

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

ETNIA Y GÉNERO EN EL MERCADO LABORAL ECUATORIANO: CUATRO APLICACIONES EMPÍRICAS PARA LA DESCOMPOSICIÓN SALARIAL

**TESIS PREVIA A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN ESTADÍSTICA
APLICADA**

FRANCISCO ARTURO PÉREZ MOGOLLÓN

francisco_perezxxi@yahoo.es

DANIEL IVÁN TORRESANO MELO

dtorresano@gmail.com

DIRECTOR: Dr. NELSON EDMUNDO SUBÍA CEPEDA

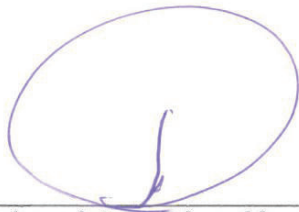
nelsonsubiac@yahoo.es

Quito, marzo 2015

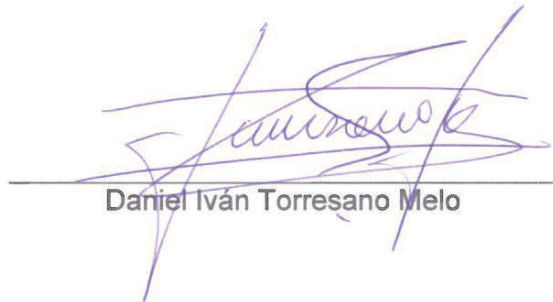
DECLARACIÓN

Nosotros, Francisco Arturo Pérez Mogollón y Daniel Iván Torresano Melo, declaramos que el trabajo aquí descrito es de nuestra autoría; que no ha sido previamente presentada para ningún grado o calificación profesional y que hemos consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.




Francisco Arturo Pérez Mogollón



Daniel Iván Torresano Melo

CERTIFICACIÓN

Yo, Nelson Edmundo Subía Cepeda, certifico que el presente trabajo fue desarrollado por Francisco Arturo Pérez Mogollón y Daniel Iván Torresano Melo, bajo mi supervisión.

A handwritten signature in purple ink, reading "Nelson Edmundo Subía Cepeda". The signature is written in a cursive style with a large initial "N".

Nelson Edmundo Subía Cepeda, PhD

DIRECTOR

AGRADECIMIENTOS

Agradecemos a nuestro director de tesis Nelson Subía por su valioso apoyo para el desarrollo de esta investigación.

DEDICATORIA

A Dios, nuestra familia y amigos.

ÍNDICE DE CONTENIDO

LISTA DE GRÁFICOS	8
LISTA DE TABLAS	9
LISTA DE ANEXOS	10
RESUMEN	11
ABSTRACT	13
CAPÍTULO I	15
INTRODUCCIÓN	15
1.1. OBJETIVOS DE LA INVESTIGACIÓN	20
1.1.1 OBJETIVO GENERAL.....	20
1.1.2 OBJETIVOS ESPECÍFICOS	20
1.2. JUSTIFICACIÓN	21
1.3. HIPÓTESIS DE TRABAJO	28
CAPÍTULO II	29
REVISIÓN DE LA LITERATURA	29
2.1. ALGUNOS CONCEPTOS GENERALES ACERCA DE LA DESCOMPOSICIÓN SALARIAL	29
2.2. METODOLOGÍAS PARA EL ANÁLISIS EMPÍRICO DE LA DESCOMPOSICIÓN SALARIAL	31
2.2.1 Primer método de descomposición de salarios de Blinder y Oaxaca (O-B).....	32
2.2.2 Segundo método de descomposición de salarios de Neumark.....	36
2.2.3 Tercer método de descomposición de salarios de Machado-Mata (M-M)	39
<i>Método de Machado-Mata (M-M) para la estimación de los diferencias salariales</i>	42
2.2.4 Cuarto método de descomposición de salarios de Juhn, Murphy, y Pierce (JMP) comparación en dos momentos del tiempo.....	52
<i>Enfoque de JMP para la descomposición salarial</i>	55
CAPÍTULO III	57
RESULTADOS	57
3.1. DATOS Y ESPECIFICACIÓN DE VARIABLES DE CONTROL	57
3.2. ANÁLISIS DESCRIPTIVO	63
3.3. DISCRIMINACIÓN SALARIAL POR GÉNERO Y ETNIA	67

3.3.1.	Estimación de ecuación de ingreso controlada por género y etnia.....	67
3.3.2.	Descomposición de las diferencias de ingresos por género y etnia en el Ecuador	70
3.4.	DISCRIMINACIÓN SALARIAL POR GÉNERO Y ETNIA A TRAVÉS DE LA METODOLOGÍA MACHADO-MATA (M-M).....	73
3.4.1.	Regresiones cuantílicas de salarios.....	77
3.4.2.	Descomposición de salarios de (Machado & Mata, 2005).....	80
3.5.	DESCOMPOSICIÓN DE DIFERENCIAS SALARIALES EN EL TIEMPO	85
3.5.1.	Enfoque de (Juhn, Murphy, & Pierce, 1991).....	85
CAPÍTULO IV.....		92
CONCLUSIONES Y RECOMENDACIONES		92
4.1.	CONCLUSIONES.....	92
4.2.	RECOMENDACIONES	93
BIBLIOGRAFÍA.....		95
ANEXOS.....		97

LISTA DE GRÁFICOS

Gráfico 1. Clasificación de la Población por Actividad - INEC	64
Gráfico 2. Densidad de los salarios de mujeres y hombres	74
Gráfico 3. Diferencial salarial por percentil de hombres y mujeres	75
Gráfico 4. Densidad de los salarios de indígenas y no indígenas	76
Gráfico 5. Diferencial salarial por percentil por etnia	77
Gráfico 6. Diferencia del logaritmo del salario por hora entre hombres y mujeres entre los percentiles 1 a 99	83
Gráfico 7. Diferencia del logaritmo del salario por hora entre no indígenas e indígenas entre los percentiles 1 a 99	85
Gráfico 8. Diferencial salarial por percentil, 2007-2013.....	87

LISTA DE TABLAS

Tabla 1. Características de la población del Ecuador por Género y Etnia en el año 2013	66
Tabla 2. Estadísticos descriptivos de la población ocupada de 15 a 65 años y estimación de ecuación de ingreso controlada por género y etnia 2013	69
Tabla 3. Descomposición de salarios por los métodos O-B y Neumark por género y etnia 2013	72
Tabla 4. Pruebas de normalidad de los logaritmos del salario por hora por género ...	74
Tabla 5. Pruebas de normalidad de los logaritmos del salario por hora por etnia	77
Tabla 6. Estimación de regresiones cuantílicas de ingresos a nivel poblacional, por género y etnia 2013.....	80
Tabla 7. Descomposición de ingresos de Machado y Mata por género y etnia 2013.	81
Tabla 8. Evolución del salario real (base del IPC 2004), 2007-2013	86
Tabla 9. Descomposición de los cambios en la brecha salarial por género y etnia, 2007-2013	91

LISTA DE ANEXOS

ANEXO- A Validación del modelo	98
ANEXO- B Estimaciones de ecuaciones de ingreso a nivel poblacional, por género y etnia 2013.....	108
ANEXO- C Promedio de las variables explicativas para el modelo JMP.....	109
ANEXO- D Regresiones de Mínimos Cuadrados Ordinarios de salarios	110
ANEXO- E Brecha salarial por género y etnia por percentiles de la distribución para los años 2007 y 2013.Variable dependiente: logaritmo del salario por hora	111

RESUMEN

En esta investigación se analiza el fenómeno de la discriminación salarial por género y etnia en el Ecuador a través de la aplicación de cuatro metodologías empíricas. Para la implementación de estos métodos cuantitativos se utilizó la Encuesta de Empleo y Desempleo Urbano y Rural (ENEMDUR) para los años 2007 y 2013 de representatividad nacional y se consideró la muestra de trabajadores entre los 12 y 65 años; adicionalmente, para la estimación y cálculo de dichos métodos se utilizó el software estadístico STATA. Los principales resultados de la implementación la primera metodología, relacionada con la estimación de modelos de ingreso “aditivos y multiplicativos” para el año 2013, muestran la existencia de discriminación salarial por género y etnia. En este sentido, se estimó que las mujeres e indígenas ganan aproximadamente un 13% y 12% menos que los hombres y no indígenas, respectivamente solo por el hecho de pertenecer a dichos grupos poblacionales. En segundo lugar, a través de la utilización de las metodologías de descomposición de salarios de (Oaxaca, 1973) – (Blinder, 1973) y (Neumark, 1988) para el año 2013 se estimó que la brecha salarial entre los hombres y mujeres es de aproximadamente 0.078 puntos logarítmicos; asimismo, el componente atribuido a la discriminación salarial, estimado a través de la descomposición de salarios de (Oaxaca, 1973) – (Blinder, 1973) representa más del 100% de dicha diferencia, lo que es similar a los resultados obtenidos a través de la aplicación de la metodología (Neumark, 1988). Desde la perspectiva étnica, los resultados son relativamente distintos pues a través de la aplicación de las metodologías antes mencionadas el componente atribuido a la discriminación en contra de los indígenas es relativamente menor y representa alrededor del 40% de las diferencias salariales entre los grupos indígenas y no indígenas. En tercer lugar, a través de la utilización del método (Machado & Mata, 2005), que permite realizar un análisis de las brechas salariales entre dos grupos de trabajadores en los distintos cuantiles de la distribución de sus ingresos, se obtuvo que, para el año 2013, existen diferencias salariales de diferentes dimensiones en

los distintos cuantiles de la distribución particularmente en la parte más alta y más baja de dicha distribución de los grupos de trabajadores por género y etnia. En este sentido la discriminación salarial es significativamente diferente en las distintas partes de la renta, con lo cual el análisis cuantílico y de las brechas en los mismos resulta de trascendental importancia. Finalmente, a través de la descomposición de los salarios del año 2007 y 2013 utilizando la metodología (Juhn, Murphy, & Pierce, 1991), se obtuvo que la brecha salarial tanto por género como por etnia se redujo en los años mencionados, y el componente atribuido a la discriminación tuvo un comportamiento similar.

Palabras clave: discriminación salarial, género, etnia.

ABSTRACT

In this research the phenomenon of wage discrimination by gender and ethnicity in Ecuador through the application of four empirical methodologies are analyzed. Survey of Employment and Unemployment Urban and Rural (ENEMDUR) was used for the years 2007 and 2013 nationally representative for the implementation of these quantitative methods and the sample of workers between 12 and 65 years were considered; additionally, for the estimation and calculation of such methods STATA statistical software was used. The main results of implementing the first approach, related to the estimated revenue models "additive and multiplicative" 2013, showing the existence of wage discrimination by gender and ethnicity. In this sense, it is estimated that women and indigenous earn approximately 13% and 12% less than men and non-indigenous, respectively only by the fact of belonging to these populations. Secondly, through the use of methodologies for decomposing wages (Oaxaca, 1973) - (Blinder, 1973) and (Neumark, 1988) by the year 2013 it was estimated that the wage gap between men and women is of about 0.078 log points; also attributed to wage discrimination component, estimated by decomposing wages (Oaxaca, 1973) - (Blinder, 1973) represents more than 100% of the difference, which is similar to the results obtained through the application of the methodology (Neumark, 1988). From an ethnic perspective, the results are relatively different for through the application of the methodologies mentioned above attributable to discrimination against indigenous component is relatively small and accounts for about 40% of the wage gap between indigenous groups and nonindigenous. Third, through the use of the method (Machado & Mata, 2005), which allows an analysis of the wage gap between two groups of workers in different quantiles of the income distribution, it was found that, for the 2013, wage differentials exist in different dimensions at different quantiles of the distribution particularly in the highest and lowest of the distribution of the groups of workers by gender and ethnicity part. Thus wage discrimination is significantly different in different parts of the rent, which the

cuantílico and gaps in the same analysis is of paramount importance. Finally, through the decomposition of wages in 2007 and 2013 using the methodology (Juhn, Murphy, & Pierce, 1991), it was found that the wage gap by gender and by ethnicity was reduced in the years mentioned, and attributed to discrimination component had a similar behavior.

Keywords: wage discrimination, gender, ethnicity.

CAPÍTULO I

INTRODUCCIÓN

A nivel mundial, ha sido ampliamente aceptado que la pobreza tiene diferentes dimensiones de análisis. Para (Sen, 2001) por ejemplo, dicha variable tiene componentes relacionados con la inequidad en diferentes dimensiones y entre una de ellas la inequidad en los logros profesionales. Este último fenómeno es estudiado generalmente a través del análisis de las brechas o diferencias salariales entre grupos laborales, particularmente a través del análisis de las diferencias de ingresos entre los grupos laborales por género y etnia. Existen diferentes teorías que tratan de explicar las brechas de ingresos entre grupos por género. Una corriente de análisis enfatiza que las diferencias salariales por género están asociadas a la formas de distribución del tiempo y trabajo de los diferentes componentes del hogar. Particularmente refieren que los hogares especializan y optimizan sus tiempos de ocio y trabajo de tal forma que la mujer usualmente debe dedicar parte de su vida laboral a la realización de actividades del hogar por las habilidades innatas de dicho grupo. En tal sentido existe una depreciación de las habilidades laborales del grupo femenino por las interrupciones que estas tiene durante su carrera laboral, en tal sentido estos motivos afectan a las diferencias salariales entre hombres y mujeres (Mincer & Polachek, 1974) - (Becker, 1985).

