

CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONOMICAS, A.C.



BRECHA SALARIAL POR GÉNERO CASO MÉXICO 2010

**TESINA PARA OBTENER EL GRADO DE:
MAESTRO EN ECONOMÍA**

PRESENTA:

LEONARDO IGNACIO ARROYO TREJO

ASESOR:

DRA. EVA OLIMPIA ARCEO GÓMEZ

JUNIO 2012

Agradecimiento

A Dios por todo.

Índice

1. Introducción	1
2. Revisión de literatura	3
3. Datos	8
4. Modelo	11
4.1 Primera fase: Corrección por selección	11
4.2 Segunda fase: Corrección por sesgo por ausencia de soporte común	14
5. Resultados	18
6. Conclusiones	28
7. Referencias	29
8. Anexos	32

Brecha Salarial por Género Caso México 2010

Leonardo Ignacio Arroyo Trejo

Asesor: Eva Olimpia Arceo Gómez

CIDE/ME

Mayo 2012

1. Introducción

En las últimas tres décadas la brecha salarial entre hombres y mujeres ha disminuido considerablemente. En contra de la idea popular de que la diferencia de salarios es por un hecho meramente discriminatorio, diversos estudios han mostrado evidencia que ésta diferencia se debe a más factores. Entre los factores más importantes se encuentran: la preferencia de las mujeres por algunos trabajos, las diferencias biológicas, la segregación ocupacional tanto vertical como horizontal, el costo de oportunidad de un empleador y/o simplemente el aumento en la fuerza laboral femenina.

En México la evolución de la brecha salarial, al igual que en otros países, ha ido a la baja. A principios de la década de los noventa esta brecha se calculaba alrededor del 20% (Brown *et al*, 1999), mientras que para el 2004 la brecha es estimada en cerca del 12% (Calónico y Ñopo, 2008). Es importante recalcar que una cantidad significativa de estudios, hechos para México, se han enfocado a este fenómeno dejando de lados dos fuentes muy importantes de sesgo: La primera fuente de sesgo surge a partir de la existencia de truncamiento incidental en las observaciones, es decir, una causa de sesgo en las estimaciones es la imposibilidad de observar el salario de reserva de los individuos que no están participando en la fuerza laboral. La otra fuente de sesgo viene dada por la comparación entre no comparables, es decir, si bien es cierto que las diferencias en los ingresos, pueden en parte explicarse a partir de características observables, Ñopo (2007) argumenta que estas mediciones pueden

tener un sesgo debido a las diferencias en el soporte de las distribuciones empíricas de las características individuales entre hombres y mujeres. Menciona que hay combinaciones de características individuales en las cuales es posible encontrar hombres en la fuerza laboral pero no mujeres.

Este documento analiza las brechas salariales corrigiendo por estas dos fuentes de sesgo. La estimación se realiza en dos fases: La primera fase contempla la corrección por truncamiento incidental (o corrección por selección) mediante el modelo de dos etapas de Heckman (1979). Posteriormente la segunda fase busca establecer un soporte común por medio de *Propensity Score Matching* para entonces calcular las brechas salariales entre hombres y mujeres.

Entre los principales resultados se observa que existe evidencia del truncamiento incidental, el cual es corregido al pronosticar los estimados del log salario. Se obtiene un balance del 81% en las variables utilizadas para el cálculo del *propensity score* y se calcula que la brecha salarial está entre -14% y 24% y que en promedio es de 6% la cual es menor que la presentada en otros estudios recientes. Asimismo, alternativamente se estima la brecha salarial por medio de regresión cuantil, obteniendo resultados muy similares.

El documento está organizado de la siguiente forma: En la segunda parte, se hace una revisión de literatura para conocer la situación de la brecha salarial en otros países y como se ha estimado. También se revisa este hecho para México. En la tercera sección se hace una breve descripción de los datos y como es que fueron construidas las variables de interés. En la sección 4 se describe la metodología econométrica de dos fases: la primera

fase es el modelo de dos partes de selección de Heckman y la segunda fase se refiere al cálculo de las brechas a partir de *propensity score matching*. En la sección 5 se muestran los resultados de cada una de las estimaciones por cada sexo. Por último en la sección 6, se presentan las conclusiones.

2. Revisión de literatura

En la literatura de brechas salariales se encuentran distintas metodologías de medición y factores que causan la brecha salarial. En Europa se estudia este fenómeno en 24 Estados miembros de la unión Europea por Christofides, Polycarpou y Vrachimis (2010). Ellos usan la descomposición de Oaxaca y Ransom (1994) corrigiendo por el procedimiento de Heckman (1979). Encuentran que la brecha salarial es positiva y significativa en la mayoría de los países. Encuentran que gran parte de la brecha salarial no es explicada por las características observables. Incorporan un índice de políticas de reconciliación trabajo-familia y encuentran que conforme sean más generosas estas políticas el salario de las mujeres permanece más bajo.

Los pisos de cristal y techos de cristal (*glass floors and glass ceilings*) son abordados por Dolado y Llorens (2004) en España. Por un lado, ellos encuentran que en el grupo con educación terciaria (licenciatura o mayor) la brecha salarial por género se amplía conforme se está en la parte superior de la distribución aunque se controle por covariables. Las diferencias pueden estar dadas porque las mujeres son promovidas menos frecuentemente dado que tienen trabajos con menos oportunidades de obtener un puesto más alto, es decir, las mujeres eligen o son relegadas a empleos donde el costo de interrupciones de carrera son bajos. Otra razón puede ser que las mujeres tienen menos probabilidades de ser promovidas en empleos con grandes responsabilidades.

Por otro lado, las mujeres en el grupo con educación baja pueden sufrir frecuentes interrupciones en su carrera laboral, debido a la discriminación social de las tareas familiares, y los empleadores entonces usan esta característica como una señal de baja productividad y en consecuencia bajan sus salarios. Sin embargo, conforme su permanencia en el empleo se expande, las mujeres son más fiables a los ojos de los empleadores y sus salarios convergen a los salarios de los hombres con las mismas características.

Por otra parte, la brecha salarial entre hombres y mujeres puede estar explicada, en alguna fracción, por diferencias biológicas, esto es lo que analizan Ichino y Moretti (2009). En el estudio empírico se concentran en el ausentismo y observan que en Europa las mujeres toman 7.6 días por enfermedad al año más que los hombres de su misma edad, ocupación y nivel educativo. En Estados Unidos y Canadá esto corresponde a 3.1 y 5.2 días respectivamente. Aunque estas diferencias son relativamente pequeñas tienen consecuencias no triviales en las carreras e ingresos de las mujeres. Encuentran evidencia de que las ausencias están relacionadas con el ciclo menstrual y que en general, el ciclo explica un tercio de la brecha de género en los días de ausencia, y dos tercios de la brecha de género en el número de episodios de ausencias. Se encuentra que la relación entre ingresos y ausentismo cíclico es negativa. Las mujeres de la muestra tienen ingresos inferiores a los de los hombres en un 13.5%. Además condicionadas sobre sus características demográficas se encuentra que el ausentismo cíclico explica un 14.1% de la brecha salarial.

En general estas causales de la brecha se resumen en un estudio realizado por el Banco Mundial (2011) donde se observan dos hechos estilizados: El primero de ellos es que después de haber controlado por características observables una fracción significativa de la brecha salarial permanece inexplicado; y en segundo lugar, que los salarios en sectores dominados por las mujeres son menores que los salarios en los sectores dominados por los hombres. Se han ofrecido dos explicaciones: por un lado la discriminación, y por otro lado, la selección voluntaria de hombres y mujeres en los diferentes sectores y ocupaciones. Se arguye que interpretar la brecha salarial como indicador de discriminación debe tomarse con precaución, ya que dicha brecha puede deberse a diferencias no observadas entre hombres y mujeres, y aunque se esté controlado por diversas características aún se encuentra una brecha salarial del 12 al 20 por ciento. En cuanto a la segregación ocupacional, se dice que las mujeres tienden a elegir vacantes que ofrecen más flexibilidad que los hombres, ya sea en horarios laborales o en la entrada y salida del mercado laboral y que, por lo tanto, no requieren de una inversión grande y/o continua, ya que esto les permitiría afrontar un costo de oportunidad menor ante su eventual salida del mercado.

Castro y Huesca (2007) en su revisión de literatura encuentran evidencia débil de que la creciente participación femenina en el mercado laboral produjo una reducción en el salario de los hombres vía sustitución de trabajo y que las mujeres han contribuido a la desigualdad entre los salarios de hombres calificados y no calificados en Estados Unidos.

