizando el primer período en que cada individuo en situación de desocupación es encuestado, a partir del cual se utiliza sus diversas características observables para construir las variables predictoras. De esta forma, cada persona que entró a la encuesta en un período de tiempo determinado es asignada a la misma cohorte. Adicionalmente, se debe saber en qué situación laboral se encuentra cada entrevistado o entrevistada en un período siguiente, a aquél utilizado para construir los regresores de la estimación. Debido a que se requiere que una persona sea seguida al menos algún trimestre del siguiente año, se toma como referencia el último período en el cual el individuo es seguido, es decir, aquel que se produce en el cuarto período (de cumplirse el seguimiento completo) o el tercer período que se relevó a dicha persona (en caso de que salga antes del seguimiento completo).

Cabe destacar que esta estrategia podría subestimar o sobreestimar las predicciones de transición, ya que al tomar los períodos extremos no se estaría contemplando lo que ocurre en períodos intermedios, es decir, en los trimestres que el individuo es encuestado, pero no se considera como información relevante, y en los trimestres que la persona sale del panel. Por lo que ese sesgo puede estar dado por transiciones simétricas durante el período no considerado (conseguir un empleo y volver a ser desempleado).

Para la condición de desempleo y empleo se utiliza la definición del INDEC (2011) que sigue las normativas de la Organización Internacional del Trabajo. Así, se consideran ocupadas a todas las personas que tengan cierta edad especificada (10 años o más) y que durante un período de referencia (una semana) hayan trabajado al menos una hora<sup>6</sup>.

En el caso de los desocupados, los mismos se conforman de todas aquellas personas que sin tener trabajo se encuentran disponibles para trabajar y han buscado activamente una ocupación en un período de referencia determinado. La EPH considera como desocupados abiertos a todas las personas de 10 años o más que no trabajaron en la semana de referencia, estaban disponibles y buscaron activamente trabajo en algún momento de los últimos treinta días a la entrevista (a través de currículums, respondiendo avisos de diarios/internet, consultando a parientes o amigos, etc.) <sup>7</sup>.

Con el fin de maximizar la cantidad de observaciones disponibles se construye un pool con todos los paneles de transiciones laborales para los años comprendidos entre 2003-2019 identificando las cohortes a la que cada individuo pertenece, dando un total de 659346. Adicionalmente, para poder capturar de manera adecuada las probabilidades de inserción ocupacional de trabajadores y trabajadoras se restringe la muestra a las personas de entre 15 y 59 años de edad, totalizando en 398096 observaciones. Posteriormente, se procede identificar a las personas desocupadas en el momento, resultando en 15268 observaciones <sup>8</sup>.

## Metodología

En primer lugar, se modela la relación entre  $p(X) = \Pr(Y = 1|X)$  y  $X$ . Importar imagen a través de la función logística

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

A partir de la última ecuación se puede obtener la cantidad odds,

$$\frac{p(X)}{1 - p(X)} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}$$

La cual puede tomar valores entre cero y  $\infty$ . Valores cercanos a cero y a  $\infty$  indican muy baja y muy altas probabilidades de ocurrencia de transición del desempleo al empleo, respectivamente.

Por último, tomando logaritmo de ambos lados se obtiene,

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Conocida como log-odds o logit, donde un incremento de  $X$  en una unidad cambia el log-odds en  $\beta_p$  unidades, o equivalentemente en  $e^{\beta_p}$ . Sin embargo, el efecto marginal aquí dependerá de un valor específico de  $x$ . Así, el modelo permitiría evaluar en primera instancia el efecto de cada regresor en la variable predicha a través del signo, es decir, un signo positivo en una variable predictora estaría asociada a un incremento en la variable  $p(X)$ .

Para ajustar el modelo se utiliza el método de máxima verosimilitud, la cual busca los  $\beta_p$  de manera que la probabilidad de transición  $\hat{p}(x_i)$  de cada individuo sea la más cercana posible de la observada. Formalmente, los parámetros estimados se obtienen maximizando la función log-likelihood (Basto et al., 2016):

$$\ell(\beta_p) = \sum_{i=1}^n \left[ y_i \log\left(\frac{\pi_i}{1 - \pi_i}\right) + \log(1 - \pi_i) \right] = \sum_{i=1}^n [y_i x_i \beta_p - \log(1 + e^{x_i \beta_p})]$$

El método logístico suele ser muy adecuado para realizar predicciones donde la variable dependiente es binaria, como es el caso, presentando estimadores cuyo límite probabilístico es igual al parámetro que se desea estimar. Sin embargo, el mismo puede presentar problemas cuando las variables predictoras están altamente correlacionadas, y/o la especificación presenta muchos regresores. En este contexto, el estimador es consistente pero con alta varianza, lo cual afecta al error de predicción.

Adicionalmente, al realizar predicciones existe un trade-off entre varianza y sesgo. Varianza se refiere al monto en que el predicho varía según el conjunto de datos utilizado como entrenamiento. En cuanto al sesgo, se refiere al error que se introduce por estimar problemas reales con modelos estadístico más sencillos. A pesar de que estos componentes teóricamente pueden separarse, generalmente esto no es posible en la práctica.