Por otra parte, existe otra corriente de análisis de los diferenciales salariales entre grupos que establece que dichas brechas son el resultado de la asimetría de la información en el mercado laboral que provoca que los empresarios tomen decisiones condicionadas a la información disponible respecto de las características de ciertos tipos de trabajadores la cual establece o define además diferencias en los niveles de productividad lo que genera que individuos con idénticas habilidades pero pertenecientes a diferentes grupos no tengan una senda profesional equivalente debido al promedio de calidad, real o imaginaria del grupo del cual provienen por lo que las diferencias salariales se explican por el comportamiento discrecional y

discriminatorio en el proceso de selección de los trabajadores al momento de ser contratados (Arrow, 1971) y (Phelps, 1972).

En este sentido los estudios se han enfocado fuertemente al análisis de los diferenciales salariales entre cada uno de los grupos laborales. Las metodologías sobre las cuales se han realizado dichos estudios generalmente están asociadas a los denominados métodos convencionales que comúnmente analizan las diferencias salariales a través de su descomposición en dos componentes asociados a las características observables y no observables de los diferentes grupos laborables (Oaxaca, 1973) - (Blinder, 1973) (O-B). Sin embargo en la literatura empírica es bien conocido que en la implementación de dichas metodologías se asume que los diferentes grupos laborales tiene similares características entre sí cuando posiblemente dentro de los mismo existe una heterogeneidad por lo que se han desarrollado otras metodologías distintas a las convencionales que estudian los diferenciales salariales a través de los diferentes percentiles de la renta (Machado & Mata, 2005) (M-M). Finalmente en la literatura empírica también se ha desarrollado otra corriente que sostiene que los diferenciales salariales efectivamente pueden estar constituidos por dos componentes pero que la dinámica en el tiempo de dichas brechas salariales depende de estos pero además de los efectos de los cambios en la distribución de la renta en el tiempo (Juhn, Murphy, & Pierce, 1991) (JMP).

En esta investigación se analiza las diferencias salariales por género y etnia en el mercado laboral ecuatoriano a través de la utilización de las metodologías mencionadas anteriormente y que serán explicadas con mayor extensión en párrafos siguientes. La principal motivación por la que se realiza esta investigación está relacionada con las características de dichos grupos y sus diferencias salariales observables. En esta línea vale la pena primeramente mencionar que la pobreza total (urbana y rural) medida por ingresos en Ecuador cayó de 36,7% en diciembre de 2007 a 25,6% en el mismo mes del 2013¹. Es decir, aproximadamente 4 de cada 10 ecuatorianos eran pobres en 2007 y 3 de cada 10 en 2013. Por otro lado, la

¹ Estos datos fueron tomados del Instituto Nacional de Estadística y Censos (2014).

concentración del ingreso del país, que puede ser medida a través del coeficiente de Gini², también ha mejorado levemente al pasar de 0,55 en diciembre de 2007 a 0,48 en diciembre de 2013. Sin embargo, en el mercado laboral ecuatoriano, tal como se describe a continuación, parece que existen algunas desigualdades económicas, particularmente de ingresos, entre los distintos grupos laborales.

En este sentido, utilizando los datos de la “Encuesta Nacional de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDUR)” realizada en diciembre del 2013 por el Instituto Nacional de Estadísticas y Censos (INEC), se muestra que mientras los hombres ganan en promedio aproximadamente 360 dólares mensuales, las mujeres perciben un promedio de 288 dólares por los que las diferenciales salariales es de 62 dólares o el grupo masculino gana un 11% más que el femenino. Por otra parte, las diferencias salariales mencionadas anteriormente, también se presentan entre los grupos laborales por etnia en el período antes mencionado. Particularmente, mientras el grupo de trabajadores no indígenas ganan aproximadamente en promedio un monto de 320 dólares mensuales el grupo de indígenas percibe promedio mensual de 192 dólares lo que significa un 40% menos.

En resumen, resulta de trascendental importancia el análisis empírico de las diferencias salariales y sus componentes observables³ y no observables⁴ (discriminación salarial), los cuales pueden ser estudiados utilizando algunas técnicas estadísticas y econométricas relacionadas con el análisis de la descomposición salarial de (O-B), (Neumark, 1988), (M-M) y (JMP), mismas que abordan el problema de la descomposición salarial en un momento del tiempo, sobre la distribución de ingresos de los salarios y en distintos instantes del tiempo, respectivamente. El periodo de análisis para la aplicación del primer grupo de

² El coeficiente de (Gini, 1913) es una medida de concentración del ingreso de una sociedad. Mientras más cercano a 0 el ingreso está distribuido en la misma proporción en toda la población, en tanto que si éste es 1 esto significa que 1 familia o individuo tiene todo el ingreso.

³ Generalmente asociados con características observables individuales de cada una de las personas de los diferentes grupos como: el género, etnia, edad, educación, experiencia, estado civil y condiciones laborales, sector económico de trabajo, entre otras.

⁴ El componente no observable de las diferencias salariales entre los diferentes grupos por género y etnia generalmente está relacionado desde una perspectiva empírica con la discriminación salarial. Al respecto véase (Oaxaca, 1973), (Blinder, 1973), (Cain, 1986), (Atonji & Blank, 1999), entre otros.

metodologías es el año 2013 en tanto que para la aplicación de la última metodología se considera dos períodos el año 2007 y 2013 que permite tener un tiempo de análisis de las diferencias salariales relativamente significativo. Los datos que se utilizan en este estudio están relacionados con las “Encuesta Nacional de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDUR)” de los períodos mencionados.

En el Ecuador, el análisis de las diferencias salariales por género en el mercado laboral, y sus componentes ha sido abordado en algunas investigaciones. En este sentido, (Larrea, 2004) a través de la descomposición de O-B (y estimando primeramente las ecuaciones de ingresos en las que la variable dependiente fue el logaritmo de los salarios reales y variables independientes utilizadas fueron la escolaridad, experiencia, y variables dicotómicas de sector económico de actividad laboral, jefatura de familia, horas trabajadas, variables de control de tiempo y diversas interacciones) y con la utilización de las Encuestas de hogares del sector urbano de las tres principales ciudades del Ecuador para el periodo 1998-2003, se obtuvo como principal resultado que existen diferencias salariales entre el grupo de hombres y mujeres a favor del primer grupo; adicionalmente, se concluyó que el 54% de los diferenciales salariales corresponde al denominado componente no observable generalmente asociado con la discriminación salarial en contra de la mujer; así mismo, se mostró que el porcentaje adicional (46%) de las diferencias entre los ingresos de los dos grupos se asignaron a las diferencias en las dotaciones o características observables.

Por otra parte, (García- Aracil y Winter, 2006), a través de la descomposición O-B para la estimación de las diferencias salariales y sus componentes (utilizando previamente las estimaciones de las ecuaciones de ingreso en las que se incluyó como variable dependiente al logaritmo del ingreso mensual y como variables independientes a las características individuales como el género, educación, experiencia potencial, las horas trabajadas por mes, localización geográfica; dichas ecuaciones fueron estimadas a través de método de Heckman (1979) para corregir

posibles problemas de sesgo de selección) obtuvo como principal resultado que: (i) la brecha de ingresos entre la población no indígena / indígena fue del 104,2%; (ii) utilizando a la estructura de pagos de los indígenas como no discriminatoria la contribución a las brechas salariales por parte de las características observables fue del 63,7%, y el porcentaje asignado a discriminación fue del 36,3%; al considerar a la estructura de ingresos de la población no indígena como no discriminatoria el porcentaje imputado a las diferencias en las características observables fue del 43,7% y el correspondiente a la discriminación fue del 56,3%.

En resumen, tal como se mencionó anteriormente en esta investigación se busca analizar la descomposición salarial de los diferentes grupos salariales por género y etnia en el mercado laboral ecuatoriano en un momento del tiempo, a través de la distribución de la renta y en diferentes momentos del tiempo. Para cumplir con este objetivo este estudio se divide en tres partes. En la primera se establece todo el marco conceptual requerido para el entendimiento del fenómeno en cuestión y las metodologías de estudio. En el segundo se realiza a la aplicación de las metodologías mencionadas y finalmente se emiten algunas conclusiones y recomendaciones acerca de los resultados obtenidos los cuales apuntan a demostrar que existen diferencias salariales significativas desde la perspectiva estadística, entre los grupos laborales por género y etnia en el mercado laboral ecuatoriano en el año 2013 y que dichas diferencias obedecen fundamentalmente a las características no observables o discriminación salarial en contra de las mujeres y los indígenas; además parece que a lo largo de la distribución de la renta de los trabajadores de los diferentes grupos por género y etnia en el mismo período también existen diferencias salariales las cuales se ven explicadas nuevamente por la discriminación salarial, sin embargo no necesariamente dichas diferencias son las mismas a lo largo de toda la distribución de ingresos por lo que sus componentes difieren en los distintos percentiles de la renta; finalmente parece que en el año 2007 y en el 2013 persiste las diferencias salarial y los componentes que explican dichas también está relacionado con la parte no observable o discriminación salarial.

1.1. OBJETIVOS DE LA INVESTIGACIÓN

1.1.1 OBJETIVO GENERAL

Realizar la descomposición de las diferencias salariales entre los grupos de trabajadores por género y etnia en el mercado laboral ecuatoriano utilizando cuatro técnicas cuantitativas: (i) la descomposición salarial de (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) y (Blinder, 1973) (O-B); (ii) (Neumark, 1988), (iii) el método (Machado & Mata, 2005) (M-M) para la estimación de las diferencias salariales entre grupos por género y etnia a través de toda la distribución de la renta, y (iv) método (Juhn, Murphy, & Pierce, 1991) (JMP) para la descomposición de las brechas salariales en el tiempo. Las tres primeras metodologías serán aplicadas para el año 2013 y la última será aplicada para los periodos 2007 y 2013.

1.1.2 OBJETIVOS ESPECÍFICOS

1. Calcular las diferencias salariales y sus componentes observables y no observables (discriminación) por género y etnia en el mercado laboral ecuatoriano a través de la utilización de dos modelos y métodos denominados convencionales: la metodología de descomposición salarial de O-B y Neumark para el año 2013.
2. Estimar las diferencias salariales y sus componentes observables y no observables en cada punto de las distribuciones de la renta de los grupos laborales por género y etnia a través de la utilización de la metodología M-M para el período 2013.

3. Calcular la persistencia en el tiempo de las diferencias salariales por género y etnia en el mercado laboral ecuatoriano del 2007 al 2013 a través de la aplicación de la metodología JMP.

1.2. JUSTIFICACIÓN

Justificación Teórica

El Instituto Ecuatoriano de Estadísticas y Censos (INEC) con el objetivo de hacer una medición de los niveles de empleo y desempleo en el Ecuador en las áreas urbanas y rurales realiza en la “Encuesta Nacional de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDUR)”. Dicha encuesta dirigida a hogares y a las personas que en ellos habitan muestra que en el año 2013, la población de Ecuador fue de 15,9 millones de habitantes, de los cuales el 50,3% correspondió a mujeres y el 49,7% a hombres. Así mismo, la población que se consideró “No indígena” alcanzó un 93,7%, en tanto que el grupo que se auto-identifica como “Indígena” fue de 6,3%

Adicionalmente, en dicha encuesta se muestra que la mayoría de hombres y mujeres viven en la zonas urbanas del país (68.5% de hombres habitan en zonas urbanas y 70.5% de las mujeres en dichas zonas), además dichos grupos son relativamente jóvenes (pues aproximadamente el 40% de la población se encuentra entre 15 y 30 años) y con un nivel de instrucción mayoritariamente secundario (49% de los hombres tiene un nivel de instrucción secundario mientras en las mujeres aproximadamente el 46% se caracterizan por un nivel de instrucción secundario). Sin embargo de lo anterior, la condición de actividad muestra que la mayoría de mujeres son inactivas (49,0%), o subempleadas (29,1%), en tanto que los hombres son también subempleados (37,8%) pero a diferencias del grupo femenino el grupo masculino está más asociado con una condición de ocupación plena (40,0%). Por otra parte, los puestos de trabajo de mujeres en un alto porcentaje están asociados a actividades como “Trabajadoras no calificadas” (24,8%), y como “Trabajadoras de los servicios y comerciantes” (32,1%); en tanto que los hombres también fueron categorizados en su gran mayoría como “Trabajadores no calificados” (25,5%), y

como “Oficiales, operarios y artesanos” (18,2%). Pero una de las características en más diferenciadoras de estos dos grupos laborales es el ingreso mensual. En este sentido, los resultados de la encuesta muestra que mientras los hombres ganan en promedio aproximadamente 360 dólares mensuales las mujeres tiene un ingreso mensual promedio de 288 dólares por lo que los diferenciales salariales entre estas dos poblaciones es de 62 dólares o los hombres ganan un 11% más que el grupo femenino con lo cual considerando que algunas de las características entre los hombres y las mujeres parecen ser similares parece que existieran diferencias salariales explicadas por la discriminación salarial en el mercado laboral con lo cual el estudio de estos fenómenos es de extrema importancia desde una perspectiva empírica.