Para el caso de México, se ha observado que la diferencia salarial se da por factores tales como: la educación, la segregación ocupacional, así como cuestiones del tamaño de firma y formalidad de estas. También se menciona, sin un análisis profundo, que las diferencias

biológicas y sociales pueden ser un componente de la brecha. Castro y Huesca (2007) encuentran que la disparidad en salarios se debe a que las mujeres se insertan en actividades de baja productividad, establecimientos de tamaño micro, jornadas inferiores a 35 horas semanales y empleos de corta duración.

Brown, Pagan y Rodríguez-Oreggia (1999) estudian la brecha salarial en el periodo 1987-1993. Calculan que la participación de la fuerza laboral en este periodo creció un 5%. Estiman que la brecha salarial para 1987 fue de 20.8% y que para 1993 incrementó a 22%. Este incremento se debe a tres factores: los niveles relativos de educación de los hombres aumentan, por cambios en la estructura regional de salario y cambios en las horas semanales por género. Además observan que las mujeres están sub-representadas en el ámbito profesional, de producción de precisión y operadores profesionales y sobre-representada en educación, en puestos de ocupaciones administrativas, ventas, entre otros.

Pagan y Ulibarri (2000) plantean que en países como México la brecha de salarios por género es heterogénea si se le agrupa según la condición socioeconómica, es decir, la relación salario-género difiere de acuerdo a la educación, al sector de empleo, región, y edad, entre otros. Mencionan que durante 1980 y 1990, el país experimentó un alto crecimiento en la población y en la fuerza laboral, y una sustancial caída en los ingresos reales tanto de los hombres como de las mujeres. Por último observan que la brecha salarial por género está inversamente relacionada con el tamaño de la firma y que es mayor para los empleos en el sector privado y en el sector informal. Además se observa una mayor desigualdad en los Estados del Norte de México que se atribuye a la segregación ocupacional experimentada por las mujeres en esta área.

Mayer y Cordourier (2001) con datos de 1996 encuentran que en el decil más bajo los hombres son objeto de discriminación de casi 12% y que se debe principalmente a sus deficiencias educativas, mientras que en los deciles 3 a 9 las mujeres experimentan una discriminación que implica una brecha de 15% y 18% menos que los hombres igualmente calificados. Aseveran que la brecha salarial se puede dar por tres factores: i) la productividad de cada género, ii) diferencias adicionales que se generan en los papeles biológico y social y iii) a factores de discriminación. Denominan a las dos primeras como diferenciación económica inmediata y diferenciación económica estructural, ambas tiene en común una base económica mientras la discriminación no la tiene.

En un estudio de Calónico y Ñopo (2008) estiman la brecha salarial con una metodología de *matching* con datos de 2004. Entre los principales resultados encuentra una brecha salarial del 12%. Por un lado, al simular la reducción de la segregación jerárquica (el hecho que los hombres ocupen puestos gerenciales y las mujeres puestos subordinados) reduciría la brecha salarial en un 5% y por otro lado, el hecho de reducir la segregación ocupacional en general (hombres y mujeres tienden a agruparse en diferentes ocupaciones) aumentaría la brecha en 6%.

En este estudio se pretende incorporar simultáneamente la corrección de dos posibles fuentes de sesgo. Por un lado, al igual que Págan y Ulibarri (2000) se corrige por selección por cada sexo, ya que los hombres y mujeres tienen diversas causas por las que entran o no al mercado laboral (Killingsworth y Heckman, 1986 y Pencavel, 1986). Por otro lado, se miden las brechas salariales por estrato a partir de *propensity score matching*, para así corregirse el problema de soporte común descrito por Ñopo (2008).

3. Datos

Los datos usados para este estudio provienen de la muestra del Censo de Población y Vivienda 2010 elaborados y publicados por el INEGI.¹ La información está disponible para los 32 estados de México y abarca tanto zonas rurales como urbanas. La muestra censal reporta los datos de características personales tales como la edad, la educación, el estado civil, el número de hijos, así como algunas características de empleo: ingresos, tipo de ocupación, goce de prestaciones, entre otras.

En lo que respecta a la selección de la muestra se toman en cuenta: trabajadores asalariados,² la edad mínima de evaluación es 21 años porque gran parte de los menores a esta edad son estudiantes. La edad máxima incorporada en el análisis es de 65 años ya que según la ley del IMSS el trabajador requiere dicha edad para gozar de la pensión por cesantía por edad avanzada. Se consideran sólo las localidades urbanas, es decir, localidades con más de 2,500 habitantes. Además se seleccionaron los individuos que no tienen discapacidades para no tener sesgos por este rubro. Se llevaron a cabo dos estimaciones incluyendo y excluyendo los individuos que seguían estudiando, los resultados son parecidos en ambos casos, en el documento se presentan los resultados que incluyen a los estudiantes.³ Se trabaja con el salario por hora y para calcularse se toma el criterio siguiente: el salario por hora es igual ingreso por salario mensual entre las horas

¹ La muestra del censo consta de 11,938,402 individuos.

² Se excluyen, los empleadores y trabajadores por cuenta propia.

³ Los estudiantes representan el 5.68% de la muestra utilizada para hacer las estimaciones, esto es, 131,529 individuos.

totales laboradas en el mes.⁴ En cuanto a la situación demográfica en la que los individuos se desenvuelven se construyeron cuatro categorías a partir de las 56 zonas metropolitanas del país: el Valle de México, las que tienen más de un millón de habitantes, las que tienen más de 500 mil habitantes y las zonas que tienen entre 100 mil y 500 mil habitantes. Para el caso de la estimación del *p-score* se tomaron en cuenta algunos criterios adicionales tales como la generación de polinomios e interacciones entre las variables.

Entre las variables más relevantes están: la edad, en diversos estudios se encuentra evidencia de que existe una relación en forma de U invertida entre la probabilidad de participación en el mercado laboral y la edad, conforme se aumenta la edad se presenta un incremento en la participación pero se llega a un máximo y después decrece dicha participación (esta condición se presenta con la combinación de edad y edad al cuadrado). Situación de pareja, la evidencia sugiere que la participación en la fuerza laboral y horas de trabajo de mujeres casadas responde a cambios en los salarios de los esposos. También el hecho de ser casada está asociado con una disminución en la probabilidad de participación en el mercado laboral. Fertilidad, es probable que el tener hijos menores de 6 años incremente el salario de reserva de modo que la fertilidad reduce en promedio la probabilidad de la incorporación a la fuerza laboral, este fenómeno probablemente es debido a que las preferencias de las madres se ven alterados cuando los niños son pequeños y necesitan de cuidados de tiempo completo de calidad. La educación, se ha encontrado que mayor escolaridad se asocia a mayor probabilidad en la participación de la fuerza laboral.⁵ Los ingresos que no son obtenidos por el trabajo también pueden alterar la probabilidad de

⁴ Las horas totales laboradas en el mes se calculan a partir de las horas trabajadas en la semana anterior a la encuesta multiplicada por 4.3 que es un factor del promedio de semanas por mes.

⁵ Heckman y Killingsworth (1986)

incorporarse al mercado laboral de forma negativa, en este rubro se encuentran las transferencias del gobierno, remesas, ingreso de otro miembro del hogar, entre otros.

Tabla 1. Promedio de características observables por género 2010				
Variable	Hombre	Mujer	Diferencia	t
Edad	36.9022 (0.0119)	37.7684 (0.0103)	-0.8662	-55.1
Escolaridad acumulada	9.4286 (0.0046)	8.7246 (0.0041)	0.7040	113.7
Pareja	0.7201 (0.0005)	0.7029 (0.0004)	0.0172	28.5
Trabaja	0.8018 (0.0004)	0.3330 (0.0004)	0.4688	797.6
Ingreso mensual	6049.2490 (11.8631)	5094.1240 (12.3671)	955.1250	52.2
Horas trabajadas	49.9069 (0.0208)	41.4536 (0.025)	8.4534	253.1
Hijos menores de 6	0.1891 (0.0004)	0.2955 (0.0004)	-0.1064	-190.0
Hijos menores de 6 trabaja	0.2149 (0.0005)	0.2437 (0.0006)	-0.0287	-36.7
Hijos menores de 6 no trabaja	0.0847 (0.0006)	0.3213 (0.0005)	-0.2366	-220.0
Tamaño de muestra	980,112	1,336,854		

Notas: (1) Desviaciones estándar entre paréntesis. (2) Los valores de la t se obtuvieron sin tomar en cuenta el factor de ponderación

En la Tabla 1 se reportan las principales características observables de hombres y mujeres en México, de la muestra sobre la cual se hacen las estimaciones (2,316,966 individuos). Las mujeres en promedio son un poco mayores respecto a los hombres en la muestra. Se observa que los hombres tienen 0.7 más años de educación que las mujeres. Cabe resaltar

que un 80% de los hombres está activo en la fuerza laboral mientras que el 33% de las mujeres pertenece a la fuerza laboral. Esto coincide con lo estimado por Calónico y Ñopo (2008). Con respecto a las horas trabajadas los hombres trabajan 8.4 horas a la semana más que las mujeres. En general las mujeres declaran tener más hijos menores de 6 años y se observa que la porción de madres con hijos menores de 6 años es mayor cuando no se trabaja, lo que es consistente con la preferencia del cuidado de los hijos.