Aquí las estimaciones mediante regresión ridge y lasso son útiles, ya que regularizan los coeficientes compensando un pequeño aumento en el sesgo con una mayor reducción en la varianza de la predicción. Por lo tanto, dichos métodos manejan el problema de la multicolinealidad y muestran las propiedades ideales para minimizar la inestabilidad numérica que puede ocurrir debido al sobreajuste (Pereira, Basto y da Silva, 2016).

Así, la regresión ridge busca maximizar la siguiente ecuación:

$$\ell_{\lambda}^R(\beta_p) = \sum_{i=1}^n [y_i x_i \beta_p - \log(1 + e^{x_i \beta_p})] - \lambda \sum_{j=1}^p \beta_j^2$$

Donde  $\lambda$  es el parámetro de regularización, el cual tiene el efecto de contraer los  $\beta_p$  estimados, reduciéndolos hacia cero. Cuando  $\lambda = 0$  el término de regularización o penalidad no tiene efecto, y así ridge producirá estimadores tipo logit tradicional. Sin embargo, cuando  $\lambda \rightarrow \infty$  el impacto de la penalidad crece y los coeficientes estimados tenderán a cero 9. Por lo tanto, es clave elegir un valor óptimo de  $\lambda$  10

Otra alternativa que se utiliza en el trabajo es la regresión lasso, cuya versión penalizada en la función log-likelihood es la siguiente:

$$\ell_{\lambda}^L(\beta_p) = \sum_{i=1}^n [y_i x_i \beta_p - \log(1 + e^{x_i \beta_p})] - \lambda \sum_{j=1}^p |\beta_j|$$

En lasso se utiliza una regularización que no solo contrae los coeficientes hacia cero, sino selecciona coeficientes, de manera de forzarlos hacia exactamente cero (Basto et al. 2016; James et al. 2013) 11. Por ultimo, en cuanto al trade-off entre varianza y sesgo, tanto la regresión ridge como lasso, procuran reducir una excesiva varianza a expensas de un aumento en el sesgo, con el objetivo de aumentar la precision de la predicción.

Siguiendo a Pereira, Basto y da Silva (2016), la regresión lasso tiene una ventaja respecto a ridge. Debido a la posibilidad de seleccionar regresores, el modelo final podría involucrar solo a un grupo de predictores, lo cual mejora su interpretabilidad. En cuanto al desempeño en la predicción, la ventaja que tendrá uno sobre el otro depende de la cantidad de predictores que tienen coeficientes sustanciales: cuando solo un pequeño número de predictores tienen coeficientes de magnitud considerable, uno puede esperar que lasso funcione mejor, mientras que cuando todos los coeficientes son aproximadamente del mismo tamaño, uno espera un mejor rendimiento en la regresión ridge.

Siguiendo a Varian (2014) a los efectos de evaluar la capacidad predictiva de cada uno de los modelos se debe comparar los resultados por fuera de la muestra, ya que al ajustar un modelo con la totalidad de datos con que el investigador cuenta, puede que este caiga en el sobreajuste, es decir el error de predicción puede estar subestimado. Para esto se divide el total de datos en una base de entrenamiento para estimar el modelo, una de test para escoger el modelo y una de validación para conocer la performance del modelo elegido. En el caso de la totalidad de datos se toma una cohorte aleatoria de cada año, totalizando en 4268 observaciones la base de validación. Las restantes 11000 observaciones, se divide en una de entrenamiento y en una de testeо, teniendo la primera el 80 por ciento y la ultima el restante 20 por ciento de las observaciones, ambas escogidas aleatoriamente. Por último, se aclara que en promedio se cuenta con 288 observaciones por cohorte12, adicionalmente, el número promedio de observaciones por año13 es de 1018 y en cuanto al número de observaciones promedio por trimestre 14 es de 255.

Para seleccionar el valor de  $\lambda_{opt}$  se utiliza 10-cross validation, aquí los datos son particionados en subconjuntos de igual tamaño aproximado, en el cual se toma uno de los subconjuntos para utilizarlo de grupo de testeо, es decir, medir la precisión del modelo estimado, mientras que los restantes grupos son utilizados para “entrenar” el modelo, es decir, ajustar las regresiones ridge y lasso a estas observaciones. Este procedimiento es repetido veces, cambiando en cada uno el grupo test. Así, el valor óptimo de lambda es el que maximiza la función log-likelihood de validación cruzada (cross validated log-likelihood)15. En este caso específico, se procura el valor óptimo de lambda en base al menor desvío binomial, medida utilizada para medir el error en variables de respuesta dicotómicas. Esta se expresa como:

$$D = 2 \sum o_i \log\left(\frac{o_i}{e_i}\right),$$

En donde  $o_i$  es el observado,  $e_i$  el valor esperado y la sumatoria se hace sobre la cantidad de aciertos y no aciertos, siendo el valor ideal el desvío binomial D igual a cero.

Con el fin de encontrar un  $\lambda$  óptimo, primeramente se genera un vector de 101 valores, partiendo de un valor de 0 y finalizando en 1x1010, conjunto de valores que es seleccionado a criterio de los investigadores.

