

Aplicaciones Económicas del Concepto Identidad

Ricardo Héctor Trepát

Universidad del CEMA

Junio de 2022

DIRECTOR DE TESIS

Dr. Jorge Miguel Streb _____

JURADO DE TESIS

Dr. Pablo Garófalo (externo) _____

Dr. Julio Jorge Elías _____

Dr. Esteban Maioli _____

Dedicatoria

A mi esposa Erica y a mis hijas Paula, Juana y Malena, por su paciencia, cariño y apoyo.

A mis padres, Dorita y Pedro, que me vieron comenzar este esfuerzo años atrás y que no pudieron estar para ver completa mi trayectoria académica con la defensa de la tesis de doctorado.

Contenido

Contenido	iv
Listado de Figuras, Gráficos y Tablas	viii
Prefacio	xi
1. El Concepto de Identidad.	1
I. Indagaciones Preliminares.	1
II. Problema de investigación y justificación.	2
A. Criterio de relevancia metodológica.	3
B. Criterio de relevancia social.	3
III. Marco Teórico.	4
A. Enfoque de la Utilidad.	4
B. Retorno a la Educación.	5
C. Enfoque de Utilidad bajo la mirada del otro.	5
D. Mecanismos internos y externos y significantes de identidad social.	6
E. Enfoque de la Psicología Social.	7
F. Sobre normas fuertemente observadas y normas vacuas.	8
G. Variable identidad. Efecto de acentuación.	9
H. Abordaje sociológico al concepto de identidad.	9
I. Otros significados de la variable identidad.	9
J. Posición conceptual elegida para la variable identidad.	10
K. Teoría de Economía de la Identidad.	11
2. Elección de Identidad y Participación en el Mercado Laboral Evidencia empírica con datos de Estados Unidos de América.	12
I. Resumen.	12
II. Introducción.	13

III. Función de utilidad general con Identidad.	17
IV. Disonancia cognitiva y determinantes exógenos de la identidad.	19
V. Discriminación y comportamiento concomitante de la minoría.	20
VI. Hipótesis a validar.	21
VII. Análisis Empírico.	23
VIII. Resultados.	24
IX. Conclusiones.	29
3. Aplicación del modelo de Economía de la Identidad a la decisión de sufragar.	30
I. Resumen	30
II. Introducción.	31
III. Propósito de la investigación.	33
IV. Datos y variables.	33
V. Particularidades de los grupos respecto a registración y voto.	38
VI. El modelo económico.	41
A. Juego secuencial de decisiones identitarias.	41
B. Función de utilidad general con Identidad.	41
VII. El modelo econométrico.	43
A. Modelo Probit Bivariado.	43
B. Estimación de Máxima Verosimilitud.	44
VIII. Interpretación de la correlación.	44
A. Resultados de las regresiones para los coeficientes de interés.	45
B. Correlación entre entre residuos del modelo.	48
IX. Aplicación del modelo prototipo de Economía de la Identidad de Akerlof y Kranton a la decisión de ejercer el derecho al voto.	54
A. Posibles Equilibrios de Nash Perfectos en Subjuegos.	56
B. Subjuego electoral dentro de juego de elección identitaria.	58
4. Aplicación de la Teoría de Economía de la Identidad al grupo de Aborígenes en Argentina.	60

I.	Resumen	60
II.	Teoría de fusión de datos.....	61
III.	Datos del INDEC.....	62
	A. Fracción del grupo minoritario por subdominio geográfico.....	64
	B. Distancia de Hellinger (HD) entre las variables de cruce.	67
IV.	Método no paramétrico propuesto para estimar coeficientes de grupos no identificados unívocamente.	67
V.	Propiedades de los coeficientes OLS del pooling.	69
	A. Representatividad del diseño muestral y de la expansión muestral.....	71
	B. Selección de la expansión muestral.	73
VI.	Resultados.....	75
	A. Regresiones de Mincer y Heckman por sexo para toda la muestra.	75
	B. Regresiones imputando la categoría de aborigen en función de la probabilidad.....	76
	C. Poolings donde se controla el porcentaje de aborígenes.	76
VII.	Conclusiones.....	78
VIII.	Anexo Estadístico.	79
	A. Estimación retorno a la educación aborígenes en Estados Unidos.....	79
	B. Resultados con grupos perfectamente identificados.....	79
	C. Identificación con método no paramétrico.	80
	D. Generación de expansiones muestrales.	80
	E. Interpretación de resultados.....	81
	F. Selección de la expansión muestral.	82
	G. Comparación con el uso de un método paramétrico y selección de registros por “ <i>matching</i> .”	83
	H. Expansiones muestrales y pobreza.	85
	I. Conclusiones.....	87
5.	Resumen y Conclusiones.....	89

I. Aportes de la variable identidad al análisis del comportamiento.	89
II. Aportes de esta tesis a la Economía de la Identidad.....	90
III. Próximas oportunidades de investigación.	91
Referencias Bibliográficas.....	92

Listado de Figuras, Gráficos y Tablas

Figura 1 - Ejemplo de áreas PUMAs empleadas en la encuesta ACS para agrupar los datos PUMS usados en las regresiones.....	23
Tabla 2 - Regresión de Heckman - estadísticos seleccionados para hombres (males) de 16 años o más.	25
Tabla 3 - Regresión de Heckman - estadísticos seleccionados para mujeres (<i>females</i>) de 16 años o más.	26
Tabla 4 - Ciudadanos Registrados para la Votación y Fuera de la Fuerza de Trabajo....	34
Tabla 5 - Ciudadanos que efectivamente votaron.....	34
Tabla 6 - Ciudadanos registrados por año.	35
Tabla 7 - Registración y participación en la fuerza de trabajo por grupo étnico.	36
Tabla 8 – Áreas metropolitanas más segregadas y menos segregadas.	37
Tabla 9- Regresión biprobit por grupo y por género, coeficientes de interés.....	47
Tabla 10- Regresión biprobit por grupo y por género con segregación, coeficientes de interés.	48
Tabla 11 – Regresión biprobit para varones afroamericanos sin regresor de segregación – reporte de Stata.	50
Tabla 12 – Regresión biprobit para varones afroamericanos con regresor de segregación – reporte de Stata.	51
Tabla 13 - Correlación residual en el modelo bivariado por grupo y género.	52
Tabla 14 – Porcentajes de aborígenes en la población por subdominio.	65
Tabla 15 – Variables de cruce entre el cuestionario ampliado y la EAHU.....	67
Tabla 16 - Ejemplo encuesta con 2 registros con variable de expansión muestral.....	68
Tabla 17 - Ejemplo encuesta con 2 registros con variable de expansión muestral modificada.	68
Tabla 18 – Expansiones muestrales generadas con el método no paramétrico sobre datos de la EAHU.	74
Tabla 19 – Regresión de Heckman para varones sobre expansión muestral aborpondera603	77
Tabla 20 – Descomposición de Oaxaca Blinder del gap de ingresos de varones aborígenes en Argentina, comparado con la restante población de aborígenes.....	77

Tabla 21 – Regresión de Heckman para varones sobre expansión muestral aborpondera603	78
Tabla 22 - Descomposición de Oaxaca Blinder del gap de ingresos de mujeres aborígenes en Argentina, comparado con la restante población de aborígenes.....	78
Tabla 23– Regresión OLS de Mincer para toda la población con expansión muestral pwgtp - Output de Stata	79
Tabla 24 – Regresión de Mincer para aborígenes con expansión muestral fwgtp - Output de Stata	80
Tabla 25 – Expansiones muestrales generadas con el método no paramétrico sobre datos PUMS de la ACS.	82
Tabla 26 – Resultados comparados de la regresión de Mincer con expansión muestral 8 y la misma regresión de Mincer exclusivamente para aborígenes con variable pwgtp original de la ACS, empleando <i>frequency weights [fw]</i> en lugar de <i>sampling weights [pw]</i>	83
Tabla 27 – Probabilidad logit de que un registro de la ACS corresponda a un aborígen (racaian=1) empleando las variables de cruce.	84
Tabla 28 – Probabilidades calculadas de pertenecer al grupo aborígen.	84
Tabla 29 – Resultados Regresión de Mincer sobre registros con probabilidad no paramétrica mayor a 0,8.	85
Tabla 30 – Cambio en la composición por deciles de la variable <i>povpip</i>	86
Gráfico 1 - Comportamiento de autoexclusión/autoinclusión en función del perjuicio/beneficio pecuniario resultante de la discriminación/nepotismo.	27
Gráfico 2 - Registración y voto en elecciones presidenciales.	39
Gráfico 3 - Registración y voto en elecciones de medio término.....	39
Gráfico 4 - Diferencias en votación, elecciones presidenciales.	40
Gráfico 5 - Diferencia en votación, elecciones de medio término.	40
Gráfico 6 – Representación del juego de Akerlof y Kranton en forma extensiva.	55
Gráfico 7 – Identificación de juego y subjuegos en la representación extensiva.	56
Gráfico 8 - representación del juego en forma normal con los diferentes ENPS.	57
Gráfico 9 – Representación extensiva de los cuatro posibles ENPS.	57
Gráfico 10 – Evolución del coeficiente b^{**}_{yearse} en función del % del grupo minoritario en el pooling - datos de la Tabla 18.	75
Gráfico 11 - distribución de <i>povpip</i> en deciles para <i>racaian</i> =0.	86

Gráfico 12 - distribución de *povpip* en deciles para *raaian=1*. 87

Prefacio

La intención inicial de esta tesis fue la aplicación de la teoría de la Economía de la Identidad al análisis del nivel de acumulación de capital humano y participación en el mercado laboral de los aborígenes en Argentina relativo al resto de la población, mediante un estudio sincrónico a partir del Censo y de la Encuesta Anual de Hogares Urbanos del año 2010 del Instituto Nacional de Estadísticas y Censos (INDEC).

Las comunidades aborígenes, una parte significativa de la población argentina, fueron pobremente relevadas estadísticamente, o estudiadas con modelos teóricos de economía. La elección de identidad, siendo la elección más importante que realiza una persona (Akerlof & Kranton, *Economics and Identity*, 2000), significa para los miembros de los pueblos aborígenes la disyuntiva entre adoptar la identidad hegemónica, integrada a la sociedad, o permanecer inmóvil, encerrado en sus costumbres ancestrales, en el seno de su comunidad aborigen, postergado, adoptando una identidad antagonista o marginal.

La identidad hegemónica argentina tiene características del tipo europeo, como explícitamente es el deseo de los constitucionalistas al fomentar la inmigración de ese origen. (Congreso General Constituyente, 1853).

Lo que para un ciudadano argentino típico significa una inversión con un retorno equivalente cierto y conocido tendrá un menor atractivo como elección de vida (elección por la identidad hegemónica) para un aborigen, dada por su condición de sujeto discriminado. Un menor retorno al capital humano de por sí resulta en una menor inversión de equilibrio en años de estudio. Es decir que lo que es virtuoso y deseable para un individuo representativo de la media de la población argentina, en el caso del sujeto discriminado puede ser sustancialmente menos atractivo.

Entonces, la discriminación por gustos “*taste based*” y el análisis económico iniciado por Gary Becker (1993, pág. 385) sobre aspectos sociales previamente considerados fuera de los límites de la Ciencia Económica son centrales para explicar el menor nivel de desarrollo de los pueblos aborígenes.

Esta tesis integra el cuerpo de literatura económica sobre discriminación iniciada por Becker (Becker, *The Economics of Discrimination*, 1957) y comprobado empíricamente por más de medio siglo de análisis econométricos (Heckman, Lochner, & Todd, *Fifty Years of Mincer Earnings Regressions*, 2003) con el abordaje novel de Akerlof y

Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) sobre diversos objetos de estudio que son grupos sociales minoritarios. Y realiza validaciones empíricas sobre modelos económicos que, como indican Iversen et al (*Caste Dominance and Economic Performance in Rural India*, 2014) conjugan tanto los mecanismos externos como los mecanismos internos que vinculan el bienestar económico con significantes de identidad social.

Al incorporar la variable identidad se busca explicar situaciones de desarrollo personal fallido aún más marcadas que las explicadas por la discriminación por gustos. Como la inversión en capital humano es vista como un intento de integrarse al cuerpo social hegemónico, y ese esfuerzo debilita la identificación del sujeto con su grupo de pertenencia, existirá para el individuo una pérdida de utilidad como resultado de su pérdida de identificación, que en este trabajo se usa como sinónimo de pérdida de identidad¹.

Para la teoría de la disonancia cognitiva la persona es sensible al hecho de que los propios actos sean contrarios a las normas y creencias propias, y el reconocimiento de esta disonancia motivará al individuo a tomar acciones para resolverla, ya sea cambiando su marco de valores y creencias ó cambiando sus acciones ó, más sutilmente, cambiando la percepción de sus acciones (Festinger L. , 1957). La *desutilidad* por identidad se explica por la disonancia cognitiva a nivel de la psiquis del individuo entre sus decisiones de comportamiento y las normas y prescripciones de pertenencia al grupo. También proviene de actitudes de escarmiento externo por parte de otros miembros de su comunidad; escarmiento por pretender ser distinto, querer sobresalir, por ser orgulloso y aspirar a ser una persona por encima de sus pares, con características de capital humano propias del grupo hegemónico. Esta *desutilidad* por escarmiento se verá atenuada en la medida en que el individuo que intenta acumular mayor capital humano lo haga fuera del alcance de la vista de otros individuos de su comunidad. Es decir, la *desutilidad* será menor si de alguna forma se perfecciona “aislado”; contrariamente, la *desutilidad* será mayor bajo la mirada reprobadora de su comunidad que verá al individuo como alguien asimilado al grupo social hegemónico. En este último caso el individuo se arriesga a convertirse en un cipayo, traidor, vendido y, por lo tanto, rechazado por su comunidad por no seguir las normas de su identidad aborigen y discriminado desde la población mayoritaria por considerarlo un inferior que aspira a comportarse como un igual pero que nunca podrá serlo.

¹ Stuart Hall, en *Introduction: Who Needs Identity?* menciona que identificación es un concepto preferible a identidad (Hall, *Questions of Cultural Identity & Paul du Gay*, 2011)

Identificadas las fuentes de información para Argentina provistas por el INDEC y dadas las peculiaridades de las mismas, en el contexto del Seminario de Proyecto de Tesis el autor y los directores de esta tesis discutieron y concluyeron que el plan de trabajo requeriría primero testear las hipótesis en ámbitos en los que existen datos de uso directo. Luego replicar los modelos con los datos argentinos empleando la técnica de *matching* estadístico específicamente propuesta que incluye el uso de variables de expansión muestral como pesos en las regresiones.

Esto explica el recorrido de la investigación iniciando con datos de encuestas de Estados Unidos en las que se relevan conjuntamente la condición racial, de ingresos y la registración política.

La primera aplicación estudió el comportamiento alrededor de una decisión identitaria, la de participar o no en el mercado laboral, mientras que la segunda aplicación incorporó una segunda decisión identitaria, la de registrarse o no para votar. Dos decisiones identitarias requirieron de un modelo econométrico bivariado, lo cual enriqueció el análisis.

En el tercer paper finalmente se retomó el objeto original de estudio, los aborígenes en Argentina, replicando el análisis del primer paper donde la decisión identitaria es la participación en el mercado laboral; como en Argentina la participación en las elecciones federales bienales es obligatoria, incluso a nivel de primarias, no se dan las condiciones para replicar el segundo paper con datos locales.

Finalmente, desearía agradecer a mi Director de Tesis, el profesor Jorge Miguel Streb, a los profesores Julio Jorge Elías, Esteban Maioli y Pablo Garófalo, miembros del Jurado de Tesis, y también a la profesora Mariana Conte Grand, por las útiles indicaciones y sugerencias recibidas para el desarrollo de esta tesis doctoral.

1. El Concepto de Identidad.

El concepto de identidad ha sido recientemente un objeto de estudio revisitado por diversas disciplinas y a la vez sujeto de crítica en la literatura de las ciencias sociales con el propósito de reemplazar la idea de una identidad integral, originaria y única. (Hall, Introduction: Who Needs Identity?, 1996) Para la Economía, ciencia que lo incorpora como concepto nuevo principalmente con el trabajo seminal de Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000) la idea de identidad se introduce como una elección racional, la más importante que realiza el individuo. De modo que el concepto de identidad se agrega al cuerpo de la teoría de decisión racional a la que previamente Gary Becker incorporó otros temas de relevancia social, entre ellos la discriminación. (Becker, Nobel Lecture: The Economic Way of Looking at Behavior, 1993)

En este primer capítulo se expone la Teoría de la Economía de la Identidad (Akerlof & Kranton, Economics and Identity, 2000) que motiva esta Tesis, se brinda un marco teórico relacionando el concepto de identidad con la Sociología, la Antropología y la Psicología, se justifica la relevancia del trabajo, se fijan los objetivos de investigación y se plantean las hipótesis a explorar. Especial importancia para la Tesis tiene el concepto de disonancia cognitiva (Festinger L. , 1957) proveniente de la Psicología Social, que es compatible con el comportamiento de autoexclusión revelador de la pérdida de utilidad por identidad de individuos de grupos minoritarios al intentar integrarse a las normas y costumbres hegemónicas.

I.Indagaciones Preliminares.

George Akerlof, premio nobel de Economía por su teoría sobre información asimétrica y su efecto sobre el comportamiento económico, introduce con Rachel Kranton el concepto de identidad para la interpretación económica del comportamiento humano. (Akerlof & Kranton, Economics and Identity, 2000) Su nuevo abordaje rescata el impacto de las categorías que definen a la identidad, como son las normas e ideales que las personas pertenecientes a esas categorías deben respetar. Las ganancias o pérdidas de utilidad que se obtienen por actuar conforme o no a las propias identidades explican hechos económicos que contradicen la teoría tradicional preexistente, como son el persistente y

marcado retraso en el que viven grupos minoritarios sin que se verifique convergencia en su nivel de desarrollo.

La economía de la identidad ve al individuo no como un sujeto meramente pasivo que acepta del exterior una identidad según le corresponda, sino como una persona que elige qué identidad adoptará. Las características innatas no definen una identidad irrevocable para el individuo, sino que existe la alternativa de una decisión de identificación de la persona con un grupo social, una decisión de integrarse a la identidad hegemónica, de adoptar una identidad de “*insider*”, ó de marginarse a una identidad antagonista o de “*outsider*”; este mecanismo interno de identificación, a diferencia del mecanismo externo de discriminación, generará una polarización y la desaparición del continuo de equilibrios intermedios, identidades híbridas, como una opción racional. (Iversen, Kalwij, Verschoor, & Dubey, 2014)

II. Problema de investigación y justificación.

Como ya se expresó en el Prefacio de esta tesis, la elección de identidad, siendo la elección más importante que realiza una persona (Akerlof & Kranton, 2010) significa para los miembros de los pueblos aborígenes la disyuntiva entre adoptar la identidad hegemónica, integrada a la sociedad, ó permanecer “inmóvil”, encerrado en sus costumbres ancestrales, en el seno de su comunidad aborígen, postergado, adoptando una identidad “antagonista” o “marginal.”

La identidad hegemónica tiene características del tipo europeo, como explícitamente es el deseo de los constitucionalistas al fomentar la inmigración de ese origen. (Constitución de la Confederación Argentina, Artículo 25, 1853).

Lo que para un ciudadano argentino típico significa una inversión con un retorno equivalente cierto y conocido tendrá un menor atractivo como elección de vida (elección por la identidad hegemónica) para un aborígen, dada por su condición de sujeto discriminado. Un menor retorno al capital humano de por sí resulta en una menor inversión de equilibrio en años de estudio. Es decir que lo que es virtuoso y deseable para un individuo representativo de la media de la población argentina, en el caso del sujeto discriminado puede ser sustancialmente menos atractivo.

Por lo tanto el problema de investigación puede definirse como el análisis de las causas del endémico atraso en el que los aborígenes de Argentina se encuentran respecto al resto de la sociedad.

A. Criterio de relevancia metodológica.

La situación problemática es relevante para la disciplina de la Economía, ya que ofrece la oportunidad de aplicar un nuevo abordaje, la Economía de la Identidad, para explicar diferencias sociales persistentes en un objeto de estudio, las etnias aborígenes, no abordado desde la Ciencia Económica para el ámbito de Argentina.

B. Criterio de relevancia social.

La Economía de la Identidad permite comprender elecciones de grupos minoritarios cuyo comportamiento es el de “*outsiders*” del sistema, quienes eligen o son relegados a una identidad antagonista a la hegemónica del grupo “*insider*.” Estas realidades son vistas como disfuncionales o autodestructivas por el grupo hegemónico, y por aquellos que fijan normas de conducta, patrones e ideales requeridos para ser “*insiders*.”

A partir de la adopción de su Constitución la Argentina comenzó a transitar el período de organización nacional. Aún sin reconocer la preexistencia étnica y cultural de estos pueblos, la Constitución Nacional vigente desde 1853 establecía:

“Proveer a la seguridad de las fronteras, conservar el trato pacífico con los indios y promover la conversión de ellos al catolicismo.” (Congreso General Constituyente, 1853)

Por lo tanto la Asamblea General Constituyente parece ser un hito importante para la definición de la identidad argentina, ya que se menciona a los indios (sic) como parte de Argentina, con una pretensión de fundar una Nación, incluyendo a personas de comunidades aborígenes.

Benedict Anderson (1993) provee una definición de nación como una comunidad política imaginada, inherentemente limitada y soberana. Es imaginada porque es imposible que en una nación todos sus miembros se conozcan mutuamente. Es limitada conceptualmente por sus fronteras, no abarcando a todos los seres humanos, a diferencia de las religiones, como el cristianismo, el hinduismo o el confucianismo, e ideologías, como el marxismo o el liberalismo. Es soberana porque es libre, sin dependencia de ninguna

otra nación; y si de algo dependiera, sería sólo y directamente de Dios. Es una comunidad por ser concebida como una profunda camaradería horizontal, de hermanos; el sentimiento nacionalista justifica el sacrificio máximo de dar la vida por la nación.

Comprender mejor los procesos que llevan a miembros de las comunidades aborígenes a adoptar una identidad antagonista a la identidad de la población argentina en general, la identidad nacional “*mainstream*” o de “*insider*”, tendrá implicancias para entender las circunstancias estructurales de los aborígenes en Argentina, y de su endémico bajo desarrollo o de la situación de miseria que típicamente caracteriza a estas comunidades.

También tendrá implicancias en el desarrollo de programas a medida para estas comunidades a ser implementados por los actores políticos argentinos.

III.Marco Teórico.

A.Enfoque de la Utilidad.

Las curvas de utilidad son la herramienta para expresar las motivaciones de las personas en la Ciencia Económica y su origen proviene del siglo XIX. El proceso de ajuste de la toma de decisiones personales a estas motivaciones no exige una actitud consciente y racional, lo cual no va en detrimento de su idoneidad. Las curvas de utilidad son válidas por su poder explicativo sobre decisiones del sujeto, independiente de lo consciente que el sujeto esté sobre la aplicación de las mismas.

El uso original de las curvas de utilidad se dio en mercados de bienes y servicios para determinar la combinación óptima en la que el consumidor maximizará su utilidad dada una restricción presupuestaria. Se demuestra que esta combinación óptima está en el punto donde se igualan las tasas marginales de sustitución entre todos los productos y servicios y la pendiente de la curva presupuestaria, dada por la razón de los precios.

Gary Becker (1993) amplió el uso de las curvas de utilidad para modelar todo tipo de preferencias de las personas, más allá de productos y servicios. Extendió la teoría tradicional de la elección racional individual hacia aspectos sociales, más allá de los previamente considerados dentro de las fronteras del poder explicativo de la Ciencia Económica, como discriminación contra minorías, crimen y castigo, capital humano y formación, disolución y estructura de las familias (Becker, Nobel Lecture: The Economic Way of Looking at Behavior, 1993).

B.Retorno a la Educación.

La inversión en educación y entrenamiento en el trabajo son variables explicativas del modelo de ingresos formulado originalmente por Mincer (1958) y sus variantes (Heckman, Lochner, & Todd, 2008). Los empleadores pueden tener una menor predisposición a pagar por el trabajo del sujeto discriminado; alternativamente, otros trabajadores pueden exigir un mayor pago por trabajar junto a una persona a la que discriminan. El residuo no explicado por el modelo de Mincer se lo atribuye a la “discriminación por gustos”.

Arrow (1973) y Phelps (1972) aportaron otra explicación económica de la discriminación, que no recurre a gustos o preferencias, sino que se basa en una toma de decisiones racional pero estereotipada sobre grupos minoritarios de personas, por parte de grupos mayoritarios o influyentes. El resultado es el mismo que en el modelo de discriminación por preferencias, y los prejuicios estereotipados pueden volverse profecías autocumplidas. Los miembros de las minorías presentarán una brecha negativa en inversión en capital humano, ya que su retorno por el esfuerzo será menor (resultado del estereotipo), y finalmente el estereotipo inicial se transformará en la situación objetiva final.

“This analysis [stereotyped reasoning or statistical discrimination] suggests that the beliefs of employers, teachers, and other influential groups that minority members are less productive can be self-fulfilling, for these beliefs may cause minorities to underinvest in education, training, and work skills, such as punctuality. The underinvestment does make them less productive” (Becker, Nobel Lecture: The Economic Way of Looking at Behavior, 1993, pág. 388)

C.Enfoque de Utilidad bajo la mirada del otro.

Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000) llevan el uso de las curvas de utilidad a un conjunto nuevo de comportamientos que no se dan en el contexto estrictamente individual sino que surgen de las interacciones sociales y que tienen sentido cuando hay otro que observa, y juzga, la conducta del individuo. Estos comportamientos tienen en común el concepto de identidad ya estudiado extensamente en otras disciplinas, y que pueden asociarse a preferencias de segundo orden, relacionadas más con la autorrealización que con la subsistencia.

El uso de las curvas de utilidad aportado por Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) permite modelar el impacto de las categorías que definen la identidad, las normas que las personas que pertenecen a esas categorías deben respetar, las utilidades o pérdidas de utilidad que obtienen las personas por actuar conforme o no a sus identidades, y finalmente la explicación de hechos económicos que contradicen la teoría tradicional preexistente.

Desde el punto de vista del individuo, comportarse de acuerdo a su identidad deseada es una preferencia de segundo orden más cercana a la autorrealización que a la subsistencia.

El poder explicativo de este nuevo uso de las curvas de utilidad alcanza el comportamiento de personas pertenecientes a un grupo social que no tomarán decisiones óptimas desde el punto de vista esperable por las normas e ideales aceptados por el grupo social dominante si experimentan una pérdida por desviarse de la identidad de su grupo.

D.Mecanismos internos y externos y significantes de identidad social.

Iversen et al (2014) califica a la discriminación basada en actitudes, preferencias y cálculos desarrollada por Becker (1993) o la discriminación estadística desarrollada por Phelps (1972) y Arrow (1973) como mecanismos externos que vinculan el bienestar económico con significantes de identidad social. Son mecanismos externos, porque el individuo discriminado es sujeto pasivo de preferencias o prejuicios que llevan a otros a tomar, inicialmente, acciones discriminatorias. En el caso de los gustos, llevan a esperar un mayor retorno económico por tener que interactuar con quien se discrimina, se rechaza y se quiere evitar. En el caso de los estereotipos, llevan a tomar una cierta característica del grupo discriminado como señal de su nivel de habilidad. Finalmente, el mecanismo externo se torna en una profecía auto cumplida, a través de un mecanismo interno de ajuste de la inversión en capital humano.

Iversen, Kalwij, Verschoor y Dubey (*Caste Dominance and Economic Performance in Rural India*, 2014) identifican también dos mecanismos internos que explican variaciones en el grado de bienestar. Los basados en la autoimagen personal propuestos por Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) y los basados en la inercia colectiva propuesto por Young (2001).

La Economía cuenta entre los mecanismos internos para la explicación del comportamiento humano a las preferencias personales de segundo orden, que pueden ser sustantivamente diferentes entre distintos grupos.

Young (2014) emplea teoría de juegos para mostrar cómo normas sociales y patrones de comportamiento se convierten en auto exigibles en un contexto de grupo; todos quieren cumplir cuando se espera que todos los demás vayan a cumplir. Este fenómeno ha sido descrito en la disciplina de Comportamiento Organizacional como resultado de un alto compromiso en el ámbito de un proyecto u objetivo compartido. Quienes comparten la toma de decisiones pueden incurrir en una aberración denominada Pensamiento de Grupo o “*group-think*”; en este contexto, el fuerte compromiso interno resulta en la censura y el rechazo de cualquier disenso, aunque este pueda ser lícito y merezca tenerse en cuenta².

E.Enfoque de la Psicología Social.

La pérdida de utilidad asociada a la identidad personal es la explicación económica de un fenómeno extensamente aceptado en la psicología: la ansiedad que una persona experimenta por violar las reglas internalizadas. Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000), citando a Thomas (1996), explica que la identidad personal, o ego, debe ser continuamente defendido de la ansiedad, con el fin de limitar la disrupción y mantener el sentido de unidad. Este concepto de identidad también es consistente con los mecanismos propuestos por Young (2001).

Cabe destacar que la pérdida de utilidad derivada del desvío de la persona respecto al comportamiento esperable de su grupo se produce tanto en la persona como en el grupo que la rodea. Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000) proveen un modelo simple de juego secuencial, que muestra cómo la identidad afecta las interacciones individuales.

Por caso, se puede considerar lo ofensivo que puede resultar para un grupo el estilo de vida de una persona, como podría ser la poligamia; esto frecuentemente resulta en acciones punitivas del grupo sobre el individuo para restaurar su utilidad perdida, y en lo posible obligar a la persona singular a no desviarse de las normas de su grupo.

² Sobre *Group-think* ver “Final Voyage of the Challenger” (Hauptman & Iwaki, 1991)

F. Sobre normas fuertemente observadas y normas vacuas.

¿Qué relevancia puede tener incorporar las normas sociales como motivaciones de la conducta? Al fin y al cabo, reza la sabiduría del pueblo que las normas están hechas para ser rotas. Sin embargo, rara vez la derogación de una norma coloca al espíritu humano en un Edén de libertad absoluta, sino que la anterior norma es reemplazada por otra, muchas veces tanto o más rígida y coercitiva que la anterior.³

Nuevamente desde el campo del Comportamiento Organizacional, Chatman (2014) diferencia entre el contenido de la cultura (sus normas) y la intensidad de adhesión a las mismas (qué tan fuerte es el sentimiento de las personas si alguien viola una norma central de la cultura). En caso de existir consenso sobre las normas y baja adhesión, se generan culturas vacuas en las organizaciones. Por transitividad, lo mismo ocurre en marcos sociales más amplios que las organizaciones, donde antiguas normas pierden su fuerza coercitiva; la mayoría de los actores sociales adoptan un comportamiento permisivo, desde el punto de vista de aquellos que aún adhieren fervorosamente a las viejas normas, y no se genera una reacción enérgica de parte del cuerpo social frente a su transgresión. Por el contrario, una cultura fuerte implica la no indiferencia de parte de los miembros del grupo frente a la transgresión de la norma, sino una reacción proporcionada al nivel de compromiso, como el descrito por Young (2001).