4. Modelo

Para medir la brecha salarial se procede en dos fases. En la primera fase se corrige el problema de truncamiento incidental por medio del modelo de dos etapas de Heckman (1979), ya que en las estadísticas no se observa el salario de reserva. En la segunda fase se utiliza el modelo de *propensity score (p-score) matching* para el cálculo de brechas salariales por estrato.

4.1. Primera fase: Corrección por selección

Se utiliza la metodología de selección muestral⁶ debido a que los salarios tienen un problema de truncamiento incidental, es decir, el salario se observa sólo si la persona participa en la fuerza laboral. Se hace el siguiente supuesto paramétrico:

$$w_i^0 = \exp(x_{i1}\beta_1 + u_{i1}), \quad (1)$$

$$w_i^r = \exp(x_{i2}\beta_2 + \gamma_{i1}a_i + u_{i2}), \quad (2)$$

Donde w_i^0 es el salario observado, w_i^r es el salario de reserva, x_{i1} son las covariables del salario observado, a_i es el ingreso no laboral y x_{i2} son las covariables del salario de reserva. Se supone que (u_{i1}, u_{i2}) es independiente de (x_{i1}, x_{i2}, a_i) . Los parámetros del

⁶ El modelo de selección muestral se tomó de Wooldridge (2002)

modelo están denotados por $\beta_1, \beta_2, \gamma_{i1}$. Dado w_i^o es observado sólo para personas que trabajan, y esto es sólo si $w_i^o \geq w_i^r$ entonces al obtener logaritmos y sustraer ambas ecuaciones tenemos que:

$$\log(w_i^o) - \log(w_i^r) = x_{i1}\beta_1 - x_{i2}\beta_2 - \gamma_{i1}a_i + u_{i1} - u_{i2} \equiv x_i\delta_2 + v_{i2} > 0$$

Ahora bien, sea y_1 el logaritmo de w^0 y sea y_2 un indicador binario de la participación en la fuerza laboral, toma valor de 1 si trabaja y 0 si no trabaja. El modelo puede ser descrito como:

$$y_1 = x_1\beta_1 + u_1 \quad (3)$$

$$y_2 = 1[x_i\delta_2 + v_{i2} > 0] \quad (4)$$

Supuestos: (a) Para enfatizar el problema de selección muestral se asume que (x_i, y_2) son observadas siempre, y_1 es observada solo cuando $y_2 = 1$; (b) un supuesto estándar para exogeneidad de x_i es que (u_1, v_2) son independientes de x_i con media cero; (c) se asume que la varianza es 1 sin pérdida de generalidad porque y_2 es una variable binaria $v_2 \sim N(0,1)$; (d) Por ultimo requiere linealidad en la regresión poblacional $E(u_1|v_2) = \gamma_1 v_2$.

Dado que y_1 es observada solo cuando $y_2 = 1$, lo que podemos estimar entonces es $E(y_1|x, y_2 = 1)$ junto con $P(y_2 = 1|x)$ y bajo los supuestos hechos:

$$E(y_1|x, v_2) = x_1\beta_1 + E(u_1|x, v_2) = x_1\beta_1 + E(u_1|v_2) = x_1\beta_1 + \gamma_1 v_2. \quad (5)$$

Nótese que si $\gamma_1 = 0$, lo cual implica que u_1 no está correlacionada con v_2 , entonces $E(y_1|x, v_2) = E(y_1|x) = E(y_1|x_1) = x_1\beta_1$ como y_2 es una función de $(x|v_2)$, se sigue de inmediato que $E(y_1|x, y_2) = E(y_1|x_1)$. Es decir, en este caso no hay problema de selección muestral y β_1 puede ser consistentemente estimado por mínimos cuadrados ordinarios.

Cuando $\gamma_1 \neq 0$, usando expectativas iteradas sobre la ecuación

$$E(y_1|x, y_2) = x_1\beta_1 + \gamma_1 E(u_1|x, v_2) = x_1\beta_1 + \gamma_1 h(x, y_2) \quad (6)$$

Si conocemos $h(x, y_2)$ podemos estimar β_1 y γ_1 como la muestra con selección tiene $y_2 = 1$, necesitamos encontrar $h(x, 1)$. Pero $h(x, 1) = E(v_2|v_2 > -x_i\delta_2) = \lambda(x_i\delta_2)$, donde λ es la razón inversa de Mills. Y entonces podemos escribir

$$E(y_1|x, y_2 = 1) = x_1\beta_1 + \gamma_1\lambda(x_i\delta_2) \quad (7)$$

Siguiendo a Heckman (1979), la ecuación anterior se realizará en dos partes, en primer lugar se tomará como variable dependiente la oferta laboral y se estimará con un modelo *probit*:

$$P(y_2 = 1) = \Phi(\beta_2 x_2 + v_2) \quad (8)$$

Donde y_2 es una variable dummy igual a 1 si el individuo reporta que trabaja y 0 en otro caso y x_2 son las variables control: educación, edad, una variable dummy que toma el valor de 1 cuando el individuo reporta tener hijos vivos menores de 6 años, tipo de población (conjunto de variables que indican si el individuo vive en la zona metropolitana de Valle de México, o en zonas metropolitanas con más de un millón de habitantes, entre 500 mil y un millón de habitantes, o entre 100 mil y 500 mil habitantes) y el ingreso no laboral ya sea por una transferencia del gobierno o de los ingresos salariales de los integrantes del hogar. Entonces se calcula la razón inversa de Mill con los resultados obtenidos en la estimación:

$$\lambda_i = \frac{\phi(\cdot)}{\Phi(\cdot)}$$

En la segunda parte, se estimará el log de los salarios incluyendo la razón inversa de Mills como una variable explicatoria (ecuación 7) por medio de mínimos cuadrados ordinarios.

Las covariables usadas en esta estimación son: edad, edad al cuadrado, educación, la dummy hijos menores de 6 y las variables indicadoras de zona metropolitana donde habita. Este procedimiento se realizará por cada género por separado.

4.2. Segunda fase: Corrección por sesgo por ausencia de soporte común⁷

Como señala Ñopo (2008), surge un problema de sesgo por ausencia de soporte común cuando se pasa por alto las diferencias en las características individuales entre hombres y mujeres.. Lo ideal es comparar a hombres y mujeres que tengan la misma edad, vivan, relativamente, en el mismo lugar, tengan el mismo número de hijos, hayan cursado los mismos años de escolaridad, tengan puestos de trabajo relativamente iguales, entre otras características observables.

La aproximación *matching* es una posible solución para este problema. La idea básica es encontrar un grupo grande de hombres quienes son similares a las mujeres en todas sus características relevantes y observables.⁸ Sin embargo, ante esta solución surge el problema de la multidimensionalidad⁹ del vector de covariables. Para resolver la cuestión de la multidimensionalidad Rosenbaum y Rubin (1983) sugieren el uso del llamado *balancing score* $b(X)$, esto es, funciones de las variables observables relevantes X tales que la distribución condicional de X dado $b(X)$ es independiente de asignar el tratamiento.

⁷ Esta sección esta basada en Caliendo y Kopeinig (2008)

⁸ A esto se le denomina pareo exacto.

⁹ Si se tienen n variables dicotómicas el número de posibles vectores de las covariables X , será de 2^n . Por ejemplo, si sólo se tiene una covariable, digamos estado civil, la cual es dicotómica (1 para casado, 0 para soltero) el matching se generaría entre casados y solteros. Pero si tenemos dos covariables, digamos estado civil y educación (supongamos que ambas dicotómicas también) tendremos casados que estudiaron, casados que no estudiaron, solteros que estudiaron y solteros que no estudiaron. Entonces el problema de la alta dimensionalidad se eleva no sólo con el número de variables sino también que muchas variables no son dicotómicas sino incluso continuas.