Por último, se necesita una medida de bondad de ajuste o métrica para comparar la precisión de cada modelo a la hora de ver su performance predictiva en observaciones nuevas o no utilizadas, de manera tal de computar los coeficientes de cada regresión para calcular los valores predichos para las observaciones no utilizadas y compararlos con los valores reales de estas, a partir de dicha métrica. Para esto, se recurre al promedio de aciertos de cada modelo, esto es, el ratio de predicciones acertadas sobre el total de las observaciones nuevas utilizadas, tomando como punto de corte 0.5 para poder clasificar a un valor predicho como 1 (hacer la transición al empleo) o 0 (no hacerla).

## Resultados

Debido a que se trata de un ejercicio de predicción, las características de cada individuo corresponden al período presente (momento  $t$ ), mientras que la transición al empleo se produce a partir de su estado ocupacional futuro (momento  $t+1$ ), descartando del análisis toda característica futura del encuestado.

En primer lugar, en la tabla 1 se muestra la proporción desempleados en el período inicial que se insertan mediante la obtención de un empleo o, en caso contrario, continúan siendo desempleados.

TABLA N° 1:  
Inserción ocupacional

Transición	Varones	Mujeres
Continua desempleado	31% (50%)	39% (50%)
Obtener empleo	69% (59%)	61% (41%)

**Nota:** ratios construidos a partir de la participación de mujeres y varones en cada categoría sobre el total de cada sexo. Ratio fila entre paréntesis.

elaboración propia en base a datos de la EPH.

Como se puede ver en la tabla, del total de mujeres que estaban desempleadas en el período inicial, el 61 por ciento obtiene efectivamente un empleo, mientras que el restante 39 por ciento no logra emplearse en el siguiente período. Si bien en los varones la distribución es similar, el porcentaje que consigue empleo es 8 puntos porcentuales mayor que el de las mujeres<sup>16</sup>. Ahora, si se compara las categorías de inserción ocupacional entre varones y mujeres se observa que el porcentaje de mujeres que logra obtener un empleo es 18 puntos porcentuales menos que la proporción de varones.

Por otro lado, se predice las transiciones a la ocupación mediante distintos modelos logit, lasso y ridge. En el caso de la estimación con logit, se compara los resultados con conjunto de datos completos y el de testeo. Luego, se compara los resultados de predicción de los modelos ridge y lasso con respecto al de logit. En la regresión logit se incorporan variables sociodemográficas como sexo, edad, nivel educativo, status marital, sociolaborales como número de asalariados registrados y la ocupación anterior del desocupado, efectos fijos

por región (dummies para las regiones del NEA, NOA, Cuyo, Pampeana, Patagonia y CABA-CBA) y año (2003-2018).

Luego, se mide la capacidad de predicción del modelo para las transiciones al empleo o en su defecto, la no transición, tanto para varones como mujeres. Los resultados se encuentran en la siguiente tabla:

TABLA N° 2  
matriz de confusión regresión logit en base total y de testeo

Base Total y Testeo		Logit - Base Total		Logit – Testeo	
		Observado	Observado	Observado	Observado
	Predicho	No Transición	Transición	No Transición	Transición
Mujeres	No Transición	3,97%	1,81%	8,90%	4,40%
	Transición	96,03%	98,18%	91,10%	95,60%
	Total	2395	4019	369	527
Varones	No Transición	0,15%	0,14%	1,30%	1,50%
	Transición	99,84%	99,85%	98,70%	98,50%
	Total	2581	6271	395	909
		Precisión	68%	Precisión	65%

elaboración propia en base a datos del EPH

En la tabla superior se calcularon las columnas totales, ya que el eje de discusión es la capacidad predictiva del modelo para los distintos estados de transición. Analizando el modelo logit con la base completa, del total de mujeres que obtuvieron empleo, el 98,18 por ciento de las observaciones fueron predichas correctamente.

Por otro lado, del total observado de las mujeres que no transitaron al empleo, solamente el 4 por ciento fue predicho correctamente, siendo el porcentaje de predicción mucho menor para las no transiciones en relación a las transiciones. Lo mismo ocurre con las transiciones estimadas de los varones, en el que los aciertos de predicción fueron del 99,85 por ciento de las transiciones y una estimación del 0,15 por ciento para las no transiciones.