La conclusión lógica de este punto de vista es que esperamos encontrar mayor variedad y riqueza social, y menos individuos que adopten personalidades antagonistas o de *outsiders*, en sociedades más permisivas y menos rígidas. Desde los más tempranos estudios antropológicos se observa que el costo social y económico que deben pagar aquellas sociedades donde está ausente el factor cambio es la ausencia de progreso (O'Toole, 1995).

³ La posición elegida en este paper está en línea con las ideas posmodernas sobre opresión y violencia de las estructuras planteadas por Michael Foucault y Jacques Derrida quienes coinciden en que no existe tal cosa como la liberación, ya que una estructura que se destruye será inevitablemente reemplazada por otra de poder y autoridad también hegemónica y opresiva.

Edward Saïd propone una tesis alternativa afirmando que la autenticidad puede ser recuperada exponiendo, deconstruyendo y desmantelando las verdades hegemónicas de, por ejemplo, el colonialismo, el imperialismo, el racismo, el paternalismo, la masculinidad tóxica, el "*Whiteness*" y otras, legitimando así las guerras culturales y la cultura de la cancelación. (Saïd, 1978)

G.Variable identidad. Efecto de acentuación.

Tajfel (1959) presentó originalmente el efecto de acentuación donde individuos que se convencen en base a cierta dimensión que una categoría social les corresponde adoptan acciones que refuerzan el vínculo con esa categoría y los diferencian de las otras. Este mecanismo confirma el efecto que en los individuos tiene la conciencia de su propia identidad social para construir su personalidad. Individuos con muy similares características pueden diferir en el grado en el que se perciben como pertenecientes a cierto grupo social.

H.Abordaje sociológico al concepto de identidad.

El abordaje sociológico del concepto de identidad lo origina Mead (1934) y lo desarrolla Stryker (1980). Para la sociología existe una influencia mutua entre el ego y las estructuras sociales.

“El supuesto central del abordaje sociológico es que los individuos ocupan posiciones dentro de las estructuras sociales, y que la identidad social se explica por cómo ellos son influenciados por y también cómo negocian estas estructuras sociales.

Entonces, más que usar la idea del abordaje de identidad social de correspondencia o *matching* entre categorías sociales dadas y características individuales para tratar la identidad como algo que el individuo posee, el abordaje sociológico de la identidad usa la idea de posiciones sociales negociadas resultantes de la interacción entre individuos en estructuras sociales para tratar la identidad social como algo que producen los individuos.” (Davis, *Social Identity Strategies In Recent Economics*, 2006, pág. 376)

Stryker (1980) asume que las estructuras sociales son relativamente estables.

I.Otros significados de la variable identidad.

Otros abordajes permiten abordar la identidad como un fenómeno de grupos.

Un enfoque antropológico de la identidad relevante para la Economía es la determinación de las fronteras que delimitan la identidad étnica y de grupo.

John B. Davis (Davis, Akerlof and Kranton on identity in economics: inverting the analysis, 2007) en su lectura crítica de Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) remarca que los autores toman para su concepto de identidad:

1° el abordaje de identidad social de la Psicología que explica que las personas se identifican con estereotipos sociales preexistentes.

2° la teoría psicodinámica de la personalidad, que explica cómo la identidad se relaciona con el comportamiento, a través de reglas internalizadas, en un proceso que genera la identidad o sentido de sí mismo del individuo, para disminuir la ansiedad o “disonancia cognitiva.”

John B. Davis (2007) remarca la diferencia entre la función I_j como colección de diferencias (*mismatch*) entre las características individuales y las del estereotipo externo de identidad, o imagen social del individuo, y la función U_j que corresponde al sentido de uno mismo como un todo *per se*, como unidad indivisible. Por eso echa en falta una relación reflexiva del individuo consigo mismo; resultado de esa relación reflexiva surge el concepto de identidad (imagen) personal, o personalidad. Ignorar este paso de construcción de la personalidad lleva a estereotipar a los individuos, según Turner (1985; 1987), no rescatando su status como entidades independientes.

J. Posición conceptual elegida para la variable identidad.

Una decisión relevante para este marco teórico será, entonces, la posición a tomar sobre el concepto de identidad para la Economía de Identidad: si mantener el elegido por George Akerlof y Rachel Kranton (Akerlof & Kranton, *Economics and Identity*, 2000), o tomar el abordaje sociológico de John Davis (Davis, 2007).

¿Es el abordaje sociológico una crítica válida al abordaje de psicología social como sustento de la teoría de *Identity Economics*? ¿O ya está resuelta por el original abordaje de Akerlof y Kranton (2000)?

El aporte de Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) aun cuando supone la existencia de estereotipos provenientes de normas e ideales, permite y da ejemplos de elección de la propia identidad. La elección se da a través de un proceso reflexivo y evaluativo, y de modificación de los estereotipos, a través de una negociación y acción social colectiva. La elección de la propia identidad se presenta en casos en los que un individuo debe elegir si adhiere o no a las normas requeridas y se transforma, o intenta transformarse, en una *persona integrada*. O, por el contrario, si renuncia a guiarse por las normas requeridas y elige una *identidad antagonista*.

Un caso paradigmático de individuo que a pesar de los dificultades y costos personales elige su camino desde un hogar monoparental de muy bajos ingresos ubicado en un

ghetto afroamericano de Washington DC a una universidad *Ivy-League*, rompiendo el estereotipo de varón afroamericano, es el de Cedric Jennings, relatado en la etnografía de Ron Suskin. (*A Hope in the Unseen: An American Odyssey from the Inner City to the Ivy League*, 1998)

El caso de interacción entre individuos y estructuras para producir la identidad, Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) lo grafican a través de la acción de líderes positivos, por caso Ghandi y Martin Luther King, quienes resignificaron la identidad india o afroamericana, cambiando normas autodestructivas que las regían, elevando a sus comunidades. (Akerlof & Kranton, 2010, págs. 113-120)

En conclusión, pensamos que el abordaje de Akerlof y Kranton (2000; 2010) admite no sólo los estereotipos que normalmente están asociados a normas e ideales, sino también contemplan un proceso reflexivo de elección de identidad, y un proceso de negociación sobre el contenido de esa identidad. Muchos individuos quedarán atrapados en el estereotipo, pero la identidad como elección seguirá existiendo como oportunidad de elección personal.

K. Teoría de Economía de la Identidad.

Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) postulan una función de utilidad con identidad:

$$U_j = U_j(\mathbf{a}_j, \mathbf{a}_{-j}, I_j)$$

Donde la utilidad del individuo j depende de su identidad o ego I_j , sus acciones \mathbf{a}_j y las acciones de los demás \mathbf{a}_{-j}

$$I_j = I_j(\mathbf{a}_j, \mathbf{a}_{-j}, \mathbf{c}_j, \boldsymbol{\varepsilon}_j, \mathbf{P})$$

A su vez la identidad de la persona es una función de sus acciones y de las de los demás, de su categoría social asignada \mathbf{c}_j y del “*mismatch*” entre las propias características personales y las de su categoría social asignada $\boldsymbol{\varepsilon}_j$ indicadas por las prescripciones \mathbf{P} .

A los aumentos o caídas de Utilidad U_j debida a I_j se las llama ganancias o pérdidas de utilidad por identidad.

2. Elección de Identidad y Participación en el Mercado Laboral Evidencia empírica con datos de Estados Unidos de América.

I. Resumen.

El objetivo de este trabajo es aplicar la teoría de la Economía de la Identidad en grupos minoritarios agrupados por su etnia o por otra característica común, como su país de origen. Se revela la variable identidad como efecto de la mirada del otro afectando la decisión de la persona entre integrarse socialmente de acuerdo a las normas hegemónicas o autoexcluirse, renegando de las normas y dando la espalda a lo que la sociedad espera de ella.

Para algunos grupos sujetos a discriminación laboral la decisión de las personas puede consistir en adoptar una identidad antagonista entrando en conflicto con la norma hegemónica de la ética del trabajo, el mandato social de que el individuo adulto debe trabajar para ganar su sustento. Este hecho es narrado en diversas etnografías (Liebow E. , 1967) (Suskin, 1998) y también es el objeto de estudio de economistas que diferencian entre “*lowerclass*” y “*underclass*.” (Oxoby, 2004). La “*lowerclass*” está conformada por individuos desaventajados, pero que se integran al sistema; en cambio la “*underclass*” la conforman individuos que no se resignan a ocupar la posición de discriminación que la sociedad les depara, reniegan de la ética del trabajo e incluso pueden buscar alcanzar status con comportamientos vistos por la mayoría como propio de holgazanes, como autolesivos o como directamente ilícitos.

Los datos PUMS (*Public Use Microdata Sample*) empleados son los de la encuesta ACS (*American Community Survey*) del Census Bureau de los Estados Unidos (United States Census Bureau, s.f.), cubren un cuatro por ciento de la población de Estados Unidos y fueron relevados anualmente entre los años 2012 y 2015. Incluyen variables categóricas para identificar indoamericanos, afroamericanos, iberoamericanos y diversos grupos de asiáticos, como grupos con un denominador común racial; inmigrantes, como grupo de referencia para el caso en el que no existe una identidad común; y el grupo mayoritario de raza blanca.

II.Introducción.

Gary Becker en su disertación doctoral (Becker G. S., *The Economics of Discrimination*, 1957) introduce la discriminación como un comportamiento objeto de estudio de la Economía en lo que sería la primera entre muchas preferencias y actitudes que, explicadas desde un marco teórico micro, tienen implicancias macro grupales. (Becker, 1993). El coeficiente de discriminación captura el efecto de los gustos por discriminación en los costos y en los precios, como los salarios. Si existen gustos por discriminación en los costos y en los precios, como los salarios. Si existen gustos por discriminación contra una minoría numérica, la siguiente relación es condición necesaria y suficiente para que la discriminación sea efectiva: el ingreso agregado de la mayoría (los “*race White*”) debe ser mayor que el ingreso agregado de la minoría discriminada (los “*race Black*”, “*race American Indian*”, “*Immigrants*”, “*Iberoamericans*”, etc.) (Becker G. S., *The Economics of Discrimination*, 1957, pág. 19)

En los Estados Unidos, dado que la mayoría blanca dispone de aproximadamente el 80% del trabajo agregado y de una fracción aún mayor del capital agregado, la condición previa implica que cualquier minoría hacia la cual exista un gusto por discriminación será efectivamente discriminada. (Becker G. S., 1957, pág. 28)

Diferentes gustos por discriminación hacia un grupo aumentan el costo π_j de empleo de miembros de un grupo (Becker G. S., *The Economics of Discrimination*, 1957, pág. 84):

$$\pi_j \times (1 + d_a + d_b + d_c) = \pi_j \times (1 + d_1)$$

Donde d_a es el coeficiente de discriminación de la clase de empleadores, d_b es el coeficiente de discriminación del gobierno y d_c es el coeficiente de discriminación de los sindicatos; finalmente d_1 es el coeficiente de discriminación combinado.

En un mercado de trabajo competitivo y en equilibrio, el grupo discriminado recibirá un salario reducido en comparación con otros grupos.

$$\pi_{-j} = \pi_j \times (1 + d_1)$$

En conclusión, el modelo de Becker de Economía de la Discriminación atribuye las diferencias persistentes que se observan en los salarios de los trabajadores a discriminación por gustos, en un contexto de equilibrios con pleno empleo⁴. Sin embargo los datos muestran diferencias marcadas de participación en el mercado laboral entre grupos de la población.

⁴ En la sección cuarta se presentan las alternativas que para la minoría discriminada Becker ya identificó como respuestas a la discriminación.

George Akerlof y Rachel Kranton (2000) introducen el concepto de Identidad en el análisis del comportamiento humano, concepto ya sobradamente explorado por otras ciencias sociales, como la Psicología Social, la Antropología y la Sociología (Hall, Introduction: Who Needs Identity?, 1996) (Tajfel H. , 1981).

La función de utilidad con identidad propuesta es la siguiente:

$$U_j = U_j(\mathbf{a}_j, \mathbf{a}_{-j}, I_j)$$

La utilidad individual es resultado de la propia identidad o ego I_j del individuo j , de sus acciones \mathbf{a}_j y de las acciones de los otros \mathbf{a}_{-j} .

$$I_j = I_j(\mathbf{a}_j, \mathbf{a}_{-j}, \mathbf{c}_j, \boldsymbol{\varepsilon}_j, \mathbf{P})$$

Por otro lado, la identidad individual es función de las propias acciones, de las acciones de los otros, de su categoría social asignada \mathbf{c}_j y del “*mismatch*” entre sus propias características y aquellas de su categoría social asignada $\boldsymbol{\varepsilon}_j$ dictada por prescripciones \mathbf{P} .

Aumentos o disminuciones de la utilidad U_j debidas a I_j son denominadas ganancias o pérdidas de utilidad por identidad.

La potencia implícita del concepto de identidad en el análisis económico queda revelada, por ejemplo, por su capacidad de caracterizar situaciones como la asociación entre empleos y géneros, es decir la existencia de algunos trabajos considerados masculinos y de otros trabajos considerados femeninos. (Akerlof & Kranton, 2010, págs. 84-85).

Robert Oxoby (Cognitive Dissonance, Status and Growth of the Underclass, 2004) presenta un modelo de adaptación cognitiva del individuo con una función de utilidad que tiene componentes tanto pecuniarios como no pecuniarios. (Oxoby, 2004, pág. 735)

$$U(x, e, \lambda, \bar{x}, \bar{e}) = x - c(e) + \delta[\lambda s_x + (1 - \lambda)\alpha s_e] - (1 - \lambda)\gamma$$

Donde x es el ingreso pecuniario y $c(e)$ es el costo asociado al esfuerzo; δ representa el valor relativo del status en la función de utilidad total; λ es la actitud del individuo respecto del status. Status que puede provenir del consumo $s_x = x - \bar{x}$ o del esfuerzo $s_e = \bar{e} - e$ (Oxoby, 2004, pág. 733). α es una constante que tiene en cuenta diferencias en la medición del status de cada posible fuente. El status total de un individuo surge del diferencial de consumo respecto del promedio, del diferencial de esfuerzo respecto del promedio, y de los pesos relativos entre estos diferenciales.

En el último término del modelo de Oxoby el parámetro γ representa el costo de desviarse de la norma que genera status. Para un individuo orientado al consumo $\lambda=1$, y para un individuo orientado al ocio $\lambda=0$. (Oxoby, 2004, pág. 734)

Si alcanzar status a través de un salario alto está vedado para el individuo, es factible que cambie sus preferencias hacia otras fuentes de status que no provienen de los incentivos tradicionales, como es el éxito laboral, sino de aquellas que puede proveer el ocio. El parámetro λ sintetiza la elección del individuo, y el término $(1 - \lambda)\gamma$ es el costo de esa elección; dado que el status resulta de la combinación lineal $\lambda s_x + (1 - \lambda)\alpha s_e$ entonces ningún individuo debería elegir un λ diferente de 0 o 1; las elecciones intermedias desaparecen como combinaciones que maximizan utilidad. (Oxoby, 2004, pág. 734)

El modelo de Oxoby permite endogeneizar una elección de identidad a través de cambiar la valoración subjetiva sobre qué es valioso para el individuo como generador de “status”; γ es el costo que el individuo enfrenta por cambiar su actitud desde el status orientado al consumo, que es la norma, hacia el status orientado al ocio, que es el propio de la “underclass”. (Oxoby, 2004, pág. 738)

La ecuación de utilidad de Oxoby se encuadra en el modelo general de Akerlof & Kranton siendo una aplicación específica que resalta la elección de identidad para minimizar la desutilidad por disonancia cognitiva de un trabajador discriminado.

En un primer término se formula el beneficio pecuniario que recibe un individuo, es decir el ingreso afectado por el coeficiente de discriminación d_1 menos el costo asociado al esfuerzo; es un componente de la utilidad que depende de la acción $a_j(e)$ del individuo primariamente en cuanto a la decisión de cuánto esfuerzo dedicar a la generación de un ingreso, pero también de acciones \bar{a}_{-j} de otros, como pueden ser las acciones de discriminación “*taste-based*” que describe Becker (The Economics of Discrimination, 1957) y que afectan su ingreso, pero que la persona j no puede modificar con sus acciones.

En los otros dos términos están los componentes no pecuniarios de la utilidad.

$\lambda = \lambda(I) = \lambda(a_j, a_{-j}, c_j, \epsilon_j, P)$ es la actitud del individuo respecto al status, donde $P = \bar{P}(\bar{c}_j)$ las prescripciones invariantes en el corto y mediano plazo (Kranton, 2016) que combinada con la categoría asignada, que para el caso estudiado por Oxoby $c_j = \bar{c}_j = \text{trabajador}$ queda definida como $P = \bar{P}(\text{trabajador}) = \text{ética del trabajo}$.

Es decir que del tipo de persona que estudia Oxoby se espera que cumpla con un ideal de trabajador y que gane un ingreso a través de su empleo.

ϵ_j es el mismatch entre la categoría asignada y las características específicas del individuo; en el caso de no encajar dentro de la categoría asignada de trabajador promedio por pertenecer a una casta inferior ó grupo discriminado, sufrirá una desutilidad por identidad. Por lo tanto a mayor ϵ_j , menor status logrará el individuo por vía pecuniaria y por lo tanto más inclinado estará a lograr status proveniente del ocio, dedicándole menos esfuerzo al trabajo, lo que significa que la actitud óptima frente al trabajo tenderá a cero $\lambda^*(a_j) \rightarrow 0$.

Además $\gamma = \bar{\gamma}$ el costo del individuo por cambiar su actitud hacia el trabajo de *lowerclass* a *underclass* es invariante en el corto plazo. Con las consideraciones anteriores quedan definidas las dos posibles identidades.

$$I_j = I_j(a_j = \textit{esfuerzo en el trabajo}, \bar{a}_{-j}, \bar{c}_j, \bar{\epsilon}_j, \bar{P}) = \textit{"lowerclass"}$$

$$I_j = I_j(a_j = \textit{holgazanería ó delito}, \bar{a}_{-j}, \bar{c}_j, \bar{\epsilon}_j, \bar{P}) = \textit{"underclass"}$$

La identidad I_j del individuo es una decisión racional que tiene en cuenta el costo γ por cambiar su actitud respecto al ocio, y los beneficios ó pérdidas pecuniarios y de status en que ese cambio de actitud le redundarán.

Oxoby limita su modelo al plano subjetivo de la elección de identificación del individuo, con qué grupo prefiere identificarse, dejando el funcionamiento del modelo en el ámbito individual y descartando explícitamente los costos que imponen otros sobre el individuo. El análisis de Oxoby se “abstrae de cualquier castigo social o presión de los pares que pueda existir.” (Oxoby, 2004, pág. 733, traducción del autor.) Es decir que en el modelo de Oxoby se descarta la externalidad provocada por el entorno social, tanto de los individuos que son pares del mismo grupo como de los individuos que no son pares, el resto de la sociedad.

A diferencia de Oxoby, el presente trabajo le dá a la dimensión social y al costo impuesto por los otros en el individuo un rol central. El individuo elige su identidad no sólo por un mecanismo de autoestima; si nadie relevante observa su conducta entonces el individuo sufrirá una menor pérdida de utilidad por identidad.

La segunda sección del trabajo presenta la función de utilidad con un componente pecuniario y otro componente que lo determina la decisión de identidad endógena del individuo; en la sección tercera se desarrolla el concepto de disonancia cognitiva, el motor

psicológico detrás de la decisión personal de identidad en el modelo y también se discuten los determinantes exógenos de la identidad; en la sección cuarta se exploran los tipos de comportamiento que pueden explicarse al introducir la identidad al análisis de discriminación; en la sección quinta se aborda la metodología empírica para evaluar las hipótesis estudiadas y en la sección sexta se presentan los resultados, que se detallan en tablas y en un gráfico.

III. Función de utilidad general con Identidad.

La función de utilidad del individuo j perteneciente al grupo k es aditiva y separable entre un componente pecuniario y uno no pecuniario:

$$U_{jk} = \textit{Utilidad por ingreso monetario} + \textit{Utilidad por factores no pecuniarios}$$

Cada componente se detalla aplicando la teoría de la Economía de la Discriminación de Gary Becker (1957) en el primer término y la Economía de la Identidad de Akerlof y Kranton (2000) en el segundo término. De esta forma se tiene una ecuación más completa que conjuga los mecanismos externos que vinculan el bienestar económico con significantes de identidad social y los mecanismos internos de disminución de la disonancia cognitiva propuestos por Akerlof y Kranton, tal como Iversen et al define. (2014)

$$U_{jk} = [(\textit{salario de mercado}_j \times \textit{esfuerzo}_j) \times (1 + \textit{CD}_k)] + [\delta_j \times \epsilon_j \times \textit{presión_social}_k]$$

donde:

$\textit{esfuerzo}_j$ es la cantidad de horas que el individuo j dedica al trabajo remunerado con el $\textit{salario de mercado}$, de acuerdo a sus dotaciones de capital humano y de tiempo.

\textit{CD}_k es el coeficiente de discriminación de Becker consolidado de empleadores, empleados, el gobierno y los sindicatos sobre el grupo k al que pertenece j .

δ_j es el valor relativo de la utilidad por identidad en la función de utilidad general, similar al valor relativo del status presentado por Oxoby (Cognitive Dissonance, Status and Growth of the Underclass, 2004, pág. 735). Este parámetro captura la heterogeneidad de los individuos dentro de su grupo.

ϵ_j es la medida del *mismatch* de Akerlof y Kranton entre las características personales de j y aquellas propias de la categoría social de su propio grupo; en este trabajo ϵ_j se reduce a cuánto se apega el comportamiento del individuo j a la “*Ética del trabajo.*”

$\textit{presión_social}_k$ es la medida de la intensidad de la mirada del otro perteneciente al mismo grupo k sobre el individuo j , expresa el “*neighborhood effect*” en el modelo.

La elección de identidad del individuo es binaria; o bien elige ser un “*insider*” o bien adopta una identidad “*outsider*”, lo cual intencionalmente descarta identidades intermedias.

Adhiriendo a la posición de Akerlof y Kranton, los individuos del grupo hegemónico son “*insiders*” por definición (Akerlof & Kranton, 2010, pág. 103) y δ_j es siempre igual a cero, no existiendo incongruencia, sin importar lo que hagan. Por el contrario, los individuos de las minorías deben tomar una opción clara de adopción de la norma social mayoritaria para convertirse en “*insiders*”; la acción que los convierte en “*insiders*” es ejercer esfuerzo en un empleo, incorporándose al mercado de trabajo, siguiendo la norma social hegemónica, en cuyo caso $esfuerzo_j > 0$ y $\delta_j = 1$; en este caso opera el segundo término de la función de utilidad total.

Para el individuo promedio de una minoría discriminada, δ_j es esperable que sea negativo ya que ejercer esfuerzo vía participación en la fuerza de trabajo implica comportarse como un individuo blanco, “*acting white*” (Austen-Smith & Fryer, 2005), pero también implica ser visto por otros de su propio grupo como acomodaticio a la posición de inferioridad social que se le asigna a la minoría discriminada, lo que conlleva un sufrimiento para el individuo, resultado de experimentar disonancia cognitiva (Akerlof & Kranton, 2010, pág. 104).

Alternativamente, para algunos grupos aventajados δ_j es esperable que sea positivo, resultado de un “*feedback effect*” de ganancia de utilidad por identidad que el individuo verifica al participar en el mercado laboral y que le confirma su propia superioridad. Aún cuando una persona del grupo aventajado no sea particularmente entusiasta por dedicar su tiempo a una carrera laboral, ella “preferirá que sus elecciones sean confirmadas por sus pares [...] apartarse le conllevará desaprobación y ostracismo, si sus pares han hecho diferentes elecciones de identidad.” (Akerlof & Kranton, 2010, p. 104, traducción del autor). Los individuos obtienen más estima de sus pares cuando se conforman, adaptan, a su elección de valores; en el caso de las personas aventajadas, conformarse también significa mayor utilidad por el premium, el coeficiente de discriminación mayor que cero, en el componente pecuniario del ingreso. (Akerlof R. J., Value Formation: The Role of Esteem, 2016)

En este modelo simplificado para que exista utilidad por identidad la “*presión_social_k*” debe ser mayor que cero, la conducta del individuo debe estar expuesta a la mirada

de sus pares, otros del mismo grupo, siendo esta una externalidad, más específicamente un “*neighborhood effect*” económico.

IV. Disonancia cognitiva y determinantes exógenos de la identidad.

Como se mencionó en el prefacio, al incorporar la utilidad por identidad en el modelo se busca explicar situaciones de desarrollo personal fallido aún más marcadas que las explicadas por la discriminación por gustos. Comportarse de acuerdo a la norma mayoritaria es visto como un intento de integrarse al cuerpo social hegemónico, y ese esfuerzo afecta la identificación del sujeto con su grupo de pertenencia, pudiendo resultar para el individuo en una pérdida de utilidad como resultado de su pérdida de identificación, que en este trabajo se usa como sinónimo de pérdida de identidad.

Para la teoría de la disonancia cognitiva la persona es sensible al hecho de que los propios actos sean contrarios a las normas y creencias propias, la persona sufre por adoptar un comportamiento que da la espalda a las normas de identidad de pertenencia a su grupo; el reconocimiento de esta disonancia motivará al individuo a tomar acciones para resolverla, ya sea cambiando su marco de valores y creencias o cambiando sus acciones o, más sutilmente, cambiando la percepción de sus acciones (Festinger, *A Theory of Cognitive Dissonance*, 1957) (Akerlof & Dickens, *The Economic Consequences of Cognitive Dissonance*, 1982). La pérdida de utilidad por identidad proviene de la disonancia cognitiva entre las decisiones de comportamiento del individuo y las normas de pertenencia al grupo, pero también proviene de actitudes de “escarmiento” externo por parte de otros miembros de la propia comunidad. Escarmiento por pretender ser distinto, querer sobresalir, por ser orgulloso y aspirar a ser una persona por encima de sus pares del grupo desaventajado, con un comportamiento propio del grupo hegemónico. Pero en el caso del grupo aventajado el escarmiento puede provenir por no estar a la altura, no adherirse a aquellas normas que confirman la superioridad del propio grupo.

Esta pérdida de utilidad por escarmiento se verá atenuada en la medida en que el individuo que intenta cumplir la norma hegemónica lo haga fuera del alcance de la vista de otros individuos de su comunidad. La pérdida de utilidad será mayor bajo la mirada reprobadora de su comunidad que verá al individuo como alguien comportándose conforme a lo que demanda el grupo social hegemónico, aquel que lo discrimina; por ejemplo para la minoría afroamericana, en el entorno del “*inner city*” de Washington D.C. es normal que una persona pobre de raza negra encuentre de sus pares una fuerte presión contra

su esfuerzo de acumular capital humano completar un nivel de educación medio como se relata en la etnografía de Ronald Suskin (*A Hope in the Unseen: An American Odyssey from the Inner City to the Ivy League*, 1998). En este último caso, el individuo se arriesga a ser en un cipayo, traidor, vendido y, por lo tanto, rechazado por su comunidad por no seguir las normas de su identidad de grupo, y además a ser discriminado por el grupo hegemónico que lo considera un inferior con aspiraciones a comportarse como un igual, pero que nunca podrá serlo.

Los gustos por discriminación de Becker se consideran exógenos y fijos en el corto plazo y están relacionados a categorías sociales asignadas y prescripciones sociales, que también son exógenas en el modelo.

“En un modelo de Economía de la Identidad, en el corto plazo las personas eligen sus acciones, como el esfuerzo que le dedican al trabajo, tomando como dadas las normas y categorías sociales así como también sus identidades y las de los otros, ya sea en la escuela o en la empresa. En el mediano plazo, los individuos tienen la posibilidad de elegir sus identidades y tomar acciones para cambiar las categorías y las normas en un contexto limitado.

En el largo plazo, nada está fijo ni puede tomarse como dado. Las acciones de las personas [...] afectan las normas y las categorías. Las normas sociales y las divisiones sociales son endógenas.” (Kranton, 2016, traducción del autor.)

V. Discriminación y comportamiento concomitante de la minoría.

Becker en su disertación doctoral plantea tres alternativas para el grupo discriminado. La primera es adoptar un comportamiento de represalia, discriminando a quienes lo discriminan. (Becker G. S., *The Economics of Discrimination*, 1957, pág. 22). Este comportamiento dañaría primeramente al propio grupo ya que reduciría aún más los beneficios de interactuar económicamente con el grupo mayoritario, que suele ser aquel que cuenta con mayor dotación de capital.

La segunda alternativa es la segregación; este comportamiento puede remediar la discriminación entre empleados, permitiendo reducir el costo de emplear a personas de la minoría discriminada, siempre que no exista discriminación desde los empleadores. En este caso las firmas se integran en gran parte por empleados de un mismo grupo a quienes se les puede pagar sueldos superiores que los que conseguirían en el mercado, sin ser perjudicados por el coeficiente de discriminación. Este comportamiento también puede

verse en el caso de profesionales (por ejemplo abogados, contadores y médicos) que buscan clientes primordialmente de su propia minoría.

Casos extremos de segregacionismo como el movimiento “*Back to Africa*” o el “*Apartheid*” sudafricano resultaron muy lesivos para el grupo discriminado. (Becker G. S., *The Economics of Discrimination*, 1957, pág. 22).

La tercera alternativa es aceptar, soportar, resignarse a las condiciones de mercado desfavorables; este comportamiento que resulta en el retorno pecuniario reducido por el coeficiente de discriminación, pero más alto que las opciones anteriores (Becker G. S., *The Economics of Discrimination*, 1957).

Akerlof & Kranton introducen una cuarta alternativa: la elección entre una identidad de “*insider*” y una de “*outsider*” (2010, págs. 102-104). Los “*outsiders*” se autoexcluyen del mercado de trabajo y soportan un costo de oportunidad pecuniario igual al salario que dejan de percibir. Los “*insiders*” aceptan las condiciones del mercado de trabajo, participan de él y perciben un salario de mercado reducido por el efecto de la discriminación por gustos explicada por Becker. Adicionalmente soportan una pérdida de utilidad por Identidad que es una externalidad negativa, resultado de la falta de autorrespeto de parte de ellos mismos y de otros. Los “*outsiders*” son menos sumisos, no se acomodan a lo que se espera de ellos y no sufren; se ahorran esta falta de autorrespeto que es sinónimo de pérdida de utilidad por identidad. (Akerlof & Kranton, 2010, pág. 103)

La elección de Identidad es un comportamiento racional que considera la utilidad general del individuo, no sólo la resultante de la maximización pecuniaria, y que incluye la reducción consciente o inconsciente de la disonancia cognitiva.