Es importante anotar que este método de pareo es usado ampliamente en la literatura de evaluación de programas ya que fue diseñado para “eliminar” el sesgo de selección entre individuos expuestos a un tratamiento (tratados) e individuos no expuestos (controles). En particular en este documento se establece un tratamiento artificial que depende del género. Se supone que el conjunto de mujeres es el grupo de tratamiento y que el conjunto de hombres es el grupo de control.

El parámetro de evaluación que nos interesa destacar es el llamado efecto de tratamiento promedio sobre los tratados (ETPT), es decir conocer el impacto sobre salarios ya que el individuo es mujer. Tal efecto puede expresarse como:

$$ETPT = E(\tau|D = 1) = E(Y(1)|D = 1) - E(Y(0)|D = 1)$$

Donde $Y(1)$ es el resultado potencial del individuo a quien se le dio el tratamiento, i.e. el salario dado que es mujer, y $Y(0)$ es el resultado potencial del individuo a quien no se le dio el tratamiento, que es el salario dado que es hombre. En términos generales el efecto de tratamiento promedio de los tratados es igual al valor esperado resultado que tuvo el individuo dado que se le aplicó el tratamiento menos el valor esperado del individuo si no se le aplico el tratamiento dado que se le aplicó el tratamiento.

Cómo la media del contrafactual para los tratados no es observable, se tiene que elegir un sustituto propio en orden de estimar el ETPT.

$$E[Y(1)|D = 1] - E[Y(0)|D = 0] = ETPT + E[Y(0)|D = 1] - E[Y(0)|D = 0]$$

La diferencia entre el lado izquierdo de la ecuación es llamado el sesgo de selección de modo que el verdadero parámetro ETPT es sólo identificado si:

$$E[Y(0)|D = 1] - E[Y(0)|D = 0] = 0$$

Cabe mencionar que en los experimentos se asegura que el efecto de tratamiento está identificado. Pero en los estudios no experimentales se recurre al supuesto de identificación que resuelve el problema de selección. El método que se usa en este documento es el *propensity score matching*, este método controla por una función particular de covariables, en este caso en específico por la probabilidad de ser mujer $\Pr[D_i = 1|X_i]$.

Los supuestos del *propensity score matching* son:

- i. *Soporté común*. Asegurarse que las personas con los mismos valores de X tengan una probabilidad positiva de ser participantes y no participantes

$$0 < P(D = 1|X) < 1$$

Este supuesto asegura que haya hombres emparejados con mujeres que tengan características observables comparables.

- ii. *Supuesto de valor estable de unidad de tratamiento*. El supuesto del SUTVA (por sus siglas en inglés) es un supuesto de la no existencia de efectos de equilibrio general, el cual implica que el tratamiento no afecta indirectamente a las observaciones de los no tratados (Cameron y Trivedi, 2005).

- iii. *Independencia condicional*. El supuesto de independencia condicional (CIA por sus siglas en inglés) implica que las diferencias sistemáticas en los resultados entre los individuos tratados y controles con los mismos valores de covariables son atribuibles al tratamiento.

$$Y(0), Y(1) \perp D|X$$

Dado un conjunto de covariables X las cuales no son afectadas por el tratamiento, los resultados potenciales son independientes de la asignación del tratamiento. Esto implica que todas las variables que influyen en la asignación del tratamiento y resultados potenciales simultáneamente tienen que ser observados por el investigador. Sería también claro que condicionar sobre todas las variables relevantes es limitado en el caso de un vector multidimensional X . Para lidiar con este problema Rosenbaum y Rubin (1983) sugieren usar el llamado *balancing score*. Muestran que si los resultados potenciales son independientes del tratamiento condicional sobre covariables X , también son independientes del tratamiento sobre el *balancing score* $b(X)$. El *p-score* $P(D = 1|X) = P(X)$, esto es la probabilidad para un individuo de participar en un tratamiento dadas sus covariables observadas X , es un posible *balancing score*.

$$Y(0), Y(1) \perp D | P(X)$$

iv. *Condición de Balance*. La condición de balance se refiere a conocer si cada uno de los valores del *p-score* tiene la misma distribución tanto en el grupo de tratamiento como en el de control (Wang, 2006), es decir:

$$D \perp X | P(X)$$

La idea de la condición de balance es verificar si el *p-score* es un adecuado *balancing score*. En lo que respecta a la estrategia de estimación: por un lado, se estima el *propensity score* usando un modelo *probit*¹⁰ y así se reduce la dimensión de condicionamiento. De este modo también se estima la probabilidad de ser mujer versus ser hombre. Por otra parte, se elige el modelo por intervalos y estratificación: La idea de la estratificación es particionar el soporte común del *p-score* en un conjunto de intervalos (estratos) para calcular el impacto

¹⁰ Junto con el modelo logit son los modelos paramétricos que son usados convencionalmente y sus resultados no difieren significativamente.

dentro de cada intervalo tomando la diferencia media de los resultados entre las observaciones de los tratados y controles. La elección del número de estratos estará dada por las condiciones del balance del *propensity score* o las covariables dentro de cada estrato.

Para checar el soporte común se utiliza tanto un análisis visual de la distribución de la densidad del *propensity score* en ambos grupos, como el criterio de mínimo y máximo que consiste en eliminar todas las observaciones cuyo *p-score* es más pequeño que el menor y más grande que el máximo.

Para la estimación de la brecha salarial, ya que se obtuvo un *p-score* bien balanceado el método que se utiliza es el de estratificación y consiste en la siguiente regresión:

$$Y_i = \alpha + \theta T_i + \gamma p(x) + \pi T(p(x) - \mu) + \varepsilon_i$$

Donde Y_i es el salario, T_i es el tratamiento (en este caso ser mujer), $p(x)$ es el valor del *p-score*, y ε_i es el error. Esta regresión se hará para cada uno de los estratos.

5. Resultados

En esta sección se exponen los resultados del modelo de Heckman que servirán como insumos para el cálculo de la brecha salarial que se obtendrá a través del *p-score*. En el Anexo 1, se presentan las estimaciones de la primera etapa referentes a la participación en el mercado laboral para ambos sexos. Para lograr la identificación del modelo, la ecuación de selección incluye variables que no impactan directamente al salario pero sí a la probabilidad del individuo de incorporarse al mercado laboral. Dichas variables son: el salario de la pareja y el salario de otros miembro de la familia, una variable dummy que

indique tener hijos menores de 6 años, las transferencias del gobierno y una variable que indique si se es jefe de familia.

En lo que respecta al salario de la pareja, los resultados son los que se esperaban, ya que se encuentra una relación negativa entre salario y probabilidad de trabajar, es decir, un aumento en el salario de la pareja reduce la probabilidad de entrar al mercado laboral. El impacto de tener hijos menores de 6 años es consistente con la teoría, ya que en general reduce la probabilidad de que una mujer ofrezca su trabajo. También se observa que las transferencias del gobierno desincentivan a los individuos para entrar al mercado laboral. En todos los casos, ser jefe de familia incrementa la probabilidad de que el individuo trabaje.¹¹

En la Tabla 2 se muestra la estimación del modelo de selección para hombres y mujeres respectivamente. En el primer par de columna se encuentran las estimaciones sobre todas las observaciones, en el segundo y tercer par de columnas se encuentran los coeficientes de las regresiones sólo tomando en cuenta las muestras por género. En la Tabla 3 se selecciona la muestra por el género y la situación de pareja. En las estimaciones de ambas tablas, también se estimaron los coeficientes de mínimos cuadrados ordinarios (MCO).

Se obtuvo el parámetro del coeficiente de Mills estadísticamente significativo para todas las especificaciones (véanse Tablas 2 y 3) lo que sugiere que existe evidencia de truncamiento incidental. De igual importancia, se observa que sí hay diferencias en los coeficientes estimados, corregidos por selección en comparación con los de MCO dado que, en un

¹¹ El estadístico *Chi cuadrado* de todas las estimaciones es significativo, por lo tanto, se rechaza la hipótesis nula de que los coeficientes son iguales a cero.

intervalo de 95% de confianza, los resultados de ambas especificaciones son estadísticamente diferentes.