Por otra parte, se estima la regresión ridge, incluyendo 162 variables disponibles para estimar la transición de los individuos, la cual incluye las variables utilizadas en la regresión logit y otras variables relativas a tiempo de métodos y tiempo búsqueda laboral, sector al que pertenece el individuo, decil de ingresos, categoría ocupacional<sup>17</sup>, tamaño del establecimiento en el que trabajaba, razones por la que dejó su trabajo, otras fuentes de ingreso, características de la vivienda, rango etario, presencia de menores de 10 años y número de personas en el hogar. A continuación, se presenta los resultados de la validación cruzada:

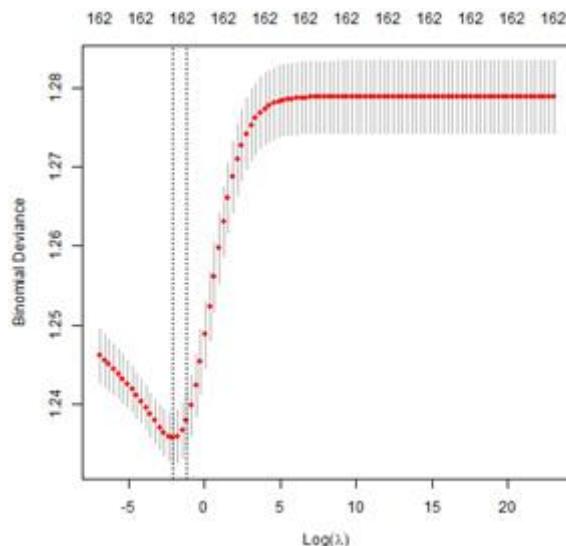


GRÁFICO N° 1:  
Desvío binomial en función de la regresión ridge.  
elaboración propia en base a datos del EPH.

En el gráfico N°1 se aprecia el desvío binomial en función del logaritmo de  $\lambda$ , y en la parte superior está la cantidad de variables por cada lambda18.

De los distintos valores del parámetro regularización, se encuentra que el mínimo es  $\lambda=0.12618$  ( $\log 0.12618 = -2.07$ ). Una vez obtenido el valor del parámetro de tuning, se estima el error de predicción en la base de testeo utilizando los regresores obtenidos con la base de entrenamiento y utilizando el lambda óptimo. Los resultados se detallan a continuación:

TABLA N° 3:  
Matriz de confusión de regresiones logit y ridge.

Base de testeo		Logit		Ridge $\lambda=0.12$	
		Observado		Observado	
	Predicho	No Transición	Transición	No Transición	Transición
Mujeres	No Transición	8,90%	4,40%	11,92%	6,64%
	Transición	91,10%	95,60%	88,07%	93,35%
	Total	369	527	369	527
Varones	No Transición	1,30%	1,50%	3,03%	1,43%
	Transición	98,70%	98,50%	96,96%	98,56%
	Total	395	909	395	909
		Precisión	65%	Precisión	66%

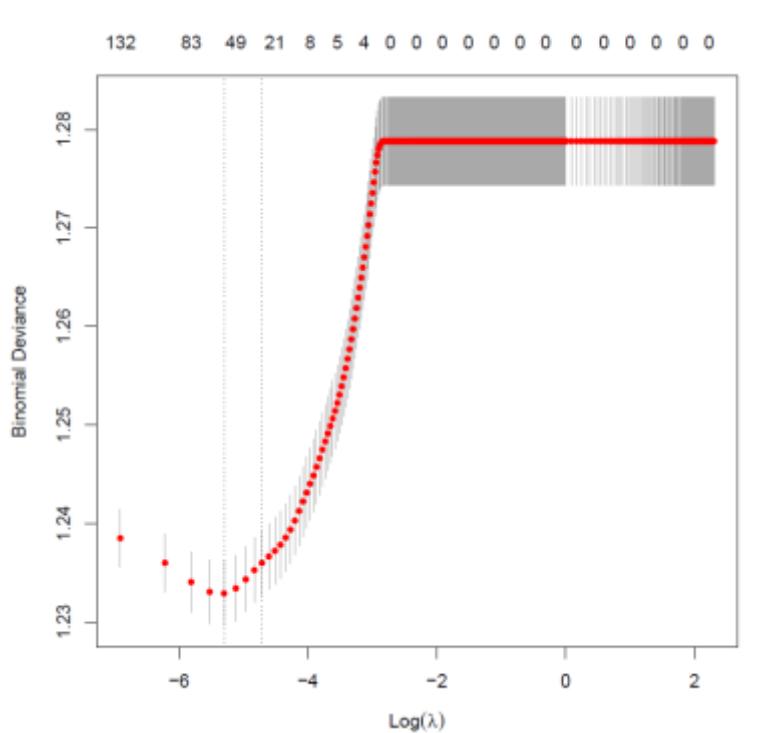
elaboración propia en base a datos del EPH

En la tabla superior, las diferencias en los errores de predicción tanto para transición como no transición son menores al 2 por ciento para sus respectivos cuadrantes, tanto para varones como mujeres.

Por otro lado, se puede notar diferencias en los porcentajes de aciertos entre las estimaciones de ridge y logit. Se puede observar que en la tabla N°3 el cuadrante de la no transición para las mujeres tiene un menor porcentaje (8.9 por ciento) de aciertos en relación al porcentaje de no transición de la regresión ridge (11,92 por ciento). Adicionalmente, lo mismo sucede con los varones, ya que el porcentaje de aciertos del modelo logit de testeo es de 1,3 por ciento siendo menor al de ridge (3,03 por ciento). En contraposición a lo mencionado anteriormente, el porcentaje de aciertos de la regresión logit es levemente menor para varones con un 98,50 por ciento de aciertos en comparación a ridge (98,56 por ciento). En el caso de las mujeres,

el modelo logit tiene una mejor performance en la predicción de las transiciones con un 95,6 por ciento de aciertos en comparación a la estimación de ridge (93,35 por ciento).