VI.Hipótesis a validar.

El comportamiento de autoexclusión estudiado necesita de un contexto social para que pueda darse la externalidad negativa. La forma de controlar por pérdida de utilidad por identidad es incluir en el modelo la presión social del grupo relevante, usando el porcentaje del propio grupo en el total de la población del área donde vive el individuo como variable proxy de la presión social de los pares, el “*neighborhood effect*” en la ecuación de selección.

El a_j en la ecuación de utilidad individual con identidad de Akerlof y Kranton (Akerlof & Kranton, *Economics and Identity*, 2000) incluye las acciones de

discriminación de otros grupos, así como los sentimientos de desprecio por parte de pares del mismo grupo. La acción a_j es la participación en el mercado laboral.

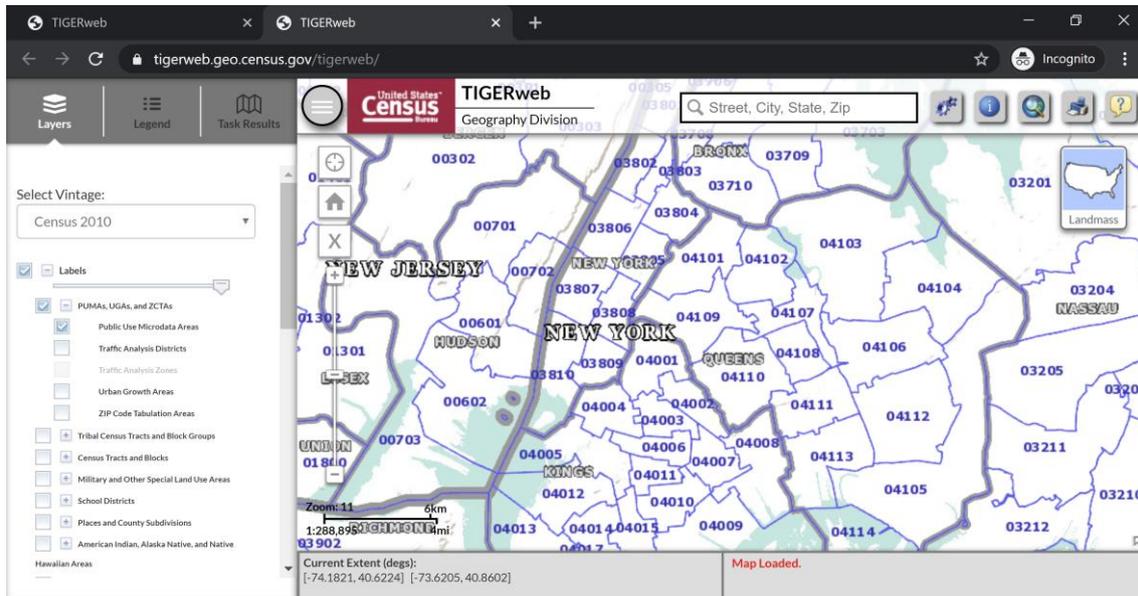
Condición necesaria para que la teoría de la Economía de la Identidad sea válida es que se verifiquen “*feedback effects*” (Akerlof & Kranton, 2010, pág. 105): que la persona se comporte de modo no indiferente entre ignorar abiertamente las normas de identidad del propio grupo, es decir a plena vista de otros de su minoría y seguir el comportamiento que de él esperan sus pares. La hipótesis nula consiste en que la identidad es irrelevante; la hipótesis alternativa consiste en que la identidad importa en la toma de decisiones individual. Si no hay afectación de la decisión individual por el grado de exposición del individuo a la mirada del otro, la externalidad, entonces no se puede rechazar la hipótesis nula de que la identidad del individuo en lo relativo con su grupo de pertenencia es irrelevante en la toma de decisiones.

El grado de identificación de un individuo con un grupo es una característica subjetiva difícil de observar directamente. Pero la decisión del individuo de integrarse o de autoexcluirse socialmente puede observarse por su decisión de participar en el mercado laboral. Para implementar empíricamente esta medición necesitamos una variable de control que afecte a la utilidad por identidad y también contar con supuestos sobre las normas específicas de identificación particular de cada grupo, para poder interpretar los resultados en el sentido de una mayor autoexclusión o mayor integración.

La probabilidad de participar en el mercado laboral del individuo promedio es un parámetro que varía entre grupos siendo la mirada del otro, efecto de vecindad, o presión social de su mismo grupo un factor determinante; la decisión de identidad entonces se manifiesta en la decisión de participación o de autoexclusión. El coeficiente de interés es la tasa de cambio en la probabilidad de participar del mercado laboral con el cambio en el porcentaje del grupo relevante en el área geográfica donde se encuentra el individuo.

Las áreas geográficas empleadas para el análisis son las PUMAs, acrónimo de “*Public Use Microdata Areas*”, empleadas en la ACS (United States Census Bureau, s.f.), cuyos bordes pueden consultarse en <https://tigerweb.geo.census.gov/tigerweb/>.

Figura 1 - Ejemplo de áreas PUMAs empleadas en la encuesta ACS para agrupar los datos PUMS usados en las regresiones.



VII. Análisis Empírico.

El capital humano que posee un individuo tiene su retorno en la forma de ingreso obtenido por el trabajo. La regresión de Mincer en su forma tradicional permite calcular retornos promedios al capital humano modelizado por años de educación y aprendizaje en el trabajo (Mincer J. A., 1974) (Heckman, Lochner, & Todd, Earnings functions and rates of return, 2008).

Con la descomposición de Oaxaca-Blinder (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) se explica el diferencial de ingresos entre diferencia de dotaciones de capital humano, por años de educación formal y de aprendizaje en el trabajo, y diferencia de coeficientes atribuible a discriminación pura, asociable al coeficiente de discriminación de Becker. También puede calcularse el costo pecuniario de la discriminación en unidades monetarias para el individuo promedio.

Con el método de dos pasos de Heckman (Heckman J. J., Sample selection bias as a specification error, 1979) se corrige el sesgo de selección en los retornos de la ecuación de Mincer al tener en cuenta las observaciones censuradas de individuos que se autoexcluyen del mercado de trabajo de acuerdo a la ecuación de selección especificada bajo el supuesto que sigue una distribución Probit.

Por último, el análisis de los efectos marginales de la regresión Probit de la primera etapa provee los efectos dF/dx que son las tasas de aumento en la probabilidad de participar del mercado laboral con el aumento del porcentaje del grupo de interés en el área.

Para los diferentes grupos unívocamente identificados en la encuesta ACS se empleará la regresión de Heckman con ecuación de selección Probit donde se incluye como regresor el porcentaje de la población del grupo de pertenencia del individuo para capturar el “*neighborhood effect*” sobre la decisión del individuo; también se incluye el porcentaje del grupo mayoritario de raza blanca, como variable proxy de la pujanza económica del área donde se encuentra el individuo. Los demás regresores de participación son años de educación, número de niños en el hogar, y variables categóricas para identificar la condición de discapacidad, la condición de retirado, si el idioma empleado en el hogar es el inglés, si la persona cambió de estado civil de soltero a casado u otro y la ordenada al origen (PUMS Technical Documentation). Como se realizan regresiones separadas para cada género, no es necesario el empleo de la variable categórica sexo.

En la regresión de retorno al capital humano la formulación es la típica de Mincer con aprendizaje en el trabajo; los regresores son años de educación, experiencia, el cuadrado de la experiencia y la ordenada al origen. Las regresiones se calculan para cada género por separado, en lugar de utilizar la variable categórica de género en una misma regresión, lo que permite obtener los parámetros de procesos generadores de ingreso específicos por sexo y que también es preferible para realizar la descomposición de Oaxaca-Blinder, tal como lo explica Myoung-Jae (Lee, Reference parameters in Blinder-Oaxaca decomposition: Pooled-sample versus intercept-shift approaches, 2015).

VIII.Resultados.

Los resultados son consistentes con estudios de retorno a la educación que emplean regresiones de Mincer. (Psacharopoulos, 1981) (Psacharopoulos, 1993) (Psacharopoulos & Patrinos, 2004). A modo de ejemplo, para los varones blancos el retorno promedio por año de educación es de 12,62%, el coeficiente de experiencia es de 10,28% por año de antigüedad en el trabajo y el coeficiente de experiencia al cuadrado es de -0,001651 y para las mujeres blancas es de 15,56%, 8,03% y -0,00131 respectivamente.

En las Tablas 2 y 3 se presentan los principales estadísticos, como el tamaño del grupo en la muestra, su salario ajustado promedio y qué porcentaje de personas del grupo en la muestra no participan del mercado laboral. También se muestran los efectos

pecuniarios de la discriminación y el cambio en la probabilidad de participar en el mercado laboral, dF/dx , con el porcentaje del grupo en la PUMA para el individuo promedio. La Tabla 2 corresponde a varones y la Tabla 3 a mujeres. Los estimadores dF/dx resultaron estadísticamente significativos con intervalos de confianza al 95% para todos los grupos, menos para los vietnamitas varones y las coreanas mujeres, en cuyos casos no se puede rechazar la hipótesis nula de que el porcentaje de individuos del mismo grupo (vietnamitas y coreanos respectivamente) es irrelevante en su decisión de participar en el mercado laboral.

Tabla 2 - Regresión de Heckman - estadísticos seleccionados para hombres (males) de 16 años o más.

	<i>Salario ajustado promedio para el grupo</i>	<i>Salario ajustado promedio para el resto</i>	<i>dF/dx</i>	<i>Tamaño del grupo en la muestra</i>	<i>% fuera de la fuerza de trabajo</i>	<i>Efecto pecuniario del CD</i>
Chinese_m	\$78.883,15	\$56.308,00	2,24454 (.1631346)	36.898	33,3%	\$ 13.835
Asian Indians_m	\$99.087,86	\$56.206,84	4,82854 (.3610225)	26.046	24,1%	\$ 16.001
Vietnamese_m	\$53.066,32	\$56.520,35	0,50664 † (.4698296)	17.986	32,1%	-\$ 3.227
Koreans_m	\$73.191,94	\$56.459,89	3,47131 (.7537362)	11.137	31,3%	\$ 6.810
Filipinos_m	\$53.086,36	\$56.529,44	1,49168 (.1430819)	28.088	29,6%	-\$ 13.812
Japanese_m	\$72.235,79	\$56.476,34	- 1,41663 (.1266977)	8.562	39,3%	\$ 5.172
Other Asians_m	\$53.125,71	\$56.544,55	0,58325 (.0464156)	44.609	28,3%	-\$ 1.443
Amer Indians_m	\$38.617,38	\$57.908,69	- 0,20088 (.0168425)	72.732	42,2%	-\$ 12.214
Blacks_m	\$38.224,86	\$59.456,79	- 0,11392 (.0112)	381.750	40,2%	-\$ 16.284
Iberoamericans_m	\$41.695,81	\$57.997,68	- 0,37052 (.0081524)	377.019	31,3%	-\$ 5.560
Whites_m	\$59.790,23	\$46.069,57	0,16349 (.0023879)	3.320.445	34,7%	\$ 8.337
Immigrants_m	\$43.552,67	\$57.588,05	0,69329 (.0096321)	278.744	20,7%	-\$ 7.031

† para Vietnamese_m el estimador dF/dx no es estadísticamente distinto de cero. dF/dx es la tasa de cambio de la variable participación en el mercado laboral “*in_labor_force*” respecto al porcentaje del grupo de referencia en el área PUMA, obtenida con el comando `dprobit`.

Los inmigrantes varones (`immigrants_m`) fueron identificados por ser individuos que no cuentan con ciudadanía, ningún otro grupo racial incluye personas inmigrantes.

Para los grupos de Indios Americanos (`Amer Indians_m`), Negros (`Blacks_m`), Iberoamericanos (`Iberoamericans_m`) y Blancos (`Whites_m`), la descomposición de Oaxaca Blinder se realizó entre los miembros de un grupo dado, excluyendo inmigrantes, y el resto de los individuos en la muestra, excluyendo inmigrantes. Para los demás grupos se mantuvo a los inmigrantes en el grupo de comparación.

Tabla 3 - Regresión de Heckman - estadísticos seleccionados para mujeres (*females*) de 16 años o más.

	<i>Salario ajustado promedio para el grupo</i>	<i>Salario ajustado promedio para el resto</i>	<i>dF/dx</i>	<i>Tamaño del grupo en la muestra</i>	<i>% fuera de la fuerza de trabajo</i>	<i>Efecto pecuniario del CD</i>
Chinese_f	\$57.026,26	\$36.737,83	1,79329 (.1662381)	44.072	40,6%	\$ 16.403
Asian Indians_f	\$58.709,79	\$36.807,55	8,05538 (.4380634)	25.740	38,9%	\$ 12.810
Vietnamese_f	\$39.019,92	\$36.928,36	1,13117 (.5550414)	19.517	39,1%	-\$ 2.092
Koreans_f	\$50.164,92	\$36.893,56	- 0,36469† (.6592693)	16.044	45,3%	\$ 10.727
Filipinos_f	\$49.817,67	\$36.816,77	1,25290 (.1365957)	38.125	34,2%	\$ 6.082
Japanese_f	\$52.318,55	\$36.907,26	- 2,06620 (.1121218)	11.453	54,4%	\$ 9.653
Other Asians_f	\$37.983,15	\$36.925,46	0,98579 (.0557324)	48.470	37,3%	-\$ 3.770
Amer Indians_f	\$29.121,36	\$37.581,21	- 0,12395 (.0173809)	82.882	48,8%	-\$ 3.950
Blacks_f	\$32.220,52	\$38.118,92	- 0,08184 (.0104875)	508.370	42,6%	-\$ 3.901
Iberoamericans_f	\$30.390,49	\$37.617,39	- 0,07723 (.0088697)	432.080	42,3%	-\$ 953
Whites_f	\$37.863,57	\$35.598,70	0,01185 (.0025025)	3.678.621	45,4%	\$ 544
Immigrants_f	\$28.397,06	\$37.434,61	0,19243 (.0125852)	293.363	50,9%	-\$ 665

† para Korean_f el estimador dF/dx no es estadísticamente distinto de cero. dF/dx es la tasa de cambio de la variable participación en el mercado laboral “*in_labor_force*” respecto al porcentaje del grupo de referencia en el área PUMA, obtenida con el comando dprobit.

Las inmigrantes mujeres (*immigrants_f*) fueron identificadas por ser individuos que no cuentan con ciudadanía, ningún otro grupo racial incluye personas inmigrantes.

Para los grupos de Indios Americanos (*Amer Indians_f*), Negros (*Blacks_f*), Iberoamericanos (*Iberoamericans_f*) y Blancos (*Whites_f*), la descomposición de Oaxaca Blinder se realizó entre los miembros de un grupo dado, excluyendo inmigrantes, y el resto de los individuos en la muestra, excluyendo inmigrantes. Para los demás grupos se mantuvo a los inmigrantes en el grupo de comparación.

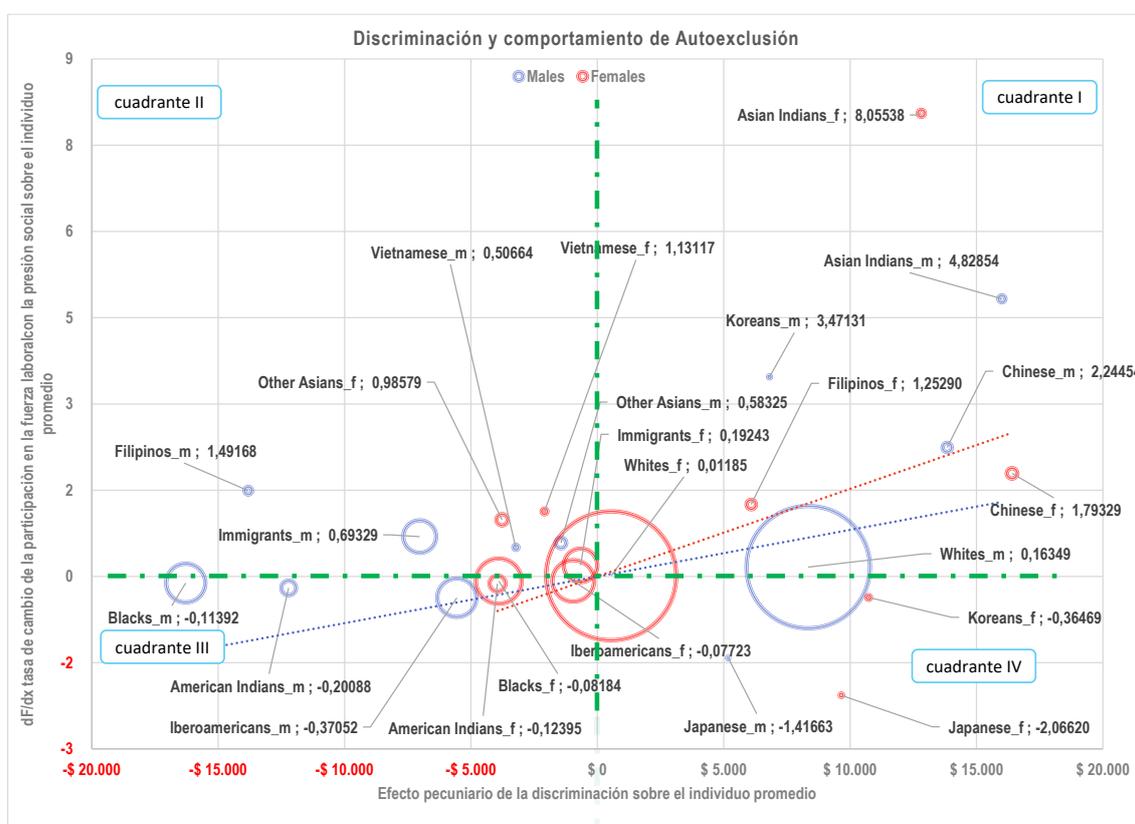
Efecto pecuniario del CD se obtiene multiplicando la diferencia entre el Salario ajustado promedio para el grupo y el Salario ajustado promedio para el resto y la fracción que la descomposición de Oaxaca-Blinder explica como atribuible a discriminación..

Efecto pecuniario del coeficiente de discriminación (CD) se obtiene multiplicando la diferencia entre el salario ajustado promedio para el grupo y el salario ajustado promedio para el resto y la fracción que la descomposición de Oaxaca-Blinder explica como atribuible a discriminación.

Con los datos de las tablas se construye el Gráfico 1, donde se muestra la principal relación del modelo: si existe un perjuicio (beneficio) pecuniario por discriminación, entonces los diferentes grupos mostrarán una menor (mayor) disponibilidad a participar en el mercado de trabajo en la medida que estén más expuestos a la mirada del otro: el

individuo promedio experimenta un “neighborhood effect” y eso influye en su decisión de participar del mercado de trabajo. Notoriamente para las minorías de Afroamericanos, Aborígenes Americanos e Iberoamericanos, que son un 18,8% de la población en la muestra. Todos estos grupos sufren discriminación efectiva en el mercado de trabajo, que en promedio es de \$11.065 para los hombres y de \$2.660 para las mujeres. Todos estos grupos se comportan de acuerdo a lo previsto por el modelo, con un $\delta < 0$ ya que son discriminados.

Gráfico 1 - Comportamiento de autoexclusión/autoinclusión en función del perjuicio/beneficio pecuniario resultante de la discriminación/nepotismo.



Nota: Para cada grupo, el área del círculo que lo representa es proporcional a su tamaño en la muestra y el centro del círculo es el punto de coordenadas (efecto pecuniario de la discriminación relativo al promedio; dF/dx). Las líneas de trazos rectas tienen la pendiente de la recta de mejor ajuste de los datos de las dos tablas, por sexo, y son sólo a título ilustrativo.

Los grupos del cuadrante I son los privilegiados que muestran autoinclusión. Los grupos del cuadrante II son los discriminados que muestran autoinclusión. Los grupos del cuadrante III son los discriminados que muestran autoexclusión. Finalmente los grupos del cuadrante IV son los grupos privilegiados que muestran autoexclusión.

De gran interés es el caso de los inmigrantes que cubren un 5,8% de la muestra y que claramente eligen asimilarse y conformarse al sistema aun cuando por discriminación

los varones reciben \$7,031 menos de lo que les correspondería por su capital humano. Los inmigrantes, aun contando en promedio con claras características de “*outsiders*” como su menor dominio del idioma inglés y su acento no nativo, sus antecedentes académicos y sus pocas o nulas referencias laborales, eligen claramente una identidad “*insider*”. El valor +0,69 del parámetro dF/dx para los inmigrantes varones es atribuible a que los inmigrantes se concentran donde hay más chances de trabajo para ellos, bajo el supuesto que la decisión de inmigrar se basa en las chances de mayores ingresos. Pero para el caso de los inmigrantes, dada su enorme variedad étnica, de lengua materna, de país de origen, su condición de no ciudadanos de los Estados Unidos y su reducido tiempo en el país (se entiende que con el tiempo aumentan las chances de un inmigrante de cambiar de status migratorio, obtener la ciudadanía y pasar a integrar uno de los otros grupos), ellos carecen de un grupo de pares con una identidad común capaz de influir sobre ellos y presionarlos a permanecer como “*outsiders*”. Para estos grupos $presión_social_k=0$ por lo que no sufren pérdida de utilidad por identidad.

Además de los varones blancos, que son “*insiders*” por definición, otros grupos privilegiados son los indoasiáticos, los chinos, los varones coreanos y las mujeres filipinas; todos ellos muestran mayor predisposición a participar del mercado laboral bajo el “*neighborhood effect*” de un mayor porcentaje de su grupo de pertenencia en su área geográfica. Lo cual también es un resultado esperado del modelo, con un $\delta > 0$ ya que son privilegiados.

Los varones filipinos reciben un ingreso en promedio \$13.812 menor atribuible a discriminación, son un grupo de ciudadanos que cubren el 0,7% de la población; este grupo es una anomalía ya que muestra una actitud conformista ($\delta > 0$) aún siendo fuertemente discriminado. Otra anomalía se encuentra entre los japoneses, que son un 0,2% de la muestra; los japoneses, siendo un grupo claramente privilegiado en su ingreso pecuniario, elige un comportamiento de autoexclusión ($\delta < 0$).

Los fenómenos de más fuerte discriminación y marcada autoexclusión se dan entre grupos de varones. Los varones blancos reciben en promedio \$8.337 más que lo que les correspondería por su capital humano, y ese valor para las mujeres es sólo de \$544 por año, lo cual es mucho menos perceptible para la persona. Para iberoamericanos, indoamericanos y negros el comportamiento de autoexclusión es también mucho más marcado entre varones que entre mujeres.

IX. Conclusiones.

El grupo de raza blanca abarca el 71% de la población en la muestra con 16 o más años; este grupo fija las normas de identidad y siempre son “*insiders*.”

El modelo de utilidad con identidad se comprueba en 9 de cada 10 individuos que no son de raza blanca:

- El 20,5% sigue la regla de mayor auto exclusión de la fuerza laboral como comportamiento concomitante a su realidad de grupo discriminado ($\bar{\delta}_j < 0$), y mayor auto inclusión siguiendo su buena fortuna de disfrutar un mayor sueldo que el que corresponde a su capital humano acumulado ($\bar{\delta}_j > 0$), existiendo la externalidad de **presión_social_k>0**.
- El 5,8% de la muestra son los inmigrantes, mayoritariamente hombres, que siendo claramente discriminados adoptan una identidad de “*insiders*”. En el caso de inmigrantes varones estos muestran la mayor predisposición a trabajar con sólo un 20,7% del grupo no participando del mercado laboral, el porcentaje más bajo de todos los grupos. El comportamiento de los inmigrantes de fuerte adopción de la ética del trabajo del país que los recibe subraya la importancia del “*neighborhood effect*” en la decisión de identidad de un grupo discriminado, como caso emblemático de lo que ocurre cuando no existe la presión de la mirada del otro sobre el individuo.
- Para un 2,3% de la muestra se verifica un comportamiento no explicado por el modelo de utilidad con identidad, ya que son situaciones de grupos que se auto-excluyen aún ganando por encima de lo que explica su capital humano, o situaciones de auto-inclusión concomitante a una realidad de discriminación.

3. Aplicación del modelo de Economía de la Identidad a la decisión de sufragar.

I. Resumen

El objetivo de este trabajo es aplicar la teoría de la Economía de la Identidad (Akerlof & Kranton, Economics and Identity, 2000) a la decisión de los ciudadanos de participar en el proceso político de elección de autoridades federales en Estados Unidos. Se identifican cinco grupos con un común denominador étnico y se analiza cómo influyen diversas variables en la probabilidad de que una persona se registre como votante, siendo una de estas la variable que captura la presión social del mismo grupo sobre el comportamiento de adhesión a normas identitarias del individuo.

Al existir otra variable negativamente correlacionada con la participación en las elecciones que es la autoexclusión del mercado laboral, también explicada por la decisión de identidad de la persona, se emplea una regresión biprobit especificada como “*seemingly unrelated bivariate probit regression*” que permite evaluar en Stata la validez del modelo de comportamiento.

Identificarse con el propio grupo en un cierto aspecto y darle la espalda en otro aspecto no es un comportamiento de costo cero, sino que conlleva una pérdida de utilidad para el individuo resultado de tener un comportamiento no íntegro, no honesto, no “*real*”, algo que podría considerarse una falsedad. Este aspecto de coherencia entre las decisiones identitarias se analiza observando la correlación de los residuos en el modelo bivariado.

En la Introducción se presenta el objeto de estudio en el marco de la Economía Laboral y de la Economía Política, y se introducen los conceptos de Economía de la Identidad. En el Modelo Económico se plantea una ecuación de utilidad con todos los componentes que explican la decisión de la persona endogeneizando la elección de identidad. El Modelo Económico evalúa la validez del modelo Económico con los datos del Census Bureau para diferentes grupos. Al final del trabajo se modeliza la decisión identitaria con un modelo de juego secuencial, primero de la decisión de participación política, y luego inscribiendo esta última decisión como un subjuego de un juego más amplio que comienza por la decisión laboral.

II. Introducción.

La identidad de las personas, o más propiamente la decisión personal de identificación con un grupo (Akerlof R. J., *Value Formation: The Role of Esteem*, 2016), es una variable relevante en la Economía del Comportamiento ya que influye tanto en decisiones con resultado pecuniario directo (Economía Laboral) como en decisiones sin resultado pecuniario directo (Economía Política). La persona, como agente racional que busca maximizar su bienestar, elige entre sus prioridades privadas (aquellas que prefiere si está aislada, no sujeta a la mirada del otro) y otras alternativas que sí tienen en cuenta un “*neighborhood effect*” que en este caso es la utilidad por identidad.

Para la Economía Laboral el estudio del efecto de la discriminación sobre el ingreso de las minorías tuvo su trabajo seminal en la disertación doctoral de Gary Becker. (*The Economics of Discrimination*, 1957) Son varias las minorías que sufren discriminación laboral, medida por menores ingresos para igual puesto y calificación, siendo esta discriminación más marcada entre los diversos grupos de varones que entre los diversos grupos de mujeres, aun cuando la discriminación por género sigue verificándose en una brecha de ingresos entre mujeres y hombres que no declina con el tiempo. (Mincer J. A., 1974) (Psacharopoulos, 1981) (Psacharopoulos, 1993) (Psacharopoulos & Patrinos, 2004) Y es comprensible que individuos de los grupos étnicos sujetos a discriminación se rebelen frente a la posición de desventaja a la que se los relega. Esta rebeldía se manifiesta a nivel de grupo en un comportamiento de mayor autoexclusión del mercado laboral y conlleva una mirada reprobadora hacia el propio (afroamericano, aborigen, latino) que acepta la posición de desventaja, lo que peyorativamente se designa como “*acting white*”. (Austen-Smith & Fryer, 2005) Por otro lado, una minoría aventajada, denominada “*model minority*”, como es el caso de las etnias asiáticas, experimenta lo contrario, que Becker califica de nepotismo (Becker G. S., 1957, pág. 15), participando sus individuos del mercado laboral a tasas aún más altas en la medida en que el individuo esté expuesto a la presión social, la mirada de su propio grupo.

La Economía Política también presenta objetos de estudio donde es posible introducir la identidad como variable explicativa del comportamiento político del individuo; cada persona tiene gustos propios sobre si participar o no del proceso político, participación que requiere incurrir un costo individual en registrarse y luego en sufragar a pesar de lo irrelevante que es la participación individual en el resultado de una votación. Pero a nivel de grupo, la participación mayor o menor en una elección sí es relevante para

definir al ganador porque la inclinación de una minoría tiene el potencial de definir una elección. En la medida en que la decisión de registrarse, que es el primer paso necesario para eventualmente emitir el sufragio, cambia en la medida en que el individuo está más ó menos expuesto a la mirada de sus pares de grupo, hay un efecto de la identidad sobre el comportamiento político del individuo. Es un efecto de la presión social del grupo sobre la persona, el “*neighborhood effect*” que se comprueba incluso cuando el sufragio es secreto.

Desde el punto de vista de la teoría económica una persona puede modelizarse como un conjunto peculiar de gustos y preferencias; las decisiones de ocio, estudio, trabajo, consumo y ahorro se toman respetando restricciones presupuestarias y de saciedad, y principios como el de “más es mejor que menos”. (Becker G. S., 2009) Sin embargo la introducción de la variable identidad, como espejo contra el cual la persona juzga cada una de sus elecciones, relaciona estas diferentes decisiones, dado que lo que busca el individuo, desde el punto de vista de Psicología Social, es la reducción de la ansiedad, de la disonancia cognitiva, del sufrimiento experimentado por no adherirse plenamente al deber ser y el temor a las consecuencias inmediatas y futuras en que esto puede resultar. (Akerlof & Kranton, Economics and Identity, 2000) (Akerlof & Dickens, 1982) (Festinger, A Theory of Cognitive Dissonance., 1957)

Vale aclarar que el hecho de que cada persona tenga gustos y preferencias individuales y de que esa variedad sea un componente básico del modelo económico de este trabajo, se descarta la existencia de los estereotipos raciales; de los ciudadanos tomados individualmente y aislados de cualquier externalidad se espera que en promedio tengan la misma predisposición a participar del proceso político o del mercado laboral, independientemente de su raza.

En promedio, las diferencias en ingreso entre personas con igual nivel de capital humano se explican por discriminación por gustos o discriminación estadística y no porque las diferencias de género ó de fenotipo expliquen diferencias en la productividad individual. (Becker, 1993)

De forma similar, en promedio personas de cualquier grupo tendrán igual interés en participar en el proceso político, no explicándose las diferencias por características de género o fenotipo.