Se observa que los retornos a la educación son más grandes para los hombres que para las mujeres, con excepción de la muestra que contempla a los individuos que tienen pareja donde se revierten los retornos. Las mujeres con pareja (véase Tabla 3) tienen una compensación más alta en sus ingresos que las solteras. Esto puede ser debido a que las mujeres casadas que están empleadas pueden autoseleccionarse positivamente en términos de salario y tener ingresos relativamente altos (Pagán y Ulibarri, 2000).

Con respecto a la edad, se obtuvieron los signos que se deseaban para ambos sexos. Se observa que los hombres casados tienen un ciclo de vida salarial más amplio pues su tasa decreciente es más pequeña (0.0002) en cambio, los hombres solteros y las mujeres casadas alcanzan su salario máximo antes de la edad promedio de ingresos máximos. En general se observa que los hombres ingresan más que las mujeres en las zonas metropolitanas. Además en las zonas metropolitanas de más de un millón de habitantes existe un mayor salario en comparación con las zonas con menos de 100 mil habitantes.

Tabla 2. Estimación del modelo de Heckman: Variable dependiente log salario por hora.

	Mujeres		Hombres		Todos	
	MDE*	MCO	MDE*	MCO	MDE*	MCO
Edad	0.0572 (0.001)	0.1427 (0.0007)	0.0471 (0.0006)	0.1766 (0.0008)	0.0484 (0.0005)	0.1570 (0.0006)
Edad al cuadrado	-0.0006 (0.0000)	-0.0017 (0.0000)	-0.0004 (0.0000)	-0.0023 (0.0000)	-0.0005 (0.0000)	-0.0019 (0.0000)
Escolaridad acumulada	0.0710 (0.0007)	0.1132 (0.0003)	0.0788 (0.0003)	0.0711 (0.0003)	0.0827 (0.0002)	0.1074 (0.0002)
Situación de pareja	0.2805 (0.0063)	-0.8119 (0.0027)	-0.1855 (0.0052)	0.6382 (0.0034)	0.0805 (0.0018)	-0.1843 (0.0024)
Valle de México	0.1678 (0.0043)	0.0813 (0.0036)	0.1426 (0.0034)	0.1989 (0.0041)	0.1740 (0.0024)	0.1285 (0.0031)
Más de un millón de hab.	0.1251 (0.0046)	0.1270 (0.0037)	0.1568 (0.0035)	0.2944 (0.0042)	0.1808 (0.0025)	0.2107 (0.0032)
Más de 500 mil hab.	0.0762 (0.0051)	0.1948 (0.0043)	0.1265 (0.0039)	0.2554 (0.0048)	0.1439 (0.0028)	0.2366 (0.0036)
Más de 100 mil hab.	-0.0188 (0.0056)	0.1182 (0.0046)	0.0657 (0.0042)	0.1786 (0.0052)	0.0645 (0.003)	0.1653 (0.0039)
Constante	1.4116 (0.0242)	-2.1867 (0.0137)	1.5674 (0.0137)	-1.9316 (0.0158)	1.2130 (0.0102)	-2.2346 (0.0118)
Coeficiente de Mills	-0.6395 (0.0116)		-0.9678 (0.0134)		-0.3064 (0.0026)	

Nota: Entre paréntesis se presentan los errores estándar.

*Estimaciones con el modelo de dos etapas de Heckman.

Tabla 3. Estimación del modelo de Heckman: Variable dependiente log salario por hora.

	Con pareja				Solteros			
	Mujeres		Hombres		Mujeres		Hombres	
	MDE*	MCO	MDE*	MCO	MDE*	MCO	MDE*	MCO
Edad	0.0613 (0.0016)	0.0916 (0.0008)	0.0362 (0.001)	0.1347 (0.001)	0.0511 (0.0011)	0.2360 (0.0013)	0.0499 (0.0014)	0.2217 (0.0018)
Edad al cuadrado	-0.0007 (0.0000)	-0.0011 (0.0000)	-0.0002 (0.0000)	-0.0018 (0.0000)	-0.0004 (0.0000)	-0.0029 (0.0000)	-0.0005 (0.0000)	-0.0027 (0.0000)
Escolaridad acumulada	0.0929 (0.0012)	0.1217 (0.0003)	0.0723 (0.0005)	0.0898 (0.0003)	0.0737 (0.0006)	0.0967 (0.0005)	0.0798 (0.0006)	0.0288 (0.0007)
Valle de México	0.1899 (0.0065)	0.0031 (0.0042)	0.0856 (0.0052)	0.1970 (0.0046)	0.1569 (0.0054)	0.2404 (0.0067)	0.2526 (0.007)	0.2165 (0.0084)
Más de un millón de hab.	0.1373 (0.0066)	0.0638 (0.0042)	0.1230 (0.0054)	0.2794 (0.0046)	0.1420 (0.0061)	0.2842 (0.0075)	0.2061 (0.0077)	0.3210 (0.0092)
Más de 500 mil hab.	0.1196 (0.0073)	0.1709 (0.0049)	0.0873 (0.0061)	0.2509 (0.0053)	0.0719 (0.0066)	0.2571 (0.0083)	0.1843 (0.0085)	0.2559 (0.0103)
Más de 100 mil hab.	0.0135 (0.0081)	0.0914 (0.0052)	0.0378 (0.0064)	0.1790 (0.0056)	-0.0138 (0.0072)	0.1821 (0.009)	0.0964 (0.0094)	0.1709 (0.0114)
Constante	1.1436 (0.0447)	-2.1203 (0.0166)	1.6477 (0.0205)	-0.5714 (0.0189)	1.3506 (0.0256)	-3.6700 (0.0253)	1.4779 (0.0284)	-2.3901 (0.0332)
Coeficiente de Mills	-0.3768 (0.0154)		-1.3748 (0.0272)		-0.5458 (0.0141)		-0.8671 (0.0211)	

Nota: Entre paréntesis se presentan los errores estándar.

*Estimaciones con el modelo de dos etapas de Heckman.

Realizada la estimación por el modelo de dos etapas de Heckman se pronostican los salarios para cada individuo, considerándose el género y la situación de pareja, es decir, se pronostican a partir de las estimaciones del modelo de dos etapas que se presentan en la Tabla 3.¹²

Una vez hecho el pronóstico se calcula el *p-score*, con el objeto de conocer los hombres y mujeres que se parecen en probabilidad en términos de sus características observables tales como el nivel educativo, la situación familiar, la edad, entre otras. Para el cálculo del *p-score* se toman en cuenta variables tales como: la escolaridad por nivel académico (primaria o menos, secundaria, preparatoria o formación técnica, licenciatura y maestría), la situación familiar: si se tienen hijos menores de 6, igual con una variable dicotómica y otra variable indicadora de pareja. También se generó la variable que interactúa entre la situación de pareja e hijos. Para aproximar la edad se construyeron tres variables que indican las etapas del ciclo de vida, cuyas edades para la primera etapa es de 21 a 35 años, para la segunda de 36 a 50 años y la última etapa de 51 a 60 años.¹³ Se hacen interacciones entre las variables indicadoras de las etapas del ciclo de vida y la escolaridad.

La Tabla 4 incluye los resultados de la estimación del *propensity score*. La prueba conjunta *chi cuadrada* muestra que todas las variables utilizadas para calcular el *p-score* son significativas. Entre los resultados más significativos se observa que los niveles educativos muestran una relación positiva con dicha probabilidad. El tener hijos menores de 6 años también tiene una relación positiva como era de esperarse.

¹² Las estimaciones alcanzaron un 0.93 de bondad de ajuste.

¹³ Borjas (2008) cap. 3.