Como se mencionó anteriormente, la penalidad de ridge contrae todos los coeficientes asintóticamente hacia cero (a menos que ), lo cual no debe generar problemas para ver la precisión en la predicción del modelo. Sin embargo, puede crear un desafío en la interpretación de los coeficientes, de hecho, en el modelo se cuenta 162 regresores. Por lo tanto, la alternativa de lasso podría ser útil, ya que su parámetro de penalidad tiene el efecto de forzar a un grupo de coeficientes estimados hacia exactamente cero, en aquellos casos donde el parámetro lambda es lo suficientemente grande. De esta forma, la regresión lasso también desarrolla una selección de variables, no solo las penaliza (es un sparse model). Los resultados de la regresión lasso se pueden ver en el siguiente gráfico:



## GRÁFICO N° 2: Desvío Binomial para diferentes de la regresión lasso. elaboración propia en base a datos del EPH.

Al observar el gráfico desde la derecha hacia la izquierda se observa que la performance mejora, esto es debido a que para valores de  $\lambda$  muy grandes la cantidad de variables se vuelve cero, mientras que pasado cierto umbral del parámetro, el modelo selecciona al menos una variable y el ajuste predictivo mejora, lo que sugiere mayores ganancias en precisión para valores cercanos a cero de nuestra grilla de Importar imagen.

Sin embargo, la forma del ajuste para cada valor de lambda tiene forma de U desde que el modelo selecciona al menos una variable para predecir, con un mínimo para valores positivos de  $\lambda$ . Esto sugiere que, a partir de la existencia de un valor de regularización positivo podría encontrarse una mejora en la predicción de la transición laboral bajo estudio.

En la siguiente tabla se encuentran los resultados de la predicción lasso:

TABLA N° 4:  
Matriz de confusión de regresiones logit y lasso en base de testeо.

Grupos test		Lasso con $\lambda^{opt} = 0,05$		Logit	
		Observado		Observado	
	Predicho	No Transición	Transición	No Transición	Transición
Mujeres	No Transición	11,65%	6,07%	8,90%	4,40%
	Transición	88,35%	93,93%	91,10%	95,60%
	Total	369	527	369	527
Varones	No Transición	4,56%	1,54%	1,30%	1,50%
	Transición	95,44%	98,46%	98,70%	98,50%
	Total	395	909	395	909
		Precisión	66%	Precisión	65%

elaboración propia en base a datos del EPH

La tabla anterior muestra la performance predictiva del modelo lasso con el valor igual a 0,05. Como puede verse, este valor es bajo, pero distinto a cero, lo cual sugeriría una mejora respecto al modelo logit. De todas formas, para verlo se compara con los resultados del modelo logit utilizado en el apartado anterior. Lo que se puede ver es una mejora muy baja respecto al modelo logístico, reflejado en mayor medida por la precisión ganada en lasso para predecir aquellos varones y mujeres que no obtuvieron un empleo en el período siguiente. De todas formas, se remarca la dificultad de los modelos de acertar de manera más adecuada en situaciones donde un grupo de observaciones es considerablemente más chico, como es el caso del grupo de no transición versus el de transición.

Una de las ventajas de lasso respecto a ridge es que la primera puede seleccionar variables, dejando aquellas que son relevantes para realizar la predicción. Por lo que, de 162 variables el estimador lasso dejó 64 variables, es decir, seleccionó menos de la mitad del total de predictores. Entre las variables seleccionadas se encuentran las variables de tiempo en el que la persona estuvo buscando trabajo, las cuales sugieren que a mayor tiempo en el que la persona estuvo buscando trabajo tiene menos probabilidades de conseguir un empleo en el siguiente año. Además, a mayor tiempo transcurrido desde que terminó la ocupación anterior la probabilidad de obtener un trabajo es menor.

En el caso de las mujeres, la forma en que buscan un trabajo le ofrece distintas probabilidades de encontrarlo en el siguiente año: aquellas mujeres que se presentan en establecimientos, envían cvs, o se anotó en bolsas, listas, planes de empleo, etc. tienen menos probabilidades de conseguir empleo con respecto a la alternativa de hacer contactos o entrevistas. Mientras que la búsqueda independiente, entendiéndose a esta como la posibilidad de hacer algo para emprender por su cuenta, conlleva una mayor probabilidad de conseguir un empleo respecto a la búsqueda a través de contactos o entrevistas.

Por el lado de la situación conyugal, los varones casados o unidos tienen mayor probabilidad de conseguir un trabajo en el siguiente año, mientras que las mujeres casadas tienen menores chances de obtener un empleo en el siguiente período. Esto podría deberse a lo que se dijo antes, la considerable intensidad de búsqueda de los varones casados en comparación a los solteros, lo cual genera este resultado, mientras que las dificultades explicadas antes en la búsqueda laboral femenina se ven reflejado también aquí. En cuanto a la jefatura de hogar, las probabilidades de encontrar un trabajo aumentan si se trata de la cabeza de familia, independientemente del género.