Por otra parte, desde la perspectiva étnica la encuesta muestra que el grupo de personas auto-identificadas como “No Indígenas” viven en su mayoría en las zonas urbanas (72,4%), en tanto que el grupo de indígenas se encuentran ubicados mayoritariamente en zonas rurales (77,3%). Los dos grupos son relativamente jóvenes pero su nivel de educación difiere, ya que gran parte de la población no indígena alcanza una instrucción secundaria (47,9%), en tanto que aproximadamente el 50,9% de la población indígena apenas alcanza la instrucción primaria. La condición de actividad de los dos grupos evidenció que un alto porcentaje de “No indígenas” son inactivos (35,0%), o subempleados (32,0%), en tanto que los indígenas también son altamente subempleados (56,4%) e inactivos (25,4%). En cuanto a las características de los puestos de trabajo dentro del establecimiento donde laboran los grupos, se observó que los “No indígenas” en un alto porcentaje son categorizados como “Trabajadores no calificados” (24,4%), y como “Trabajadores de los servicios y comerciantes” (22,9%); los “Indígenas” también son categorizados en su gran mayoría como “Trabajadores no calificados” (36,3%), y como “Trabajadores calificados agropecuarios y pesqueros” (30,0). Sin embargo una de las características más diferenciadores entre estos dos grupos laborales son los niveles de ingresos mensuales promedio de los mismos. En este sentido, mientras el

grupo de trabajadores no indígenas ganan aproximadamente en promedio un monto de 320 dólares mensuales el grupo de indígenas percibe un ingreso promedio mensual de 192 dólares lo que significa un 40% menos. En conclusión las brechas salariales entre estos dos grupos es más amplia lo que podría ser resultados de las diferencias en sus características observables descritas anteriormente y de las características no observables relacionadas con la discriminación salarial. Por lo que al igual que en el caso anterior, el estudio de las diferencias salariales y sus componentes, particularmente el relacionado con la discriminación salarial, es de extremo interés desde una perspectiva empírica.

Hasta el momento la literatura empírica en el Ecuador, como se mencionó en la sección anterior, referente al análisis de las diferencias salariales en el mercado laboral ecuatoriano, han aportado al estudio de la descomposición salarial y por ende sus componentes por género y etnia a través de la utilización de metodologías “convencionales”, como la denominada descomposición de O-B. Sin embargo, reconocidos investigadores a nivel internacional definen nuevas aproximaciones al estudio del fenómeno de las diferencias salariales entre grupos laborales, a través de la aplicación de nuevas metodologías que analizan los efectos de la distribución de los salarios (y los cambios de la misma) y el tiempo sobre las diferencias salariales por género y etnia en un país determinado; por lo que, en este trabajo se abordarán dichas metodologías como un aporte al estudio de la descomposición salarial en el mercado laboral ecuatoriano.

Justificación Práctica

En el Ecuador la evidencia empírica desarrollada respecto de la descomposición salarial ha sido abordada como se mencionó anteriormente a través de los denominados métodos convencionales. Sin embargo en este documento se pretende estudiar de una forma más extensiva dicho fenómeno a través de la aplicación práctica de cuatro metodologías. Las primeras dos (O-B) y Neumark que son las aplicaciones empíricas desarrollados a inicios de los años setentas que facilita la cuantificación de las diferencias salariales y su descomposición tal como

se mencionó. La principal ventaja de (O-B) es que permite dividir las diferencias en dos partes una observable y otra no observable. En la primera parte se cuantifica las diferencias de las características observables valoradas con los rendimientos de los hombres y/o mujeres. La segunda parte permite definir el porcentaje o proporción de las diferencias salariales generadas por las características no observables o rendimientos de las dotaciones individuales de los diferentes grupos (O-B, 1973). Sin embargo dichas metodologías poseen algunas debilidades relacionadas con los supuestos en la estimación. Particularmente, se asume que las características observables de los diferentes grupos en análisis son las mismas lo cual es definitivamente un problema pues generalmente existe una gran heterogeneidad entre los grupos laborales; en este sentido por ejemplo en la práctica se observa que cada grupo de trabajadores posee segmentos o subgrupos relacionados por algunas características observables (el sector económico en el cual laboran por ejemplo) lo que generalmente es muy heterogéneo.

Además se supone a priori que existe un grupo mayoritario (o no discriminado) y otro minoritario (o discriminado) de mayor y menor salario respectivamente, lo cual posiblemente sea factible; sin embargo, en la práctica suponer a priori dicha particularidad generalmente cambia los resultados del análisis y las cuantificaciones de los componentes. En este sentido, el suponer que los hombres son el grupo mayoritario que tiene mayor salario implica posiblemente la sobre valoración del componente observable más aún si dicho grupo tiene fuertes diferencias en sus características respecto del grupo femenino (Cahuc & Zylberberg, 2004).

Para contrarrestar el supuesto (Neumark, 1988) a través de diferentes derivaciones concluyó que a priori no es necesario suponer un grupo mayoritario o estructura salarial no discriminatoria por lo que es posible calcular las diferencias y sus componentes con un estimador de los rendimientos de las diferentes características observables que sea igual al obtenido de estimar la muestra conjunta de los grupos en análisis sea la mejor opción para la valoración de los diferencias entre las características y por lo tanto para el cálculo de los componentes. La principal

desventaja de esta metodología es que al igual que el caso anterior, se presupone, para el cálculo de las diferencias, que las características grupales son similares.

En este punto es importante mencionar que el principal aporte que se brindará a la comunidad es la posibilidad de tener una comparación entre las primeras dos metodologías para definir los mejores resultados. Para lograr dicha comparación la metodología que se utiliza implica primeramente estimar ecuaciones de ingresos por cada grupo de trabajadores, las cuales serán contrastadas con los principales estadísticos de análisis de regresión lineal. Posteriormente, se estiman las diferencias salariales por BO y Neumark. Para cada parte de este proceso se utiliza pruebas estadísticas acerca de la significancia de las diferencias salariales entre los grupos y la de sus componentes.

Aun cuando los avances empíricos han sido significativos y han permitido estudiar las diferencias salariales, los resultados descritos muestran algunas limitaciones generalmente asociadas al supuesto de que la diferencia salarial y sus componentes estimados son los mismos en cada parte de la distribución de la renta de los diferentes grupos de análisis. En tal sentido, se supone que el diferencial salarial medio obtenido, sus componentes asociados, es un valor único en toda la distribución de ingresos de los diferentes grupos. Por lo que, tomando en consideración el problema mencionado, (Machado & Mata, 2005) puntualizan que el marco de análisis de oferta y demanda laborales típicamente empleado para analizar los cambios en la estructura salarial, implícitamente asume que para cada tipo de trabajo (calificado y no calificado) y después de controlar por medidas de características, hay un salario único, en contraposición con la existencia de una distribución de salarios. Particularmente, el análisis mencionado se enfoca únicamente en el estudio del salario promedio, lo que es consistente con la estimación de las ecuaciones de salarios mincerianas por el método de mínimos cuadrados ordinarios, que proveen estimaciones de los efectos de la educación y otras variables en la media condicional de la distribución de salarios. En esta misma línea, otros autores han llegado a afirmar que la razón por el interés en los efectos

de la oferta (laboral) radica en la esperanza de que la inversión en capital humano mitigue la futura inequidad. Sin embargo, recientes investigaciones, empleando métodos estadísticos distintos a mínimos cuadrados ordinarios, han revelado que la educación tiene un mayor efecto en los salarios de los individuos, en la parte superior de la distribución de salarios que en la parte inferior de dicha distribución.

En este sentido otro aporte significativo de esta investigación es la utilización de la metodología (M-M) que facilita descomponer las diferencias salariales y sus componentes en cada punto de la distribución de la renta. Dicho método está basado sobre la estimación de la función de densidad marginal de los salarios en un año determinado, que implica la generación de simulaciones de distribuciones de algunas o todas las características observables de los diferentes grupos, para la valoración de los componentes de las diferencias salariales y sus componentes en los diferentes cuantiles de la distribución de ingresos.

En este trabajo, se sigue los pasos propuestos por (Galvis, 2010), para la aplicación de M-M mediante la siguiente metodología: (i) simular m realizaciones de los cuantiles q que se obtienen de una distribución uniforme en el intervalo (0,1); (ii) estimar m coeficientes de la regresión por cuantiles q simulados previamente empleando los datos de la muestra de hombres. Esto nos daría las tasas de retorno de los atributos o dotaciones en diferentes puntos de la distribución condicional de los salarios de los hombres; (iii) muestrear aleatoriamente y con reemplazo un conjunto de atributos de la matriz X pertenecientes a una muestra de m mujeres; (iv) construir una realización de la distribución contrafactual obtenida de multiplicar los atributos seleccionados en (iii) por los retornos a dichos atributos estimados en (ii). Esta simulación corresponde, en consecuencia, a la distribución de salarios que se observaría si los individuos tuviesen los atributos X de las mujeres, X_m , pero fuesen remunerados con los retornos obtenidos por los hombres, β_h ; (v) generar una distribución de las realizaciones obtenidas en (iv) repitiendo los anteriores pasos un número k de veces para obtener intervalos de confianza para los efectos dotación y remuneración. Posteriormente para conseguir simular el componente observable y

no observable de las diferencias salariales para cada uno de los grupos en análisis se estima una ecuación de ingresos para cada uno de los grupos laborales utilizando los pasos (i) al (v) utilizando atributos y sus respectivos retornos para cada uno de los géneros, para obtener la distribución de las mujeres $Q_{\theta}(X'_m\beta(\theta)_m)$ y la distribución de los hombres $Q_{\theta}(X'_h\beta(\theta)_h)$. Finalmente, para calcular los errores estándar de las regresiones cuantílicas y de la descomposición de M-M se utiliza el procedimiento bootstrap que consiste básicamente según (Guan, 2003) en una técnica no paramétrica que permite a través de la muestra de datos original obtener estimaciones empíricas en este caso de los errores estándar de los parámetros calculados, los pasos son los siguientes: i) extraer muestras aleatorias con reemplazo en varias ocasiones de la muestra de datos original; ii) estimar el estadístico deseado de las muestras generadas del paso anterior; y, iii) calcular la desviación estándar de la muestra de la distribución muestral.

La ventaja de esta la metodología descrita es que utiliza métodos no paramétricos para la estimación de las diferencias salariales y sus componentes que a su vez se constituye en un aporte significativo en el análisis de este fenómeno en el Ecuador. Esta metodología es comparada con las dos primeras descritas anteriormente, debido a que permite obtener las diferencias salariales y sus componentes en cada una de los puntos de la distribución de la renta. A priori, se supone que los resultados obtenidos de O-B deben coincidir con los estimados en el percentil 50 de la metodología M-M.

Finalmente, la literatura empírica en el ámbito de la Economía del mercado laboral coinciden que posiblemente las diferencias salariales entre distintos grupos de trabajadores en el tiempo varíen significativamente y por lo tanto los componentes de la misma. Por ello y considerando que en el país todavía no se ha realizado ninguna aplicación para el mercado laboral ecuatoriano, en esta investigación también se desarrolla la metodología JMP que significa analizar las diferencias salariales en dos instantes del tiempo

Para cumplir con los objetivos antes mencionados, se utilizarán las encuestas de empleo y desempleo urbano y rural (ENEMDU) para el período 2007-2013, de representatividad nacional desarrolladas por el Instituto Nacional de Estadística y Censos del Ecuador (INEC), debido a que en el período indicado, las encuestas son comparables ya que contienen las mismas variables, además de ser significativas a nivel poblacional (aun cuando no conforman un panel de datos) lo que facilitará el trabajo de descomposición de salarios en un año dado o en el tiempo.

1.3. HIPÓTESIS DE TRABAJO

- Existen diferencias salariales entre los grupos laborales por género y etnia en el mercado laboral ecuatoriano, las cuales son significativas desde la perspectiva estadística para el año 2013. El principal componente de dichos diferenciales están asociados al componente no observable o discriminación por lo que parece existir discriminación salarial en contra de las mujeres y el grupo laboral indígena para el período antes mencionado.
- Existen diferencias salariales en cada punto de la distribución de la renta de los grupos laborales por género y etnia en el Ecuador para el año 2013. El principal componente de estas diferencias en la distribución mencionada está asociada al componente no observable o discriminación con lo cual parece existir discriminación salarial en contra las mujeres, y no indígenas en el período analizado y a través de la los diferentes puntos de la distribución de la renta.
- Existen diferencias salariales en el año 2007 y 2013 entre los grupo de laborales por género y etnia el mercado laboral Ecuatoriano; dichas diferencias son persistentes entre los dos instantes del tiempo y su componente principal está asociado a las características no observables con lo cual parece que la discriminación salarial entre los grupos mencionados es significativa y persiste en el tiempo.

CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1. ALGUNOS CONCEPTOS GENERALES ACERCA DE LA DESCOMPOSICIÓN SALARIAL

Previo a la estimación de las diferentes técnicas para el análisis de la descomposición salarial, se debe definir algunos conceptos importantes para el entendimiento del fenómeno en estudios. En este sentido, se debe aclarar que previo a la estimación de las diferencias salariales y sus componentes es necesario de primeramente estimar las denominadas ecuaciones de ingresos para cada uno de los grupos lo que facilita la obtención del perfil de ingresos estimado de trabajadores controlado por diferentes características observables de cada uno de los individuos que conforman los grupos de análisis. En este sentido, al conjunto de características observables mencionadas también se le denomina “estructura salarial” la cual puede diferir dependiendo del grupo. Así mismo, se debe mencionar que al conjunto de coeficientes resultado de la estimación de las denominadas ecuaciones de ingresos se les denomina “rendimientos” de las características observables y se conoce también como características no observables asociadas a la denominada discriminación que será definida más adelante. Un segundo paso, luego de las estimaciones mencionados, se puede obtener las diferencias salariales entre grupos laborales y esta manera descomponer dichas diferencias salariales o brechas de ingresos.