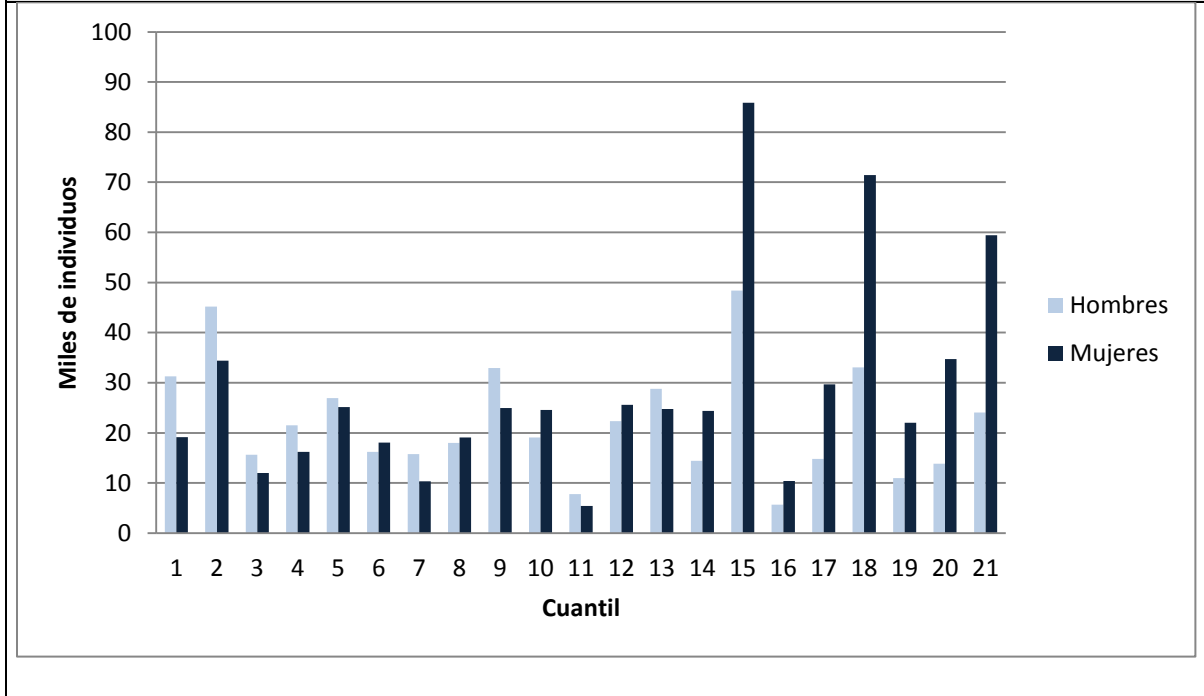
Tabla 4. *Propensity Score*

Variable	Coficiente	Error estándar	z
Escolaridad			
Primaria o menos	0.7001	0.0317	22.09
Secundaria	0.4720	0.0317	14.88
Preparatoria	0.5486	0.0318	17.27
Licenciatura	0.3704	0.0318	11.63
Maestría	0.2181	0.0338	6.45
Familia			
Hijos menores de 6	0.5171	0.0022	235.22
Situación de pareja	-0.2321	0.0020	-114
Etapas del ciclo de vida			
Etapas de 21 a 35 años	0.1309	0.0515	2.54
Etapas de 50 a 65 años	-0.1662	0.0513	-3.24
Zona donde habita			
Valle de México	0.0181	0.0025	7.27
Más de un millón de habitantes	-0.0177	0.0026	-6.82
Más de 500 mil habitantes	-0.0150	0.0029	-5.1
Más de 100 habitantes	-0.0214	0.0032	-6.74
Interacciones			
Etapas 1 e hizo la primaria	-0.4001	0.0516	-7.75
Etapas 3 e hizo la primaria	0.2338	0.0514	4.55
Etapas 1 e hizo la secundaria	-0.2986	0.0516	-5.79
Etapas 3 e hizo la secundaria	0.1513	0.0516	2.93
Etapas 1 e hizo preparatoria	-0.3582	0.0517	-6.94
Etapas 3 e hizo preparatoria	0.3291	0.0517	6.36
Etapas 1 e hizo licenciatura	-0.1869	0.0517	-3.61
Etapas 3 e hizo licenciatura	0.0225	0.0518	0.43
Etapas 1 e hizo maestría	0.0049	0.0543	0.09
Etapas 1 e hizo maestría	-0.0174	0.0561	-0.31
Constante	-0.2275	0.0317	-7.19

Nota: Las variables excluidas para no causar un problema de multicolinealidad perfecta, fueron la variable dummy de la segunda etapa del ciclo de vida y sus respectivas interacciones con la escolaridad.

En la Figura 1, se muestra el histograma de los valores del *p-score* estimado dividido en 21 cuantiles. Como se esperaba, se presentan más mujeres conforme la probabilidad se va acercando a 1 y viceversa para el caso de los hombres.

Figura 1. Histograma de *Propensity Score* estimado



Fuente: Elaboración propia.

Para probar el soporte común se realiza el *test* de balance utilizando el algoritmo de Dehejia y Wahba (2002). En lo que respecta a la condición de balance, se encuentra que los atributos especificados están balanceados en 81% en promedio, en el rango de 50 a 100%.

Tabla 5. Condición de Balance, entre hombres y mujeres

Porcentaje de balance en el cuantil	1	50%	Porcentaje de balance en el cuantil	11	100%
Porcentaje de balance en el cuantil	2	100%	Porcentaje de balance en el cuantil	12	60%
Porcentaje de balance en el cuantil	3	100%	Porcentaje de balance en el cuantil	13	100%
Porcentaje de balance en el cuantil	4	100%	Porcentaje de balance en el cuantil	14	60%
Porcentaje de balance en el cuantil	5	60%	Porcentaje de balance en el cuantil	15	70%
Porcentaje de balance en el cuantil	6	90%	Porcentaje de balance en el cuantil	16	100%
Porcentaje de balance en el cuantil	7	60%	Porcentaje de balance en el cuantil	17	80%
Porcentaje de balance en el cuantil	8	60%	Porcentaje de balance en el cuantil	18	70%
Porcentaje de balance en el cuantil	9	60%	Porcentaje de balance en el cuantil	19	80%
Porcentaje de balance en el cuantil	10	100%	Porcentaje de balance en el cuantil	20	100%
			Porcentaje de balance en el cuantil	21	100%
Promedio		81%			

Fuente: Elaboración propia.

Al estimarse la brecha por cada uno de los bloques balanceados¹⁴ (véase Anexo 2) se calcula que en promedio la brecha salarial es de 6% (Tabla 6) lo cual es inferior al 12% que encuentran Calónico y Ñopo (2008). Al realizarse una estimación sin considerar la serie que corrige el truncamiento incidental, es decir, con la serie original que contempla a los individuos que son parte de la fuerza laboral, se obtiene un 10.9% de brecha salarial (véase Anexo 3) lo cual se aproxima a la estimación de los autores antes mencionados.¹⁵

Es importante señalar que la brecha salarial fluctúa entre un 14% a favor de las mujeres a un 24% en contra, lo cual indica que existen grupos de mujeres que tienen mayores ingresos salariales que los hombres pero las diferencias salariales no son tan amplias como en el caso de los hombres, es decir, el cuantil donde la brecha salarial es más amplia a favor de los hombres en comparación con el cuantil donde la brecha salarial es más amplia a favor de las mujeres, aún es mayor en un 10%.

En el cuantil 15 en el cual la brecha es más amplia, en contra de las mujeres, se encuentran en su mayoría, individuos que cursaron primaria o menos y que tienen edades entre 36 y 50 años. Las características que son preponderantes en el cuantil donde la brecha esta a favor de las mujeres son individuos sin hijos menores de 6 años que tienen licenciatura o preparación técnica y tienen entre 21 y 35 años de edad. Cabe señalar que aquellos cuantiles donde la brecha es casi nula los individuos están perfectamente pareados.

¹⁴ Aunque existen algunos cuantiles donde hay un balance perfecto, también hay un considerable número de cuantiles donde el balance es inferior al 70% lo que causa que las diferencias en el promedio de las características no sean significativas.

¹⁵ La diferencia entre la estimación de Calónico y Ñopo y el presente estudio puede ser causada a las distintas bases de datos así como el periodo en que son hechos ambos análisis.

Tabla 6. Brecha salarial por estrato del *p-score*

Cuantil	Diferencial log salario	Brecha salarial	Cuantil	Diferencial log salario	Brecha salarial
block 1	-0.1504 (0.0028)	-13.96%	block 11	-0.1573 (0.0019)	-14.56%
block 2	0.0468 (0.0012)	4.80%	block 12	-0.1117 (0.0013)	-10.57%
block 3	0.1302 (0.0019)	13.91%	block 13	-0.0735 (0.0018)	-7.09%
block 4	0.0877 (0.0015)	9.17%	block 14	-0.1861 (0.0022)	-16.98%
block 5	-0.0012 (0.001)	-0.12%	block 15	-0.2785 (0.0011)	-24.31%
block 6	-0.2167 (0.0009)	-19.48%	block 16	0.0017 (0.001)	0.17%
block 7	-0.0759 (0.0018)	-7.30%	block 17	0.0070 (0.0015)	0.70%
block 8	-0.1558 (0.002)	-14.43%	block 18	-0.0304 (0.0012)	-2.99%
block 9	-0.1487 (0.0021)	-13.82%	block 19	0.0801 (0.0016)	8.34%
block 10	-0.0693 (0.0012)	-6.70%	block 20	0.0100 (0.0012)	1.01%
Promedio	-6.03%		block 21	-0.1331 (0.0014)	-12.46%

Nota: Los errores estándar para la brecha salarial son calculados usando *bootstrap* con 200 replicas.