Hasta aquí se ha comparado el modelo tradicional y los modelos de aprendizaje supervisado ridge y lasso mediante un enfoque de validación cruzada, donde se entrenó cada modelo y comparó su performance fuera de la muestra, mediante la construcción de un grupo de entrenamiento o train en el cual corremos cada regresión (80 por ciento del dataset) y un grupo de testeо (20 por ciento del dataset), donde se evaluó qué tan bien se desenvuelve cada estimador con observaciones no utilizadas dentro de la muestra a través de la

comparación mediante el porcentaje de aciertos de cada predicho estimado respecto al valor observado. El enfoque de entrenamiento y testeo plantea la separación aleatoria de un grupo de observaciones para no ser utilizadas hasta el momento de comprobar el poder predictivo de los estimadores fuera de la muestra. Sin embargo, el proceso de aleatorización de entrenamiento y testeo podría no estar totalmente separado del grupo de entrenamiento, por lo que antes de iniciar las pruebas se separó un grupo adicional de validación, el cual se compone de una cohorte por año seleccionada aleatoriamente, es decir, se toma para cada año un mismo grupo de personas que entró por primera vez a la encuesta, representando aproximadamente el 25 por ciento de nuestra base. La idea detrás de esta estrategia es evaluar los estimadores para grupos completos que no han estado antes en el conjunto de datos, buscando reproducir rendimientos de los estimadores para observaciones completamente nuevas.

**TABLA N° 5:**  
Matriz de confusión de regresiones logit y ridge en base de validación.

Base de Validación		Logit		Ridge	
		Observado		Observado	
	Predicho	No Transición	Transición	No Transición	Transición
Mujeres	No Transición	7,54%	3,75%	12,67%	9,1%
	Transición	92,46%	96,25%	87,33%	90,90%
	Total	584	1176	584	1176
Varones	No Transición	1,22%	1,23%	1,67%	0,92%
	Transición	98,78%	98,97%	98,33%	99,08%
	Total	658	1850	658	1850
		Precisión	71%	Precisión	70%

elaboración propia en base a datos del EPH.

En las tablas N°8 y 7 se aprecia que el acierto total de cada modelo es aproximadamente igual, siendo del 71 por ciento para logit y 70 por ciento ridge y lasso.

A pesar de que los aciertos totales de cada modelo son similares, existen diferencias entre cada uno de ellos. En el caso del modelo Ridge el porcentaje de aciertos de las transiciones y no transiciones es mayor a la del modelo logit para los varones.

**TABLA N° 6**  
Matriz de confusión de la regresión lasso en base de validación

Base de Validación		Lasso con $\lambda^{\text{opt}} = 0,05$	
		Observado	
	Predicho	No Transición	Transición
Mujeres	No Transición	10,45%	8,59%
	Transición	89,55%	91,41%
	Total	584	1176
Varones	No Transición	2,58%	1,46%
	Transición	97,42%	98,54%
	Total	658	1850
		Precisión	70%

elaboración propia en base a datos del EPH.

En el caso de las mujeres solo se puede ver una mejora en la predicción de la no transición. En cuanto a la estimación del modelo lasso, el porcentaje de aciertos de no transición es mayor a la del modelo logit para el caso de los varones como de las mujeres, pero por otro lado el porcentaje de aciertos de la no transición es menor para ambos sexos. Por lo tanto, se puede decir que los modelos de ridge y lasso tienen mayor poder de predicción de las no transiciones en relación al modelo logit. Aun así, se remarca que el bajo porcentaje de acierto en la predicción en los modelos de ridge y lasso puede darse por el bajo número de personas que no hicieron la transición, alguna forma funcional que no se haya testeado, o el faltante de algún grupo de variables en el modelo.

## CONCLUSIÓN

Si bien existieron grandes avances en los últimos años en materia de género, las brechas entre varones y mujeres son una problemática que continúa en la actualidad. La participación de las mujeres en la fuerza laboral es considerablemente menor, inclusive, de ingresar al mercado laboral la posibilidad de encontrar efectivamente un trabajo es también menor a la chance que tienen los varones de hacerlo. De hecho, en el primer trimestre de 2018, el 50 por ciento de varones encuentra un trabajo en ese mismo trimestre del siguiente año, mientras que ese porcentaje para mujeres cae al 32 por ciento.

El interés por encontrar respuestas a las distintas transiciones del desempleo que enfrentan varones y mujeres, junto a la posibilidad de acceder a otras herramientas de predicción, nos llevó a preguntarnos si estos estimadores podrían ser más certeros para predecir transiciones laborales del desempleo al empleo para mujeres y varones del mercado laboral argentino, para de esta forma realizar una caracterización más acertada acerca de los determinantes más importantes para pronosticar inserciones ocupacionales en un período de dos años.

De los distintos modelos evaluados en este trabajo, se midió el porcentaje de aciertos como medida de bondad de ajuste, encontrándose que la capacidad predictiva de los modelos lasso y ridge son mejores en las no transiciones, aunque la diferencia con el modelo de logit en puntos porcentuales es bajo, destacándose este último por ser el modelo con menor cantidad de variables.