En este sentido una de las primeras definiciones que requieren ser aclaradas es el concepto desde la perspectiva teórica y empírica de las denominadas “ecuaciones de ingreso” o “ecuaciones mincerianas de ingresos” ampliamente utilizadas en la economía laboral para el estudio de los perfiles de ingresos de los trabajadores. El estudio de estas ecuaciones fue desarrollado por (Mincer, 1958) el cual desarrolló

un marco conceptual y empírico que demuestra que los perfiles de ingresos de los individuos en el mercado laboral, corresponde a un proceso de optimización dinámica en el cual los agentes minimizan los costos y a la vez, maximizan los ingresos de la educación. Desde una perspectiva empírica, la aplicación de dicha teoría implicó el planteamiento y estimación de una ecuación denominada “minceriana” o de ingresos, que es una función semi-logarítmica de ingresos que expresa la relación lineal entre el logaritmo de ingresos de los individuos, su nivel de educación (nivel de escolaridad observable-6 años) y el nivel de experiencia laboral (observable y su cuadrado lo que demuestra una relación no necesariamente lineal entre el ingreso y la experiencia). Luego de dicha aplicación, en el marco de la Economía de la educación y del mercado laboral, se han desarrollado diferentes versiones de las ecuaciones de ingresos, evolucionado hacia ecuaciones más complejas, que además de contener controles de educación y nivel experiencia de los trabajadores como determinantes del ingreso, han introducido otras variables independientes como las relacionadas con el género, la etnia, localización geográfica, sectoriales y otras características observables de los individuos, para de esta manera analizar desde otra perspectiva fenómenos como los de la discriminación salarial en el mercado laboral.

Al igual que en el caso anterior, otro de los conceptos que se requieren aclarar previo al estudio de las diferencias salariales y sus componentes entre grupos laborales es el de la denominada discriminación salarial. Este fenómeno, ampliamente estudiado en la economía laboral, teóricamente es entendido como una situación en la cual un grupo de trabajadores tiene un mayor salario en relación a otro que posee las mismas características observables como niveles de educación, experiencia laboral, entre otras condiciones socio económicas. En este sentido es importante decir que desde una perspectiva empírica la discriminación salarial puede ser calculada a través de las estimaciones de ecuaciones de ingresos controladas por género o etnia entendiendo a la discriminación salarial como el coeficiente positivo o negativo que acompaña a dichos controles y que debe ser significativo desde la perspectiva estadística. Adicionalmente, otra forma de abordar la discriminación salarial es a

través de la denominada descomposición salarial. Este es un método de estimación de las brechas salariales a través de la descomposición de las mismas en dos partes: un componente asociado a las características observables y otro a las no observables. Luego del estudio de las diferencias o brechas salariales y de evidenciar su existencia entre dos grupos de trabajadores, y además de obtener sus componentes también es factible el cálculo de la discriminación salarial ya que a este fenómeno generalmente se lo asocia desde una perspectiva empírica al llamado componente no observable de la diferencia salarial.

2.2. METODOLOGÍAS PARA EL ANÁLISIS EMPÍRICO DE LA DESCOMPOSICIÓN SALARIAL

En este punto se resume los cuatro métodos de descomposición salarial⁵ que se utilizarán para el estudio de las diferencias salariales por género y etnia, y por tanto para cuantificar la parte de estas diferencias salariales atribuida a discriminación salarial. En primer lugar, se describe el método de Oaxaca-Blinder (O-B), en segundo lugar, el método de Neumark, los dos se basan en modelos de regresión lineal, el tercer método es el de Machado-Mata (M-M), para el que se utilizan regresiones cuantílicas; los tres métodos citados anteriormente se utilizan para un momento del tiempo; y, finalmente, el cuarto método de descomposición, denominado JMP (Juhn, Murphy, & Pierce), que permite mediante regresiones lineales obtener la descomposición salarial comparando dos momentos del tiempo.

Para estimar los dos primeros métodos, O-B, Neumark; y, el cuarto JMP, se utilizaron en una primera instancia modelos de regresión lineal de salarios en función de las características observables, a estos modelos se les aplicó varias pruebas y procedimientos para corregir: puntos atípicos, influyentes y palanca,

⁵ El término de descomposición salarial se lo puede entender como a la parte de la diferencia salarial entre un grupo de altos y bajos ingresos y como se descompone esta diferencia en una parte explicada por los factores observables como educación, experiencia laboral, y otra parte por factores no observables entre estos la discriminación salarial.

heteroscedasticidad; también se realizaron pruebas de linealidad del modelo, prueba de multicolinealidad, y pruebas de normalidad de los residuos; para corregir este último problema se aplicó el método de bootstrap⁶, para obtener mejores errores estándar e intervalos de confianza de las regresiones de salarios, los procedimientos aplicados se pueden observar en el Anexo A.

2.2.1 Primer método de descomposición de salarios de Blinder y Oaxaca (O-B)

Modelo de regresión lineal de salarios

Para realizar la descomposición salarial se utilizan las ecuaciones de salarios mincerianas, las cuales tienen como variable dependiente al logaritmo del salario por hora y como variables independientes las características observables de los trabajadores (educación, experiencia laboral, estado civil, jefe de hogar, migración, y variables de localización geográfica como área rural, entre otras variables que se las puede obtener de las encuestas de empleo como: género, etnia⁷; y, otras como el ratio de Mills). El modelo descrito puede ser expresado como:

$$w_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ji} + \mu_i \quad (1)$$

Donde: w_i es el logaritmo del ingreso por hora del individuo $i = 1, 2, \dots, n$ y X_{1i}, \dots, X_{ki} son las k características individuales observables para explicar w_i , μ es el término de error que sigue una ley normal $\mu \sim N(0, \sigma^2)$ y no está auto-correlacionado, $E(\mu_i \mu_j) = cov(\mu_i, \mu_j) = 0$; donde β_1, \dots, β_k son los coeficientes a estimarse en función de X_{1i}, \dots, X_{ki} ; y, β_0 es el intercepto del modelo a estimarse por el método de Mínimos Cuadrados Ordinarios.

⁶ Según (Guan, 2003) en una técnica no paramétrica que permite a través de la muestra de datos original obtener estimaciones empíricas en este caso de los errores estándar de los parámetros calculados, los pasos son los siguientes: i) extraer muestras aleatorias con reemplazo en varias ocasiones de la muestra de datos original; ii) estimar el estadístico deseado de las muestras generadas del paso anterior; y, iii) calcular la desviación estándar de la muestra de la distribución muestral.

⁷ Correspondientes generalmente a grupos de ingresos más bajos en relación a hombres y/o no indígenas. En el caso de que los coeficientes asociados fueran negativos y estadísticamente significativos, se podría decir que parecería haber discriminación salarial en contra de las mujeres o indígenas, debido a que por el hecho de ser mujer recibirían un menor salario por hora en comparación a hombres y no indígenas.

Inclusión del ratio de Mills en la regresión de salarios y solución del problema de sesgo de selección muestral

Otra observación relativa a la regresión de salarios es que se excluye de la muestra a los trabajadores que no participan en el mercado laboral, sin embargo para (Madden, 2000), si la discriminación también aplica en términos de acceso al mercado laboral, entonces este fenómeno debería ser tomado en cuenta cuando se trata de construir un escenario más completo de la discriminación mercado laboral. Esto supone la adición de una variable de control como un regresor en la función de esperanza condicional de los salarios.

Lo anterior quiere decir que puede existir el caso en que la muestra de ocupados hombres y mujeres ó indígenas o no indígenas, no sea una muestra aleatoria, en vez de eso podría sufrir un sesgo de selección. Entonces, si las características inobservables que influyen en el acceso al mercado laboral están correlacionadas con los factores inobservables que influyen en los salarios, entonces hay una relación entre el proceso determinante de la participación en el mercado laboral y el proceso determinante de los salarios. Si a su vez, los factores inobservables están correlacionados con los factores observables en la ecuación de salarios, entonces, no tener en cuenta esto conduciría a un sesgo en los coeficientes de la ecuación de salarios.

La forma estándar de tratar con este problema es adicionar un término extra, la inversa del ratio de Mills (IRM) (Heckman, 1979), a la ecuación estándar de salarios. Para lo cual se implementa el siguiente procedimiento (Baltagi, 2008): i) se estima un modelo probit de participación en el mercado laboral: $y_i^* = \gamma_0 + \sum_{j=1}^k \gamma_j Z_{ji} + v_i$, $i = 1, 2, \dots, n$, donde $v \sim N(0, \sigma^2)$; $y_i^* = \{1 \text{ ocupado}; 0 \text{ no ocupado}$ y dicha probabilidad de estar ocupado está condicionada a un conjunto de características observables individuales Z_i : sexo, indígena, edad, años de escolaridad, jefe hogar, estado civil, ingresos recibidos por: bono de desarrollo humano, remesas, capital; y, variables de localización geográfica como área urbana; ii) se calcula el ratio de Mills y se lo

incorpora en la regresión de salarios, para esto se supone que μ_i y v_i están las dos normalmente distribuidas con media cero y varianza y covarianza $var\begin{pmatrix} \mu_i \\ v_i \end{pmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$, así el ratio de Mills es $\rho\sigma v_i$ donde v_i es reemplazado por su media condicional $E[v_i/y_i = 1] = \phi(\gamma_0 + \sum_{j=1}^k \gamma_j Z_{ji}) / \Phi(-(\gamma_0 + \sum_{j=1}^k \gamma_j Z_{ji}))$; donde $\phi(\cdot)$ es la función de densidad de la normal estándar y $\Phi(\cdot)$ es la normal estándar de la función de distribución acumulada con varianza de v_i normalizada a uno. Entonces el modelo de regresión de salarios quedaría

$$w_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ji} + \rho\sigma \frac{\phi(\gamma_0 + \sum_{j=1}^k \gamma_j Z_{ji})}{\Phi(-(\gamma_0 + \sum_{j=1}^k \gamma_j Z_{ji}))} + \eta_i ; i = 1, 2, \dots, n; \text{ donde } \eta \sim N(0, \sigma^2); \text{ iii)}$$

si el ratio de Mills es estadísticamente significativo implica que si existe un problema de sesgo de selección y este término debe ser incluido en el modelo, caso contrario se lo elimina de la regresión.

Modelo de Blinder y Oaxaca para la descomposición salarial⁸

Por otra parte, si se está interesado en comparar dos grupos demográficos de altos y bajos ingresos (sean estos por ejemplo, hombres y mujeres ó indígenas y no indígenas), se realiza la estimación de la ecuación (1) para cada grupo:

$$w_i^A = \beta_0^A + \sum_{j=1}^k \beta_j^A X_{ji}^A + u_i^A \quad (2)$$

$$w_i^B = \beta_0^B + \sum_{j=1}^k \beta_j^B X_{ji}^B + u_i^B \quad (3)$$

Donde el superíndice A indica al grupo de altos salarios y el superíndice B indica el grupo de bajos ingresos. Adicionalmente, en este modelo, se supone que las esperanzas de los términos de error condicionadas a las características observables de las ecuaciones (2) y (3) de cada uno de los grupos es cero o $\mathbb{E}(u_i^A | X_i) = 0$ y $\mathbb{E}(u_i^B | X_i) = 0$, y X es el vector de variables independientes. Esto último es

⁸ Este apartado es un resumen de (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) y (Blinder, 1973).

concordante con los supuestos del modelo de Mínimos Cuadrados Ordinarios (MCO), por lo cual adicionalmente al supuesto antes mencionado en esta investigación, se validan todas las hipótesis establecidas en el modelo MCO.

Dado (2) y (3), es importante calcular el diferencial: $\sum_j \beta_j^A \bar{X}_j^A - \sum_j \beta_j^B \bar{X}_j^B$, adicionalmente se puede conocer la parte de la diferencia relacionada a los coeficientes β_j^A y β_j^B ; y a las diferencias en los promedios de las características, \bar{X}_j^A y \bar{X}_j^B . En particular, la brecha salarial entre el grupo de alto y bajos ingresos puede expresarse como:

$$\sum_j \beta_j^A \bar{X}_j^A - \sum_j \beta_j^B \bar{X}_j^B = \sum_j \beta_j^A (\bar{X}_j^A - \bar{X}_j^B) + \sum_j \bar{X}_j^B (\beta_j^A - \beta_j^B) \quad (4)$$

Pero adicionalmente, (4) también puede representarse como:

$$\sum_j \beta_j^A \bar{X}_j^A - \sum_j \beta_j^B \bar{X}_j^B = \sum_j \beta_j^B (\bar{X}_j^A - \bar{X}_j^B) + \sum_j \bar{X}_j^A (\beta_j^A - \beta_j^B) \quad (5)$$

El primer sumando de (4) o (5) corresponde a la parte de la diferencia de los salarios debido a las diferencias en las características observables de los grupos A y B. El segundo término, es la parte de la brecha salarial debido a las diferencias en los coeficientes relacionados a cada una de las estimaciones de las ecuaciones de ingresos de los grupos mencionados, comúnmente denominado o asociado a la "Discriminación salarial". En este sentido, para (Neumark, 1988), si en ausencia de discriminación del grupo A y B pudiera recibirse idénticos retornos para las mismas características, y las diferencias en los salarios pudieran deberse sólo a las diferencias en los pagos relacionados a las características, entonces el segundo término puede ser interpretado como la parte de la brecha salarial debido a la discriminación.

Como se puede apreciar (4) y (5) involucran un problema relacionado con la elección de la estructura salarial no discriminatoria, es decir los resultados de la descomposición de las brechas salariales de los grupos de altos y bajos ingresos

depende de los precios (o coeficientes) que sirven para valorar las características observables y por ende, para la valoración de las diferencias salariales tal como lo detallan en el método O-B.