El deber ser son las normas y prescripciones exigidas por el ideal de identidad peculiar de cada grupo con un común denominador étnico. (Kranton, 2016) Algunos grupos pueden ser muy celosos respecto a la observancia de comportamientos religiosos,

políticos o laborales requeridos por su identidad de grupo, mientras que otros pueden ser muy laxos ó ignorar completamente algunos aspectos en la formación de la identidad. Cómo se generan esas identidades está fuera del alcance de este trabajo pero sí es relevante para el presente análisis el hecho de que los gustos personales puedan estar más o menos alineados con el deber ser del individuo, siendo que una norma puede estar siendo celosamente tutelada a nivel social y su cumplimiento obligado, o puede llegar al individuo como una tradición, como una costumbre, como lo que se considera el comportamiento apropiado, un punto de vista que de ningún modo esté soportado por la coerción de una obligación legal.

Las elecciones relativas a la participación en el mercado laboral y a la participación política no deben suponerse completamente independientes sino condicionales a la identidad elegida por el individuo. Identificarse con el propio grupo en un cierto aspecto y darle la espalda en otro no es un comportamiento de costo cero, sino que conlleva una pérdida de utilidad para el individuo resultado de tener un comportamiento no íntegro. (Akerlof & Kranton, 2010)

III. Propósito de la investigación.

¿Cuáles son las variables que explican la decisión de los ciudadanos de registrarse para sufragar? ¿Puede explicarse el comportamiento de participación electoral de un grupo con un común denominador étnico por la decisión del individuo en su elección de identidad? De ser así, ¿cómo se comporta esta variable de participación en elecciones conjuntamente con otra variable también plausible de ser explicada por una elección de identidad?

IV. Datos y variables.

La base de datos usada es la *Current Population Survey* (CPS) (United States Census Bureau, 2021) y sus suplementos sobre registración y voto entre los años 1976 y 2020. CPS cuenta con variables de identificación racial, participación en la fuerza laboral, estado civil, nivel de estudios alcanzado y condición de retirado. Para cada año par, la encuesta CPS incluye en el mes de noviembre un suplemento específico sobre registración y voto. (United States Census Bureau, 2020 y años previos)

En el diseño del suplemento sobre registración y voto (Current Population Survey, November 2020. Voting and Registration Supplement. Technical Documentation, 2020

y años previos, pp. 8-1) primero se le pregunta a la persona si es mayor de 18 años (los que no lo son se consideran “fuera del universo”), luego se pregunta si votó y sólo a aquellas que no votaron se les pregunta sobre registración. El dato de quién votó es de lectura directa, pero el dato de quién se registró surge de sumar los registros de quienes votaron y de quienes se registraron y no votaron. Así se genera la variable binaria que captura la decisión de registrarse en la última elección de noviembre.

Tabla 4 - Ciudadanos Registrados para la Votación y Fuera de la Fuerza de Trabajo.

Registrados para la votación del último noviembre	Frecuencia	Porcentaje	Acumulado
No registrados	674.659	30,52%	30,52%
Registrados	1.535.629	69,48%	100,00%
Total	2.210.288	100,00%	
En la fuerza de trabajo	Frecuencia	Porcentaje	Acumulado
En la fuerza de trabajo	1.690.869	76,50%	76,50%
Fuera de la fuerza de trabajo	519.419	23,50%	100,00%
Total	2.210.288	100,00%	

Nota: confeccionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

En toda la muestra considerada, entre 1976 y 2020 inclusive, un 69,48% de las personas se registraron para votar. Por otro lado, un 81% de las personas que se registraron responden que votaron. Estas cifras son coherentes con los análisis propios del Census Bureau. En la tabla 6 se muestran series históricas de porcentajes de registración y de voto abiertos por grupo étnico, para elecciones presidenciales (*Presidential*) y de medio término (*Congressional*).

Tabla 5 - Ciudadanos que efectivamente votaron.

Votaron en la elección de noviembre más reciente	No Registrados	Registrados	Total
No votaron	488.628 72,43%	276.318 17,99%	764.946 34,61%
Votaron	0 0,00%	1.243.520 80,98%	1.243.520 56,26%
Rehusaron a votar	14.942 2,21%	871 0,06%	15.813 0,72%
No saben	28.574 4,24%	13.945 0,91%	42.519 1,92%
No contestan	142.515 21,12%	975 0,06%	143.490 6,49%
Total	674.659 100,00%	1.535.629 100,00%	2.210.288 100,00%

Nota: confeccionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

Tabla 6 - Ciudadanos registrados por año.

Año de la encuesta	No Registrados	Registrados	Total
1976	29.454	59.627	89.081
1978	34.668	66.728	101.396
1980	34.962	84.065	119.027
1982	32.247	73.672	105.919
1984	29.066	76.001	105.067
1986	32.639	70.619	103.258
1988	36.380	70.088	106.468
1990	34.036	69.236	103.272
1992	26.214	73.504	99.718
1994	30.162	65.110	95.272
1996	24.110	60.424	84.534
1998	26.600	57.476	84.076
2000	24.325	58.103	82.428
2002	31.464	67.174	98.638
2004	25.642	70.810	96.452
2006	29.054	65.041	94.095
2008	25.660	66.700	92.360
2010	31.771	62.437	94.208
2012	26.144	68.167	94.311
2014	32.843	63.424	96.267
2016	26.913	66.881	93.794
2018	28.400	60.349	88.749
2020	21.905	59.993	81.898
Total	674.659	1.535.629	2.210.288

Nota: confeccionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

Tabla 7 - Registración y participación en la fuerza de trabajo por grupo étnico.

Grupo	Registración	En la fuerza de trabajo	Fuera de la fuerza de trabajo	Total
Hombres afroamericanos	No registrados	23.681 26,44%	10.200 11,39%	33.881 37,83%
	Registrados	44.499 49,68%	11.188 12,49%	55.687 62,17%
Mujeres afroamericanas	No registrados	23.526 19,25%	15.497 12,68%	39.023 31,93%
	Registrados	60.119 49,20%	23.059 18,87%	83.178 68,07%
Hombres aborígenes	No registrados	2.143 30,46%	904 12,85%	3.047 43,31%
	Registrados	3.177 45,15%	812 11,54%	3.989 56,69%
Mujeres aborígenes	No registrados	1.823 22,35%	1.179 14,45%	3.002 36,80%
	Registrados	3.691 45,24%	1.465 17,96%	5.156 63,20%
Hombre hispanos	No registrados	27.135 38,08%	6.630 9,30%	33.765 47,39%
	Registrados	32.442 45,53%	5.047 7,08%	37.489 52,61%
Mujeres hispanas	No registrados	21.582 26,77%	14.295 17,73%	35.877 44,50%
	Registrados	31.979 39,66%	12.775 15,84%	44.754 55,50%
Hombres asiáticos	No registrados	8.067 38,71%	1.637 7,85%	9.704 46,56%
	Registrados	10.057 48,26%	1.080 5,18%	11.137 53,44%
Mujeres asiáticas	No registrados	8.051 33,87%	2.791 11,74%	10.842 45,61%
	Registrados	10.743 45,19%	2.188 9,20%	12.931 54,39%
Hombres Blancos	No registrados	198.825 23,75%	44.061 5,26%	242.886 29,01%
	Registrados	510.807 61,02%	83.484 9,97%	594.291 70,99%
Mujeres Blancas	No registrados	163.397 17,77%	86.850 9,45%	250.247 27,22%
	Registrados	483.185 52,55%	186.075 20,24%	669.260 72,78%
Total		1.668.929	511.217	2.180.146

Nota: la tabla no incluye 30.142 registros correspondientes a personas que no se identifican ninguno de los grupos. Las minorías aborígenes y asiáticas comenzaron a reportarse separadamente en las encuestas a partir de 1986.

Confeccionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

También se emplean datos del nivel de segregación elaborado por *Othering and Belonging Institute* de *UC Berkeley*. (Othering and Belonging Institute, 2022) La definición de segregación empleada es la separación espacial de uno o más grupos de personas en base a su identidad de grupo (Oxford English Dictionary, 2021), y la métrica preferida y por defecto es el “*divergence index*” que provee un valor holístico del nivel de segregación de cualquier comunidad en Estados Unidos. (Othering and Belonging Institute, 2022) En esta etapa se emplean los datos reportados para las 250 regiones metropolitanas con más de 200,000 habitantes en 2019 (Other and Belonging Institute - Most to Least Segregated Metro Regions, 2022), que son las mismas áreas que emplea el Census Bureau en la CPS, y que se muestran en la Tabla 8 para los valores más altos y más bajos.

Tabla 8 – Áreas metropolitanas más segregadas y menos segregadas.

Metro	Divergence Index	Segregation Category
Chicago-Joliet-Naperville, IL-IN-WI	0,3985	High
New York-Northern New Jersey-Long Island, NY-NJ-PA	0,3919	High
Milwaukee-Waukesha-West Allis, WI	0,3504	High
Detroit-Warren-Livonia, MI	0,3381	High
Trenton-Ewing, NJ	0,3354	High
Miami-Fort Lauderdale-Pompano Beach, FL	0,3272	High
Los Angeles-Long Beach-Santa Ana, CA	0,3175	High
Cleveland-Elyria-Mentor, OH	0,302	High
Beaumont-Port Arthur, TX	0,2994	High
Atlanta-Sandy Springs-Marietta, GA	0,2923	High
Fort Collins-Loveland, CO	0,0244	Low-Medium
Eugene-Springfield, OR	0,0221	Low-Medium
Spokane, WA	0,0218	Low-Medium
Kingsport-Bristol-Bristol, TN-VA	0,0157	Low-Medium
Laredo, TX	0,011	Low-Medium
Palm Bay-Melbourne-Titusville, FL	0,00679	Integrated
San Luis Obispo-Paso Robles, CA	0,00619	Integrated
Jacksonville, NC	0,00586	Integrated

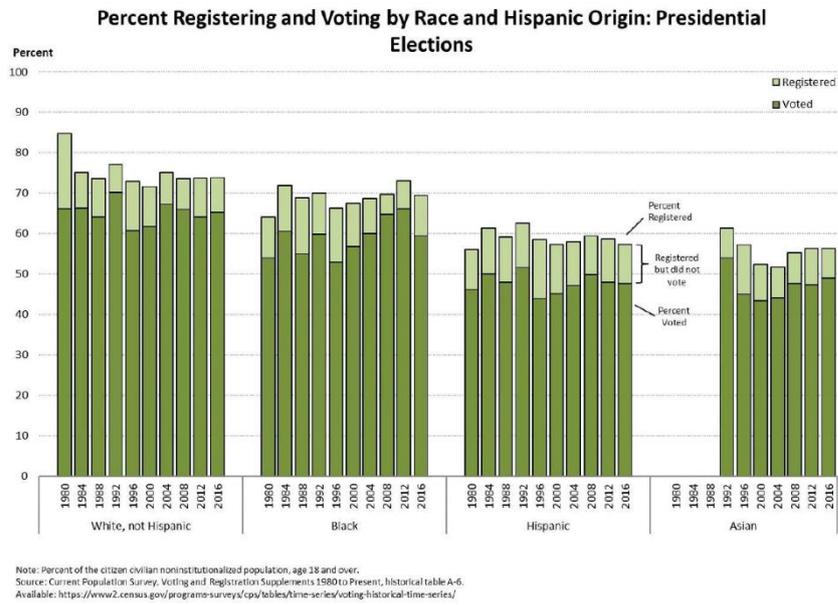
Fuente: “*Most to Least Segregated Metropolitan Regions in the US, 2019*” (Other and Belonging Institute - Most to Least Segregated Metro Regions, 2022)

V. Particularidades de los grupos respecto a registraci3n y voto.

Los grupos de blancos y de afroamericanos muestran valores de registraci3n mucho m3s altos que los grupos de asi3ticos, hispanos o abor3genes. En general la registraci3n entre blancos es m3s alta que entre afroamericanos, aunque en las 3ltimas elecciones presidenciales esta diferencia se redujo e incluso se revirti3 para la elecci3n y la reelecci3n de Barack Obama. El comportamiento de votaci3n difiere en algo del de registraci3n, mostrando para los afroamericanos picos en las elecciones de 2012 y 2016 superando incluso a los blancos. Para el grupo de hispanos, los porcentajes de votaci3n son mucho menores que los de registraci3n, y decaen con el tiempo para las elecciones de medio t3rmino; algo similar ocurre para el grupo de asi3ticos.

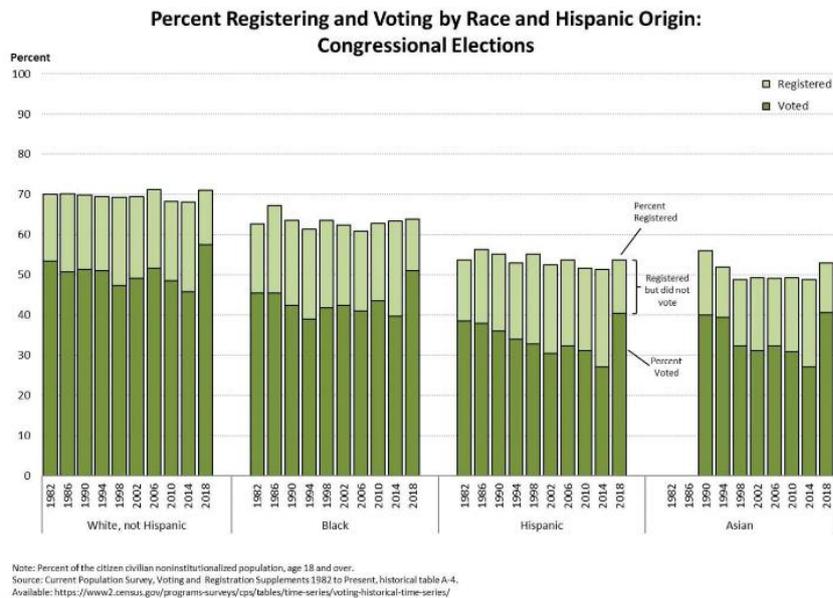
Las elecciones presidenciales son m3s convocantes que las intermedias y la mayor registraci3n y voto se acent3a para los grupos minoritarios. Las diferencias entre personas registradas y personas que efectivamente votan, y la ampliaci3n de esas diferencias para ciertos grupos y entre elecciones presidenciales y de medio t3rmino son coherentes con la idea de que sufragar es una decisi3n con un costo individual relevante que determina el comportamiento del individuo el d3a de la elecci3n, que puede variar entre individuos por factores personales en el momento de ir a sufragar; tambi3n son indicativas de las diferencias entre grupos en sus normas y prescripciones identitarias y de decisiones personales de identificaci3n con el grupo.

Gráfico 2 - Registración y voto en elecciones presidenciales.



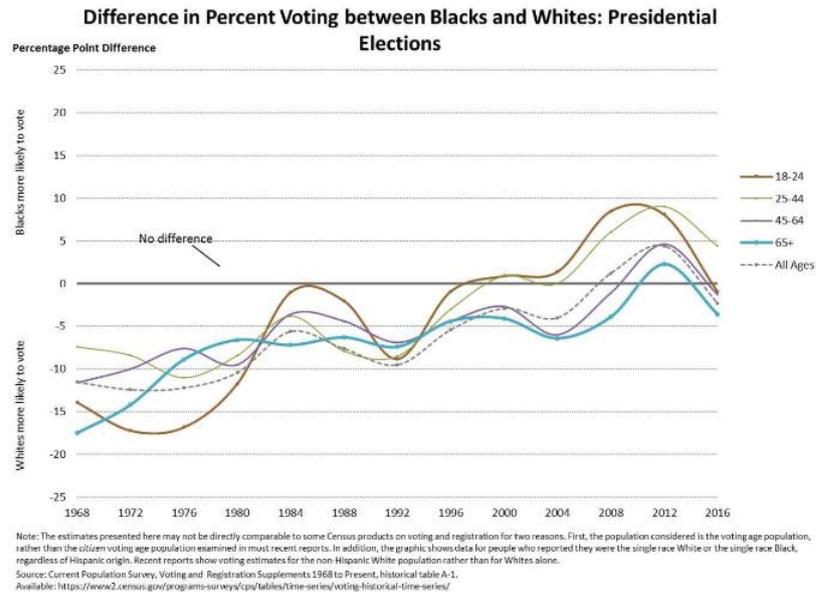
Nota: gráfico obtenido de <https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/a6-presidential.jpg>

Gráfico 3 - Registración y voto en elecciones de medio término.



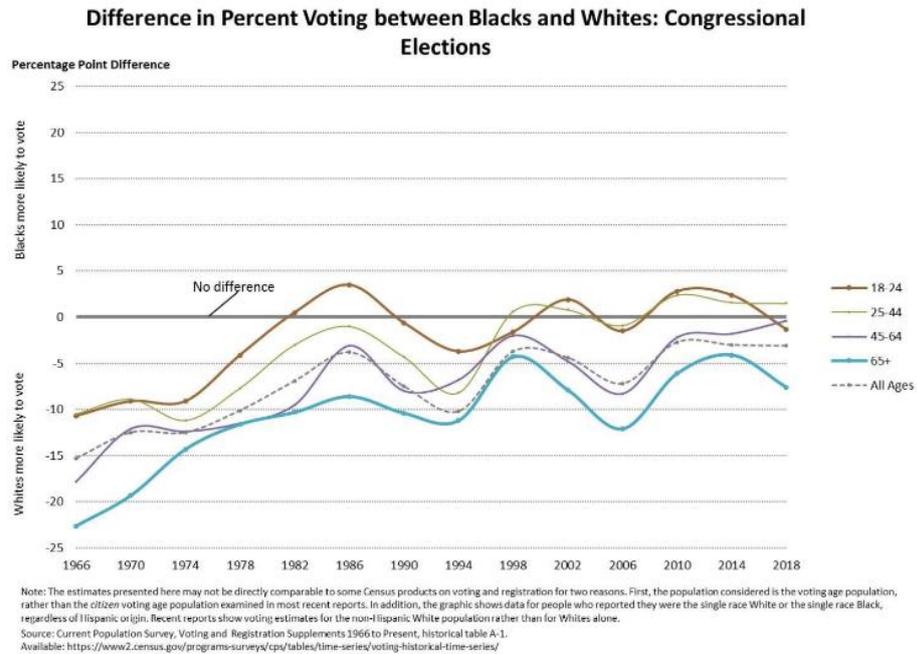
Nota: gráfico obtenido de <https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/a4-congressional.jpg>

Gráfico 4 - Diferencias en votación, elecciones presidenciales.



Nota: gráfico obtenido de <https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/a1-presidential2.jpg>

Gráfico 5 - Diferencia en votación, elecciones de medio término.



Nota: gráfico obtenido de <https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/a1-congressional2.jpg>

VI.El modelo económico.

A. Juego secuencial de decisiones identitarias.

La persona toma dos decisiones sucesivas de identificación con su grupo que afectan las elecciones laborales y políticas; es un objetivo del estudio entender cómo se comportan las dos variables que capturan ambas decisiones en forma conjunta. La decisión relativa a la participación en el mercado laboral se toma o se revierte en cualquier momento a partir de que la persona tiene edad suficiente para trabajar legalmente; la decisión de registrarse para sufragar sólo puede tomarse una vez cada dos años, luego de que la persona alcanza la edad legal requerida. Se supone que la decisión de registrarse para la elección es posterior a una decisión de participar en el mercado laboral.

El otro jugador es su propio grupo, que puede imponer un costo, de forma más o menos directa, en función de las decisiones de la persona, y cuya externalidad será más costosa en la medida que el grupo sea más numeroso alrededor de la persona.

B. Función de utilidad general con Identidad.

La función de utilidad del individuo j perteneciente al grupo k es aditiva y separable entre un componente pecuniario y uno no pecuniario:

$$U_{jk} = \textit{Utilidad por ingreso monetario}_{jk} + \delta_j \times \textit{Utilidad por factores no pecuniarios}_{jk}$$

Para poder sumar las utilidades pecuniarias y no pecuniarias se emplea un coeficiente δ_j que es el valor relativo de la utilidad por identidad en la función de utilidad general, similar al valor relativo del status presentado por Oxoby. (Oxoby, 2004, pág. 735) Este parámetro captura la heterogeneidad de los individuos dentro de su grupo.

La utilidad por factores no pecuniarios corresponde a los “*amenities*” ó “*disamenities*” relacionados con participar o no de la fuerza laboral, con participar o no de una elección. Por ejemplo, asociado con el trabajo una persona puede valorar estar en contacto con gente con intereses similares, y asociado a ejercer el voto una persona puede satisfacer un deseo íntimo de premiar o castigar a un gobierno.

$$\begin{aligned} \textit{Utilidad por factores no pecuniarios}_{jk} = & \textit{Utilidad por amenities}_j \\ & - \textit{Desutilidad por disamenities}_j + \textit{Utilidad por identidad}_{jk} \end{aligned}$$

Dentro de estos factores no pecuniarios se encuentran la utilidad por identidad que surge de la adición de tres factores:

Utilidad por identidad $_{jk}$ = Utilidad por lo que "j" percibe sobre sus decisiones $_{jk}$
 + Utilidad por cómo los demás en "k" perciben a "j" sobre sus decisiones $_{jk}$
 + Utilidad que le genera a "j" la coherencia entre sus decisiones $_j$

La Utilidad por Identidad no es una amenidad más, ya que no se limita a la satisfacción de un gusto propio, cómo se siente el individuo j sobre si sus decisiones se adhieren o no al “deber ser” requerido para identificarse con el grupo k , sino que también captura en el segundo término el aspecto social de cómo influye la mirada del otro en la decisión de identificación del individuo con normas y prescripciones (percibiendo el sentido de satisfacción o de desaprobación de otros hacia su persona). La utilidad por identidad puede resultar en que la persona se desvíe de sus propias preferencias por las consecuencias que tiene para él la mirada del otro cuando está expuesta a ella. (Akerlof & Dickens, 1982)

En los primeros dos términos entra el coeficiente ϵ_{jk} que es la medida del “mismatch” de Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000) entre las características personales de j y aquellas propias de la categoría social de su grupo étnico k ; en este trabajo el efecto de ϵ_{jk} se reduce a cuánta ansiedad le provoca a la persona lo que está eligiendo respecto a lo que tendría que elegir (por ejemplo lo que hace un afroamericano íntegro es no aceptar un trabajo si le pagan menos que a un blanco o lo que hace un afroamericano es votar por un demócrata) y el juicio de los demás sobre sus decisiones; en síntesis, la decisión identitaria realizada en un ámbito social, no aislado, conlleva una externalidad o “neighborhood effect” presente en el modelo. (Akerlof & Kranton, Economics and Identity, 2000)

El tercer término aparece cuando se considera más de una decisión generadora de utilidad por identidad, tal como una decisión laboral y otra política. Este término captura el efecto que sobre la utilidad de la persona produce adoptar decisiones identitarias coherentes ya sea adhiriéndose consistentemente a las normas de grupo ó rechazando consistentemente esa adhesión.

La medida de incoherencia del individuo en sus decisiones de elección de identidad, el tercer término de la ecuación, se asocia a la variable Θ_j . En la medida que exista coherencia en la elección de identidad, Θ_j es cero, y si hay incoherencia es uno; no hay niveles de coherencia intermedios.

VII.El modelo econométrico.

Los datos empleados permiten identificar los siguientes cinco grupos: afroamericanos, aborígenes, hispanos, asiáticos y blancos.

Se analiza la variable binaria “*registered*” que captura la decisión de registrarse o no en la última elección de noviembre para cada grupo y género por separado empleando las siguientes variables: edad; educación: desde niveles iniciales hasta posgrados universitarios; lugar de residencia: si vive en los suburbios (1) en contraposición a si vive en el “*inner-city*” o fuera de un área metropolitana (0); tamaño de la población en la que vive la persona en saltos de 250.000 habitantes; condición de persona retirada; variables binarias que indican que en una elección presidencial participó un candidato determinado (J.Carter, R.Reagan, G.Bush, B.Clinton, G.W.Bush, B.Obama, D.Trump, J.Biden); variables binarias que indican la pertenencia a un grupo etario (generación silenciosa, baby boomers, generación X, generación Y); variable binaria “*not in labor force*” (*nilf*) cuyo valor es 1 si la persona no participa en el mercado de trabajo y que es la variable dependiente de la segunda ecuación del modelo bivariado; y porcentaje correspondiente al grupo de referencia de la persona dentro de la población del área.

La segunda decisión binaria relacionada con la identidad del individuo es la participación ó no en la fuerza de trabajo, capturada por la variable *nilf*. Esta variable se explica para cada uno de los grupos mediante las siguientes variables independientes: edad; años de educación; condición de retirado; estado civil; porcentaje correspondiente al grupo de referencia de la persona dentro de la población del área y nivel de segregación en el área metropolitana medido por la variable “*divergence index*.” (Othering and Belonging Institute, 2022)

A.Modelo Probit Bivariado.

Se especifica el siguiente modelo recursivo de dos ecuaciones simultáneas con perturbaciones correlacionadas para evaluar el modelo económico de utilidad con identidad: (Greene, Econometric Analysis, 2002, p. 710)

$$y_1^* = X_1' \beta_1 + \gamma y_2 + \varepsilon_1,$$

$$y_2^* = X_2' \beta_2 + \varepsilon_2$$

$$E[\varepsilon_1 | X_1, X_2] = E[\varepsilon_2 | X_1, X_2] = \mathbf{0}$$

$$VAR[\varepsilon_1 | X_1, X_2] = VAR[\varepsilon_2 | X_1, X_2] = \mathbf{1}$$

$$COV[\varepsilon_1, \varepsilon_2 | X_1, X_2] = \rho$$

Basado en desarrollos de Maddala (Limited-dependent and qualitative variables in econometrics, 1983, p. 123) se demuestra que la naturaleza endógena de y_2 en el miembro derecho de la primera ecuación puede ignorarse cuando se formula la verosimilitud logarítmica. (Greene, Econometric Analysis, 2002, p. 715)

B. Estimación de Máxima Verosimilitud.

Como antecedente de un abordaje similar al que se empleará en este trabajo, Burnett plantea un modelo bivariado para la presencia de un curso en Economía de Género dada la oferta de un programa de Estudios de la Mujer en un campus (Burnett, 1997) y Greene comenta este trabajo criticando el abordaje econométrico original, en el que se empleó mínimos cuadrados ordinarios (OLS), y proponiendo otro abordaje que obtiene los coeficientes del modelo recursivamente con estimadores de máxima verosimilitud (MLE.) (Greene, 1998)

$$Prob(y_1 = 1, y_2 = 1 | X_1 X_2) = \Phi_2(X_1' \beta_1 + \gamma y_2, X_2' \beta_2, \rho)$$

$$Prob(y_1 = 1, y_2 = 0 | X_1 X_2) = \Phi_2(X_1' \beta_1, -X_2' \beta_2, -\rho)$$

$$Prob(y_1 = 0, y_2 = 1 | X_1 X_2) = \Phi_2(-(X_1' \beta_1 + \gamma y_2), \beta_2' X_2, -\rho)$$

$$Prob(y_1 = 0, y_2 = 0 | X_1 X_2) = \Phi_2(-X_1' \beta_1, -X_2' \beta_2, \rho)$$

Análogamente al problema de Burnett, en este trabajo se plantea un modelo bivariado que explica conjuntamente las dos variables dependientes de decisiones identitarias. Un detalle de interés del modelo de Burnett es la inclusión entre los regresores del porcentaje de mujeres en el programa y en el campus, del todo similares al regresor de porcentaje del grupo de interés aquí empleado.

VIII. Interpretación de la correlación.

Se realizan para grupos étnicos y por género regresiones “*Seemingly unrelated bivariate Probit*” que comparan dos regresiones Probit univariadas no correlacionadas contra un modelo bivariado donde sí se permite que la correlación sea distinta de cero. El *rho* (ρ) reportado es el que resulta si existe un patrón en los residuos no explicado por los modelos de distribuciones marginales, luego de que la influencia de las variables explicativas fue tomada en cuenta.

The correlation coefficient measures the correlation between the disturbances in the equations, the omitted factors. That is, ρ measures (roughly) the correlation between the outcomes after the influence of the included factors is accounted for. (Greene, Econometric Analysis, 2002, p. 717)

A.Resultados de las regresiones para los coeficientes de interés.

En las Tablas 9 y 10 se muestran los coeficientes de interés de la regresión biprobit por grupo y por género, que a continuación se analizan.

El factor que más impacto tiene en la decisión de registrarse para cualquier grupo y género es la participación del mercado laboral (*ilf*) con un coeficiente siempre positivo y estadísticamente relevante.

Para las dos ecuaciones del modelo bivariado, la inclusión o no de las variables identitarias en los modelos ("*percentage_X*", "*divergence index*", "*ilf*") tiene mínimo efecto en el valor de los coeficientes de las variables de control (educación, edad, condición de retirado, condición de veterano militar, estado civil, dummies de candidatos a presidente y de elección de medio término ó de generación a la que pertenece la persona), sugiriendo que son ortogonales y que no están determinadas conjuntamente. En la tabla 9 se muestra la regresión bivariada sin el regresor *divergence* y en la Tabla 10 se muestra la regresión bivariada con el regresor de segregación en la ecuación *ilf*.

Para la decisión de registrarse, el coeficiente del porcentaje del grupo en el área metropolitana es positivo y estadísticamente relevante para los afroamericanos (+0,4704954 varones y +0,3946397 mujeres), los varones aborígenes (+0,7949553), los hispanos (+0,1527695 varones y +0,1447622 mujeres) y los blancos (+0,3569642 varones y +0,3302829 mujeres), y no es estadísticamente diferente de cero para las mujeres aborígenes y para los asiáticos. Lo anterior sugiere que existe presión social hacia la participación política en los primeros siete grupos y no en los últimos tres.

En la decisión de participar en el mercado laboral (*ilf*) las dos variables para las que se evalúa su influencia sobre la decisión identitaria son el porcentaje del propio grupo en el área metropolitana y el "*divergence index*".

Para el grupo de afroamericanos varones incluyendo el regreso "*divergence index*" su coeficiente (-0,5015583) es negativo y el de "*percentage_black*" (-0,1914903) es positivo. Para esta minoría estas dos variables están altamente correlacionadas (correlación de 0,4832), sugiriendo que existe una fuerte relación: la segregación en áreas metropolitanas ocurre donde hay una importante congregación de afroamericanos; adicionalmente

los afroamericanos tienen pocas chances de ser aceptados por la mayoría blanca en otras zonas de la ciudad. El comportamiento de autoexclusión queda capturado por el coeficiente de la variable de segregación (*divergence index*). Vale notar que si no se incluye el regresor de segregación, el coeficiente de la variable de presión social (*percentage_black*) tendrá coeficiente negativo, tal como se obtuvo en el primer paper (capítulo II). La interpretación es que personas afroamericanas, específicamente varones, segregados geográficamente en “*inner-cities*” ó “*ghettos*”, interactúan desproporcionadamente con otras personas de su grupo y tienen mayor probabilidad de comportarse como outsiders del mundo exterior al ghetto, que incluye el mercado laboral y su ética de trabajo; si salen del *ghetto*, la presión social de otros afroamericanos se invierte a “virtuosa” y aumenta su probabilidad de elegir una identidad de insider, participando del mercado laboral y asimilándose socialmente.