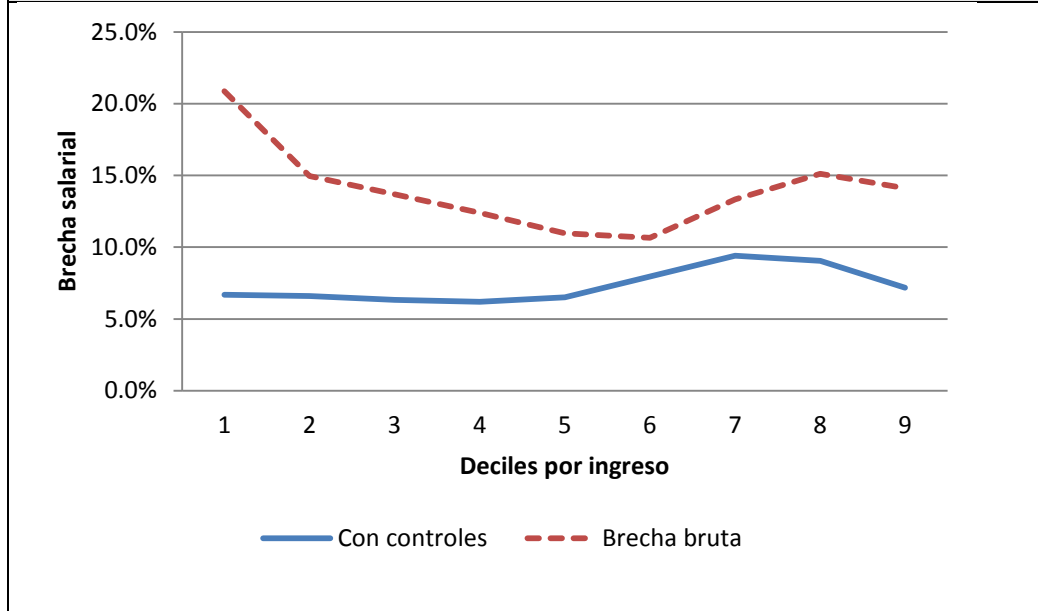
Adicionalmente se especifica la regresión cuantil,¹⁶ de la cual se realizan dos estimaciones:

$$\sum_{i=y \geq \beta}^N q|y_i - x' \beta| + \sum_{i=y \leq \beta}^N (1 - q)|y_i - x' \beta|$$

Por un lado, sólo se toma en cuenta el sexo del individuo obteniendo la brecha bruta según estrato de ingreso. Por otro lado, la especificación considera el sexo y como covariables el *p-score*, la educación y la edad. Estas especificaciones ayudan a observar la brecha que aún no es explicada y que puede ser atribuida a otros aspectos fuera de las características individuales. En la siguiente gráfica se presentan los resultados (véase Anexo 4).

¹⁶ Las estimaciones se hacen a partir de una muestra de 132,336 datos con soporte común.

Figura 2. Regresión cuartil, por grupos de salario



Fuente: Elaboración propia.

En promedio en la especificación sin controles, se obtiene una brecha salarial bruta de un 14%, cercano a lo obtenido sin la corrección por selección. En el mismo sentido, al controlar por las covariables antes descritas se encuentra una brecha de 7.3%.

Como se muestra en la Figura 2 en el decil más bajo de ingresos, la diferencia entre las brechas salariales es más amplia, la cual va reduciendo hacia el centro de la distribución de ingresos para después ensancharse un poco más al final de la distribución. Esta situación muestra evidencia de que en los estratos de ingresos bajos se pueden encontrar brechas más amplias, mientras que en estratos medios de ingreso esta brecha se reduce.

6. Conclusiones

En este estudio se analiza la brecha salarial por género para el año 2010. Dado que las comparaciones previamente hechas con estimadores *matching* no contemplan la corrección por la existencia de truncamiento incidental en los salarios (no se observan los salarios de

reserva), a este estudio se le incorpora la corrección por esta fuente de sesgo. Por otro lado, la falta de soporte común (entendida como que haya combinaciones de características individuales en las cuales es posible encontrar hombres pero no mujeres), es ajustada mediante la metodología de *propensity score matching*, cuya idea general es conjuntar las particularidades de un individuo en un solo “índice”, de modo que todos los que tengan aproximadamente el mismo “índice” son comparables.

A partir de los datos del Censo de Población y Vivienda 2010, se obtienen las brechas salariales por género, usando una metodología que se realiza en dos fases: En la primera fase, se estiman los modelos de Heckman, de forma separada, por situación de pareja y por género y se pronostican los salarios, corrigiendo de este modo el problema de truncamiento incidental. En la segunda fase, se estima el *p-score*, se elige el soporte común y se prueba su balance para obtener al final las brechas salariales.

Entre los principales resultados se encuentra que la brecha promedio es de 6% la cual es menor al 12% reportado por Calónico y Ñopo (2008). Debido a este resultado se realizó otra estimación que contemplara la serie de salarios no corregida por selección tal como fue considerada por dichos autores. En promedio se calcula que la brecha salarial es de 10.9% la cual se aproxima con lo obtenido por Calónico y Ñopo, por lo que se puede decir que el 4.9% de la brecha se da por un problema de truncamiento incidental. Este hallazgo muestra evidencia de que las brechas salariales en estudios anteriores pueden estar sobrestimadas tanto por truncamiento incidental como por la comparabilidad entre no comparables.

7. Referencias

Banco Mundial (2011). “World development report: gender equality and development”

<http://www.app.collinsindicate.com/worldbankatlas-gender/en>

Brown, Cynthia J., Pagan, Jose A., Rodriguez-Oreggia, Eduardo (1999). “Occupational Attainment and Gender Earnings Differentials in Mexico”. *Industrial and Labor Relations Review*, 53(1), 123-135.

Caliendo, Marco y Kopeining, Sabine (2008). “Some practical guidance for the implementation of propensity score matching”. *Journal of Economic Surveys*, 22 (1), 31-72.

Calónico, Sebastian y Ñopo, Hugo (2008). “Gender Segregation in the Workplace and Wage Gaps: Evidence from Urban Mexico 1994-2004”. Working Papers 4579. *Banco Interamericano de Desarrollo*.

Cameron, A. Colin y Trivedi, Pravin K. (2005). *Microeconometrics: Methods and applications*, Cambridge University Press.

Cameron, A. Colin y Trivedi, Pravin K. (2009). *Microeconometrics using Stata*, Stata Press.

Castro Lugo, David y Huesca Reynoso, Luis (2007). “Desigualdad salarial en México: una revisión. *Papeles de Población*”. 13(54), 225-264.

Christofides, Louis N., Polycarpou, Alexandros y Vrachimis, Konstantinos (2010). “The Gender Wage Gaps, ‘Sticky Floors’ and ‘Glass Ceilings’ of the European Union”. *Forschungsinstitut zur Zukunft der Arbeit*. Discussion Paper No. 5044.

Dehejia, Rajeev H. y Wahba, Sadek (2002). “Propensity score-matching methods for nonexperimental causal studies”. *The Review of Economics and Statistics*, 84(1), 151-161.

Dolado, Juan José y Llorens, Vanesa (2004). “Gender wage gaps by education in Spain: Glass floors versus glass ceilings”. *Centre for Economic Policy Research*. Discussion Paper No. 4203 www.cepr.org/pubs/dps/DP4203.asp

Greene, William H. (1998). *Análisis económico* (3a ed.). España. Prentice Hall Iberia.

Heckman, James J. (1976). “The common structure of statistical model of truncation, sample selection and limited dependent variables and a simple estimator for such models”. *Annals of economic and social measurement*, 5(4), 120-137.

Ichino, Andrea y Moretti, Enrico (2009). “Biological Gender Differences, Absenteeism, and the Earnings Gap”. *American Economic Journal: Applied Economics*, 1(1), 183–218.

Killingsworth, Mark y Heckman, James J. (1986). "Female labor supply: A survey".

Ashenfelter, O. Layard, R. Ed. *Handbook of Labor Economics*, 1, 103-204.

Machado José A. F. y Mata, José (2005). "Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression". *Journal of Applied Econometrics*, 20(4), 445-465.