Respecto a los coeficientes de las estimaciones la regresión logit arrojó como resultado coeficientes los signos esperados, donde la probabilidad de conseguir empleo en el siguiente periodo es menor para las mujeres que para los varones.

Por otra parte, a partir de la regresión lasso se encontró que la probabilidad de obtener un empleo disminuye para las mujeres casadas en relación a las solteras, pero aumenta para los varones, los cuales podrían reflejar la presencia de roles tradicionales dentro de un hogar, que repercuten en intensidades de búsqueda laboral heterogéneas a favor del varón. De hecho, la estimación lasso sugiere que a mayor tiempo de búsqueda laboral, menor probabilidad de conseguir un empleo en el siguiente período, como también, un mayor tiempo transcurrido desde que un desempleado terminó su último trabajo, las probabilidad de conseguir un trabajo en el siguiente año disminuyen. También la forma de búsqueda influye en la probabilidad de inserción por parte de las mujeres, teniendo mayores probabilidades de encontrar un trabajo buscando a través de contactos o entrevistas, que realizando la búsqueda a través de envío de currículums o anotándose en bolsas de trabajo, o planes de empleo, mientras que tienen más chances de insertarse en una ocupación al siguiente año si lo buscan de manera independiente, es decir, emprendiendo por su cuenta.

## RESUMEN BIBLIOGRÁFICO

Agustín Staudt

Es Licenciado en Economía por la Universidad Nacional de Misiones (UNaM). Asesor en el Ministerio de Desarrollo Productivo de la Nación, realizando investigaciones en temáticas de mercado laboral y género.

**Juan Luis Heredia**

Es Licenciado en Economía por la Universidad Nacional de Misiones (UNaM). Actualmente es investigador independiente en MisES Consulting, realizando indicadores económicos e investigación de distintas temáticas de economía argentina.

## REFERENCIAS

- Banco De Desarrollo de América Latina (CAF). (2019). Brechas de Género en América Latina. Un Estado de Situación. CAF.
- Beccaria, L., Maurizio, R., Trombetta , M., & Vázquez, G. (2016). Una evaluación del efecto scarring en Argentina. Buenos Aires: Revista Desarrollo y Sociedad.
- Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento (CIPPEC); Organización de las Naciones Unidas . (2019). El Género del Trabajo. Buenos Aires: Fundación CIPPEC.
- Cerruti Marcela. (2000). Determinantes de la participación intermitente de las mujeres en el mercado de trabajo del Área Metropolitana de Buenos Aires. Buenos Aires: Desarrollo Económico, Vol. 39, No. 156 (Jan. - Mar., 2000), pp. 619-638.
- Fabrizi, E., & Mussida, C. (2009). The Determinants of Labour Market Transitions. Giornale degli economisti e annali di economia.
- Favata, F. (2020, junio). Duración del desempleo en Argentina (2003-2019).
- Freeman, R., & Ballen , J. (1986). Transitions between Employment and Nonemployment.
- Friedman, J., Hastie, T., Narasimhan, B., Tay, K., Simon , N., & Qian, J. (2020). Package ‘glmnet’: Lasso and Elastic-Net Regularized Generalized Linear Models.
- Goldin, C. (2006). The Quiet Revolution That Transformed Women’s Employment, Education, and Family. American Economic Review, 96(2), 1–20.
- Instituto Nacional de Estadísticas y Censos . (2003). La Nueva Encuesta Permanente de Hogares de Argentina. . Buenos Aires: INDEC.
- 11.Instituto Nacional de Estadísticas y Censos . (2011). Encuesta Permanente de Hogares. Conceptos de Condición de Actividad, Subocupación Horaria y Categoría Ocupacional. Buenos Aires.
- Iturriza , A., Bedi, A. S., & Sparrow, R. (2008). Unemployment Assistance and Transition to Employment in Argentina. IZA.
- James, G., Witten, D., Hastie, T., & Tibshirani. (2013). An Introduction to Statistical Learning. Springer.
- Kütük, Y., & Güloglu, B. (2019). Prediction of Transition Probabilities from Unemployment to Employment for Turkey via Machine Learning and Econometrics: a Comparative Study . Istanbul: İktisat Araştırmaları Dergisi • Journal of Research in Economics; ss/pp. 58-75.
- Ministerio de Trabajo, Empleo y Seguridad Social . (2018). Las Mujeres en el Mundo del Trabajo. Buenos Aires: Presidencia de la Nación Argentina.
- Mullainathan , S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach . Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 87–106.
- Mussida, C., & Fabrizi , E. (2009). The Determinants of Labour Market Transitions.
- Pereira, J., Basto, M., & Ferreira da Silva, A. (2016). The logistic lasso and ridge regression in predicting corporate failure. Procedia Economics and Finance 39 634 – 641.
- Russell, H., & O'Connell, P.J. (2001). Getting a Job in Europe: The Transition from Unemployment to Work among Young People in Nine European Countries. The Economic and Social Research Institute, 1-24.