Para el grupo de blancos los dos coeficientes de “*percentage_white*” y “*divergence index*” son positivos (+0,18705 y +0,3828 para los varones y +0,4539 y +0,4443 para las mujeres respectivamente) lo que refleja el comportamiento de “*insiders*” propio del grupo hegemónico: cuanto más segregación racial hay en su área metropolitana, es decir cuanto más probable sea que un blanco interactúe más con otro blanco en su lugar de trabajo más a gusto estará el americano blanco promedio y asociado a eso aumentará su propensión a participar del mercado laboral.

Para los demás grupos los efectos no son tan marcados. En un extremo, las mujeres aborígenes e hispanas tienen ambos coeficientes no estadísticamente diferentes de cero. Para los varones aborígenes ambos coeficientes son estadísticamente relevantes pero la probabilidad de autoexclusión disminuye con el porcentaje de aborígenes (-1,271801) y aumenta en áreas metropolitanas más segregadas (+1,025354); el “*divergence index*”, que mide el nivel de segregación intra región metropolitana, no parece particularmente sensible para identificar los lugares geográficos donde se encuentran segregados los aborígenes; las comunidades aborígenes no se concentran históricamente en ghettos de grandes ciudades sino que se asientan en reservas indígenas donde la segregación por aislamiento geográfico reduce las chances tanto de interactuar económicamente con los blancos como de ser objeto de discriminación de empleadores u otros trabajadores blancos, a pesar de que ese aislamiento resulta en un menor ingreso promedio. (Becker G. S., *The Economics of Discrimination*, 1957, pág. 24) En conclusión, mover a un varón aborígen promedio de una reserva indígena a un área metropolitana típica significa que aumenta

su probabilidad de participar en el mercado laboral a la vez que disminuye el porcentaje de aborígenes y aumenta el “*divergence index*”.

Para los varones hispanos, la primera minoría en tamaño, el coeficiente en la ecuación “*ilf*” de “*percentage_hispan*” es negativo (-0,1353487) y el coeficiente de “*divergence index*” es levemente positivo (+0,0217097) y no estadísticamente relevante, posiblemente por no confinarse esta minoría tan marcadamente en *ghettos* urbanos.

Para la minoría asiática el coeficiente de “*percentage_asian*” es positivo (+0,335218 para varones y +0,4562361 para mujeres) y el de “*divergence index*” es negativo (-0,3419566 para varones y -0,592167 para mujeres) siendo los dos estadísticamente relevantes; el primer coeficiente es indicativo de un comportamiento de mayor participación en el mercado laboral concomitante al premio al retorno que esta minoría recibe a igual capital humano, lo que Becker denomina nepotismo. (Becker G. S., *The Economics of Discrimination*, 1957, pág. 15) El valor positivo del segundo coeficiente muestra que la minoría de personas asiáticas, a pesar de ser considerada la “*model minority*” resulta perjudicada por la segregación urbana.

Tabla 9- Regresión biprobit por grupo y por género, coeficientes de interés.

Grupo	Ecuación registered: coeficiente <i>percentage</i>	Ecuación <i>ilf</i> : coeficiente <i>percentage</i>
Hombres Afroamericanos	+0.4704954	-0.1566194
	(.0515739)	(.0575379)
Mujeres Afroamericanas	+.3946397	+0.1624259
	(.0468394)	(.0492104)
Hombres aborígenes	+0.7949553	-1.417203
	(.3316118)	(.3666574)
Mujeres aborígenes	+0.1012822	+0.0902174
	(.3147834)	(.3459047)
Hombre hispanos	+0.1527695	-0.1299689
	(.0276128)	(.0318904)
Mujeres hispanas	+0.1447622	+0.0468534
	(.0259674)	(.0260363)
Hombres asiáticos	+0.0481916	+0.3105833
	(.0541955)	(.0755295)
Mujeres asiáticas	+0.2112376	+0.4327348
	(.0467601)	(.0579476)
Hombres Blancos	+0.3569642	-0.0342475
	(.0164937)	(.0185747)
Mujeres Blancas	+0.3302829	+0.2035461
	(.0163414)	(.0154581)

Tabla 10- Regresión biprobit por grupo y por género con segregación, coeficientes de interés.

Nota: las correlaciones entre las variables categóricas se calculan intragrupos; los errores estándar se reportan entre

Grupo	Ecuación registered: coeficiente <i>percentage</i>	Ecuación ilf: coeficiente <i>percentage</i>	Ecuación ilf: coeficiente <i>divergence</i>
Hombres Afroamericanos	+ 0,4439134	+0,1914903	-0,5015583
	(0,0516539)	(0,0672697)	(0,0496684)
Mujeres Afroamericanas	+ 0,3923551	+0,223994	-0,0907686
	(0,046872)	(0,0568462)	(0,0418653)
Hombres aborígenes	+ 0,7504677	- 1,271801	+ 1,025354
	(0,3324939)	(0,3683283)	(0,2202645)
Mujeres aborígenes	+ 0,0979136	+0,137158	+0,3311677
	(0,315268)	(0,3469597)	(0,1902156)
Hombre hispanos	+ 0,1534691	- 0,1353487	+ 0,0217097
	(0,0276648)	(0,0344025)	(0,0522062)
Mujeres hispanas	+ 0,145117	+ 0,0436677	+ 0,0134945
	(0,0259843)	(0,0279512)	(0,0431105)
Hombres asiáticos	+ 0,048691	+ 0,335218	- 0,3419566
	(0,0541733)	(0,0762133)	(0,1036453)
Mujeres asiáticas	+ 0,029513	+ 0,4562361	- 0,592167
	(0,0535835)	(0,0632525)	(0,0846644)
Hombres Blancos	+ 0,34601	+ 0,1870592	+ 0,3828274
	(0,0165259)	(0,0240949)	(0,0267325)
Mujeres Blancas	+ 0,3387953	+ 0,4539953	+ 0,4443262
	(0,0164973)	(0,0200192)	(0,0226506)

paréntesis.

Confecionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

B. Correlación entre entre residuos del modelo.

La correlación entre “*registered*” e “*ilf*” que varía desde (-0,0459) para los varones blancos hasta (-0,1095) para los varones afroamericanos sugiere que las dos variables dependientes están conjuntamente determinadas. La interpretación de este coeficiente de correlación negativo es que una menor participación en el mercado laboral es acompañada por una menor registración para sufragar, lo cual tiene sentido ya que ejercer el derecho al voto no le resuelve las necesidades acuciantes de supervivencia que se supone tendrá una persona sin un trabajo que provea un ingreso, dado además que la participación individual es irrelevante para el resultado de la elección.

En forma más general puede pensarse que la relación entre ambas variables proviene del costo de oportunidad de sufragar que es más alto para una persona que no tiene sus necesidades mínimas cubiertas a través de un ingreso laboral.

Si en la regresión bivariada se omite a *ilf* como variable explicativa de la registración entonces los valores de correlación de los residuos de las ecuaciones, que varían desde (-0,0766372) para los varones blancos hasta (-0,142637) para varones afroamericanos, resultados muy similares a las correlaciones entre variables binarias individuales antes vistas y esperados ya que no se está incluyendo el efecto de *ilf* en *registered*.

Una vez introducida la variable *ilf* como variable explicativa de “*registered*” la correlación entre residuos del modelo varía sustancialmente evidenciando un patrón diferente en los residuos. Es el caso para el modelo econométrico planteado y resuelto con el abordaje de Maddala (Limited-dependent and qualitative variables in econometrics, 1983). En la Tabla 11 se muestra el reporte completo de la regresión bivariada para el caso de los varones afroamericanos sin regresor de segregación (*divergence index*) y en la Tabla 12 se repite la regresión esta vez incluyendo la segregación como variable explicativa.

Tabla 11 – Regresión biprobit para varones afroamericanos sin regresor de segregación
– reporte de Stata.

Seemingly unrelated bivariate probit		Number of obs = 80238			Log likelihood = -86857.445	
Wald chi2(27) = 21049.17		Prob > chi2 = 0.0000				
registered=age yearse percentage_black congressional carter reagan bush clinton wbush obama trump biden						
suburb population retired silent boomers genX genY genZ ilf						
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	+0,0190855	0,0004928	38,73	0	0,0181196	0,0200513
yearse	+0,0298962	0,0014416	20,74	0	0,0270707	0,0327218
percentage_black	+0,4704954	0,0515739	9,12	0	0,3694124	0,5715784
congressional	-0,1036525	0,0281418	-3,68	0	-0,1588095	-0,0484956
carter	-0,1247876	0,0249512	-5	0	-0,1736909	-0,0758842
reagan	+0,1340779	0,0246579	5,44	0	0,0857494	0,1824064
bush	+0,0137828	0,0264166	0,52	0,602	-0,0379927	0,0655583
clinton	-0,0368015	0,0256431	-1,44	0,151	-0,0870612	0,0134581
wbush	-0,0288395	0,0315004	-0,92	0,36	-0,0905791	0,0329001
obama	+0,083943	0,0322515	2,6	0,009	0,0207312	0,1471548
trump	+0,0064433	0,0353205	0,18	0,855	-0,0627836	0,0756701
biden	-0,0469758	0,0284806	-1,65	0,099	-0,1027968	0,0088452
suburb	-0,0822972	0,0106651	-7,72	0	-0,1032004	-0,0613941
population	-0,0034747	0,0009464	-3,67	0	-0,0053295	-0,0016198
retired	-0,4907743	0,0210821	-23,28	0	-0,5320944	-0,4494542
silent	+0,1498376	0,0505651	2,96	0,003	0,0507317	0,2489434
boomers	-0,0843405	0,0451265	-1,87	0,062	-0,1727868	0,0041058
genX	-0,1179892	0,0435348	-2,71	0,007	-0,2033158	-0,0326626
genY	-0,0739341	0,042991	-1,72	0,085	-0,1581949	0,0103266
ilf	+1,120354	0,0328556	-34,1	0	1,055959	1,18475
_cons	-1,756123	0,0532983	-32,95	0	-1,860585	-1,65166
ilf=age yearse percentage_black married retired veteran divergence						
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-0,0178002	0,00035	50,86	0	-0,0184862	-0,0171143
yearse	+0,0642507	0,0010867	-59,13	0	0,0621209	0,0663805
percentage_black	-0,1566194	0,0575379	2,72	0,006	-0,2693917	-0,0438472
married	+0,7218387	0,0115706	-62,39	0	0,6991607	0,7445168
retired	+7,443986	1003,148	-0,01	0,994	-1958,69	1973,578
veteran	-0,0201977	0,0168825	1,2	0,232	-0,0532868	0,0128914
_cons	+0,1969374	0,0228787	-8,61	0	0,1520961	0,2417788
rho	-0,5121822	0,0207156			-0,5516321	-0,4704458
Likelihood-ratio test of rho=0: chi2(1) = 441.582 Prob > chi2 = 0.0000						

Tabla 12 – Regresión biprobit para varones afroamericanos con regresor de segregación
– reporte de Stata.

Seemingly unrelated bivariate probit		Number of obs = 80238			Log likelihood = -86806.555	
Wald chi2(27) = 21252.27		Prob > chi2 = 0.0000				
registered=age yearse percentage_black congressional carter reagan bush clinton wbush obama trump biden						
suburb population retired silent boomers genX genY genZ ilf						
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	+0,0191414	0,0004927	38,85	0	0,0181757	0,0201071
yearse	+0,0291056	0,0014552	20	0	0,0262535	0,0319578
percentage_black	+0,4439134	0,0516539	8,59	0	0,3426737	0,5451531
congressional	-0,1070127	0,0281276	-3,8	0	-0,1621417	-0,0518837
carter	-0,1290123	0,0249442	-5,17	0	-0,1779019	-0,0801226
reagan	+0,1277908	0,0246606	5,18	0	0,0794569	0,1761247
bush	+0,0013145	0,0264334	0,05	0,96	-0,0504941	0,0531231
clinton	-0,0321077	0,0256427	-1,25	0,211	-0,0823665	0,0181511
wbush	-0,0267246	0,0314831	-0,85	0,396	-0,0884304	0,0349811
obama	+0,079053	0,0322412	2,45	0,014	0,0158614	0,1422446
trump	+0,0002644	0,0353091	0,01	0,994	-0,0689402	0,069469
biden	-0,0472198	0,028464	-1,66	0,097	-0,1030082	0,0085686
suburb	-0,0803921	0,0106576	-7,54	0	-0,1012807	-0,0595036
population	-0,0022682	0,0009564	-2,37	0,018	-0,0041426	-0,0003938
retired	-0,4941658	0,0210712	-23,45	0	-0,5354646	-0,4528669
silent	+0,1428269	0,0505505	2,83	0	0,0437497	0,2419041
boomers	-0,0883723	0,0450989	-1,96	0,05	-0,1767646	0,0000199
genX	-0,1196643	0,0435018	-2,75	0,006	-0,2049261	-0,0344024
genY	-0,0755978	0,0429589	-1,76	0,078	-0,1597957	0,0086
ilf	+1,124296	0,0325425	-34,55	0	1,060514	1,188078
_cons	-1,743688	0,0531255	-32,82	0	-1,847812	-1,639564
ilf=age yearse percentage_black married retired veteran divergence						
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	-0,0175883	0,0003506	50,17	0	-0,0182753	-0,0169012
yearse	+0,0689865	0,0011887	-58,04	0	0,0666567	0,0713162
percentage_black	+0,1914903	0,0672697	-2,85	0,004	0,0596442	0,3233364
married	+0,7109873	0,0116078	-61,25	0	0,6882363	0,7337382
retired	+7,507408	1000,88	-0,01	0,994	-1954,181	1969,196
veteran	-0,0220728	0,0168694	1,31	0,191	-0,0551362	0,0109906
divergence	-0,5015583	0,0496684	10,1	0	-0,5989066	-0,40421
_cons	+0,1314908	0,0238417	-5,52	0	0,084762	0,1782197
rho	-0,5150339	0,0205144			-0,5541012	-0,4737028
Likelihood-ratio test of rho=0: chi2(1) = 452.626 Prob > chi2 = 0.0000						

Lo que en el modelo econométrico se omite es una variable relacionada al término de coherencia del modelo económico. Dado que la coherencia significa realizar elecciones identitarias tal como indica la presión social del propio grupo tanto en el mercado laboral como en el proceso político, existen cuatro posibles combinaciones de decisiones que hay que considerar: ser insider participando en el mercado laboral y en el proceso político, ser un outsider no participando en ninguno de los dos, y las combinaciones cruzadas de participar en el mercado laboral y no en el proceso político y por último de autoexcluirse del mercado laboral y participar del proceso político.

Si la coherencia es relevante entonces existen costos por no ser coherente que inciden en la frecuencia relativa con que encontramos cada una de las cuatro combinaciones; si unas combinaciones de elecciones prevalecen más que otras, los factores que no fueron incluidos en el modelo econométrico se expresarán en un patrón de los residuos y en la correlación de los mismos.

Las diferencias de correlaciones residuales del modelo que se muestran en la Tabla 13 entre el grupo de varones afroamericanos (+0,5150339), los varones hispanos (+0,4950785), los varones aborígenes (+0,5137666) y los varones blancos (+0,4144303) ó asiáticos (+0,2833089), son consistentes con la existencia de un costo asociado a la falta de coherencia en las decisiones.

Tabla 13 - Correlación residual en el modelo bivariado por grupo y género.

Grupo	Varones	Mujeres
Afroamericanos	0,5150339 (0,0205144)	0,2080601 (0,0321903)
Aborígenes	0,5137666 (0,0750342)	0,3893602 (0,1161611)
Hispanos	0,4950785 (0,0228075)	0,3354138 (0,0480077)
Asiáticos	0,2833089 (0,0531264)	0,0272023 (0,1137428)
Blancos	0,4144303 (0,0071242)	-0,0709628 (0,0253263)

Nota: las correlaciones entre las variables categóricas se calculan intragrupos; los errores estándar se reportan entre paréntesis. En todos los casos hay significatividad estadística para niveles de confianza de al menos 95%.

Las correlaciones surgen de una regresión biprobit especificada en Stata como “seemingly unrelated bivariate probit regression” que permite para controlar por endogeneidad no observada; la hipótesis es que ambas variables dependientes están determinadas por la elección de identidad del individuo, y que a su vez la elección de autoexcluirse del mercado laboral impone costos adicionales en la elección de no registrarse para la elección, como se explica en el modelo de juego secuencial.

Confecionado con datos del United States Census Bureau, Current Population Survey y Voting and Registration Supplement, año 2020 y años previos.

Para los tres primeros grupos, la autoexclusión del mercado laboral es un comportamiento concomitante a la discriminación laboral y en los tres tenemos una mayor correlación positiva de residuos entre la decisión de autoexclusión y la de participación del proceso político, que es la dirección de la presión social. La decisión de coherencia a la que se ven sometidos muchos individuos de los tres primeros grupos consiste en ser un outsider en el mercado laboral y un insider en lo político.

La falta de coherencia en los planos laboral y político genera disonancia cognitiva (Akerlof & Dickens, 1982) (Festinger, A Theory of Cognitive Dissonance., 1957) y una consecuente pérdida de utilidad por identidad; hay una razón de reducción de la ansiedad y de preservación del bienestar en el individuo que lo motiva a evitar una íntima incoherencia.

La consecuencia es que para estas minorías unas combinaciones de decisiones (los equilibrios del juego secuencial presentado en la próxima sección) se vuelven más prevalentes en desmedro de otras. Los equilibrios en los que $\theta_j = 0$ (mercado laboral, $ilf=1$: outsider; proceso político $registered=1$: insider) y ($ilf=0$: insider; $registered=0$: outsider) son menos costosos desde el punto de vista de la coherencia que las alternativas donde $\theta_j = 1$ ($ilf=0$: insider; $registered=1$: insider) y ($ilf=1$: outsider; $registered=0$: outsider). El resultado esperado es un cambio en el valor de la correlación en los residuos.

IX. Aplicación del modelo prototipo de Economía de la Identidad de Akerlof y Kranton a la decisión de ejercer el derecho al voto.

Se plantea el siguiente juego adaptado del modelo original de Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000): frente al derecho de sufragio universal voluntario de los ciudadanos, tal como existe en Estados Unidos, dos son las posibles posturas: votar o no votar. Votar (incluso el hecho previo de registrarse para votar) es la elección que define a una persona como “*insider*”, el comportamiento que indica la norma social, el comportamiento que se espera de un ciudadano “comprometido” y que a nivel de grupo garantiza tener una representación y defensa a través del sistema político. Pero desde luego que hay un continuo en el gusto por votar, desde mucho a poco, y en el extremo hay quienes piensan que hacerlo es demasiado costoso a nivel individual dado que no hay beneficios directos; a estos ciudadanos podemos calificarlos de “indolentes”.

El ciudadano “comprometido” realiza una acción de identificación con una norma de social, participa en el proceso político, y por lo tanto elige una identidad de “*insider*” respecto a su grupo. Por el contrario el ciudadano “indolente” no realiza la acción que prescribe identidad de buen ciudadano y, por omisión, al no participar del proceso político, elige una identidad de “*outsider*”.

En un planteo con información incompleta, las identidades de las personas no quedan expuestas hasta que no toman una acción que los revela como “comprometidos” o como “indolentes”. En un planteo secuencial, la identidad del “comprometido” se revela al “indolente” antes de que este último tome una decisión; por ejemplo, el “comprometido” pega una calcomanía con la escarapela patria impresa en el paragolpes de su automóvil que queda a la vista de todos y que dice “orgulloso de haberme registrado para la votación”; la calcomanía la obtuvo de la autoridad electoral al momento de registrarse. En el presente planteo de juego secuencial al jugador íntimamente “comprometido” no le importa si el otro jugador íntimamente sea un “indolente” con tal que elija la identidad de “*insider*” y se registre para votar.

El modelo teórico de juego secuencial entre dos jugadores con información completa y perfecta consiste en un jugador “*j*” que mueve primero y tiene una sola acción disponible, que es votar, seguido por el jugador “*i*” que puede elegir entre votar y no votar. El juego concluye si el jugador “*i*” decide votar. Pero si el jugador “*i*” decide no votar, entonces nuevamente el jugador “*j*” tiene que decidir entre aplicar un escarmiento

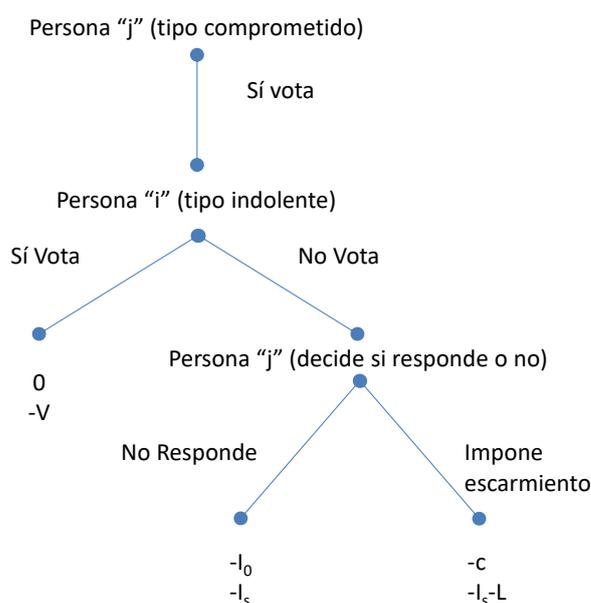
al jugador “*i*” por no haber votado, o resignarse y rumiar su descontento. Luego de esta decisión concluye el juego.

No existen beneficios pecuniarios asociados con las decisiones; como “*j*” siempre vota, normalizamos su utilidad neta por votar a cero; pero votar sí genera una utilidad neta negativa ($-V$) a “*i*” por el esfuerzo asociado a sufragar.

Adicionalmente existen dos externalidades que surgen si el jugador “*i*” elige votar: I_0 es la pérdida de utilidad que le produce a “*j*” ver que el “*i*” resulta ser un “indoleté” y que no vota; I_s es la pérdida de utilidad que le provoca a un “*i*” la mirada desaprobadora de “*j*”, haber quedado expuesto socialmente como un “indolente”.

Por último, el jugador “*j*” puede recomponer su pérdida de utilidad por identidad I_0 respondiendo a la decisión del jugador “*i*” con un escarmiento sobre “*i*” a un costo “ c ”, que le generará a “*i*” una pérdida de utilidad adicional “ L ”. Este escarmiento no modifica el hecho de que “*i*” sea un indolente y que no vote, pero puede ser algo costoso para “*j*”, como por ejemplo ser tratado con frialdad y desdén.

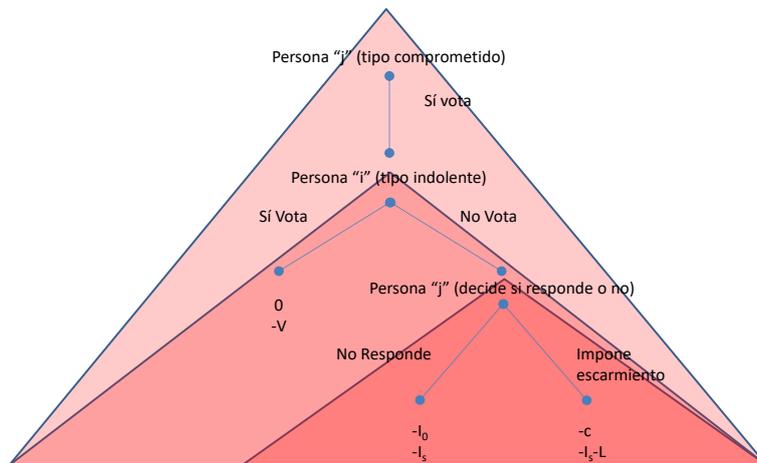
Gráfico 6 – Representación del juego de Akerlof y Kranton en forma extensiva.



Continuamos el análisis del juego identificando los posibles equilibrios de Nash perfectos en subjuegos (ENPS) que pueden darse a partir de diferentes valores de V , I_0 , I_s , c y L .

Existen dos subjuegos; comenzando en el nodo de la persona “*j*” tenemos el juego completo; comenzando en el nodo donde decide la persona “*i*” tenemos un sub juego, y comenzando en el nodo donde la persona “*j*” puede por segunda vez decidir tenemos el segundo sub juego.

Gráfico 7 – Identificación de juego y subjuegos en la representación extensiva.



A. Posibles Equilibrios de Nash Perfectos en Subjuegos.

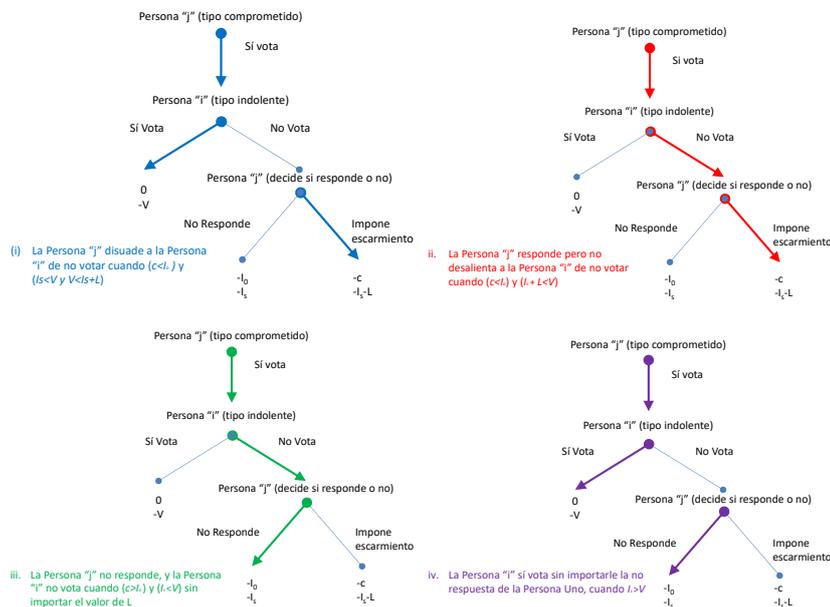
- (i) La persona "j" desalienta a la persona "i" de no votar si: $\begin{cases} c < I_0 \\ I_s < V < I_s + L \end{cases}$
- (ii) La persona "j" decide aplicar un escarmiento pero sin lograr disuadir a la persona "i" de no votar si: $\begin{cases} c < I_0 \\ I_s + L < V \end{cases}$
- (iii) La persona "j" no responde y la persona "i" se registra para votar si: $\begin{cases} c > I_0 \\ I_s < V \end{cases}$
- (iv) La persona "i" no se registra para votar independientemente de la respuesta de la respuesta de la persona "j" si: $I_s > V$

Gráfico 8 - representación del juego en forma normal con los diferentes ENPS.

		Persona "j" – Insider comprometido			
		Si vota, escarmentar; si no vota, escarmentar	Si vota, escarmentar; si no vota, no responder	Si vota, no responder; si no vota, escarmentar	Si vota, no responder; si no vota, no responder
Persona "i" – Outsider indolente	Sí vota	-V, 0 	-V, 0 	-V, 0 	-V, 0
	No vota	$-I_s-L$, -C 	$-I_s$, $-I_0$ 	$-I_s-L$, -C 	$-I_s$, $-I_0$

- (i) La Persona "j" disuade a la Persona "i" de votar cuando ($c < I_0$) y ($I_s < V$ y $V < I_s + L$)
- (ii) La Persona "j" responde pero no desalienta a la Persona "i" de votar cuando ($c < I_0$) y ($I_s + L < V$)
- (iii) La Persona "j" no responde, y la Persona "i" vota cuando ($c > I_0$) y ($I_s < V$) sin importar el valor de L
- (iv) La Persona "i" no vota sin importarle la no respuesta de la Persona "j", cuando $I_s > V$

Gráfico 9 – Representación extensiva de los cuatro posibles ENPS.



El equilibrio (iv) se dá cuando la persona "i", que es íntimamente un outsider indolente, experimenta una ansiedad " I_s " tal que la cohibe de manifestar públicamente su condición de outsider. Este es el único equilibrio de autocontrol puro.

Si la ansiedad " I_s " que "i" experimenta por quedar expuesta su condición de indolente a la mirada de "j" no es tan insoportable entonces la decisión de votar o de no votar

dependerá de qué tan fuerte sea el escarmiento del que dispone el jugador “*j*” para condicionar la conducta del jugador “*i*”.

La decisión binaria de “*i*” de registrarse o no para una elección cambia cuando $I_s + L = V$; es decir que si “*j*” puede infligir un escarmiento que provoca un daño en “*i*” $L > V - I_s$ se logrará condicionar la conducta de “*i*” y evitar que exteriorice que es un “*outsider*” indolente.

Suponiendo que el jugador “*j*” no es un individuo, sino que es un grupo de “*insiders*” comprometidos idénticos entre sí cuyo número puede variar, pero cuyo poder coercitivo “*L*” es proporcional al tamaño del grupo, entonces la conducta de “*i*” se verá condicionada por el porcentaje en toda la comunidad de individuos del tipo “*j*”, a mayor porcentaje de individuos “*j*” en la misma comunidad de “*i*”, mayor el costo “*L*” que podrán imponerle por exteriorizar su identidad indolente.

Más aún, si existe información incompleta sobre quién es en realidad el otro, si un “*outsider*” ó un “*insider*”, el control social puede ser aún más fuerte, ya que un “*outsider*” averso al riesgo puede preferir ocultar su verdadera preferencia para no soportar el riesgo de un escarmiento.

En el juego anterior lo que la comunidad “*k*” logrará es que cambie el equilibrio (ii) en el que “*i*” no se registra a la votación al equilibrio (i) en el que sí se registra para votar.

El último equilibrio a considerar en el cual “*j*” no impone un escarmiento sobre “*i*” se da cuando es muy alto el costo “*c*” para “*j*” y en consecuencia no le queda otra opción a “*j*” que quedarse sin reaccionar sufriendo la pérdida “*I_o*”. Esto puede darse si al comportamiento de “*j*” de imponer un escarmiento se le puede asociar algún tipo de penalidad social por parte de los demás outsiders, o cuando se establece una distancia prudencial entre “*i*” y “*j*” tal que impida que el escarmiento resulte efectivo, o si es el caso en el que reducir esa distancia entre “*i*” y “*j*” resulte en un aumento inversamente proporcional de “*c*”.

B.Subjuego electoral dentro de juego de elección identitaria.

Se mencionó previamente que el modelo de utilidad general propuesto prevé que la elección de identidad afecte a dos variables: la participación en el mercado laboral, que es una elección abierta a cambiarse en cualquier momento, y la participación en elecciones que puede darse los meses de noviembre de los años pares.

En un planteo de juego secuencial corresponde primero elegir si se participa en el mercado laboral y luego, cada dos años, si se vota; en un grupo en el que las dos decisiones

son identitarias, la primera elección define en qué subjuego se tomará la segunda decisión y por lo tanto los pagos.