Mayer Foulkes, David y Cordourier Real Gabriela (2001). "La brecha salarial y la teoría de igualdad de oportunidades: un estudio de género para el caso mexicano". *El trimestre económico*, 68(1), 71-107.

Ñopo, Hugo (2008). "Matching as a tool to decompose wage gaps". *The Review of Economics and Statistics*, 90(2), 290-299

Pagán, José A. y Ullibarri, Miren (2000). "Group Heterogeneity and the Gender Earnings Gap in Mexico". *Economía Mexicana. Nueva Época*. 9(1), 23-40.

Pencavel, John (1986). "Labor supply of men: A survey". Ashenfelter, O. and Card, D. Ed. *Handbook of Labor Economics*, 1, 3-102

Rosenbaum, Paul R. y Rubin, Donald B. (1983). "The central role of the propensity score in observational studies for causal effects". *Biometrika* 70(1), 41-50.

Wang-Sheng Lee (2008). “Propensity Score Matching and Variations on the Balancing Test”. *Empirical Economics*, 13(1), 1-34.

Wooldridge, Jeffrey M. (2002), *Introductory Econometrics: A Modern Approach* (4^a ed.). USA. South Western.

Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross Section and Panel Data*. USA. MIT Press.

Anexo 1. Participación de hombres y mujeres en el mercado laboral (Primera etapa, ecuación probit)							
	Con pareja		Solteros		Mujeres	Hombres	Todos
	Mujer	Hombres	Mujer	Hombres			
Edad	-0.0072 (0.0002)	-0.0297 (0.0002)	-0.0127 (0.0002)	-0.0032 (0.0003)	-0.0080 (0.0001)	-0.0196 (0.0002)	-0.0234 (0.0001)
Escolaridad acumulada	0.0976 (0.0004)	0.0284 (0.0004)	0.0432 (0.0005)	-0.0174 (0.0006)	0.0764 (0.0003)	0.0095 (0.0003)	0.0476 (0.0002)
Situación de pareja	-	-	-	-	-0.5490 (0.0029)	0.5699 (0.004)	-0.0587 (0.0021)
Hijos menores de 6 años	-0.3029 (0.0037)	0.0304 (0.0054)	0.1252 (0.0061)	0.4608 (0.0592)	-0.1913 (0.003)	0.1148 (0.0051)	-0.4101 (0.0023)
Transferencias	-0.1870 (0.0041)	-0.3402 (0.0068)	-0.5015 (0.0057)	-0.7951 (0.0098)	-0.3080 (0.0033)	-0.5269 (0.0054)	-0.6295 (0.0027)
Jefe de familia	0.4847 (0.0057)	0.3800 (0.0055)	0.4202 (0.0053)	0.3983 (0.0079)	0.4326 (0.0037)	0.3841 (0.0044)	1.2394 (0.0021)
Salario del esposo(a)	-0.0001 (0.0000)	-0.0006 (0.0001)	-	-	0.0000 (0.0000)	0.0001 (0.0000)	-0.0006 (0.0000)
Salario de otro**	0.0000 (0.0000)	0.0007 (0.0001)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0002 (0.0000)
Valle de México	0.0027 (0.0045)	0.1216 (0.0059)	0.1162 (0.0058)	0.0408 (0.007)	0.0385 (0.0035)	0.0868 (0.0045)	0.0258 (0.0027)
Millón	0.0874 (0.0045)	0.1706 (0.0062)	0.1442 (0.0065)	0.1203 (0.0078)	0.1013 (0.0037)	0.1547 (0.0048)	0.0837 (0.0028)
Más de 500 mil hab.	0.1519 (0.0051)	0.1728 (0.0071)	0.1467 (0.0071)	0.0791 (0.0087)	0.1450 (0.0042)	0.1348 (0.0054)	0.0937 (0.0032)
Menos de 100 mil hab.	0.1385 (0.0055)	0.1430 (0.0074)	0.1532 (0.0078)	0.0742 (0.0096)	0.1393 (0.0045)	0.1166 (0.0058)	0.1001 (0.0034)
Constante	-1.2193 (0.0085)	1.6745 (0.009)	-0.0004 (0.01)	0.6692 (0.0119)	-0.4687 (0.0063)	0.8565 (0.0068)	0.3083 (0.0044)
Obs. Censuradas	709375	103428	182360	90867	891735	194295	1086030
Obs. No censuradas	230336	602355	214783	183462	445119	785817	1230936
Chi cuadrada	8943.78	50924.44	26674.53	24680.86	60802.58	126213.59	223976.38

Nota: Los errores estándar están entre paréntesis

Anexo 2. Estadística descriptiva después de considerar el soporte común

	Hombre	Mujer	Diferencia
Escolaridad			
Primaria	0.3107 (0.0005)	0.3844 (0.0004)	-0.0736 (0.0006)
Secundaria	0.2827 (0.0005)	0.2520 (0.0004)	0.0307 (0.0006)
Preparatoria	0.2141 (0.0004)	0.2090 (0.0004)	0.0050 (0.0005)
Licenciatura	0.1764 (0.0004)	0.1440 (0.0003)	0.0324 (0.0005)
Ciclo de vida			
Etapa 1	0.5165 (0.0005)	0.4855 (0.0004)	0.0310 (0.0007)
Etapa 2	0.3278 (0.0005)	0.3419 (0.0004)	-0.0141 (0.0006)
Etapa 3	0.1557 (0.0004)	0.1726 (0.0003)	-0.0169 (0.0005)
Familia			
Hijos menores de 6	0.1891 (0.0004)	0.2955 (0.0004)	-0.1064 (0.0006)
Pareja	0.7201 (0.0005)	0.7029 (0.0004)	0.0172 (0.0006)
pscore	0.5574 (0.0001)	0.5912 (0.0001)	-0.0338 (0.0001)

Nota: estadística después de realizar el balance

Anexo 3. Brecha salarial por estrato del *p-score*, sin tomar en cuenta el ajuste por selección

Cuantil	Diferencial log salario	Brecha salarial	Cuantil	Diferencial log salario	Brecha salarial
block 1	0.0200 (0.0132)	2.02%	block 11	-0.0658 (0.0149)	-6.37%
block 2	-0.1500 (0.0095)	-13.93%	block 12	-0.0233 (0.0115)	-2.30%
block 3	0.0553 (0.0123)	5.69%	block 13	-0.1118 (0.0104)	-10.58%
block 4	-0.0262 (0.0127)	-2.58%	block 14	-0.1681 (0.0118)	-15.47%
block 5	-0.0089 (0.0088)	-0.89%	block 15	-0.3983 (0.0108)	-32.86%
block 6	-0.2632 (0.0123)	-23.14%	block 16	-0.0382 (0.0175)	-3.75%
block 7	-0.0275 (0.0136)	-2.71%	block 17	-0.2150 (0.0112)	-19.34%
block 8	-0.1035 (0.0109)	-9.83%	block 18	-0.2647 (0.0091)	-23.25%
block 9	-0.0332 (0.0127)	-3.27%	block 19	-0.0657 (0.0118)	-6.36%
block 10	-0.0186 (0.012)	-1.84%	block 20	-0.2023 (0.0121)	-18.31%
Promedio	-10.90%		block 21	-0.5059 (0.0146)	-39.71%

Nota: Los errores estándar están entre paréntesis

Anexo 4. Estimación cuantil de la brecha salarial

Decil	Brecha bruta		Brecha con controles		Diferencial
	log-salario	%	log-salario	%	
1	-0.0691 (0.0000)	6.7%	-0.2339 (0.0046)	20.9%	14.2%
2	-0.0682 (0.0006)	6.6%	-0.1621 (0.0048)	15.0%	8.4%
3	-0.0655 (0.0001)	6.3%	-0.1472 (0.0006)	13.7%	7.3%
4	-0.0639 (0.0006)	6.2%	-0.1324 (0.0012)	12.4%	6.2%
5	-0.0672 (0.0004)	6.5%	-0.1162 (0.0029)	11.0%	4.5%
6	-0.0828 (0.0017)	7.9%	-0.1127 (0.0046)	10.7%	2.7%
7	-0.0988 (0.0002)	9.4%	-0.1431 (0.0041)	13.3%	3.9%
8	-0.0949 (0.0008)	9.1%	-0.1638 (0.0047)	15.1%	6.1%
9	-0.0745 (0.0009)	7.2%	-0.1520 (0.0048)	14.1%	6.9%
Promedio		7.3%		14.0%	6.7%

Nota: Los errores estándar están entre paréntesis