2.2.2 Segundo método de descomposición de salarios de Neumark

Para (Neumark, 1988), en concordancia con lo mencionado en el párrafo anterior, las ecuaciones (4) y (5) pueden presentar diferentes resultados para un mismo fenómeno. En este sentido para entender la fuente de estas diferencias se deben analizar los supuestos en el uso de las ecuaciones mencionadas. Particularmente en (4) se asume que en ausencia de discriminación la estructura salarial de los hombres podría prevalecer, por lo tanto en la estimación de los diferenciales salariales que podrían existir, los coeficientes de la estructura salarial de los hombres son usadas para ponderar las diferencias en las características. Contrariamente en (5) se asume que la estructura salarial de las mujeres podría prevalecer.

En tal sentido, el autor mencionado propone la elección de la estructura salarial que no necesariamente corresponda a la del grupo mayoritario o minoritario, sino aquella que provenga de la naturaleza del comportamiento discriminatorio. Esto significa que: β^A y β^B , no son las únicas estructuras salariales discriminatorias consideradas, en tal sentido propone el siguiente formato de descomposición:

$$\sum_j \beta_j^A \bar{X}_j^A - \sum_j \beta_j^B \bar{X}_j^B = \sum_j \beta (\bar{X}_j^A - \bar{X}_j^B) + \left[\sum_j \bar{X}_j^A (\beta_j^A - \beta) - \sum_j \bar{X}_j^B (\beta_j^B - \beta) \right] \quad (6)$$

Donde β es la estructura salarial no discriminatoria. En (6) el primer término puede ser interpretado como la parte de las diferencias en los salarios debido a las diferencias en las características, así mismo, el segundo término puede ser entendido como la parte debido a la discriminación. Si se asume que en ausencia de discriminación la estructura salarial de los hombres prevalece, entonces $\beta = \beta^A$ y la descomposición (6) es idéntica a (4). Si en vez de esto se asume que $\beta = \beta^B$, entonces (6) se reduce a (5), por lo que los estimadores de O-B son dos casos especiales de este caso general de descomposición.

Una de las cuestiones que surge es cómo estimar la estructura salarial no discriminatoria β . Se propone el método de mínimos cuadrados para estimar la estructura salarial no discriminatoria de la muestra conjunta del grupo de salarios altos y bajos, o en el presente contexto (hombres y mujeres o no indígenas e indígenas). Esto significa que:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{Y}) \quad (7)$$

Donde \mathbf{X} es la matriz de observaciones de variables independiente \mathbf{Y} es el vector que contiene el logaritmo natural de los salarios por hora y $\hat{\beta}$ es el estimador de mínimos cuadrados ordinarios obtenidos de la muestra conjunta de hombres y mujeres o no indígenas e indígenas.

Estimación de las varianzas muestrales para los métodos de O-B y Neumark

Las expresiones anteriores definen las brechas salariales entre diferentes grupos de análisis, y los componentes explicativos de dichas diferencias. Sin embargo, para definir la robustez de dichos resultados es importante considerar el grado de variabilidad y significancia de los mismos. Para (Jann, 2008) mientras un importante resultado del análisis de regresión es que no importan las estimaciones de la varianza, si los regresores son estocásticos, esto no es necesariamente cierto en la descomposición O-B o de Neumark. Es decir, dicha descomposición está basada en los coeficientes de una regresión multivariante para la media de los regresores. Por lo que, si los regresores son estocásticos, entonces las medias tienen una varianza muestral. Estas varianzas son del mismo orden asintótico que las varianzas de los coeficientes. Para conseguir un error estándar para los resultados de la descomposición, se debe tomar en cuenta la variabilidad inducida por la aleatoriedad de los predictores.

En este sentido, dada la siguiente expresión

$$w_i = \sum_{j=1}^k \beta_j X_{ji} + \mu_i; i = 1, 2, \dots, n; \mu \sim N(0, \sigma^2) \quad (8.1)$$

Donde \bar{X} es un vector de medias estimado para los predictores y $\hat{\beta}$ contiene las estimaciones de los coeficientes por el método de mínimos cuadrados. Si los predictores son estocásticos entonces \bar{X} y $\hat{\beta}$ conjuntamente están sujetos a variación muestral o de muestreo. Asumiendo que estos dos términos no estén correlacionados entonces la varianza de (8.1) se puede escribir así:

$$V(\bar{X}\hat{\beta}) = E(\bar{X})V(\hat{\beta})E(\bar{X}) + E(\hat{\beta})V(\bar{X})E(\hat{\beta}) + tr\{V(\bar{X})V(\hat{\beta})\} \quad (8.2)$$

Donde $V(\bar{X})$ y $V(\hat{\beta})$ son la matriz de varianzas y covarianzas de \bar{X} y $\hat{\beta}$, respectivamente. A través de la introducción de las estimaciones de las esperanzas y las varianzas se logra conseguir el estimador de la varianza de la regresión de la siguiente manera:

$$V(\bar{X}\hat{\beta}) = \bar{X}'\hat{V}(\hat{\beta})\bar{X} + \hat{\beta}'\hat{V}(\bar{X})\hat{\beta} + tr\{V(\bar{X})V(\hat{\beta})\} \quad (8.3)$$

Donde $\hat{V}(\hat{\beta})$ es la matriz de varianzas y covarianzas obtenida del proceso de regresión. Un estimado cercano de $V(\bar{X})$ es $\hat{V}(\hat{\beta}) = \chi'\chi / \{n(n-1)\}$ donde χ es la matriz centrada de \mathbf{X}

El proceso análogo al anteriormente descrito para la definición de las varianzas en la descomposición de los diferenciales salariales de la metodología O-B puede ser derivado de forma similar. Asumiendo que los grupos en análisis son independientes, la aproximación a un estimador de la varianza para los dos términos de la descomposición de las brechas es:

$$\hat{V}\{(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A\} \approx (\bar{X}_A - \bar{X}_B)'\hat{V}(\hat{\beta}_A)(\bar{X}_A - \bar{X}_B) + \hat{\beta}_A'\{\hat{V}(\bar{X}_A) + \hat{V}(\bar{X}_B)\}\hat{\beta}_A \quad (8.4)$$

Y

$$\hat{V}\{\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B)\} \approx \bar{X}_B'\{\hat{V}(\hat{\beta}_A) + \hat{V}(\hat{\beta}_B)\}\bar{X}_B + (\hat{\beta}_A - \hat{\beta}_B)'\hat{V}(\bar{X}_B)(\hat{\beta}_A - \hat{\beta}_B) \quad (8.5)$$

En estas expresiones se utiliza la definición que establece que la varianza de la suma de dos variables aleatorias no correlacionadas es igual a suma de las varianzas individuales. Un importante hecho de estas dos últimas expresiones es que ignora la naturaleza estocástica de los predictores que afectarán la varianza del primer término de la descomposición de la descomposición (de la parte explicada).⁹

2.2.3 Tercer método de descomposición de salarios de Machado-Mata (M-M)

Regresiones cuantílicas

Para (Machado & Mata, 2005), el marco de análisis de oferta y demanda laborales típicamente empleado para estudiar los cambios en la estructura salarial (o perfil de salarios de los trabajadores), implícitamente asume que para cada tipo de trabajo (calificado y no calificado), después de controlar por medidas de características observables, hay un salario único, en contraposición con la existencia de la distribución de salarios de cada uno de los grupos de estudio. Particularmente, el análisis mencionado se enfoca en el estudio del salario promedio, lo que es consistente con la estimación de las ecuaciones de salarios mincerianas por el método de mínimos cuadrados ordinarios para todo el grupo de trabajadores, que proveen estimaciones de los efectos de la educación y otras variables en la media condicional de la distribución de salarios. La razón por el interés en los efectos de la oferta (laboral), radica en la esperanza de que la inversión en capital humano mitigue la futura inequidad. Sin embargo, recientes investigaciones, empleando métodos estadísticos distintos a mínimos cuadrados ordinarios han revelado que la educación tiene un mayor efecto en los salarios de los individuos en la parte superior de la distribución de salarios que en la parte inferior de dicha distribución.

Para (García, Hernández, & Lopez-Nicolas, 2001), desde la perspectiva de género, mujeres con características no observables con un ingreso por arriba del promedio del ingreso condicionado a sus características observadas, se compararán con hombres cuyos salarios podrían estar situados por arriba de la media de los salarios

⁹ Para más explicación acerca de este punto véase (Jann, 2008).

condicionados sobre las mismas características observadas. Por lo tanto, para resolver el problema descrito, se utiliza la comparación de los cuantiles de las dos distribuciones de salarios condicionados a los mismos conjuntos de características observables, lo que da como resultado una aproximación a una medida no observable o de discriminación salarial, lo que se representa en la siguiente ecuación de descomposición salarial por género en el cuantil (θ):

$$\widehat{\Delta}_i = \widehat{Q}_\theta(w_h|X_i) - \widehat{Q}_\theta(w_m|X_i) \quad (9)$$

Donde, $\widehat{\Delta}_i$ corresponde a las diferencias salariales en un cuantil determinado i , $\widehat{Q}_\theta(w|X_i)$ representa el θ i -ésimo cuantil de la función de densidad del ingreso condicionado a X_i ; cuya representación econométrica formal está dada por:

$$\widehat{Q}_\theta(w_h|X_i) = w_{ih}^\theta = X_{ih}\beta_\theta + e_i \quad (9.1)$$

En esta última ecuación se asume que en el i -ésimo cuantil, el término de error que contiene los efectos fijos inobservables y el puro elemento aleatorio es cero. Bajo esta representación la medida de discriminación en la ecuación (9) está definida por:

$$\widehat{Q}_\theta(w_h|X_i) - \widehat{Q}_\theta(w_m|X_i) = X_i'(\beta_{\theta h} - \beta_{\theta m}) \cong \widehat{\Delta}_i \quad (9.2)$$

Esta aproximación, requiere obtener estimaciones de las funciones cuantílicas condicionadas de los salarios para hombres y mujeres (y de la población de trabajadores indígenas y no indígenas). Dichas aproximaciones serán utilizadas en esta investigación lo que a su vez se constituirá en la primera aplicación en el país para el análisis del fenómeno de la descomposición salarial por género y etnia.

Sin embargo, previo a describir la alternativa metodológica para estimar (9.2) desarrollada por M-M, que a su vez integra el método de regresiones cuantílicas en cada una de los grupos y la metodología de descomposición O-B, es importante mencionar algunas particularidades de los cuantiles (representados en 9.1), su estimación y la motivación para su utilización. Esta metodología parte de la

elemental definición de los cuantiles de la muestra, la cual es extendida al modelo lineal.

Dado, $\{w_i: i = 1, \dots, n\}$ una muestra aleatoria de la variable aleatoria W con función de distribución F . Entonces el cuantil θ i-ésimo de la muestra, con $0 < \theta < 1$, podría ser definido como una solución al problema de minimización:

$$\min_{b \in \mathbb{R}} \left[\sum_{i:w_i \geq b} \theta |w_i - b| + \sum_{i:w_i < b} (1 - \theta) |w_i - b| \right] \quad (10)$$

Por otra parte, la observación de que los puntos atípicos son difíciles de identificar en el contexto de regresión, señala la ambigüedad esencial involucrada en la extensión al modelo lineal de las nociones ordinarias de los cuantiles muestrales basados en una ordenación de las observaciones de la muestra. Por lo que una generalización del problema de minimización planteado en (10) resuelve la ambigüedad detallada.

Denotando $\{x_i: i=1, \dots, n\}$ una secuencia de K vectores de una conocida matriz de diseño (matriz de características observables), suponiendo $\{w_i, i = 1, \dots, n\}$ una muestra aleatoria del proceso de regresión $u_i = w_i - x_i b$ con distribución F . La θ i-ésima regresión cuantílica (tal como se muestra la expresión 9.1), con $0 < \theta < 1$, es definida como la solución del problema de minimización:

$$\min_{b \in \mathbb{R}^K} \left[\sum_{i:w_i \geq b} \theta |w_i - x_i b| + \sum_{i:w_i < b} (1 - \theta) |w_i - x_i b| \right] \quad (11)$$

En el modelo de localización ($K = 1$, $x_i = 1$, para todo i) los problemas de minimización descritos en las ecuaciones (10) y (11) coinciden. Adicionalmente, el estimador del error absoluto mínimo es la regresión de la mediana que también se obtiene, en la regresión cuantílica para $\theta = 1/2$ ¹⁰.

Por otra parte para (Buchinsky, 1998), el modelo lineal análogo del θ i-ésimo cuantil es definido en una manera similar. Esto es, $\widehat{\beta}_\theta$ el estimador para β_θ en la ecuación (9.2) es resultado del siguiente problema de minimización:

¹⁰ En este sentido para (Galvis, 2010) las estimaciones por MCO son un caso especial de la estimación por cuantiles, pues la regresión por MCO daría aproximadamente los resultados de la regresión por cuantiles evaluados en el percentil 50.

$$\min_{\beta \in \mathbb{R}^K} \left[\sum_{i:w_i \geq \beta} \theta |w_i - x_i \beta| + \sum_{i:w_i < \beta} (1 - \theta) |w_i - x_i \beta| \right] \quad (12)$$

Para (Koenker & Bassett Jr, 1978) existen algunas propiedades de los elementos de $\beta^*(\theta)$, del conjunto solución $B^*(\theta)$ del problema de minimización de la regresión cuantílica. Para los autores mencionados, dado $\mathcal{N} = \{1, 2, \dots, N\}$ y \mathcal{H} el conjunto de K elementos subconjunto de \mathcal{N} . Los elementos $h \in \mathcal{H}$ tienen un complemento relativo $\bar{h} = \mathcal{N} - h$, y ambos sirven a la partición de w y X .¹¹ Finalmente dado:

$$H = \{h \in \mathcal{H} | \text{rango } X(h) = K\}$$

En tal sentido si X tiene un rango K en el conjunto de regresiones cuantílicas, $B^*(\theta)$ tiene al menos un elemento de la forma:

$$\beta^*(\theta) = X(h)^{-1}w(h) \quad (13)$$

Para algún $h \in H$. Más aun $B^*(\theta)$ es el conjunto compacto de todas las soluciones que tienen la forma de la ecuación (13). Adicionalmente, si F es continua con probabilidad 1: $\beta^* = X(h)^{-1}w(h)$ es una solución única al problema si y solo si

$$(\theta - 1)_{i_K} < \sum_{i \in h} \left[\frac{1}{2} - \frac{1}{2} \text{sgn}(w_i - x_i \beta^*) - \theta \right] x_i X(h)^{-1} < \theta_{i_K} \quad (14)^{12}$$

Las estimaciones mencionadas anteriormente serán realizadas para el análisis de la discriminación salarial por género y etnia.