La incoherencia en la segunda elección tendrá un costo individual adicional, supongamos que del doble del costo " I_s ." En el primer subjuego donde el indolente eligió no participar en el mercado de trabajo ($ilf=0$) lo que lleva a que $\theta_j = 1$ si $registered=0$. Del juego anterior se vuelven más prevalentes los equilibrios (iv) de autocontrol y (i) de disuasión por el escarmiento potencial, en detrimento de los equilibrios (ii) y (iii), en los que la persona revela sus verdaderos gustos de outsider indolente, y que tienden a disminuir. Al disminuir la frecuencia con que los equilibrios (ii) y (iii) aparecen.

En el segundo subjuego donde el indolente eligió sí participar en el mercado de trabajo ($ilf=1$), $\theta_j = 1$ si $registered=1$. En este subjuego se vuelven más prevalentes los equilibrios (ii) y (iii), en los que la persona revela sus verdaderos gustos de outsider indolente, y que tienden a disminuir.

Los resultados de los dos subjuegos en conjunto llevan a que aumenta la frecuencia de las combinaciones ($ilf=1; registered=0$) y ($ilf=0; registered=1$) y disminuya para las combinaciones ($ilf=0; registered=0$) y ($ilf=1; registered=1$). Es decir que, para el subgrupo de ciudadanos que íntimamente son indolentes, la coherencia modifica la frecuencia relativa en que encontramos las combinaciones de decisiones identitarias (1;0) y (0;1), y disminuya la de (0;0) y (1;1), de forma que para todo el grupo la correlación de residuos del modelo es más alta por este grupo de indolentes que se vé afectado por la existencia de un costo de incoherencia en sus decisiones identitarias. Por otro lado, si alguna o las dos decisiones no son identitaria, no sé dá este efecto de la coherencia modificando la frecuencia relativa de la combinación de decisiones.

4. Aplicación de la Teoría de Economía de la Identidad al grupo de Aborígenes en Argentina.

I. Resumen

En este trabajo se aplica el modelo de Economía de la Identidad al grupo de aborígenes en Argentina, con las mismas herramientas econométricas que en el primer paper se usaron para el grupo de aborígenes en Estados Unidos. La peculiaridad de este trabajo, además del cambio de objeto de estudio, es el empleo de una técnica de Statistical Matching sobre la información disponible en una encuesta y en un censo para obtener una base de datos sobre la cual realizar las regresiones de dos etapas y la descomposición de Oaxaca-Blinder.

Un hecho descrito en diversas etnografías y explicado económicamente por George Akerlof y Rachel Kranton (Akerlof & Kranton, 2010) consiste en la decisión individual de autoexclusión del mercado de trabajo como comportamiento concomitante a sufrir discriminación laboral. Este trabajo pretende comprobar la hipótesis de autoexclusión, luego de haber atribuido el diferencial de ingresos del grupo de aborígenes en Argentina entre menor dotación de capital humano y discriminación.

El Instituto Nacional de Estadísticas y Censos de Argentina releva datos sobre poblaciones indígenas pero, a diferencia del Census Bureau de Estados Unidos, no reporta en una sola encuesta todas las variables necesarias para las regresiones econométricas de este análisis, por lo que un paso previo requerido es un trabajo fusión de datos entre dos fuentes. Los datos de origen étnico se obtienen del Censo al que sólo puede accederse consultando tablas; los datos de ingreso están disponibles a través de los microdatos de una encuesta anual de hogares urbanos.

Luego de completar el anterior paso y generar una base de datos artificial con todas las variables requeridas se realizan regresiones de dos etapas de Heckman y descomposiciones de Oaxaca-Blinder por sexo para el grupo objeto de estudio y su complemento poblacional, se analizan los resultados y se comparan esos resultados con los previamente obtenidos para Estados Unidos.

II. Teoría de fusión de datos.

Cuando las variables de interés de un estudio econométrico se encuentran disponibles en múltiples fuentes estadísticas una posible herramienta a emplear es la fusión de datos, también conocida como *Statistical Matching*, que involucra técnicas diversamente aplicadas en estudios sobre preferencias de consumidores y por Institutos Nacionales de Estadística. (Räessler, 2002) El propósito es potenciar la complementariedad entre diversas fuentes que comparten variables comunes y que cumplen con algunos criterios, aplicando una diversidad de herramientas para cruzar datos desde una fuente donante a otra fuente receptora.

Aunque los registros de las dos fuentes deben provenir de la misma población no es un requisito que haya superposición de unidades, incluso puede no existir ningún individuo en común entre las dos fuentes. Es decir que a diferencia del ejercicio de *Data Merge* ó *Record Linkage* en el que se completan los datos de un individuo con datos de ese mismo individuo unívocamente identificado proveniente de otra fuente, el *Statistical Matching* comúnmente resulta en datos de un individuo que se completan con datos de otra unidad similar o próxima empleando una función de distancia calculada con variables comunes entre las fuentes. El resultado es una nueva base de datos sintética, con registros que no se corresponden con individuos reales, pero útiles para poder hacer inferencia estadística con las variables que no se observan conjuntamente.

Es deseable que la base de datos donante tenga un tamaño mucho mayor que la receptora pero son condiciones del *Statistical Matching* que los datos provengan de una misma población, que las variables de cruce estén armonizadas (lo que significa que tengan igual definición de categorías y que hayan sido relevadas de forma similar, lo cual es más probable que suceda cuando las bases son producidas por el mismo instituto) y que las distribuciones empíricas en las dos bases de datos sean muy próximas. Esto se puede verificar calculando la distancia de Hellinger, de acuerdo a la siguiente fórmula:

$$HD = \sqrt{\frac{1}{2} \times \sum_{i=1}^K \left(\sqrt{\frac{n_{D_i}}{N_D}} - \sqrt{\frac{n_{R_i}}{N_R}} \right)^2}$$

Donde K es el total de celdas en la tabla de contingencia, D es la base de datos donante, y R es la base de datos receptora. (Leulescu & Agafitel, 2013, pág. 14)

Para cada variable de cruce; el umbral aceptado es una diferencia de hasta 5%.

Finalmente para que los resultados sean válidos debe cumplirse que toda la relación entre las variables de las dos fuentes (la receptora y la que donante) esté completamente explicada por las variables de cruce; a esta condición se la llama *Conditional Independence Assumption (CIA)* y puede expresarse en el caso de variables continuas como:

$$f(x, y, z) = f(y|x) \times f(z|x) \times f(x)$$

Y en el caso de variables categóricas como:

$$P(X = i, Y = j, Z = k) = P(Y = j|X = i) \times P(Z = k|X = i) \times P(X = i)$$

Donde x ó X son las variables de cruce, y ó Y las variables exclusivas de la base receptora y z ó Z las variables exclusivas de la base donante.

En una aplicación tradicional de *Statistical Matching* donde sólo están disponibles la fuente donante y la fuente receptora y no hay disponibilidad de una tercera fuente con información de la distribución conjunta, no es posible testear explícitamente si la CIA se cumple o no. (D’Orazio, 2013)

III. Datos del INDEC.

Los datos de ingreso y de grupo étnico para el año 2010 están separados en el cuestionario ampliado del Censo 2010 y en la Encuesta Anual de Hogares Urbanos (EAHU) o en la Encuesta Permanente de Hogares (EPH). (INDEC, 2018)

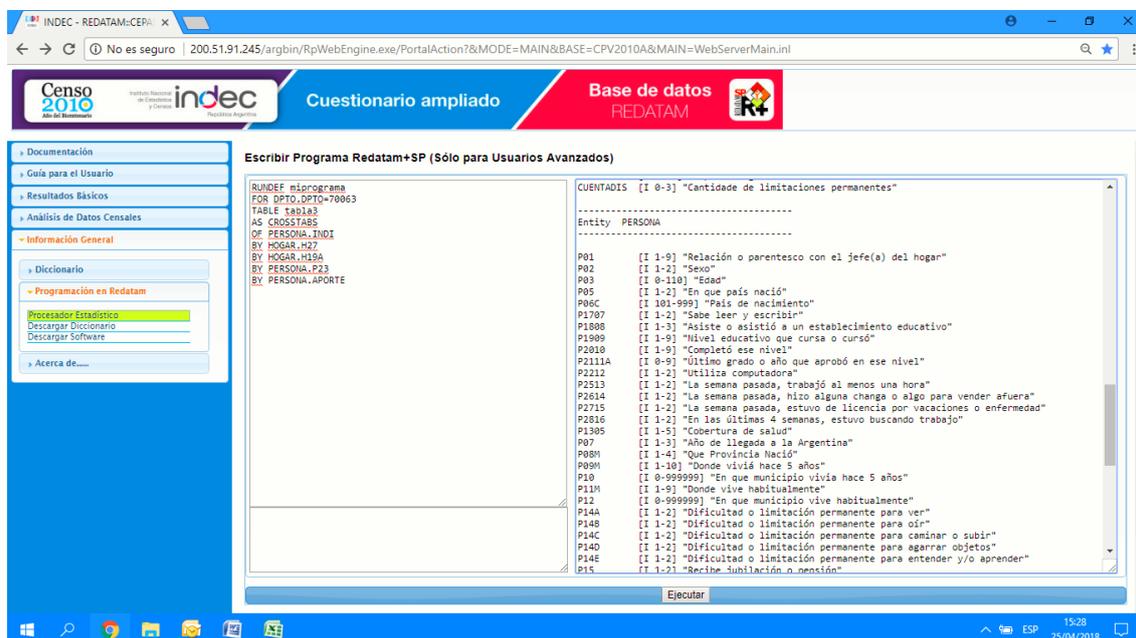
Para la EPH y la EAHU se tiene acceso a microdatos. Entre ambas encuestas se elige trabajar con la encuesta anual que “resulta de la extensión del operativo continuo “Encuesta Permanente de Hogares - 31 Aglomerados Urbanos”, a través de la incorporación a la muestra de viviendas particulares pertenecientes a localidades de 2.000 y más habitantes”, siendo el tamaño de la base aproximadamente de 46.000 hogares (INDEC, 2011). Esto garantiza que la EAHU cubre un universo de población mucho más amplio que el contemplado en la EPH, y permite pasar de una medición de poco más del 60% de población urbana al 90% de la población total del país. (www.ellitoral.com, 2011)

Respecto al cuestionario ampliado del Censo 2010, el INDEC sólo permite consultar datos a través de su página web empleando el software REDATAM y limitando el acceso al nivel de partido o departamento provincial. (INDEC, 2015). ⁵

⁵ El autor oportunamente visitó la sede del Instituto Nacional de Estadísticas y Censos, en Julio A. Roca 609, CABA, con la intención de obtener acceso a microdatos del Censo 2010 en lo relativo a poblaciones indígenas (denominación empleada por el Instituto) explicando claramente que el propósito era realizar un trabajo de investigación en el marco de una tesis doctoral. No logró que la persona que lo atendió en la biblioteca del INDEC le permitiera acceso a los microdatos, y sólo recibió la sugerencia de hacer una solicitud por email y de emplear el software REDATAM. La solicitud por email no logró un mejor

La Ilustración 1 muestra un ejemplo del código REDATAM empleado para obtener tablas del Censo para el departamento 9 de Julio de la provincia de Buenos Aires.

Ilustración 1 – Ejemplo de código REDATAM.



Una restricción adicional sobre las consultas es que sólo pueden cruzarse hasta cuatro variables. El resultado de la consulta puede descargarse en formato de planilla de cálculo Excel.

El tamaño de los datos relevados con el Censo 2010 es muy importante, como lo explica el documento Diseño de Registro del cuestionario ampliado del Censo2010:

“se aplicó a la totalidad de segmentos censales de las localidades de menos de 50.000 habitantes, según proyecciones basadas en el Censo de 2001.

En las localidades con 50.000 o más habitantes, el cuestionario ampliado se aplicó a una muestra probabilística de los segmentos censales [...]

Para mejorar la precisión de las estimaciones, ciertos segmentos con características específicas fueron incluidos en la muestra con probabilidad 1: segmentos en villas o asentamientos, segmentos en barrios cerrados, comunidades indígenas, población afrodescendiente.

[...] La tasa final de muestreo, en las localidades donde se aplicó la muestra, fue por lo tanto superior al 10%.” (INDEC, 2010 c)

resultado. De haber obtenido acceso a los microdatos del Censo 2010 se podría haber intentado realizar un *Statistical Matching* tradicional. Dicho lo anterior, el autor considera valiosa la metodología desarrollada para aplicar en otras circunstancias en las que se establecen barreras arbitrarias al acceso a la información de parte de investigadores.

A. Fracción del grupo minoritario por subdominio geográfico.

Con los datos del Cuestionario Ampliado del Censo 2010 (INDEC, 2015) se obtienen los datos de la Tabla 14 de ratio de la minoría de aborígenes en el total de la población por subdominio geográfico.

Los diez subdominios con mayor proporción de aborígenes son Resto de Chubut (11,97%), Resto de Salta (16,61%), Resto de Neuquén (11,31%), Resto de Jujuy (10,82%), Rawson-Trelew (8,65%), Rio Negro (7,23%), Resto de Formosa (7,06%), Neuquén Plottier (6,32%), Santa Rosa-Toay (5,49%) y Resto del Chaco (5,17%).

Tabla 14 – Porcentajes de aborígenes en la población por subdominio.

nombre	subdominio	aborigen	no aborigen	ratio
Gran Mendoza	subdominio 10	21777	881648	0.02410493
Corrientes Capital	subdominio 12	1921	353283	0.00540816
Gran Córdoba	subdominio 13	27703	1510611	0.01800868
Concordia	subdominio 14	509	44628	0.01127678
Formosa	subdominio 15	4543	228204	0.01951905
Neuquén Plottier	subdominio 17	22765	337059	0.06326704
Santiago del Estero - La Banda	subdominio 18	3114	403415	0.00765997
Jujuy-Palpalá	subdominio 19	14479	300790	0.04592586
Gran La Plata	subdominio 2	16762	770051	0.02130366
Río Gallegos	subdominio 20	3350	107912	0.03010911
Gran Catamarca	subdominio 22	2160	195032	0.01095379
Gran Salta	subdominio 23	12422	561301	0.02165156
La Rioja	subdominio 25	2106	177918	0.01169844
San Luis-El Chorrillo	subdominio 26	3402	198929	0.01681403
Gran San Juan	subdominio 27	4592	495867	0.00917558
Gran Tucumán	subdominio 29	10158	978443	0.01027513
Bahía Blanca-Cerri	subdominio 3	9932	287627	0.03337825
Santa Rosa-Toay	subdominio 30	6794	116796	0.05497209
Ushuaia-Río Grande	subdominio 31	3563	119522	0.02894748
CABA	subdominio 32	61876	2768940	0.02185801
Gran Buenos Aires	subdominio 33	204196	10497909	0.01907998
Mar del Plata-Batan	subdominio 34	15350	589143	0.02539318
Río Cuarto	subdominio 36	3388	239934	0.01392394
Gran Rosario	subdominio 4	20632	1316241	0.01543303
Resto de Buenos Aires	subdominio 40	61949	3256753	0.01866664
Resto de Catamarca	subdominio 41	4767	160348	0.02887079
Resto de Córdoba	subdominio 42	20051	1454834	0.01359496
Resto de Corrientes	subdominio 43	3208	626886	0.0050913
Resto del Chaco	subdominio 44	34176	626606	0.05172054
Resto de Chubut	subdominio 45	17310	86875	0.16614676
Resto de Entre Ríos	subdominio 46	9215	784827	0.01160518
Resto de Formosa	subdominio 47	19820	260808	0.07062731
Resto de Jujuy	subdominio 48	38066	313517	0.10827031
Resto de La Pampa	subdominio 49	7292	184228	0.03807435
Gran Santa Fe	subdominio 5	9081	510483	0.01747812
Resto de La Rioja	subdominio 50	1829	149821	0.01206067
Resto de Mendoza	subdominio 51	19249	798611	0.02353581
Resto de Misiones	subdominio 52	10867	758599	0.01412278
Resto de Neuquén	subdominio 53	20592	161400	0.11314783
Río Negro	subdominio 54	45375	581391	0.07239544
Resto de Salta	subdominio 55	67107	493512	0.11970162
Resto de San Juan	subdominio 56	3370	169468	0.01949803
Resto de San Luis	subdominio 57	4592	221483	0.02031184
Resto de Santa Cruz	subdominio 58	6202	144529	0.04114615
Gran Paraná	subdominio 6	278	38527	0.00716403
Resto de Santa Fe	subdominio 60	18552	1289049	0.01418781
Resto de Santiago del Estero	subdominio 61	8155	436040	0.01835905
Resto de Tucumán	subdominio 62	9159	442808	0.02026475
Posadas	subdominio 7	2139	319713	0.00664591
Gran Resistencia	subdominio 8	7128	380126	0.01840652
Comodoro Rivadavia	subdominio 9	7845	176727	0.04250374
Rawson-Trelew	subdominio 91	18124	191262	0.08655784
TOTALES		952992	38500434	

Nota: datos obtenidos a partir del Cuestionario Ampliado del Censo 2010 del INDEC.

El Censo permite obtener información a nivel de áreas (más detallada que los subdominios de la EAHU). Por ejemplo, Resto de Chubut incluye el área Gastre con un ratio de (28,28%), las áreas Iruya y San Carlos en Resto de Salta tienen ratios de (56,31%) y (45,79%) respectivamente; en Resto de Jujuy las áreas Susques, Rinconada, Santa Catalina y Cochinocha tienen ratios de (54,16%) (48,3%) (46,88%) y (44,38%)

respectivamente; en Resto de Formosa el área Bermejo tiene un ratio de (25,04%); en Resto de Chaco el área General Güemes tiene un ratio de (18,72%) y en Río Negro el área Ñorquinco tiene ratio (43,45%) Es decir que la proporción de aborígenes, lejos de ser homogénea puede llegar a superar a la mitad de la población en algunas áreas.

Con estos datos se crea una nueva base de datos con las variables “subdominio” y “evento” siendo esta última la combinación de las cuatro variables categóricas de cruce; el porcentaje de aborígenes en un subdominio dado es la frecuencia relativa de la encuesta complementaria del censo para cada uno de los 52 subdominios dada una combinación específica de variables de cruce; luego se realiza un *merge* en Stata con la base principal de la EAHU con que se venía trabajando y así se puede asociar para cada registro de la base receptora su probabilidad de corresponder a un aborigen, dada la información del Censo que es en este caso la fuente donante.

Las siguientes son las variables categóricas comunes entre el Censo y la EAHU:

- ¿cuenta con escritura de la vivienda y/o terreno donde habita? (sí=1 / no=0)
- ¿cuenta con cuarto de cocina en la vivienda donde habita? (sí=1 / no=0)
- ¿es su estado civil soltero (5), casado/unido(1), separado/divorciado(3), viudo (4)?
- ¿es su categoría ocupacional obrero/empleador (3), patrón (1), trabajador independiente (2), trabajador familiar (4), inactivo (0)?

La combinatoria resulta en 80 ($=2 \times 2 \times 4 \times 5$) posibilidades para las que pueden calcularse las probabilidades que un registro corresponda a una persona que se identifique como aborigen en cada uno de los 52 subdominios geográficos. El resultado son 4160 ($=80 \times 52$) combinaciones, de las que en los hechos 1758 tienen cero ocurrencia en el Censo (es decir, no existe ninguna persona aborigen que cumpla con los criterios en el área geográfica.)

Las variables de cruce entre el cuestionario ampliado del Censo 2010 y la Encuesta Anual de Hogares Urbanos 2010 (EAHU) se muestran en la Tabla 15. El siguiente paso para un procedimiento de *matching* estadístico es verificar que las distribuciones de estas dos variables son muy próximas en las dos bases de datos, lo que se realiza calculando las distancias de Hellinger.

Tabla 15 – Variables de cruce entre el cuestionario ampliado y la EAHU.

Variable Censo2010	Cuestionario Ampliado	Variable EAHU 2010
hogar.H27	"La vivienda está escriturada"	ii7 "Régimen de tenencia"
hogar.H19A	"Instalación de Agua en Cocina"	ii4_1 "Tiene además Cuarto de cocina"
persona.P23	"Estado civil legal"	ch07 ¿Actualmente está...(5 opciones)
persona.CONDUCT	"Condición de actividad"	cat_ocup "CATEGORÍA OCUPACIONAL (Para ocupados y desocupados con ocupación anterior)"

B. Distancia de Hellinger (HD) entre las variables de cruce.

Para cada una de las cuatro variables de cruce por sexo y por subdominio se calculó las 420 HDs (distancias de Hellinger) entre las distribuciones del Censo y de la EAHU. El promedio de las HDs está en 0,03 que es menor al umbral comúnmente aceptado de 0,05 por lo cual se consideran las distribuciones de las variables muy próximas entre sí, siendo el valor mínimo de 0,0026 y el máximo de 0,12. Los subdominios con HD mayor a 0,5 para todas las variables son *Resto de Neuquén*, *Resto de Santa Fe* y *Resto de Buenos Aires*.

IV. Método no paramétrico propuesto para estimar coeficientes de grupos no identificados unívocamente.

Provisto que las fuentes estadísticas provengan de una misma población, hayan sido armonizadas y sus distribuciones sean muy próximas medidas por la HD, el proceso de *Statistical Matching* continuaría con la elección de un abordaje para generar una base de datos sintética con atribución de valores para las variables de la fuente donante a la fuente receptora, empleando alguna alternativa como el *hot deck* o el *Propensity Score Matching*.

Como se explicó previamente los microdatos del Censo no están disponibles para el autor quien sólo pudo acceder a consultas a través del programa *Redatam* disponible en la página del procesador de datos para el cuestionario ampliado (Indec, s.f.) por lo que se requiere diseñar otro abordaje para construir la base de datos sintética sobre la que realizar las regresiones.

El camino más directo es asignar la variable categórica **aborigen** a los registros que tengan mayor probabilidad de corresponder a aborígenes, de acuerdo a las variables de cruce, de modo que alrededor de 2.895 registros de la EAHU, un 2,4% de los 120.650

registros, se les atribuya la variable categórica *aborigen* el valor “1”, consistente con la proporción de aborígenes en el Censo. Con estos datos el umbral de probabilidad estaría en aproximadamente un 15%, atribuyendo categoría de aborígen a 2.826 registros, y para ese pooling de registros la proporción esperada de aborígenes sería de 21,8%.

Adicionalmente, y abordando el problema de que el pooling sobre el que se realizan inferencias tiene una baja proporción de registros correspondientes al grupo de aborígenes objeto de este estudio, se propone una metodología consistente en generar sucesivos diseños muestrales, en los que la variable de control sea la proporción de la minoría en el pooling, y sobre ellos correr regresiones y calcular los coeficientes de interés.

Sintéticamente la metodología de generar expansiones muestrales con una mayor participación de aborígenes puede graficarse con el siguiente ejemplo:

Supongamos que contamos con la siguiente encuesta:

Tabla 16 - Ejemplo encuesta con 2 registros con variable de expansión muestral.

	“pondera”: variable de expansión muestral original de la EAHU para reflejar la población total	probabilidad “aborprob” del registro de corresponder a un individuo minoritario
individuo 1	1000	1%
individuo 2	100	10%

La expansión muestral original tiene tamaño 1100 y un valor esperado de $1000 \times 1\% + 100 \times 10\% = 20$ individuos de la minoría (1,81% del total).

Ahora se genera otra expansión muestral, empleando la probabilidad disponible:

Tabla 17 - Ejemplo encuesta con 2 registros con variable de expansión muestral modificada.

	“aborpondera3”: nueva variable de expansión muestral para aumentar el % minoritario (primera iteración)	probabilidad “aborprob” de corresponder al grupo minoritario
individuo 1	$1000 \times 1\% = 10$	1%
individuo 2	$100 \times 10\% = 10$	10%

Esta nueva expansión muestral tiene tamaño 20 y un valor esperado de $10 \times 1\% + 10 \times 10\% = 1,1$ de la minoría (5,5% del total).

Además, esta nueva expansión muestral es una submuestra de la original y también es una muestra factible del universo muestral Ω que incluye a toda la población y a cada posible muestra que se extraiga de la misma sin reposición.

Si cada grupo es internamente homogéneo este método garantiza generar un pooling donde la participación de la minoría aumenta de 1,81% a 5,5%. En el caso de partir de una encuesta como la EAHU, cuyo tamaño es de 120.650 registros, el método puede emplearse iterativamente para alcanzar participaciones de la minoría mucho mayores que la original y suficientemente alto como para realizar inferencias sobre el grupo objeto de estudio.

Finalmente, se asume que las probabilidades obtenidas con la regla de Bayes son datos sin error, ya que son derivadas de las frecuencias observadas en un Censo que cubre a toda la población. (Estrictamente, el dato de la pertenencia a una etnia se releva en el Cuestionario Ampliado del Censo 2010, que no cubre a toda la población y que sí está sujeto a error muestral.)

V. Propiedades de los coeficientes OLS del pooling.

Como punto de partida supongamos dos relevamientos del mismo cuestionario censal para dos grupos complementarios y mutuamente excluyentes. Son complementarios ya que juntos totalizan la población. Son mutuamente excluyentes porque no hay individuos que pertenezcan a los dos grupos.

La variable ingreso es el resultado de procesos diferentes correspondientes a cada grupo y que pueden ser descriptos por las siguientes ecuaciones de Mincer con aprendizaje en el trabajo:

Población no aborígen:

$$\ln(wage) = \alpha^{no-a} + \beta_1^{no-a} sexo + \beta_2^{no-a} yearse + \beta_3^{no-a} exper + \beta_4^{no-a} exper^2 + e$$

Población aborígen:

$$\ln(wage) = \alpha^{ab} + \beta_1^{ab} sexo + \beta_2^{ab} yearse + \beta_3^{ab} exper + \beta_4^{ab} exper^2 + e$$

Teniendo dos relevamientos separados, uno para cada grupo, se pueden realizar las anteriores regresiones. También se pueden integrar los datos de las dos muestras para formar una única muestra con *pooling*; la ecuación de regresión en forma matricial, de acuerdo a lo desarrollado por Leamer (1978) puede ser escrita como:

$$\begin{bmatrix} Y^{no_aborigen} \\ Y^{aborigen} \end{bmatrix} = \begin{bmatrix} X^{no_aborigen} \\ X^{aborigen} \end{bmatrix} \beta + \begin{bmatrix} u^{no_aborigen} \\ u^{aborigen} \end{bmatrix}$$

donde:

$\begin{bmatrix} Y_{no_aborigen} \\ Y_{aborigen} \end{bmatrix}$ es un vector con n^* registros de ingreso, la variable dependiente, correspondientes a no aborígenes, y n registros correspondientes a aborígenes; en total cuenta con n^{**} registros, y su dimensión es $n^{**} \times 1$.

$\begin{bmatrix} X^{no_aborigen} \\ X_{aborigen} \end{bmatrix}$ es una matriz con n^* registros de individuos no aborígenes y n correspondientes a individuos aborígenes; en total cuenta con n^{**} registros, y sus dimensiones son $n^{**} \times k$, donde k son los cinco regresores (ordenada al origen *_cons*, *yearse*, *exper*, *exper2* y variable categórica *sexo*.)

β es un vector de $k \times 1$ coeficientes para la muestra con pooling.

$\begin{bmatrix} u^{no_aborigen} \\ u_{aborigen} \end{bmatrix}$ es un vector de $n^{**} \times 1$ residuos de la regresión, los primeros n^* corresponden a registros de no aborígenes y los últimos n corresponden a registros de aborígenes.

Suponemos además que se cumplen los supuestos usuales del modelo OLS, entre ellos el que siendo la variable explicativa intrínsecamente estocástica, para los propósitos del muestreo los valores de X son fijos o conocidos, y el residuo sólo captura el comportamiento estocástico de la variable Y . (Gujarati, 2004)

El estimador OLS del coeficiente poblacional β es:

$$b^{**} = \left(\begin{bmatrix} X^* \\ X \end{bmatrix}' \begin{bmatrix} X^* \\ X \end{bmatrix} \right)^{-1} \begin{bmatrix} X^* \\ X \end{bmatrix}' \begin{bmatrix} Y^* \\ Y \end{bmatrix} = (X^{*'}X^* + X'X)^{-1}(X^{*'}Y^* + X'Y) = \\ = (X^{*'}X^* + X'X)^{-1}(X^{*'}X^*b^* + X'Xb)$$

Que se interpreta como el promedio matricialmente ponderado de los estimadores para cada grupo. (Leamer, 1978)

En nuestro caso, los estimadores desconocidos son:

$$b^* = \begin{bmatrix} b_0^* \\ b_1^* \end{bmatrix} \text{ para los } n^* \text{ registros de no aborígenes}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \text{ para los } n \text{ registros de aborígenes.}$$

El estimador conocido es el de la muestra con pooling:

$$b^{**} = \begin{bmatrix} b_0^{**} \\ b_1^{**} \end{bmatrix} \text{ que se estima directamente con la EAHU original, ó con diseños muestrales generados ad-hoc con los datos de la EAHU.}$$

El uso de pesos en los datos en un modelo correctamente especificado, es decir donde los pesos son función de las variables independientes y por lo tanto no correlacionados con el término de error, resulta en estimadores insesgados de los estadísticos

poblacionales y de los coeficientes de las correlaciones, pero con un menor nivel de eficiencia (Winship & Radbill, 1994). Dado que los tamaños de las bases de datos a emplear (INDEC, 2010 a) (INDEC, 2010 b) (United States Census Bureau, s.f.) son significativamente grandes esta pérdida de eficiencia resulta reducida.

La idea entonces es averiguar los coeficientes de las regresiones para los dos grupos por separado, a partir de muestras con pooling, para las que conocemos las participaciones de cada grupo.

Lee (2015) remarca que “un aspecto molesto de γ_l [el vector b^{**} de coeficientes OLS del pooling] es que no necesariamente se ubicará entre los componentes de γ_0 y γ_1 [vectores b^* y b de coeficientes de cada grupo por separado] a menos que ‘x’ sea un escalar [es decir que se emplee un único regresor]. Esto se debe a que ω_l [el peso ponderado matricialmente de la ecuación de Leamer] toma en cuenta la relación entre los regresores.”

En la Tabla 3 del paper de Lee (2015) se presentan resultados de una ecuación tipo Mincer para medir el gap de ingresos en el que uno de los coeficientes del pooling está por encima de los coeficientes de cada grupo por separado. En ese estudio de Economía Laboral para americanos-mejicanos y para americanos-blancos no hispanos del sur de Estados Unidos, donde uno de los regresores es la edad, se presentan los siguientes resultados: para los Hispanos el coeficiente es 0.085, para el grupo mayoritario es de 0.078 y para el pooling con 15% de Hispanos es de 0.089. Este comportamiento no convexo de parámetros del pooling se verifica en esta tesis para aborígenes en el coeficiente de años de educación en la ecuación de Mincer, tanto con datos de Argentina como de Estados Unidos.