Varian , H. (2014). Big Data: New Tricks for Econometrics . Journal of Economic Perspectives—Volume 28, Number 2—Pages 3–28.

## NOTAS

1. Las responsabilidades sociales de cuidado tienen un impacto fundamental en las posibilidades de acceso de las mujeres al mercado de trabajo (CIPPEC, 2019).
2. Dada la influencia que el uso del lenguaje tiene sobre la manera de percibir la realidad, es importante hacer uso de un lenguaje no discriminatorio y que permita visibilizar todos los géneros. Sin embargo, dado que no hay acuerdo sobre la manera de hacerlo en castellano se utiliza el criterio de CIPPEC (2019): 1) evitar expresiones discriminatorias, 2) visibilizar el género cuando la situación comunicativa y el mensaje lo requieren para una comprensión correcta, y 3) no hacerlo cuando no resulta necesario.
3. Somos conscientes que parte de la población no se identifica con los géneros tradicionales (varón y mujer) y en la actualidad no hay un consenso generalizado de cuál es la mejor forma de clasificar el género. Por otro lado, la información disponible para el análisis en este documento no cuenta con una clasificación diferente. Por lo tanto, utilizamos la división tradicional para esta investigación, aunque se remarca que en este trabajo los resultados permitirán lograr una mayor igualdad de oportunidades para ambos géneros.
4. Si bien existen una multiplicidad de enfoques metodológicos para analizar transiciones laborales, en el trabajo nos referimos a enfoque tradicional como la estimación por medio de una regresión logística binaria, la cual resulta común en este tipo de trabajos (Freeman y Ballen 1986; Russell y O'Connell 2001; Mussida y Fabrizi 2009; Kütük y Güloglu 2019; Cerruti 2000; entre otros).
5. Cabe aclarar que esta rotación en el que un porcentaje de personas ingresa a la EPH es del 25 por ciento del total de hogares encuestados en un trimestre, es decir, cada trimestre se renueva 25 por ciento de hogares y simultáneamente un 25 por ciento de hogares sale de la encuesta.
6. Se incluyen a las personas que durante un período de referencia realizaron algún trabajo de al menos una hora, hayan recibido pago o no por dicha actividad; y las personas que tienen una ocupación, pero no estaban trabajando temporalmente durante un período de referencia y mantenían un vínculo formal con su empleo.
7. Así, INDEC (2011) no considera dentro de este concepto a otras formas de empleo inadecuado tales como personas que realizan trabajos transitorios mientras buscan activamente trabajo, aquellas que trabajan jornadas involuntariamente por debajo de lo normal, los ocupados en puestos por debajo de la remuneración mínima, ni a los desocupados que han suspendido la búsqueda por falta de oportunidades visibles de empleo, etc.
8. Se aclara que, debido a la construcción del conjunto de datos por cohortes mediante el método explicado anteriormente, hay observaciones que no tienen seguimiento en el tiempo, tal es el caso de las cohortes que ingresan en el 2º trimestre de 2013, 3º y 4º de 2014 y todo 2015.
9. Sin embargo, la penalización no llevara ninguno de ellos hacia exactamente en cero.
10. Cabe aclarar que los coeficientes estimados no son invariantes a la escala en estos modelos, ya que  $x_j \beta_{j,\lambda}^R$  no solo dependerá de los valores de lambda, sino además de la escala de los predictores. Por lo que estos son estandarizados:  $t_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{162} (x_{ik} - \bar{x}_{ik})^2}}$ .
11. Cabe aclarar que queremos reducir la asociación estimada de cada variable con el *outcome*. Sin embargo, no queremos reducir el intercepto, que es simplemente una medida del valor medio de la respuesta cuando todos los regresores son exactamente cero.
12. El total de cohortes bajo análisis son 62.
13. El total de años por bajo estudio es de 19.
14. El periodo comprendido, inicia en el primer trimestre de 2003 finalizando en el tercer trimestre de 2018. Adicionalmente se aclara que no se cuenta con información del tercer trimestre de 2007, segundo trimestre de 2013, tercer y cuarto de 2014, todo 2015 y el primer trimestre de 2016. En el caso de los años 2007, 2013, 2015 y 2016, se debió a la interrupción en la publicación de la EPH. En el caso del año 2014, se debió a la falta de seguimiento en el siguiente año, de los individuos que ingresaron a la cohorte del tercer y cuarto trimestre de 2014.
15. Para realizar las estimaciones utilizamos el software R, mientras que para estimar las regresiones ridge y lasso recurrimos al paquete glmnet (Friedman et al., 2020). El paquete permite ajustar modelos lineales generalizados con diferentes penalizaciones de ridge y lasso, entre otros.
16. Cabe destacar que no contemplamos en el trabajo la posibilidad de que los varones y mujeres desempleadas pasen a la inactividad, lo cual es una problemática igual de grave y que afecta mayormente a las mujeres (CIPPEC, 2019).
17. Se refiere a los desocupados con ocupación anterior.
18. Como el método de ridge no contrae a exactamente cero los estimadores, para cada valor de lambda se encuentra la totalidad de variables que en este caso son 162.