Método de Machado-Mata (M-M) para la estimación de las diferencias salariales

Para (Machado & Mata, 2005) un total entendimiento de los cambios que han ocurrido en los salarios requiere de la desagregación de los efectos de los cambios en el stock de capital humano en la población trabajadora y de los efectos de los cambios en los retornos a los componentes de dicho capital. En este sentido, dichos investigadores proponen una metodología extendida de la tradicional de la descomposición O-B sobre la media de los salarios para toda la distribución de los salarios. Dicho método está basado sobre la estimación de la función de densidad

¹¹ Por ejemplo $w(h)$ denota el (K) - vector con elementos $\{w_i; i \in h\}$ mientras $X(\bar{h})$ denota a una matriz $(T-K) \times K$ con filas $\{x_i; i \in \bar{h}\}$.

¹² Para un análisis más exhaustivo del modelo de regresión cuantílica y sus resultados véase (Koenker & Bassett Jr, 1978).

marginal de los salarios en un año determinado, lo que implica la generación de distribuciones contra factuales o simuladas de algunos o todas las características observables.

La naturaleza del ejercicio requiere la estimación de la distribución de salarios condicional sobre las variables de interés. Para conseguir este objetivo, primeramente se utilizan regresiones cuantílicas, esto significa la estimación de modelos para los cuantiles de la distribución condicional de los salarios. Es así como, las regresiones cuantílicas capturan el impacto de los cambios en las covariantes de una distribución condicional de salarios, de la misma manera que la regresión de MCO mide el impacto de los cambios de las covariantes sobre la media de la distribución condicional de los salarios.

Sin embargo, la metodología M-M va más allá de un modelo condicional. En efecto, una distribución condicional no refleja la variabilidad de las covariantes en la población. En otras palabras, dicha distribución es la que prevalecería si todos los trabajadores tienen las mismas características observables. Por lo que el segundo paso de esta metodología propone la marginalización de la distribución condicional estimada en el primer paso, usando diferentes escenarios para la distribución de los atributos de los trabajadores.

En los siguientes párrafos se describe la metodología M-M que se resumen como: i) modelamiento de la distribución de salarios; ii) modelamiento de la distribución marginal de salarios; y, iii) descomposición de los cambios en la densidad de los salarios.

i) Modelamiento de la distribución de salarios

Dado $Q_{\theta}(w|X_i)$ para $\theta \in (0,1)$ que denota θ i -ésimo cuantil de la distribución del (logaritmo) de los salarios, dado un vector X_i de covariantes en el que se incluyen generalmente una constante, una variable dicotómica de control de género de los individuos, su nivel de escolaridad, edad, experiencia entre otras. En tal sentido, la modelación de los cuantiles condicionados está definida por:

$$Q_{\theta}(w|X_i) = X_i\beta_{\theta} \quad (15)$$

Donde β_{θ} es un vector de coeficientes de la regresión cuantílica (RC).

Particularmente, para $\theta \in (0,1)$, β_{θ} puede ser estimado por la minimización en β de

$$n^{-1} \sum_{i=1}^n \rho_{\theta}(w_i - X_i\beta)$$

Con

$$\rho_{\theta} = \begin{cases} \theta_u & \theta_u \text{ para } u \geq 0 \\ (\theta - 1)_u & \text{para } u < 0 \end{cases}$$

Para los autores, un aspecto importante a considerar de $Q_{\theta}(w|X_i)$ es que se obtiene una total caracterización de la distribución condicional de los salarios. Además, los coeficientes estimados de la RC pueden ser interpretados como los retornos (o precios) de las habilidades en el mercado laboral, en diferentes puntos de la distribución condicional de los salarios.¹³

ii) Modelamiento de la densidad marginal de los salarios

Un segundo paso de este enfoque es estimar la función de densidad marginal de los salarios, cuya dificultad es que sea consistente con la función de densidad condicional expresada en (15).

La idea básica de la estimación de la función de densidad de los salarios es: si U es una variable aleatoria uniforme en el intervalo 0 a 1 entonces $F^{-1}(U)$ tiene una función de distribución F . Entonces, si $\theta_1, \theta_2, \dots, \theta_m$ que provienen de una distribución uniforme (0,1), las correspondientes m estimaciones de la distribución condicional del salario $\{X_i\beta_{\theta}(\theta)\}_{i=1}^m$ constituyen una muestra aleatoria de las distribuciones condicionales estimadas de salarios dado los X_i .

¹³Un supuesto importante en el planteamiento de esta modelación es la linealidad del modelo de regresión cuantílica. Esto se mantendrá exactamente en el modelo condicional en el que conjuntamente la localización y su escala dependen linealmente de los covariantes.

Por otra parte, para los mismos autores, (Koenker & Bassett Jr, 1978) dadas ciertas condiciones de regularidad, la estimación de la función cuantílica condicional es un estimador fuertemente consistente de la función cuantílica poblacional, uniformemente en θ sobre un intervalo $(0,1)$.

M-M proponen que para “integrar X ” y conseguir una muestra de la función de densidad marginal, en vez de mantener fijo X , se deriva una muestra aleatoria de los covariantes de una distribución apropiada. Para entender este procedimiento y ejemplificarlo se deben seguir los siguientes pasos: primeramente se consideran a los salarios como discretos y solo una covariante (género). Luego, dado Y como el resultado del siguiente experimento aleatorio: primero se deriva un valor de X de la función de probabilidad $g(x)$ y se denota a esto como un x_0 ; y en segundo lugar se calcula un salario w de la función de distribución condicional correspondiente al valor de X previamente derivado, $f_{(w|X)}(w|X_0)$. Entonces, la probabilidad de la función de Y es $P(Y=y)=f_{(w|X)}(w|0)g(0) + f_{(w|X)}(w|1)g(1)$ lo cual obviamente es igual a $f_w(y)$ la probabilidad marginal del salario dado y .

Estimación de las funciones de densidad marginal y condicional: perspectiva empírica

Dado un $w(t)$ y $x(t)$, con $t=0,1$ que denota los salarios y las covariantes en el tiempo t (con $t=0$ en un tiempo determinado inicial y $t=1$ un tiempo definido final). Definiendo a $g(x,t)$ la densidad conjunta de las covariantes en el tiempo t . Se desea generar una muestra aleatoria de la densidad de los salarios que prevalecería en t si el modelo fuera verdadero y las covariantes fueran distribuidas como $g(x,t)$. En tal sentido se pueden describir el método en los siguientes pasos:

1. Generar una muestra aleatoria de tamaño m con distribución uniforme $U [0,1]$: $\theta_1, \dots, \theta_m$;

2. Para el conjunto de datos en el tiempo t (denotados por $x(t)$ correspondiente a la matriz de covariantes) y cada $\{u_{ij}\}$ estimar:

$$Q_{\theta_i}(w|X_i; t)$$

3. Generar una muestra aleatoria de tamaño m con reemplazo de las filas de $x(t)$ denotada por $X_i^*(t)$, usando el método de bootstrap.

4. Finalmente

$$\{w_i^*(t) = x_i^*(t) \hat{\beta}^*(\theta_i)\}_{i=1}^m$$

Que es una muestra aleatoria de tamaño m de la distribución deseada

Densidades contrafactuales

Para la aplicación de método M-M, existen básicamente dos ejercicios contrafactuales (de simulación) requeridos para definir la descomposición de los salarios y sus determinantes observables y no observables. En primera instancia se requiere estimar la función de densidad de los salarios en un momento del tiempo ($t=1$), considerando la distribución de las características observables en un tiempo atrás ($t=0$). Así mismo, en segundo lugar se requiere la función de densidad de los salarios en $t=1$, si y solo si uno de los covariantes fuera distribuido como en el tiempo $t=0$.

Para generar una muestra aleatoria de la función de distribución de salarios que podría haber prevalecido en $t=1$ si todos los covariantes han sido distribuidos como en $t=0$, se debe seguir el algoritmo antes mencionado en los pasos del 1 al 4 pero configurando una muestra aleatoria a través del método bootstrap en el tercer paso para todas las características observables en $t=0$ o $X(0)$.

Por otra parte, denotando a $y(t)$ (por ejemplo género), que describe una particular covariante de interés en el tiempo t . A continuación se debe simular la distribución que podría haber prevalecido en $t=1$, si y ha sido distribuida como en $t=0$ y los demás covariantes como en $t=1$. En este sentido es conveniente la partición del espacio $y(t)$

en J clases, $C_1(t), \dots, C_J(t)$ ¹⁴. Con estos antecedentes, el procedimiento para generar la función de densidad de los salarios en $t=1$ si y solo si uno de los covariantes fuera distribuido como en el tiempo $t=0$ pero exclusivamente para la variable género, se realiza los siguientes pasos:

1. Seguir los pasos de 1 al 4 descrito en párrafos anteriores para generar $\{w_i^*(1) = x_i^*(1) \hat{\beta}^1(\theta_i)\}_{i=1}^m$
Que es un muestra aleatoria de tamaño m de la densidad de los salarios en $t=1$.
2. Tomar una clase, llámese $C_1(t)$ (considerar género)
 - a) Dado $I_1 = \{i = 1, \dots, m | y_i(1) \in C_1\}$, seleccionar el subconjunto aleatorio de la muestra generada en paso 1 correspondiente a I_1 o $\{w_i^*(1)\}_{i \in I_1}$ (esto significa seleccionar la sub-muestra de hombres en $t=1$ de la función e densidad marginal estimada).
 - b) Generar una muestra aleatoria de igual tamaño al número de hombres en $t=0$ de la sub-muestra aleatoria del año $t=1$
3. Repetir el paso 2 para $j=1, \dots, J$ (repetir el paso 2 para las mujeres)

El procedimiento descrito pretende reproducir la distribución de los salarios que podría haber prevalecido en $t=1$ si el género se hubiera distribuido como en $t=0$. Esto es:

$$\int f_1(w|y) dF_0(y)$$

Con $f_1(w|y)$ igual a la distribución de los salarios de $t=1$ para el género o covariante y y $F_0(y)$ de la distribución de la variable género en $t=0$.

iii) *Descomposición de los cambios en la función de densidad de los salarios*

Denotando por $f(w(t))$ como un estimador de la función marginal de w (el logaritmo de los salarios) en el tiempo t basados en la muestra observada de $\{w_i(1)\}$ y por

¹⁴ Para el género o educación, estas particiones son obvias: solo pues corresponde a los dos géneros o los niveles educacionales. Para variables de continuas aleatorias como la experiencia o la edad, las clases podrían provenir de las pariciones generadas de un histogramas de cada una de los histogramas de $y(t)$

$f^*(w(t))$ un estimador de la densidad de w en t basado sobre la muestra $\{w_i^*(1)\}$, esto es lo que se ha estado denominando función marginal. En este sentido, la densidad contrafactual será denotada por $f^*(w(1); X(0))$ para la densidad que podría resultar en $t=1$ si todos los covariantes tuvieran distribución $t=0$; y , $f^*(w(1); y(0))$ para la densidad de los salarios en $t=1$ si y solo si solo el factor y estuviera distribuido como en $t=0$. Por lo que se busca es analizar los cambios entre $f(w(1))$ y $f(w(0))$ por la comparación:

- $f^*(w(1); y(0))$ con $f^*(w(0))$: la contribución de los coeficientes para el cambio total; y ,
- $f^*(w(1))$ con $f(w(1); y(0))$: la contribución de las covariantes a los cambios en la función de densidad.

Complementariamente, dado $\alpha(\cdot)$ un cuantil de la distribución de salarios, entonces la descomposición de los cambios en α es:

$$\begin{aligned} \alpha(f(w(1))) - \alpha(f(w(0))) &= \alpha(f^*(w(1); X(0))) - \alpha(f^*(w(0))) + \alpha(f^*(w(1))) \\ &\quad - \alpha(f^*(w(1); X(0))) + \text{residuo} \end{aligned}$$

En esta línea, la medida de la contribución individual de los covariantes viene definida por $\alpha(f^*(w(1))) - \alpha(f^*(w(1); X(0)))$.