Lee ofrece una sugerencia para evitar la “*molestia*” de la no convexidad de los parámetros del pooling, garantizando que cada elemento del pooling sea una combinación convexa de los elementos de cada grupo por separado; emplea en los productos matriciales sólo los elementos de la diagonal, y reemplaza por cero los demás; de esta forma artificial se determina la colinealidad entre los regresores y por lo tanto se fija una correlación igual a 1. (Lee, 2015)

A.Representatividad del diseño muestral y de la expansión muestral.

Según el INDEC la expansión muestral se justifica para evitar las fluctuaciones en las estimaciones de totales debido al error muestral, que depende del tamaño de la muestra y del dominio bajo estudio (Comari & Hoszowski, 2009).

El INDEC incluye en la base de datos de la EAHU la variable “pondera”, que es el producto de tres componentes (Piselli, 2008):

- 1) Inversa de la probabilidad con que fue seleccionada cada vivienda en la muestra.
- 2) Corrector por no respuesta a nivel de estrato.
- 3) Ajuste demográfico, se utilizan proyecciones demográficas de población elaboradas por INDEC a nivel de total de población de aglomerados.

Desde 2009 se introducen novedades metodológicas, tendientes a calibrar la variable pondera:

“[la variable pondera], si bien determina que no haya variaciones muestrales del total estimado de población, no corrige las variaciones muestrales al interior de la población, e.g. en grupos de sexo y tramos de edad. [...]. Una técnica que permite atenuar parcialmente estos efectos es la calibración. [...]. Dados ciertos totales poblacionales de subpoblaciones (e.g. total de población según sexo y tramos de edad, total de hogares unipersonales, etc.) conocidos por fuentes externas (registros administrativos, proyecciones demográficas), y denominados en general valores marginales poblacionales, se ajustan las ponderaciones para que los totales estimados coincidan con estos valores externos.” . [...].En el caso de la calibración de la EPH se optó por una postura conservadora, tomando como marginales las siguientes proyecciones:

- Total de varones.
- Total de mujeres.
- Total de población entre 0 y 14 años.
- Total de población entre 15 y 29 años.
- Total de población entre 30 y 49 años.
- Total de población entre 50 y 64 años.
- Total de población de 65 años o más.

Los tramos de edad son seleccionados en virtud de la asociación de la condición de actividad con rangos etarios.”

Es decir que la expansión muestral generada con “pondera” asegura coherencia en los informes del INDEC para los totales y para grupos etarios al interior de la población, pero puede resultar en un sobre ó sub-registro de otras variables de interés para nuestro estudio.

B. Selección de la expansión muestral.

El dato oficial INDEC del CENSO2010 a nivel país es que existe un 2,4% de población indígena ó descendiente de pueblos indígenas u originarios, para un total de 955.032 indígenas en una población total de 39.671.131 personas en viviendas particulares.

Una primera expansión muestral factible entonces es la generada con “*pondera*” con todas las consideraciones sobre su representatividad expuestas en la sección anterior. Pero con esa expansión muestral el pooling generado contiene una fracción muy baja de aborígenes, similar al 2,4%. Se pueden generar iterativamente sucesivas expansiones muestrales con el producto de “*pondera*” y potencias de “*aborprob*” que resultarán en un pooling con mayor proporción de aborígenes.

Para calcular las participaciones de aborígenes en cada diseño muestral se promedia la frecuencia relativa “*aborprob*” de aborígenes correspondiente a cada registro en la muestra por la cantidad de registros para cada expansión muestral.

En la medida que se generan sucesivas expansiones muestrales, el porcentaje del grupo de aborígenes en el pooling aumenta dado que en la ponderación sube el peso de algunos registros a expensas de otros que por redondeo pasan a tener nula ponderación; sin embargo todos los registros con probabilidad de corresponder a un aborigen tienen valor estadístico para las inferencias, ya que al menos una persona en el Censo fue registrada como aborigen para las variables de cruce correspondientes.

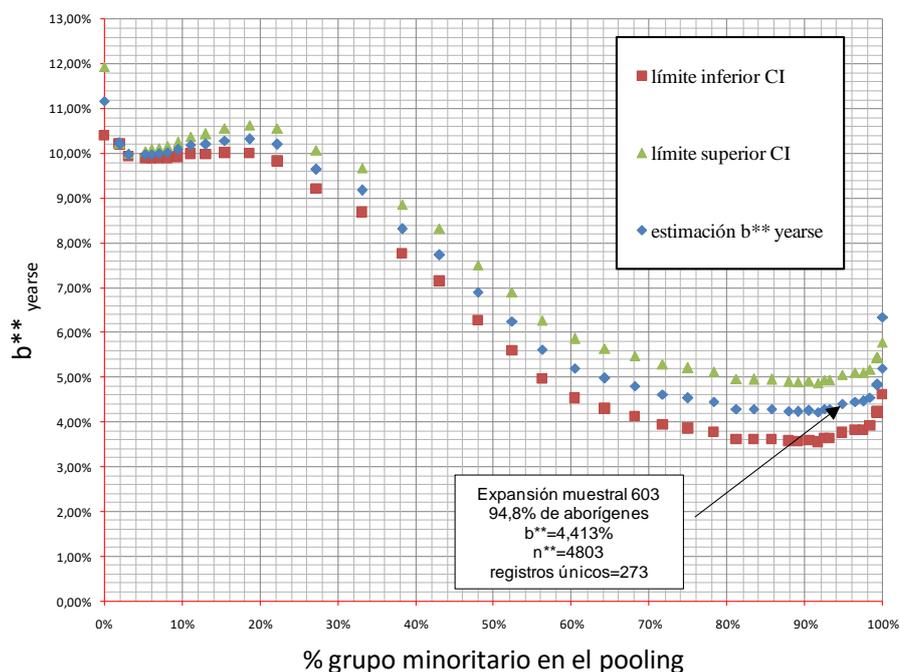
En la medida que el porcentaje del grupo de interés ya no aumenta o que en la base de datos sintética quedan muy pocos registros con alta ponderación ya no se justifica continuar iterando para generar nuevas ponderaciones.

La Tabla 18 muestra los diversos pooling que en Argentina lograron generarse, la estimación del coeficiente de retorno a la educación *yearse* y los intervalos de confianza para cada pooling, y que alcanzan un estimado de hasta 99% de aborígenes; sin embargo empleando los criterios ya presentados para el método propuesto se elige la expansión muestral “603” como representativa del grupo minoritario de aborígenes dado el número de registros únicos que involucra. Los datos se grafican en la Ilustración 2.

Tabla 18 – Expansiones muestrales generadas con el método no paramétrico sobre datos de la EAHU.

Nombre de la muestra	CI inf 95%	b**	CI sup 95%	% aborígenes	n** observaciones luego de la expansión	registros para expansión (con probabilidad>0)
EAHU original	0.097213	0.099317	0.101421	3.380%	44588	120651
"pondera00"	0.102162	0.102276	0.102390	1.900%	35282131	110303
"pondera0"	0.102100	0.102212	0.102325	1.948%	35982687	110303
1 "pondera"	0.102048	0.102160	0.102271	2.016%	36723490	110334
2 "aborpondera105"	0.099382	0.099694	0.100005	3.103%	4696922	110334
2 "aborpondera3"	0.098935	0.099736	0.100538	5.321%	740820	106161
2 "aborpondera301"	0.098900	0.099861	0.100822	6.065%	523119	98857
2 "aborpondera302"	0.098926	0.100076	0.101225	6.980%	371919	89030
2 "aborpondera303"	0.098945	0.100316	0.101688	8.096%	266385	78006
2 "aborpondera304"	0.099249	0.100880	0.102512	9.474%	191631	66784
2 "aborpondera305"	0.099931	0.101865	0.103799	11.103%	139552	57259
2 "aborpondera306"	0.099757	0.102034	0.104312	13.045%	102768	47567
2 "aborpondera307"	0.100174	0.102892	0.105609	15.446%	75635	37998
2 "aborpondera308"	0.099960	0.103148	0.106336	18.604%	55491	28529
2 "aborpondera309"	0.098318	0.102007	0.105696	22.223%	41563	20900
3 "aborpondera41"	0.092062	0.096410	0.100758	27.215%	30683	14007
3 "aborpondera401"	0.086825	0.091782	0.096738	33.114%	23118	9354
3 "aborpondera402"	0.077637	0.083161	0.088685	38.292%	18565	6885
3 "aborpondera403"	0.071457	0.077330	0.083203	43.040%	15549	5388
3 "aborpondera404"	0.062779	0.068971	0.075163	47.953%	13256	4316
3 "aborpondera405"	0.056026	0.062490	0.068954	52.372%	11598	3578
3 "aborpondera406"	0.049611	0.056209	0.062807	56.279%	10407	3075
3 "aborpondera407"	0.045347	0.052016	0.058685	60.405%	9372	2576
3 "aborpondera408"	0.043054	0.049781	0.056507	64.275%	8541	2196
3 "aborpondera409"	0.041296	0.048051	0.054807	68.106%	7849	1850
4 "aborpondera51"	0.039393	0.046173	0.052953	71.704%	7277	1539
4 "aborpondera501"	0.038622	0.045385	0.052147	74.991%	6820	1308
4 "aborpondera502"	0.037758	0.044525	0.051293	78.330%	6404	1079
4 "aborpondera503"	0.036143	0.042885	0.049626	81.189%	6080	901
4 "aborpondera504"	0.036135	0.042851	0.049568	83.400%	5847	794
4 "aborpondera505"	0.036170	0.042850	0.049529	85.737%	5613	688
4 "aborpondera506"	0.035796	0.042454	0.049112	87.884%	5413	576
4 "aborpondera507"	0.035709	0.042340	0.048971	89.126%	5297	530
4 "aborpondera508"	0.035936	0.042551	0.049165	90.528%	5173	474
4 "aborpondera509"	0.035583	0.042190	0.048796	91.663%	5078	422
5 "aborpondera61"	0.036322	0.042910	0.049499	92.511%	5005	387
5 "aborpondera601"	0.036371	0.042908	0.049445	93.170%	4951	354
5 "aborpondera603"	0.037640	0.044130	0.050620	94.799%	4820	273
5 "aborpondera605"	0.038263	0.044480	0.051019	96.483%	4695	175
5 "aborpondera607"	0.038263	0.044641	0.051019	97.476%	4622	125
6 "aborpondera71"	0.039110	0.045439	0.051769	98.358%	4554	90
7 "aborpondera81"	0.042089	0.048240	0.054391	99.305%	4472	66
7 "aborpondera801"	0.042245	0.048385	0.054526	99.312%	4471	66
7 "aborpondera802"	0.042318	0.048448	0.054578	99.332%	4469	65
8 "aborpondera9"	0.046101	0.051939	0.057777	99.917%	4413	32
sólo registros con $p=1$	-0.023299	0.063364	0.150027	100%	31	31

Gráfico 10 – Evolución del coeficiente b^{**}_{yearse} en función del % del grupo minoritario en el pooling - datos de la Tabla 18.



En el Anexo se aplica y valida el método no paramétrico propuesto con datos PUMS de Estados Unidos.

VI. Resultados.

A continuación se replican las regresiones de Heckman y la descomposición de Oaxaca-Blinder realizadas previamente con datos PUMS en el paper I de esta tesis.

A. Regresiones de Mincer y Heckman por sexo para toda la muestra.

Para toda la muestra (N=45.458) los resultados de la regresión típica de Mincer con el logaritmo del ingreso como variable dependiente y con coeficiente de ordenada al origen como regresor, muestran un retorno por cada año de educación de +0,0996169, un retorno por cada año de experiencia laboral de +0,0481058 y un coeficiente negativo para el cuadrado de la experiencia de -0,0006826, siendo todos los resultados estadísticamente significativos.

Por sexo, para los varones (N=27.325) los resultados de la regresión de Mincer muestran un retorno a la educación de +0,0944433, a la experiencia de +0,0520139 y un coeficiente negativo para el cuadrado de la educación de -0,0007446; y para las mujeres (N= 18133) el retorno a la educación de +0,1378913, a la experiencia de +0,044989 y un

coeficiente negativo para el cuadrado de la educación de $-0,0005996$ siendo todos los resultados estadísticamente significativos.

En una regresión de dos etapas tipo Heckman donde la segunda etapa es la ecuación de Mincer, y la primera es la ecuación de selección, los retornos a la educación de varones (N= 58561) caen a $+0,0902506$ y los de mujeres (N= 62089) a $+0,1374975$, donde en la primera etapa entran las variables explicativas son si la persona es estudiante, si es un trabajador familiar no remunerado y su edad.

B.Regresiones imputando la categoría de aborígen en función de la probabilidad.

Usando un abordaje más próximo a un Statistical Matching tradicional se asigna la variable categórica *aborigen=1* a los registros con una probabilidad de ser aborígenes mayor a 15%; de esta forma se identifican 2.830 registros de los 120.656 de la EAHU, y sobre ese grupo se realizan las regresiones. Para este pooling de varones con una proporción esperada de 22% de aborígenes, el retorno a la educación en el grupo de aborígenes cae a $+0,0536972$, y en la ecuación de selección el coeficiente de *percentage_abor* es fuertemente positivo e igual a $+5,101584$ lo cual indica autoexclusión en la medida que el individuo aborígen está más rodeado de otros de su mismo grupo.

C.Poolings donde se controla el porcentaje de aborígenes.

Usando la variable de expansión muestral como peso en el comando de la regresión se obtienen resultados sobre una muestra artificial en la que se aumentó la proporción del grupo de interés, en este caso de aborígenes. Se pueden emplear dos opciones de pesos; con el “*frequency weight*” [*fweight*] un mismo registro será tenido en cuenta tantas veces en la regresión como lo indique la variable de ponderación, y se modificará el número de registros pudiendo correrse la regresión sobre un mayor N; con “*sampling weights*” [*pweight*] el N se mantendrá pero los registros con mayor peso influirán más en el cálculo de los coeficientes. En los dos casos se obtiene igual valor del coeficiente, pero se prefiere el *pweight* al *fweight* ya que el segundo genera intervalos de confianza artificialmente chicos y estadísticos t artificialmente grandes, generando una certeza estadística que difícilmente pueda inferirse de los datos.

Usando la expansión muestral *aborpondera603* para una regresión de Heckman de dos etapas el retorno a la educación para los varones aborígenes es de $+0,0727816$, el retorno a la experiencia es $+0,0829766$ y en la ecuación de selección el coeficiente de *percentage_abor* es negativo e igual a $-14,16107$; todos coeficientes son estadísticamente

significativos. En esta expansión muestral el porcentaje esperado de aborígenes en el pooling de varones es de 96,87595% y N= 143 con 97 observaciones censuradas. La media de ingresos de los varones aborígenes es de \$1.082, menor a la media de \$ 2.521 del resto de la población de varones. El gap de ingresos es atribuible en un 88.2% a la discriminación.

Tabla 19 – Regresión de Heckman para varones sobre expansión muestral aborpondera603

Number of obs = 143; Censored obs = 97; Uncensored obs = 46 % esperado de aborígenes: 96,9%						
2da etapa: ecuación de Mincer; lnwage yearse exper exper2 if(sexo==0) [pweight=aborpondera603]						
lnwage	coef	robust s.e.	z	P> z	[95% coef int]	
yearse	0,0727816	0,0326283	2,23	0,026	0,0088313	0,1367319
exper	0,0829766	0,0230071	3,61	0	0,0378836	0,1280696
exper2	-0,0012962	0,0003746	-3,46	0,001	-0,0020304	-0,000562
_cons	5,169384	0,4137709	12,49	0	4,358408	5,98036
1ra etapa - ecuación de selección: lnwage>0 select(percentage_abor trabajador_familiar_no_remun estudiante ch06)						
percentage_abor	-14,16107	5,024003	-2,82	0,005	-24,00794	-4,314207
trabajador_familiar~n	-7,500984	0,2508876	-29,9	0	-7,992714	-7,009253
estudiante	-7,386353	0,2766791	-26,7	0	-7,928634	-6,844072
edad	0,0700492	0,0156361	4,48	0	0,039403	0,1006954
_cons	0,280836	0,5601115	0,5	0,616	-0,8169623	1,378634

Tabla 20 – Descomposición de Oaxaca Blinder del gap de ingresos de varones aborígenes en Argentina, comparado con la restante población de aborígenes.

Dotaciones	11,8%
Discriminación	88,2%

Para las mujeres, el retorno a la educación es de 0,0979973 y el retorno a la experiencia no es estadísticamente significativo (con el IC [-0,3489357;+0,0768192] no se puede rechazar la hipótesis nula de que es igual a cero). En la ecuación de selección el coeficiente de percentage_abor tampoco es estadísticamente significativo. El porcentaje esperado de aborígenes en el pooling de mujeres es 88,78% y N=130 con 113 observaciones censuradas.

La media de ingresos de las mujeres aborígenes es de \$925,99 menor a la media de \$1.869,86 del resto de la población de mujeres. El gap de ingresos con el resto de la población de mujeres es atribuible en un 82.7% a discriminación.

Tabla 21 – Regresión de Heckman para varones sobre expansión muestral aborpon-
dera603

Number of obs = 143; Censored obs = 97; Uncensored obs = 46 % esperado de aborígenes: 88,78%						
2da etapa: ecuación de Mincer; lnwage yearse exper exper2 if(sexo==0) [pweight=aborpondera603]						
lnwage	coef	robust s.e.	z	P> z	[95% coef int]	
yearse	0,0979973	0,2095798	0,47	0,64	-0,3127716	0,5087661
exper	-0,1360583	0,1086129	-1,25	0,21	-0,3489357	0,0768192
exper2	0,0026872	0,0010775	2,49	0,013	0,0005754	0,0047989
_cons	7,001024	4,078398	1,72	0,086	-0,9924891	14,99454
1ra etapa - ecuación de selección: lnwage>0 select(percentage_abor trabajador_familiar_no_remun estudiante ch06)						
percentage_abor	12,82616	8,031703	1,6	0,11	-2,915693	28,56801
trabajador_familiar~n	-13,39124	2,236573	-5,99	0	-17,77484	-9,007638
estudiante	-6,117448	1,202117	-5,09	0	-8,473554	-3,761341
edad	0,2219741	0,0765339	2,9	0,004	0,0719704	0,3719778
_cons	-7,081797	2,56768	-2,76	0,006	-12,11436	-2,049236

Tabla 22 - Descomposición de Oaxaca Blinder del gap de ingresos de mujeres aborígenes
en Argentina, comparado con la restante población de aborígenes.

Dotaciones	17,3%
Discriminación	82,7%

VII. Conclusiones.

De acuerdo a la teoría de Economía de la Identidad y empleando el método de expansiones muestrales, la descomposición de Oaxaca-Blinder comparando dos modelos de Heckman con autoselección verifica discriminación de ingresos hacia los aborígenes en Argentina. El grupo de varones aborígenes muestra un comportamiento de autoexclusión del mercado de trabajo, concomitante con el perjuicio que sufren por la discriminación, en la medida en la que el individuo está más expuesto a la mirada de los otros de su propio grupo, que es directamente proporcional al porcentaje de aborígenes en el subdominio. No es ese el caso de las mujeres aborígenes. Este resultado repite lo hallado en el caso de aborígenes varones y mujeres en Estados Unidos.

Por otro lado, con el método de *Statistical Matching* los resultados de las regresiones son muy diferentes en cuanto a signo y valor de los coeficientes, atribuible a que la inferencia en realidad se realiza sobre un pooling que tiene una proporción de aborígenes de menos del 25%, y por lo tanto se parece mucho más a la población complementaria que al grupo objeto de estudio.

VIII. Anexo Estadístico.

En este anexo se aplica el método propuesto de expansiones muestrales con una base de datos que contiene todas las variables de interés, y se evalúa cuánto difieren las inferencias de las que resultan empleando la variable categórica racial *raceian*, que identifica individuos aborígenes en Estados Unidos. También se compara con los resultados empleando el método de *Statistical Matching*.

A. Estimación retorno a la educación aborígenes en Estados Unidos

La oficina de Estadísticas y Censos del Gobierno Federal de Estados Unidos conduce la encuesta quinquenal de comunidades “*American Community Survey*” (ACS) donde se relevan para hogares y las familias datos de ingresos, raciales y de pobreza. Los microdatos “*Public Use Microdata Survey*” (PUMS) son accesibles en su página de Internet. (United States Census Bureau, s.f.)

La base de microdatos PUMS servirá para evaluar el método no paramétrico propuesto contra un método paramétrico de *Statistical Matching* usando la función logística.

B. Resultados con grupos perfectamente identificados.

En la Tabla 23 se muestran los resultados de la regresión OLS empleando la ecuación de Mincer con aprendizaje en el trabajo, con la variable categórica *sexo* y empleando la expansión muestral *pwgtp* que provee la encuesta ACS, el retorno a la educación para la población en general es 13,36%.

Tabla 23– Regresión OLS de Mincer para toda la población con expansión muestral *pwgtp* - Output de Stata

regress lnpincp sexo yearse exper exper2 if(raceian==1) [pw= pwgtp]						
N= 11.176.661						
lnpincp	coef.	std.err.	t	P> t	[95% Conf. Interval]	
sexo	-0,4287593	.0008509	-503.88	0,000	-.4304204	-.4270849
Yearse	0,1336571	.0001616	827.27	0,000	.1333452	.1339785
Exper	0,0612674	.0000792	773.43	0,000	.0611154	.0614259
exper2	-0,0007604	1.15e-06	-659.54	0,000	-.0007627	-.0007582
_cons	21,35932	.0026739	7987.89	0,000	21.35396	21.36445

Una observación crítica de los resultados de las regresiones empleando expansiones muestrales consiste en que los estadísticos t son artificialmente altos, producto de una inferencia con N=219.349.556 cuando la muestra tiene sólo quince millones de registros; sin embargo esa es la expansión muestral provista por el Census Bureau según sus

criterios de ajuste de las medias marginales a la población real y esa es la justificación por la cual también se procede de igual forma en esta tesis.

En la Tabla 24 se muestra el resultado de la regresión del mismo modelo de Mincer para el grupo de aborígenes (variable categórica *racaian*) el coeficiente es de 11,41% empleando la expansión muestral *pwgtp*, que es la provista en la ACS.

Tabla 24 – Regresión de Mincer para aborígenes con expansión muestral *fwgtp* - Output de Stata

regress lnpincp sexo yearse exper exper2 if(racaian==1) [fw= pwgtp]						
N= 194.794						
lnpincp	coef.	std.err.	t	P> t	[95% Conf. Interval]	
sexo	-0,2988892	.0074506	-40,12	0,000	-.3135344	-.2843284
yearse	0,1141192	.0013832	82,51	0,000	.1114188	.1168408
exper	0,0550136	.0006578	83,63	0,000	.0537269	.0563056
exper2	-0,0006379	.0000102	-62,52	0,000	-.0006579	-0,0006342
_cons	21,36606	.0211433	1010,53	0,000	21.32449	21.40737

C. Identificación con método no paramétrico.

Empleamos un método no paramétrico para asignar a los registros de la base de personas una probabilidad de corresponder a la minoría étnica aborigen. Las variables de cruce son prácticamente las mismas que están disponibles para el caso de Argentina. La Tabla 8 transcribe las categorías de estas variables en la encuesta ACS.

En función de las cuatro variables de cruce se generan 80 eventos, directamente asimilables con el caso argentino, y se calculan las frecuencias relativas de esos eventos para cada uno de los 51 estados de Estados Unidos, obteniendo 4080 probabilidades que incorporamos en la base de datos como la variable “*prob*”.

Luego de realizar un *merge* entre la base de datos de hogares y la base de datos de personas similar al realizado con datos del INDEC y asignando un valor de probabilidad a cada registro de la base de datos PUMS, se generan trece variables de expansión muestral, a partir de la variable de ponderación *pwgtp* provista.

D. Generación de expansiones muestrales.

La primera expansión muestral se obtiene como el producto de la variable de expansión muestral *pwgtp* presente en la ACS por la probabilidad calculada en el teorema de Bayes de que el registro corresponda a un aborigen [$round(pwgtp \times prob)$] y es una submuestra de la expansión *pwgtp*; iterando sucesivamente, la expansión muestral “n” se obtienen como: $round(pwgtp \times prob^n)$ y es una submuestra de la expansión muestral (n-1) que por transitividad es submuestra de la expansión *pwgtp*. Así se

demuestra que todas las expansiones muestrales sucesivamente obtenidas son factibles, es decir están incluidas en el espacio muestral Ω de todas las muestras posibles y que en sus bordes coincide con un censo completo de la población.

Es importante aclarar que “ n ” no requiere ser un entero, podría ser la potencia $2^{1/2}$, con la que se generaría una expansión muestral incluida en la muestra generada con $n=2$ y que incluiría a la muestra generada con $n=3$.

¿Cuál es el propósito de generar expansiones muestrales de esta forma? La expectativa es que cada expansión muestral sucesiva así obtenida, que es una muestra factible dentro del espacio muestral, contendrá una creciente participación de la minoría. Obtener al azar muestras grandes con una participación del grupo minoritario alto no parece factible empleando simulaciones.⁶

Con el método no paramétrico se obtienen expansiones muestrales donde la variable de control es la participación de aborígenes en la muestra. Si las probabilidades están correctamente calculadas cualquier otro parámetro de la población se modificará indirectamente por tener una mayor participación de la minoría.

Para cada expansión muestral se realiza la regresión de Mincer, con *sexo* como variable categórica, y se estima el porcentaje de aborígenes en el pooling como la media de la variable “*prob*” (primer columna % aborígenes). Al contar con la variable categórica “*racaiian*” también es posible saber el porcentaje exacto de aborígenes en el pooling para compararlo con el estimado (segunda columna % aborígenes).

Junto con las diversas estimaciones también se muestran los intervalos de confianza con un NC=95%. Los casos extremos, donde (%)=0 ó (%)=1 se calcularon a partir de registros con probabilidad igual a 0 ó 1 y en esos casos no se aplicó ninguna expansión muestral, pero estos casos no son representativos de los grupos, y por lo tanto no se deben tener en cuenta al estimar el parámetro correspondiente a la minoría aborígen.

E. Interpretación de resultados.

No todos los puntos del gráfico tienen la misma significatividad estadística, ya que a medida que aumenta la proporción de aborígenes (%) disminuye el tamaño del diseño muestral n^{**} , aumentan los errores estándares de la estimación y especialmente relevante

⁶ Como ejemplo, la probabilidad de obtener una muestra de tamaño $n^{**}=7500$, con 60% o más aborígenes, de una población de 219 millones en la que hay 3,3 millones de aborígenes puede calcularse con la distribución hipergeométrica donde $k=4,500$, $N=250$ millones, $K= 7$ millones y $n=7500$:

$$1 - P(X = k) = 1 - \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = 1 - \frac{K!}{k!(K-k)!} \times \frac{(N-K)!}{(n-k)!(N-K-n+k)!} \approx 0$$

es que disminuye el número de observaciones a partir del cual se realiza la expansión muestral. Esto resulta en que el control sobre la proporción de aborígenes (%) sea imperfecto ya que las expansiones muestrales no mantienen constante el “n”.

La tabla 25 muestra que pueden realizarse estimaciones de b^{**} con un número de registros n^{**} alto, incluso para una participación estimada de la minoría en el pooling superior al 90%. Sin embargo el mecanismo para generar expansiones muestrales no debería llevarse más allá de muestras con un estimado de más de 90% de aborígenes; el límite está dado por qué tan segregada está la minoría en la población de acuerdo a las variables que se emplean en el método no paramétrico. Si existiera una total segregación en la iteración del método se obtendría una expansión muestral con 100% de aborígenes. Si todos los registros tuviesen igual probabilidad de corresponder a un aborigen, sucesivas iteraciones no podrían aumentar en nada la participación del grupo minoritario (%).

Tabla 25 – Expansiones muestrales generadas con el método no paramétrico sobre datos PUMS de la ACS.

Nombre expansión muestral	CI inf 95%	b^{**}	CI sup 95%	% aborígenes estimado	% aborígenes real	n^{**} observaciones luego de la expansión	registros únicos
prob=0	0.1463094	0.1527649	0.1592204	0%	0%	23417	23417
PUMS original	0.1330944	0.1333255	0.1335566	2%	1.74%	11176063	11176063
pwgtp	0.1336056	0.1336571	0.1337085	2.06%	1.50%	219349556	313826
expansión 1	0.1277869	0.128257	0.1287272	11.13%	6.38%	2736264	136831
expansión 2	0.1481293	0.1500108	0.1518922	34.39%	19.47%	186626	38276
expansión 3	0.1576755	0.1612909	0.1649063	53.95%	32.85%	50511	14419
expansión 4	0.1479508	0.1530402	0.1581296	69.74%	42.92%	24842	8570
expansión 5	0.1332821	0.1394569	0.1456317	77.27%	48.55%	16438	6763
expansión 6	0.1200143	0.127212	0.1344097	83.16%	54.55%	12050	5529
expansión 7	0.1102053	0.118266	0.1263266	86.80%	58.92%	9403	4957
expansión 8	0.1022821	0.1112333	0.1201846	89%	63%	7518	4522
expansión 9	0.10066	0.1106384	0.1206169	90%	65%	6343	4213
expansión 10	0.1006417	0.1117822	0.1229226	90.67%	66.40%	5342	3911
expansión 11	0.097867	0.1097648	0.1216627	91.12%	67.50%	4661	3727
expansión 12	0.0945545	0.1075332	0.1205119	91.41%	68.27%	4054	3542
expansión 13	0.094148	0.1082149	0.1222818	91.76%	68.32%	3469	3363
prob=1	0.061262	0.3050161	0.5487701	100%	100%	18	18

Empleando la variable *racaiian* puede verificarse que el (%) de la minoría aborigen calculado como el promedio de *prob* resulta sobreestimado en el orden de un 50%. Por ejemplo para la expansión muestral#8 donde se calcula un (%)=89%, se sabe que en realidad ese diseño muestral tiene un (%) de 63%. La diferencia del 41% es importante, sin embargo no invalida el hecho que se ha logrado aumentar más de 30 veces la participación de la minoría en esa muestra, comparando con el 1,74% original.

F. Selección de la expansión muestral.

Para completar el método propuesto necesitamos responder la siguiente pregunta: ¿Cómo proceder para elegir la muestra que nos dé el estimador de b_{yearse} ?

La respuesta es establecer un criterio: avanzar sucesivamente hasta donde la pérdida de precisión dada por el aumento del intervalo de confianza no se justifique por un aumento de la participación de la minoría en la muestra.

También debe cuidarse que el tamaño de la muestra seleccionada no sea muy pequeña y que el número de registros únicos a partir del cual se genera la expansión muestral no sea menor a los 100 registros, para no perder significatividad estadística.

En este caso, entre la expansión muestral 8 y la 13 la participación del grupo minoritario (%) aumenta imperceptiblemente por lo que se preferirá la estimación de b_{yearse} que provenga de la regresión con el mayor n^{**} posible, que es $b=11,12\%$, para un $(\%)=89\%$, y que debe compararse con un valor “real” del $11,4\%$ cuando la regresión se realiza sólo sobre los registros identificados con la variable categórica *racaian*.

La tabla 26 muestra las regresiones correspondiente a la expansión muestral 8 y a la generada sólo para el grupo identificado como aborigen por la variable (*racaian*=1).

Tabla 26 – Resultados comparados de la regresión de Mincer con expansión muestral 8 y la misma regresión de Mincer exclusivamente para aborígenes con variable *pwgtp* original de la ACS, empleando *frequency weights [fw]* en lugar de *sampling weights [pw]*.

	fw=amerindianpondera8			fw=pwgtp
	Coef	[95% IC]		Coef
sexo	-0,1583586	-0,213247	-0,1034701	-0,2988892
yearse	+0,1112333	0,1022821	0,1201846	+0,1141192
exper	+0,0353688	0,0311377	0,0396	+0,0550136
exper2	-0,0002442	-0,0003124	-0,000176	-0,0006379
_cons	21,11864	20,99736	21,23992	21,36606
N	7.518			3.304.024
% american indians	89%			100%

La estimación obtenida para el coeficiente de interés “*yearse*” con la variable categórica “*racaian*” está dentro del IC al 95% de la inferencia empleando el método no paramétrico. Sin embargo, los demás coeficientes (*sexo*, *exper*, *exper2* y *_cons*) no están incluidos en los intervalos de confianza de sus estimadores.

G.Comparación con el uso de un método paramétrico y selección de registros por “*matching*.”

Para emplear esta herramienta es imprescindible tener acceso a los *microdatos* de una encuesta ó censo donde se identifiquen unívocamente los aborígenes. Es posible evaluar el desempeño del método paramétrico con los *microdatos* PUMS donde se

obtienen 3.457 valores de probabilidad para los 15.637.457 registros. La tabla 27 muestra la tabulación de la probabilidad calculada con la función logística para los siete valores más bajos y los siete más altos de probabilidad⁷:

Tabla 27 – Probabilidad logit de que un registro de la ACS corresponda a un aborígen (racaian=1) empleando las variables de cruce.

Registros con menores probabilidades		Registros con mayores probabilidades	
Probabilidad Logit de racaian=1	frecuencia absoluta	Probabilidad Logit de racaian=1	frecuencia absoluta
0	624	0,3979589	1,122
0,0032063	1	0,4032503	420
0,0033531	91	0,4114459	14
0,0034876	3	0,4167925	23
0,0035067	12	0,4304613	1
0,0035118	66	0,4388033	685
0,0035362	23	0,444237	697

Se observa que la probabilidad más alta de que un registro corresponda a un aborígen norteamericano es de 44,42%. Por otro lado la tabla 28 muestra que el método no paramétrico logra generar 18 registros con probabilidad igual a 1 y 4,222 registros con probabilidad mayor a 80%.

Tabla 28 – Probabilidades calculadas de pertenecer al grupo aborígen.

amerindianprob	racaian=0	racaian=1	Total
0,8196078	46	209	255
0,8347107	60	303	363
0,8496732	23	130	153
0,8571429	1	6	7
0,8645833	26	166	192
0,903819	136	1.278	1.414
0,9095238	38	382	420
0,9287879	47	613	660
0,9302326	3	40	43
0,9583931	29	668	697
1	0	18	18

Como las probabilidades bayesianas se calculan con los datos de la misma encuesta, las relaciones entre las columnas de la variable categórica coinciden exactamente con las de las probabilidades. Esto no necesariamente será así en el caso de cruzar dos bases de datos, como se realiza en esta tesis para el caso argentino. Con la salvedad anterior en el

⁷ Para la regresión Logit que identifique la probabilidad de un registro de corresponder a un aborígen se emplearon como regresores las mismas variables que para el método no paramétrico, por cada uno de los 52 estados, sin evaluar críticamente si esas variables, que son todas categóricas, resultan apropiadas para una estimación Logit.

caso de los microdatos PUMS el método no paramétrico demuestra ser superior generando mejores probabilidades para realizar un matching.

En la Tabla 29 se muestra la regresión de Mincer empleando registros con probabilidad generada por un modelo con probabilidad no paramétrica (*amerindianprob*) mayor al 80%. El coeficiente obtenido para *yearse* es de 10,1% con un intervalo de confianza de [9,4% - 10,8%] al 95% de NC; el valor calculado con la variable categórica es de 11,41% y cae afuera del IC.

Tabla 29 – Resultados Regresión de Mincer sobre registros con probabilidad no paramétrica mayor a 0,8.

	fw=pwgtp		
	Coef	[95% IC]	
Sexo	-0,19042	-0,230276	-0,150564
Yearse	+0,101056	0,940601	0,1080519
Exper	+0,045967	0,0427614	0,0491726
Exper2	-0,0004338	-0,000489	-0,0003786
_cons	21,21751	21,12429	21,31074
N	14.974		
% american indians	89%		

En cambio el uso del método propuesto permite generar una expansión muestral cuyo coeficiente de *yearse* es de 11,1%, con un IC de [10,2% - 12,0%] al 95% de nivel de confianza, algo más amplio que con el método de matching pero que sí incluye el valor del coeficiente calculado con la variable categórica.

¿Qué sucede? el método de matching, al truncar la muestra en probabilidades mayores al 80%, descarta casos que son relevantes para describir a la minoría aún cuando no estén tan nítidamente identificados con las variables que intervienen en la regresión Logit; el método de las expansiones muestrales logra muestras más representativas de la minoría porque la nueva información que incorpora se basa en el cambio de peso de cada registro (cuántas veces se computa el registro a través de la ponderación *frequency weight* de Stata), y sólo descarta observaciones cuando su valor de expansión muestral antes del redondeo es menor a 0,5.

H.Expansiones muestrales y pobreza.

Un último cuestionamiento al método propuesto consiste es el siguiente: el método puede estar identificando pobres, y el porcentaje de aborígenes en la muestra expandida aumenta ya que los aborígenes son pobres, pero no funciona para identificar aborígenes per se.

Para responder el cuestionamiento relativo a la capacidad de identificar aborígenes independientemente de su nivel de pobreza, se analizó para cada expansión muestral cuál es la composición en nueve clases de la variable *povpip*, que es el ratio entre ingreso y línea de pobreza como porcentaje, truncado para valores mayores a 500% para *racaian=0* primero y luego para *racaian=1*. Los resultados se muestran en la Tabla 30 y se muestran en los Gráficos 11 y 12.

Tabla 30 – Cambio en la composición por deciles de la variable *povpip*.

decil	pct	pcta0	pcta1	pcta0	pcta1	pcta20	pcta21	pcta30	pcta31	pcta40	pcta41	pcta50	pcta51	pcta60
1	70	70	43	51	25	53	15	57	7	60	0	53	0	39
2	123	124	82	95	61	97	48	105	37	110	29	106	23	99
3	175	176	120	135	91	138	75	150	60	158	50	152	47	138
4	231	233	159	180	124	185	103	201	83	213	72	208	65	191
5	293	294	206	231	161	237	136	257	111	275	93	267	86	250
6	364	366	261	291	207	298	179	325	144	342	127	335	117	308
7	451	453	329	365	266	373	232	407	196	437	167	432	148	407
8	501	501	425	467	345	477	310	501	270	501	228	501	207	501
9	501	501	501	501	485	501	446	501	402	501	340	501	307	501

pcta61	pcta70	pcta71	pcta80	pcta81	pcta90	pcta91	pcta100	pcta101	pcta110	pcta111	pcta120	pcta121	pcta130	pcta131
0	28	0	26	0	20	0	9	0	7	0	7	0	7	0
22	91	22	81	21	76	20	70	20	69	20	66	21	65	20
45	129	44	122	44	114	43	108	43	108	43	106	43	104	42
61	167	61	162	60	156	60	156	60	156	59	156	59	150	57
83	225	82	213	82	213	80	206	81	206	80	203	80	196	78
110	298	108	268	105	267	102	254	103	251	102	251	101	251	99
140	384	137	342	134	341	132	329	134	327	131	303	131	303	128
194	501	190	481	184	467	182	449	182	445	181	445	180	445	176
284	501	264	501	254	501	250	501	249	501	249	501	248	501	242

Nota: población no aborígen (columnas terminadas en 0) y aborígen (columnas terminadas en 1)

Gráfico 11 - distribución de *povpip* en deciles para *racaian=0*.

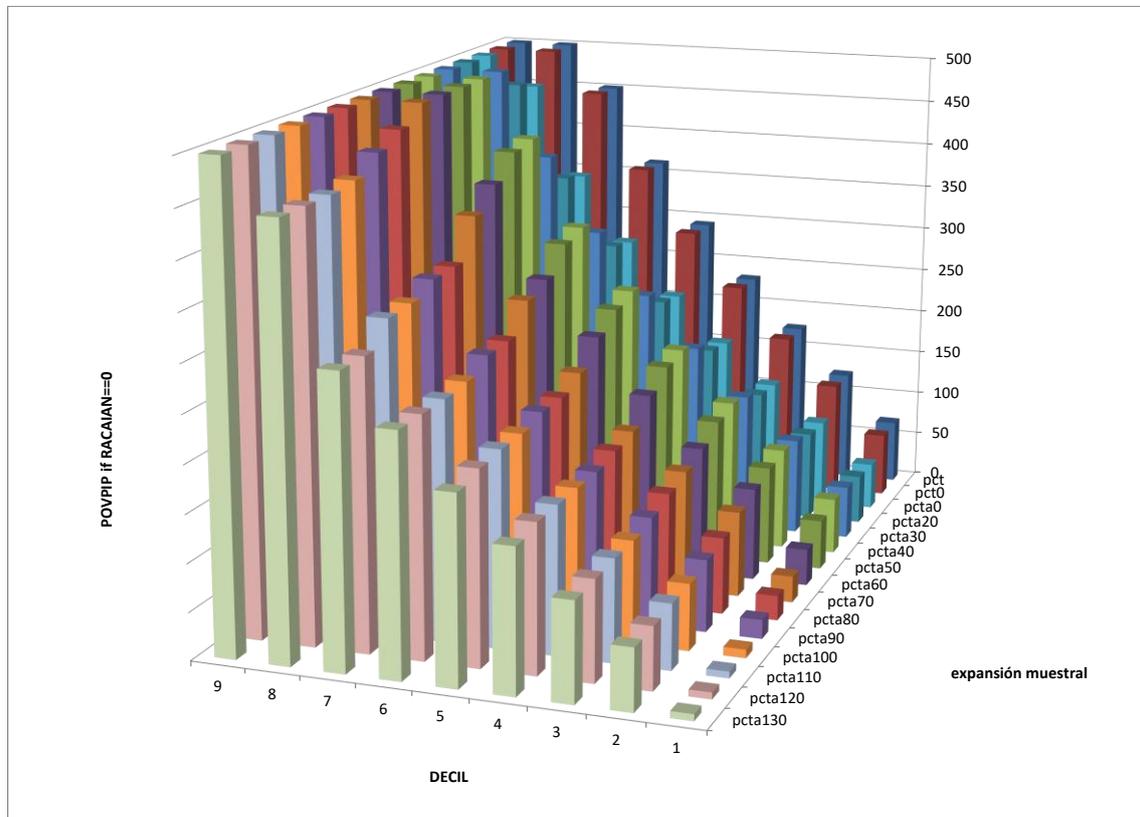
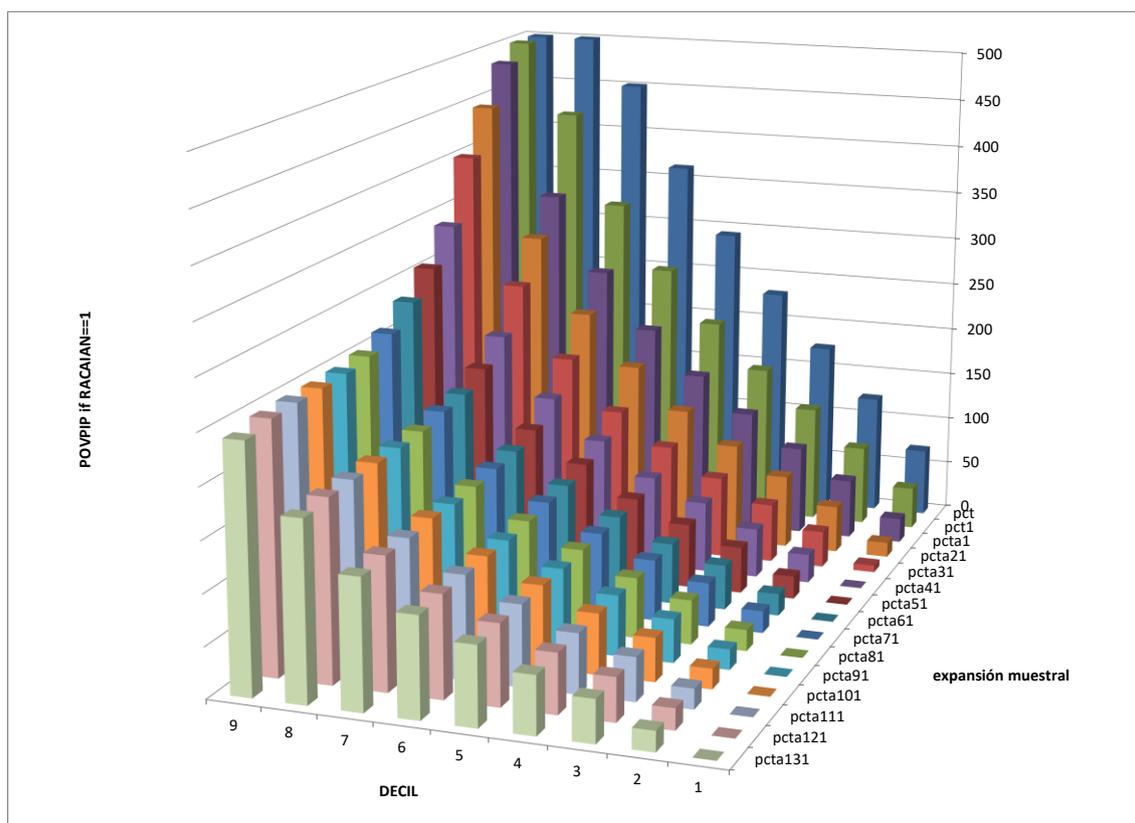


Gráfico 12 - distribución de *povpip* en deciles para *racaian=1*.



Para la minoría étnica aborigen, las sucesivas expansiones muestrales van reduciendo el *povpip* para cada uno de los deciles, pero que para la mayoría no aborigen *povpip* se reduce pero en una medida mucho menor. Para el caso de los aborígenes, la mediana está por encima de la línea de pobreza (*povpip=100*) hasta la tercera expansión muestral, mientras que para el resto de la población la mediana siempre es mayor que 100. La conclusión es que en las sucesivas expansiones muestrales aumenta el número de pobres porque aumentó el número de aborígenes, y no al revés, hecho que se confirma dado que el número de pobres aumenta en mucha menor medida para el grupo no aborigen.

I. Conclusiones.

Empleando datos del Census Bureau se compararon los resultados para validar la metodología de *Statistical Matching* propuesta. Se concluye, con datos de Estados Unidos, que la metodología no paramétrica muestra un mejor desempeño que la alternativa que emplea una función Logística para generar la base de datos.

La metodología propuesta, que implica el uso de pesos en las regresiones, dando mayor peso a unos registros y menos a otros pero usando la mayoría de ellos, también

tiene mejor desempeño que la alternativa de realizar inferencias sólo sobre los registros con probabilidad mayor a 80% de corresponder a aborígenes americanos.

También se evalúa favorablemente el efecto que las sucesivas expansiones muestrales tienen sobre la composición de pobres en la muestra, ya que los deciles de ingreso en la base de datos sintética generada no se modifican sustancialmente para el grupo de no aborígenes.

5. Resumen y Conclusiones.

I. Aportes de la variable identidad al análisis del comportamiento.

Tanto en el contexto de Estados Unidos como en el de Argentina se verificó que el modelo de utilidad con identidad es relevante para explicar decisiones como la participación en el mercado laboral, y específicamente en Estados Unidos la participación política. En términos generales, aquellos grupos de varones que sufren discriminación laboral, es decir que reciben un ingreso menor que el esperable por su dotación de capital humano, presentarán un comportamiento concomitante de mayor autoexclusión de la fuerza de trabajo en la medida que más expuestos estén a la mirada de otros de su propio grupo. Para otros grupos aventajados, que reciben un mayor ingreso que el que explica su capital humano acumulado, tienden a participar más en el mercado de trabajo en la medida en que están expuestos a la mirada de sus pares de grupo.

Entre mujeres la discriminación medida en términos pecuniarios es mucho menor que entre hombres, y es normal encontrar que la presión de su grupo lleva a una mayor participación en el mercado laboral.

En el segundo paper se incorpora adicionalmente como variable explicativa el nivel de segregación en el área metropolitana definida como la separación espacial de uno o más grupos de personas en base a su identidad de grupo. La segregación junto a la presión social de los pares, o “mirada del otro”, son las dos variables explicativas de la decisión identitaria. Es esperable una correlación importante entre las dos variables para algunos grupos, sobre todo aquellos segregados en entornos de ciudades, y que omitir una puede llevar a sesgos en los coeficientes por variable omitida en la otra.

La segregación en comunidades como afroamericanos en Estados Unidos se manifiesta en ghettos urbanos, ó “*inner cities*” donde el comportamiento individual está sobre expuesto a la mirada y evaluación de otros del mismo grupo. En el juego secuencial modelado en el segundo paper lo relevante para el grupo es que el individuo se adapte a las normas independientemente de cuáles son sus preferencias personales íntimas. Ese es el comportamiento que predice el modelo y que se testea empíricamente evaluando cómo las decisiones cambian en la medida que aumenta la presión de pares, medida por el porcentaje del propio grupo en el área metropolitana y por el nivel de segregación. Es decir que muchos individuos en soledad, fuera del ghetto, se comportarían diferente. Y también

implica que actitudes individuales de autoexclusión social y recalcitrantes que se verifican en contexto de ghettos no se darán cuando la persona está en otro contexto sin segregación.

II. Aportes de esta tesis a la Economía de la Identidad.

Siguiendo la posición antes mencionada de Hall (Introduction: Who Needs Identity?, 1996) en este trabajo no se toma la identidad como una categoría inamovible, dada por ejemplo por la etnia o el género o por otra circunstancia personal. Al contrario, se aborda la Economía de la Identidad con un concepto de identificación del individuo con un grupo, como la decisión la elección más importante que realiza una persona (Akerlof & Kranton, 2010). Las regresiones se corren para cada grupo por separado, pero teniendo en cuenta las variables que influyen en la decisión identitaria de la persona que busca maximizar su utilidad.

Las implicancias teóricas del concepto de identidad desarrolladas por el trabajo seminal de Akerlof y Kranton (Akerlof & Kranton, Economics and Identity, 2000) cubren los efectos desde el nivel individual hasta el nivel social. En los tres papers que integran esta tesis se modela y testea empíricamente el efecto de la identidad en las decisiones de las personas, al internalizarlo en gustos y preferencias, y también a través de las externalidades dadas por lo que otros piensan sobre el lugar social que le corresponde a la persona y qué esperan de su comportamiento. A diferencia de otros papers donde se descarta el efecto de la externalidad social sobre el individuo, en este trabajo se mide la misma incorporando como regresor el porcentaje del grupo de interés en el área metropolitana y el índice de divergencia.

En el segundo paper se introduce el modelo bivariado en la función general de utilidad con identidad, y se provee un modelo de juego secuencial, donde la elección de registrarse para votar es un subjuego de la decisión identitaria de participar en el mercado laboral. Los diversos equilibrios posibles de Nash en subjuegos son función de qué tan expuesta esté la persona a la mirada de los otros, del costo que esa mirada le represente, y del costo que tiene para la persona ser coherente en sus decisiones identitarias. Tanto en la función de utilidad como en el juego planteado se incorpora un costo asociado a la coherencia entre las decisiones, que es un aporte original de este trabajo.

III. Próximas oportunidades de investigación.

Queda pendiente a futuro incorporar la variable de segregación en el análisis empírico del modelo univariado de participación en el mercado laboral.

En el ámbito de Argentina sería interesante replicar el análisis para el grupo de afrodescendientes y para los inmigrantes, y compararlo con los resultados para Estados Unidos.

Fuera de Argentina y de los Estados Unidos, sería posible replicar los análisis para los demás países de Latinoamérica y comparar los resultados.

Referencias Bibliográficas.

- Akerlof, G. A., & Dickens, W. T. (June de 1982). The Economic Consequences of Cognitive Dissonance. *The American Economic Review*, 307-319.
- Akerlof, G. A., & Kranton, R. E. (August de 2000). Economics and Identity. *The Quarterly Journal of Economics*, CXV(3), 715-53.
- Akerlof, G. A., & Kranton, R. E. (2010). *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton, New Jersey: Princeton University Press.
- Akerlof, R. J. (2016). Value Formation: The Role of Esteem. *Games and Economic Behavior*.
- Anderson, B. (1993). *Imagined communities: reflections on the origin and spread of nationalism*. - rev. ed. New York: Verso.
- Arrow, K. (1973). The Theory of Discrimination. En O. Ashenfelter, & A. Rees, *Discrimination in Labor Markets* (pág. 196). Princeton, Estados Unidos: Princeton University Press.
- Austen-Smith, D., & Fryer, R. G. (2005). An Economic Analysis of Acting White. *Quarterly Journal of Economics*, 551-583.
- Becker, G. S. (1957). *The Economics of Discrimination*. The University of Chicago Press.
- Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, 101(3), 385-409.
- Becker, G. S. (2009). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
- Burnett, N. (1997). Gender economics courses in liberal arts colleges. *Journal of Economic Education*, 28 (Fall), 369-377.
- Chatman, J., Caldwell, D. J., O'Reilly, C. A., & Doerr, B. (2014). Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior*, págs. 785-808.
- Comari, C., & Hoszowski, A. (2 de diciembre de 2009).
https://www.indec.gov.ar/nuevaweb/cuadros/4/eph_metodologia_15.pdf.
Recuperado el 7 de abril de 2018, de
https://www.indec.gov.ar/nuevaweb/cuadros/4/eph_metodologia_15.pdf

- Congreso General Constituyente. (1853). *Constitución de la Confederación Argentina, Artículo 25*. Santa Fe.
- Congreso General Constituyente. (1853). *Constitución de la Confederación Argentina, Artículo 67, inciso 15*. Santa Fe.
- D'Orazio, M. (2013, November 21-22). *Statistical Matching: Methodological issues and practice with R-StatMatch*. Retrieved from https://www.eustat.eus/productosServicios/datos/Seminario_55_Slides.pdf.
- Davis, J. B. (2006). Social Identity Strategies In Recent Economics. *Journal of Economic Methodology* 13:3, págs. 371-390.
- Davis, J. B. (2007). Akerlof and Kranton on identity in economics: inverting the analysis. *Cambridge Journal of Economics*, 31, págs. 349-362.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Greene, W. H. (1998). Gender economics courses in liberal arts colleges: Further results. *The Journal of Economic Education*(29), 291-300. Obtenido de <https://doi.org/10.1080/00220489809595921>
- Greene, W. H. (2002). *Econometric Analysis* (Fifth Edition ed.). Upper Saddle River, New Jersey, Estados Unidos: Prentice Hall.
- Gujarati, D. N. (2004). *Basic Econometrics*. McGraw - Hill Companies.
- Hall, S. (1996). Introduction: Who Needs Identity? En S. Hall, & P. du Gays, *Questions of Cultural Identity* (págs. 1-17). London.
- Hall, S. (2011). *Questions of Cultural Identity & Paul du Gay*. London: SAGE Publications Ltd.
- Hauptman, O., & Iwaki, G. (11 de Junio de 1991). *Final Voyage of the Challenger*. Harvard Business Publishing.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153-162.
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003, May). Fifty Years of Mincer Earnings Regressions. *NBER WORKING PAPER SERIES*. Retrieved from <http://www.nber.org/papers/w9732>
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (February de 2008). Earnings functions and rates of return. *NBER Working Paper Series*, 1-39.
- INDEC. (2010 a). *Censo Nacional de Población, Hogares y Viviendas 2010*.
- INDEC. (2010 b). *Encuesta Anual de Hogares Urbanos*.

- INDEC. (2010 c). *Censo Nacional de Población, Hogares y Viviendas 2010. Censo del Bicentenario. Anexo metodológico. Precisión de las estimaciones*. Buenos Aires: INDEC.
- INDEC. (2011). *Encuesta anual de hogares urbanos Diseño de registro y estructura para las bases de microdatos individual y hogar. Vigente desde tercer trimestre de 2010*. Buenos Aires: Inst. Nacional de Estadística y Censos.
- INDEC. (2015). *Base de datos REDATAM - Cuestionario ampliado*. Recuperado el 2018, de <https://www.indec.gov.ar/bases-de-datos.asp?solapa=5>
- INDEC. (2018). *Instituto Nacional de Estadística y Censos. República Argentina*. Obtenido de <https://www.indec.gov.ar/>
- Indec. (s.f.). *Base de datos REDATAM*. Recuperado el 04 de 03 de 2022, de Cuestionario Ampliado Censo 2010: https://redatam.indec.gov.ar/argbin/RpWebEngine.exe/PortalAction?&MODE=MAIN&BASE=CPV2010A&MAIN=WebServerMain.inl&_ga=2.159158076.748321155.1646398692-1815844937.1644612830
- Iversen, V., Kalwij, A., Verschoor, A., & Dubey, A. (April de 2014). Caste Dominance and Economic Performance in Rural India. *Economic Development and Cultural Change*, Vol. 62(3), págs. 423-457.
- Kranton, R. E. (2016). Identity Economics 2016: Where Do Social Distinctions and Norms Come From? *American Economic Review: Papers & Proceedings*, 106(5), 405-409.
- Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inferences with Nonexperimental Data*. New York: Wiley.
- Lee, M.-J. (2015). Reference parameters in Blinder-Oaxaca decomposition: Pooled-sample versus intercept-shift approaches. *Journal of Economic Inequality*, 13, 69-82.
- Leulescu, A., & Agafitel, M. (2013). *Statistical matching: a model based approach*. Luxembourg: eurostat - Methodologies and Working Papers.
- Liebow, E. (1967). *Tally's Corner: A Study of Negro Streetcorner Men*.
- Maddala, G. (1983). *Limited-dependent and qualitative variables in econometrics*. New York, New York, Estados Unidos: Cambridge University Press.
- Mead, G. (1934). *Mind, Self, and Society*. Chicago: University of Chicago.

- Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), págs. 281–302.
- Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.
- Other and Belonging Institute - Most to Least Segregated Metro Regions. (2022, Enero 7). Retrieved from <https://belonging.berkeley.edu/most-least-segregated-metro-regions>
- Othering and Belonging Institute. (2022, Enero 9). Retrieved from <https://belonging.berkeley.edu/>
- Othering and Belonging Institute. (9 de Enero de 2022). *Othering and Belonging Institute - Technical Appendix*. Obtenido de <https://belonging.berkeley.edu/technical-appendix>
- O'Toole, J. (1995). *Leading Change*. Jossey-Bass Inc, Publishers.
- Oxford English Dictionary. (7 de Mayo de 2021). Obtenido de <https://en.oxforddictionaries.com/definition/segregation>
- Oxoby, R. J. (Oct. de 2004). Cognitive Dissonance, Status and Growth of the Underclass. (R. E. Society, Ed.) *The Economic Journal*, 114, 727-749.
- Phelps, E. S. (1972). The Statistical Theory of Racism and Sexism. *American Economic Review*, vol. 62, issue 4, págs. 659-661.
- Piselli, C. (2008). La Encuesta Permanente de Hogares:Fuente de datos socioeconómicos de Argentina. *Reunión de Discusión N° 184*. Salta: Universidad Nacional de Salta. Facultad de Ciencias Económicas, Jurídicas y Sociales. Instituto de Investigaciones Económicas. .
- Psacharopoulos, G. (1981). Returns to Education: an updated international comparison. *Comparative Education*, 17(3), 321-341.
- Psacharopoulos, G. (1993). *Returns to Investment in Education - A Global Update*. Office of the Director, Latin America and the Caribbean, The World Bank.
- Psacharopoulos, G., & Patrinos, H. A. (August de 2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2).
- PUMS Technical Documentation*. (s.f.). Obtenido de United States Census Bureau: https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMS_Data_Dictionary_2011-2015.pdf

- Räessler, S. (2002). *Statistical Matching. A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York: Springer Science+Business Media.
- Styker, S. (1980). *Symbolic Interactionism: a Social Structural Version*. Menlo Park: Benjamin/Cummings.
- Suskin, R. (1998). *A Hope in the Unseen: An American Odyssey from the Inner City to the Ivy League*. . Broadway Books.
- Tajfel, H. (1959). Quantitative Judgement in social perception. *British Journal of Psychology* vol.50, págs. 16-29.
- Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge, UK: Cambridge University Press.
- Thomas, K. (1996). The Defensive Self: A Psychodynamic Perspective. En R. Stevens, *Understanding the Self*.
- Turner, J. (1985). Social categorization and the self-concept: a social cognitive theory of group behavior. En E. Lawler, *Advances in Group Processes: Theory and Research, Vol.2* (págs. 77-122). Greenwich: JAI.
- Turner, J., Hogg, M., Oakes, P., Reicher, S., & Wetherell, M. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Blackwell.
- United States Census Bureau. (s.f.). Recuperado el 4 de Mayo de 2019, de <https://www2.census.gov/programs-surveys/acs/data/pums/2017/5-Year/>
- United States Census Bureau. (2020 y años previos). *CURRENT POPULATION SURVEY, November 2020. Voting and Registration Supplement. TECHNICAL DOCUMENTATION*. Retrieved from <https://www2.census.gov/programs-surveys/cps/techdocs/cpsnov20.pdf>
- United States Census Bureau. (2021). Obtenido de www.census.gov
- United States Census Bureau. (s.f.). *PUMS Data*. Obtenido de <https://www.census.gov/programs-surveys/acs/data/pums.html>
- Winship, C., & Radbill, L. (1994, Noviembre). Sampling Weights and Regression Analysis. *Sociological Methods & Research*, 230-257.
- www.ellitoral.com. (7 de Enero de 2011). Recuperado el 10 de Enero de 2019, de <http://www.ellitoral.com/index.php/diarios/2011/01/07/economia1/ECON-01.html>
- Young, P. (2001). The Dynamics of Conformity. En Durlauf, & Young, *Social Dynamics*. Cambridge, Massachusetts, Estados Unidos: MIT Press.

Young, P. (Octubre de 2014). The Evolution Of Social Norms. *Department Of Economics Discussion Paper Series(726)*.