Aplicación del método M-M en la estimación de los diferenciales salariales por género y etnia

Para (Albrecht, Van Vuuren, & Vroman, 2009), el método M-M es una generalización de la descomposición O-B. En este sentido, resumen el método M-M de la siguiente manera:

- considere dos grupos A y B , con características dadas por un vector estocástico X_A para el grupo A y X_B para el grupo B ; a las realizaciones de estos vectores estocásticos se las denota como x_A y x_B . Adicionalmente, se

asume que X_A y X_B tiene dimensión k y tiene funciones de distribución G_{X_A} y G_{X_B} , respectivamente. La variable endógena aplicada al caso de la descomposición de salarial se puede denotar como w_i^A y w_i^B para el grupo A y B respectivamente con una distribución incondicional F_{X_A} y F_{X_B} , correspondientemente. Además se considera a N_A y N_B los tamaños de las dos muestras de los dos grupos A y B, y finalmente se supone que los salarios y sus características son observadas para los dos grupos;

- (ii) el supuesto de M-M es que los coeficientes de la regresión cuantílica son $\beta^A(\theta)$ para el grupo A y $\beta^B(\theta)$ para el grupo B para $\theta \in [0,1]$; esto es
- $$Q_\theta(w_A | X_A = x_A) = x_A \beta^A(\theta) \quad \theta \in [0,1]$$

Y

$$Q_\theta(w_B | X_B = x_B) = x_B \beta^B(\theta) \quad \theta \in [0,1]$$

La distribución de w_A condicional a $X_A = x_A$ esta completamente caracterizada por la colección de la regresión cuantílica $\{\beta^A(\theta); \theta \in [0,1]\}$, un caso similar sucede con la distribución de w_B condicionada a $X_B = x_B$; adicionalmente se considera la variable aleatoria contrafactual w_{AB} con la propiedad de que sus cuantiles sobre x_A están dados por:

$$Q_\theta(w_{AB} | X_A = x_A) = x_A \beta^B(\theta) \quad \theta \in [0,1]$$

- (iii) El método M-M genera una muestra de la distribución incondicional de w_{AB} de la siguiente manera:
1. Generar una muestra aleatoria de tamaño m con distribución uniforme $U[0,1]$: $\theta_1, \dots, \theta_m$;
 2. Computar $\hat{\beta}^B(\theta)$, a través de la estimación de la θ i-ésima regresión cuantílica de w_B sobre x_B ;
 3. Muestrear x_A de la distribución \tilde{G}_{X_A} ;

4. Computar $\hat{w}_{AB} = x_A \beta^B(\theta)$;
5. Repetir los pasos 1 a 4 m veces.

Para (Galvis, 2010), el método M-M emplea la regresión cuantílica para ampliar el análisis de la descomposición de la brecha salarial seguido por O-B. En esta alternativa, en cada uno de los cuantiles de la distribución de los salarios se podrá evaluar cuál es la brecha salarial y cuál es el aporte de los efectos de los covariantes y coeficientes. En este sentido, dicha aplicación no es tan simple, ya que la descomposición de la brecha salarial tomando sub-muestras para distintos puntos de la distribución y aplicando luego la descomposición de O-B, lo que podría generar resultados sesgados¹⁵. Por lo que M-M tiene en cuenta esta limitación y parten de estimar $Q_\theta(w|X_i)$ donde w es el logaritmo de los salarios para el individuo i , para el cual X denota el conjunto de atributos que determinan dichos salarios. El parámetro θ es la fracción de individuos que yacen por debajo del nivel salarios del individuo i . Adicionalmente, la regresión por cuantiles para cada uno de los sexos vendría dada por las expresiones $Q_\theta(w_h|X_h) = Q_\theta(X'_h\beta(\theta)_h) + \mu_h$ para el caso de los hombres y $Q_\theta(w_m|X_m) = Q_\theta(X'_m\beta(\theta)_m) + \mu_m$ para el caso de las mujeres. En tal sentido la descomposición de la brecha salarial es:

$$\begin{aligned} & [Q_\theta(w_h|X_h) - Q_\theta(w_m|X_m)] \\ & = [Q_\theta(X'_h\beta(\theta)_h) - Q_\theta(X'_m\beta(\theta)_h)] + [Q_\theta(X'_m\beta(\theta)_h) - Q_\theta(X'_m\beta(\theta)_m)] \\ & + v \end{aligned}$$

En esta ecuación el primer término de la derecha es el efecto dotación. Este denota la contribución de las diferencias en la distribución de dotaciones existente entre los hombres (h) y las mujeres (m) evaluadas en el cuantil de la distribución. El segundo término calcula el valor simulado de la brecha salarial existente si se tiene las dotaciones de las mujeres, pero se remuneran con los retornos promedios

¹⁵ Esto porque la descomposición (Blinder, 1973)- (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) se fundamenta en la propiedad de la estimación MCO de que la media de los salarios, condicional al promedio de los atributos de los individuos es igual a la media no condicional de los salarios, lo que no se cumple en la regresión cuantílica.

observados de los hombres, con lo cual este término se refiere al denominado efecto discriminación.

Finalmente, la implementación del método M-M se lleva a cabo simulando la distribución contrafactual $Q_{\theta}(X'_m\beta(\theta)_h)$, a partir de:

1. Simular m realizaciones de los cuantiles q que se obtienen de una distribución uniforme en el intervalo (0,1).
2. Estimar m coeficientes de la regresión por cuantiles q simulados previamente empleando los datos de la muestra de hombres. Esto nos daría las tasas de retorno de los atributos o dotaciones en diferentes puntos de la distribución condicional de los salarios de los hombres.
3. Muestrear aleatoriamente y con reemplazo un conjunto de atributos de la matriz X pertenecientes a una muestra de m mujeres.
4. Construir una realización de la distribución contrafactual obtenida de multiplicar los atributos seleccionados en (3) por los retornos a dichos atributos estimados en (2). Esta simulación corresponde, en consecuencia, a la distribución de salarios que se observaría si los individuos tuviesen los atributos X de las mujeres, X_m , pero fuesen remunerados con los retornos obtenidos por los hombres, β_h .
5. Generar una distribución de las realizaciones obtenidas en (4) repitiendo los anteriores pasos un número k de veces para obtener intervalos de confianza para los efectos dotación y remuneración.

La distribución acumulada empírica para los otros componentes de la ecuación anterior pueden hallarse estimando la regresión por cuantiles para los hombres y mujeres por separado, siguiendo los pasos del 1 al 5 utilizando atributos y sus respectivos retornos para cada uno de los géneros, para obtener la distribución de las mujeres $Q_{\theta}(X'_m\beta(\theta)_m)$ y la distribución de los hombres $Q_{\theta}(X'_h\beta(\theta)_h)$.

Finalmente, para calcular los errores estándar de las regresiones cuantílicas y de la descomposición M-M se utiliza el procedimiento bootstrap que consiste básicamente, según (Guan, 2003), en una técnica no paramétrica que permite a través de la muestra de datos original obtener estimaciones empíricas en este caso de los errores estándar de los parámetros calculados, los pasos son los siguientes: i) extraer muestras aleatorias con reemplazo en varias ocasiones de la muestra de datos original; ii) estimar el estadístico deseado de las muestras generadas del paso anterior; y, iii) calcular la desviación estándar de la muestra de la distribución muestral.

2.2.4 Cuarto método de descomposición de salarios de Juhn, Murphy, y Pierce (JMP) comparación en dos momentos del tiempo

Actualmente, se han desarrollado investigaciones más complejas que han permitido desarrollar análisis de los cambios sobre el tiempo en las diferencias salariales por género y etnia, lo que facilita una comprensión más dinámica de las brechas mencionadas y sus principales factores determinantes.

Según (Atonji & Blank, 1999), dos de las principales metodologías de descomposición de salarios entre grupos y en el tiempo son el enfoque estándar y la técnica de (Juhn, Murphy, & Pierce, 1991)¹⁶. El enfoque estándar utiliza la diferencia de salarios o renta en dos momentos del tiempo entre dos grupos, diferencias que están expresadas por componentes observables y no observables. Por otro lado, en el enfoque JMP se descomponen los cambios en los salarios, poniendo énfasis en el rol de los cambios relativos a la distribución de cada grupo y se obtienen cambios en los precios de las características observables y no observables.

Enfoque estándar

La diferencia en los salarios promedio del año t pueden ser escritos como:

¹⁶ Cfr. (Atonji & Blank, 1999). Pp.3225-3227.

$$W_{1t} - W_{2t} = (X_{1t} - X_{2t})\beta_{1t} + (\beta_{1t} - \beta_{2t})X_{2t} \quad (16)$$

Donde W_{gt} representa la media de salarios del grupo g en el tiempo t (se asume que el grupo minoritario es el 2 y el mayor grupo el 1), X_{gt} son las medias de las características del grupo g las cuales afectan a los salarios, y los β_s son sus coeficientes relacionados, estimados en el tiempo t . De esta ecuación subyace una gran cantidad de trabajo empírico que trata de descomponer los salarios o las diferencias de ingresos entre los grupos en "componentes explicados" y "componentes inexplicables".

El primer término de esta descomposición representa el componente "explicado", que da diferencias en aspectos como la educación, la experiencia de los trabajadores de los grupos 1 y 2. Esta es la brecha estimada entre los grupos 1 y 2 usando al grupo mayoritario como grupo base para la comparación.

El segundo término son los componentes no explicados, y representa las diferencias en los coeficientes estimados, es decir, las diferencias en los retornos de similares características de los grupos 1 y 2.

La proporción de la brecha salarial total explicada por el segundo componente se explica por el efecto de la discriminación.

Para analizar el diferencial salarial a través del tiempo entre dos grupos, la ecuación (16) se calcula para los diferentes períodos a investigar. Dado el operador Δ representan la diferencia media entre el grupo 1 y grupo 2 en un año dado. El cambio en los diferencias salariales entre períodos de tiempo t' y t se pueden presentar como:

$$\Delta W_{t'} - \Delta W_t = (\Delta X_{t'} - \Delta X_t)\beta_{1t} + \Delta X_{t'}(\beta_{1t'} - \beta_{1t}) + (\Delta\beta_{t'} - \Delta\beta_t)X_{2t} + (X_{2t'} - X_{2t})\Delta\beta_{t'} \quad (17)$$

En la ecuación (17) el primer término representa el efecto de los cambios relativos en el tiempo de las características observadas de los dos grupos y el segundo término significa el efecto de los cambios en el tiempo de los coeficientes para el grupo 1 manteniendo las diferencias en las características observadas fijas. Estos dos componentes representan el cambio en el tiempo en la diferencia de salarios que se esperaría, dados los cambios en las características de los dos grupos y los coeficientes de dichas características en el grupo 1 en los períodos t y t' .

Los términos tercero y cuarto capturan el cambio en el componente no explicado de la brecha, $(\beta_{1t} - \beta_{2t})X_{2t}$ en la ecuación (16). El tercer término es el efecto de cambios en el tiempo en los coeficientes relativos entre los dos grupos. El cuarto término captura los cambios en el tiempo de las características del grupo 2, modifican los efectos de las diferencias en el grupo de coeficientes $(\beta_{1t} - \beta_{2t})$. Generalmente se calcula cada uno de estos términos, así como los subcomponentes correspondientes a los elementos individuales de X y β .

Una desventaja de esta descomposición es que no aporta mucho conocimiento de cómo la brecha salarial se ve afectada por los cambios en la distribución global de los salarios. Los aumentos en la dispersión de los salarios aumentará la brecha entre los salarios medios del grupo 1 y grupo 2 (si el grupo 2 está por debajo de la media y el grupo 1 está por encima de la media), incluso si estos cambios no tienen efecto sobre la ubicación de las distribuciones de los dos grupos. En el método JMP se proporciona métodos para aislar el efecto de un cambio en la dispersión de los componentes no observables salariales, afectando a ambos grupos por un cambio en la localización de la distribución de habilidad del grupo 2 respecto al grupo 1.

Enfoque de JMP para la descomposición salarial

Esta metodología permite descomponer cambios salariales, y particularmente enfatiza el rol de los cambios en la distribución relativa de cada grupo. Reescribiendo la ecuación (16) como:

$$W_{1t} - W_{2t} = (X_{1t} - X_{2t})\beta_{1t} - U_{2t} \quad (18)$$

Donde el componente inexplicable $-U_{2t}$ es $-U_{2t} \equiv (\beta_{1t} - \beta_{2t})X_{2t}$. Se puede decir que el salario del individuo i del grupo 1 en el tiempo t es $W_{1it} = \beta_{1t}X_{1it} + \mu_{1it}$, mientras que el salario del individuo j en el grupo 2 en el tiempo t es $W_{2jt} = \beta_{2t}X_{2jt} + \mu_{2jt}$, donde β_{1t} y β_{2t} ; y, se cumple que $E(u_{1it}|X_{1it}) = 0$ y $E(u_{2jt}|X_{2jt}) = 0$. Entonces la diferencia de medias se puede escribir como la ecuación (18).

Aquí se debe tomar en cuenta que μ_{1it} y μ_{2jt} además de ser de los componentes de error de las regresiones de salarios de los individuos i, j de los grupos 1 y 2 en el tiempo t , son los componentes de salario de un miembro de la población 1 y 2 que no son explicados por los grupos 1 y 2 de las regresiones. Por ejemplo: $U_{2jt} = \mu_{2jt} + (\beta_{2t} - \beta_{1t})X_{2jt}$ es el componente del salario de una persona en el grupo 2 que no es explicado por la función de regresión del grupo 1. Se puede escribir μ_{1it} como $\mu_{1it} = \sigma_t \theta_{1it}$, donde θ_{1it} es el error estándar con media 0 y varianza 1 y σ_t es la desviación estándar de μ_{1it} . Se puede también representar $\mu_{2jt} + (\beta_{2t} - \beta_{1t})X_{2jt}$ como $\sigma_t \theta_{2jt}$ donde $\theta_{2jt} \equiv U_{2jt}/\sigma_t$, es normalizada para tener una varianza de 1 (se asume que $\sigma_{1t} = \sigma_{2t} = \sigma_t$ en todos los años), se puede entonces reescribir la ecuación (17) como:

$$W_{1t} - W_{2t} = (X_{1t} - X_{2t})\beta_{1t} + \sigma_t(\theta_{1t} - \theta_{2t}) = (X_{1t} - X_{2t})\beta_{1t} + \sigma_t(-\theta_{2t}) \quad (19)$$

Donde σ_t se interpreta como el precio del componente de los salarios que no es explicado por el grupo 1 de la función de regresión y se observa que θ_{1t} y θ_{2t} son las medias para los grupos 1 y 2 de este componente. El segundo término es la brecha residual.

Los cambios entre grupos en el tiempo se expresa de la siguiente manera:

$$\Delta W_{t'} - \Delta W_t = (\Delta X_{t'} - X_t)\beta_{1t} + \Delta X_{t'}(\beta_{1t'} - \beta_{1t}) + \sigma_t(\Delta\theta_{t'} - \Delta\theta_t) + \Delta\theta_{t'}(\sigma_{t'} - \sigma_t) \quad (20)$$

Los términos tercero y cuarto son una alternativa a la descomposición de los efectos de los cambios en las características no observables sobre el cambio en la diferencia de salarios proporcionado en la ecuación (20). El tercer término representa cambios en la posición relativa del grupo 1 y el grupo 2 dentro de una distribución constante de los componentes no observables de salarios. El cuarto término representa los cambios en la distribución de los salarios.

CAPÍTULO III

RESULTADOS

3.1. DATOS Y ESPECIFICACIÓN DE VARIABLES DE CONTROL

Los datos que se utilizarán en el presente estudio se basan en la Encuesta Nacional de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDUR) elaborada por el Instituto Nacional de Estadística y Censos (INEC) de los años 2007 y 2013. Debido a que estas encuestas incluyen diferentes variables que facilitan el cálculo del ingreso laboral y permite diferenciar las características de los individuos que conforman la oferta laboral: género, etnia, jefe de hogar, estado civil, años de educación, experiencia laboral, migración, región, entre otras.

Particularmente, se considera de la encuesta a la población económicamente activa entre 15 a 65 años de edad. La razón fundamental de dicha elección es que el INEC considera a este rango de edad como la población en edad a trabajar, por lo que tomó en cuenta únicamente a dicho subconjunto poblacional. Las características generales de dicha población se analizan en los siguientes párrafos.

La Encuesta de Empleo, Desempleo y Subempleo Urbano y Rural (ENEMDUR) 2007 y 2013

Los principales objetivos de la ENEMDUR son: i) visualizar el perfil social, demográfico y económico de la población del país, a través de variables de carácter general como: sexo, edad, nivel de instrucción, condición de actividad, entre otras; ii) proporcionar información actualizada y periódica sobre la Población Económicamente Activa - PEA, con sus principales características; y, difundir datos e indicadores fidedignos del mercado laboral, relacionados con el empleo, desempleo y subempleo.

Según el (INEC, 2014) las ENEMDUR de diciembre 2007 y diciembre 2013, que se utilizan en el presente estudio, tienen cobertura nacional, se realizan en el área

urbana y en el área rural (sólo en diciembre) de las provincias de la Sierra y de la Costa incluida la Amazonía. Del ámbito de estudio se excluye la Región Insular. El INEC investigó 21768 viviendas, de 579 centros poblados urbanos y rurales, con la siguiente desagregación geográficas: 5 ciudades (Quito, Guayaquil, Cuenca, Machala y Ambato); resto sierra urbano, resto de costa urbano, amazonía urbano, sierra rural, costa rural, amazonía rural, provincial, urbano, rural y nacional.

Para la selección probabilística de las viviendas de la ENEMDU, para el período 2007 – 2013, el INEC debió construir en el año 2003 y con antecedentes del Censo de Población y Vivienda de 2001, un Marco de Muestreo¹⁷, que permita ubicar e identificar un conjunto de áreas geográficas divididas en conglomerados de viviendas llamadas “sectores” (32129) en que se dividió el territorio nacional, de modo que todas las viviendas particulares del universo que pertenecen a algún sector tienen probabilidad de ser incluidas en la muestra. En el año 2003 para la selección de la muestra se utilizó un método de selección en tres etapas; en la primera se sorteó una muestra de Unidades Primarias de Muestreo (UPM)¹⁸, en la segunda se seleccionó un tamaño muestral de sectores considerando un sector dentro de cada UPM y en la tercera, dentro de cada sector ya elegido, se sorteó una muestra de 12 viviendas. Una vivienda una vez seleccionada permanece en la muestra durante 2 años y las personas que pertenecen a la vivienda son entrevistadas 4 veces en ese período aplicándose un Cuestionario ajustado a estándares internacionales. Para evitar el cansancio de los informantes, cada período se cambia algunas viviendas de la muestra, proceso que se denomina “rotación de viviendas en la muestra”. Los resultados de la encuesta son sometidos a un algoritmo matemático, el cual entrega los resultados de los parámetros de interés, como por ejemplo, el total de desocupados y la fuerza de trabajo donde la tasa de desocupación es el cociente entre el primero y el segundo.

¹⁷ Es una lista completa, organizada en forma de base de datos que contiene a todos y cada uno de los elementos de la población de interés que participaran en cada una de las fases de selección de la muestra. El Marco también está formado por todos los mapas y planos a diferentes escalas que nos permiten identificar en forma precisa y clara los límites físicos que tienen las diferentes unidades de selección.

¹⁸ Unidad Primaria de Muestreo. Las Unidades Primarias de Muestreo están constituidas por una o parte de un Área Geo-estadística Básica (AGEB) o de varias AGEB colindantes que tienen que cumplir ciertas características básicas determinadas por el diseño muestral para su formación.

A continuación se describen las principales variables investigadas por la ENEMDUR: variables clasificadas: Población Total; Población en Edad de Trabajar, PET; Población Económicamente Activa, PEA; Población Ocupada, PO; Población Desocupada, PD; Población Económicamente Inactiva, PEI; Población Subempleada. Variables clasificatorias: sexo, edad, nivel de instrucción, condición de actividad, sector económico, rama de actividad, grupo de ocupación, categoría de ocupación, ingresos del trabajo, ingresos no del trabajo, categoría de inactividad.

Utilización de factores de expansión para la presentación de resultados

Dado que los datos provienen de una muestra compleja, se utilizó los factores de expansión para el cálculo y la presentación de los estadísticos descriptivos, mientras que, para las estimaciones de los resultados de los métodos de descomposición no se utilizaron factores de expansión debido a las siguientes consideraciones:

- a) La incorporación de los factores de expansión hace que no se verifique la hipótesis de independencia del término de error, por lo que puede hacer que los estimadores y sus varianzas sean sesgadas, lo cual tiene consecuencias para el cálculo de los intervalos de confianza y en la obtención de estadísticos de prueba para la contrastación de hipótesis (CEPAL,2012).
- b) La evidencia indica que cuando no se aplica el factor de expansión se tiende a subestimar los errores estándar de los estimadores, pero esta subestimación es mínima comparada con los resultados ponderados (CEPAL,2012). Por lo que el factor de expansión se podría omitirlo para el cálculo de los modelos de descomposición de salarios.
- c) Por otra parte, en los métodos implementados O-B, Neumark, M-M y JMP para la descomposición, no se poseen algoritmos especializados para su cálculo aplicando factores de expansión, se puede usar Mínimos Cuadrados Ponderados, pero tampoco se evaluó sea una buena alternativa, pues persisten los problemas mencionados en el punto a).

Variables utilizadas en el estudio

La variable dependiente y las variables independientes utilizadas en esta investigación se describen a continuación:

Variable dependiente:

- **Logaritmo de ingresos corrientes por hora:** es el logaritmo neperiano de la división de los ingresos mensuales laborales corrientes (de las actividades principal y secundarias)¹⁹ y las horas mensuales efectivamente trabajadas por los empleados.²⁰

Variables independientes:

De acuerdo a la literatura empírica se definen las siguientes variables de control para analizar el problema de discriminación salarial, estas variables se utilizan para el cálculo de los 4 métodos de descomposición de salarios:

- **Años de educación:** para el cálculo de esta variable se tomaron los siguientes criterios: i) primeramente se consideró los años aprobados declarados por nivel de instrucción; (ii) posteriormente se relacionó el nivel instrucción con el número de años aprobados; en tal sentido, una persona que declara tener un nivel de instrucción de primaria, el valor que le corresponde a los años completos de educación es igual a los años declarados como aprobado. Así mismo, una persona que declara tener un nivel de instrucción superior universitaria, sus años completos de educación son igual a: años declarados aprobados en su nivel de instrucción más un aproximado de 12 años que es un tiempo estimado de educación primaria y secundaria.

¹⁹ De acuerdo con el INEC los ingresos laborales corrientes, en el caso de los individuos con trabajo dependiente o asalariado, es igual a la suma de los ingresos laborales monetarios mensuales incluidos los descuentos laborales y el salario en especie o no monetario; en tanto, que para el caso de los individuos auto empleados, dichos ingresos son la suma de las remuneraciones monetarias provenientes del trabajo independiente por cuenta propia o como patrono neta de gastos más los ingresos no monetarios o para el auto-consumo y auto-suministro.

²⁰ Pero para el cálculo de esta última variable (horas trabajadas mensuales) se multiplicó las horas semanales efectivas de trabajo declaradas por 4,3 semanas del mes.

- **Experiencia:** Para el cálculo de esta variable se consideró lo definido por (Atonji & Blank, 1999), los cuales establecen como experiencia potencial al resultado de la diferencia de la edad menos los años de educación y menos 6 años de infancia en los que no se trabaja. Esta variable queda descrita como:

$$X_i = A_i - E_i - 6$$

Donde $i=1, \dots, n$; X_i es la experiencia potencial, A_i es la edad en años del i -ésimo individuo, E_i es el número de años de escolaridad completos del i -ésimo individuo.

Un aspecto importante a tomar en cuenta es que cuando la experiencia del trabajo es adquirida sin interrupción después de completada la educación formal, la experiencia potencial y actual coinciden de tal manera que esta variable que aproxima a la medición de la experiencia del trabajador, esta variable es una buena proxy de la trayectoria laboral del individuo.

- **Casado(a) / unido(a):** es una variable dicotómica igual a 1 en el caso en el que el individuo haya declarado ser casado o en unión libre y 0 si declaró ser soltero, divorciado, separado o viudo.
- **Jefe(a) de hogar:** es una variable dicotómica igual a 1 en el caso en el que el individuo sea jefe(a) de hogar y 0 en caso contrario.
- **Migración:** es una variable binaria que toma el valor de 1 en el caso de que un individuo haya declarado que ha migrado de su provincia o región de origen a otra distinta y que cuando migró tenía más de 18 años y 0 en caso contrario.
- **Área y región de residencia:** (i) para controlar el área geográfica de residencia se creó dos variables dicotómicas una para el área urbano y otra para el área rural; (ii) para controlar la región de residencia se crearon seis variables

dicotómicas: Sierra norte, Sierra centro, Sierra sur, Costa norte, Costa litoral y Amazonía, las cuales fueron agrupadas de esa manera para tener un balance en el análisis de las incidencias de la región de residencia en el nivel de ingreso laboral y las diferencias salariales; y, (iii) se creó dos variables binarias Quito y Guayaquil para controlar posible efecto que tiene los principales centros de concentración poblacional.

- **Género:** Es una variable dicotómica de género que toma el valor de 1 si el individuo es mujer y 0 si es hombre.
- **Etnia:** Hace referencia a una variable indicatriz que es igual a 1 cuando el individuo se auto identifica como indígena y 0 en caso contrario.

Para los métodos de descomposición de salarios O-B y Neumark se estimaron primeramente, por MCO de las ecuaciones de ingresos para los trabajadores ocupados; sin embargo dichas estimaciones pueden presentar un problema de sesgo de selección muestral, que se corrige a través de (Heckman, 1979), para la cual primeramente se estima un modelo probit de participación laboral para obtener la ratio de Mills que posteriormente será introducida como un nuevo regresor en la ecuación de ingresos. Para la estimación del modelo de participación laboral se consideró las siguientes variables: edad, años de escolaridad, migración, estado civil, jefe de hogar, ingresos que no provienen del trabajo (ingresos derivados del capital o inversiones, transferencias y otras prestaciones recibidas, bono de desarrollo humano; y, bono por discapacidad), área de residencia y el hecho de ser mujer o auto identificarse como indígena.

Para (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973), algunas de las variables de control como grupo de ocupación, o industria están relacionadas con la discriminación salarial, particularmente en contra de las mujeres o indígenas, por lo que la inclusión de estas variables en las ecuaciones de salarios podrían subestimar la parte de la diferencia salarial atribuida a la discriminación. En

este sentido, controlado en términos generales la definición de ocupación, se elimina de los efectos de las barreras ocupacionales como fuente de discriminación. Como resultado, probablemente se subestima los efectos de la discriminación. Por lo tanto, se sugiere estimar un conjunto de ecuaciones sin controlar por la ocupación, rama de actividad y grupo de ocupación. A este grupo de estimaciones se denominan regresiones de características personales; en tanto que, al otro grupo de estimaciones se les denomina ecuaciones de salarios a máxima escala. Para los modelos econométricos de esta investigación se utilizarán variables de características personales, para no subestimar los efectos de la discriminación salarial al utilizar variables como categoría de ocupación y rama de actividad económica.

3.2. ANÁLISIS DESCRIPTIVO

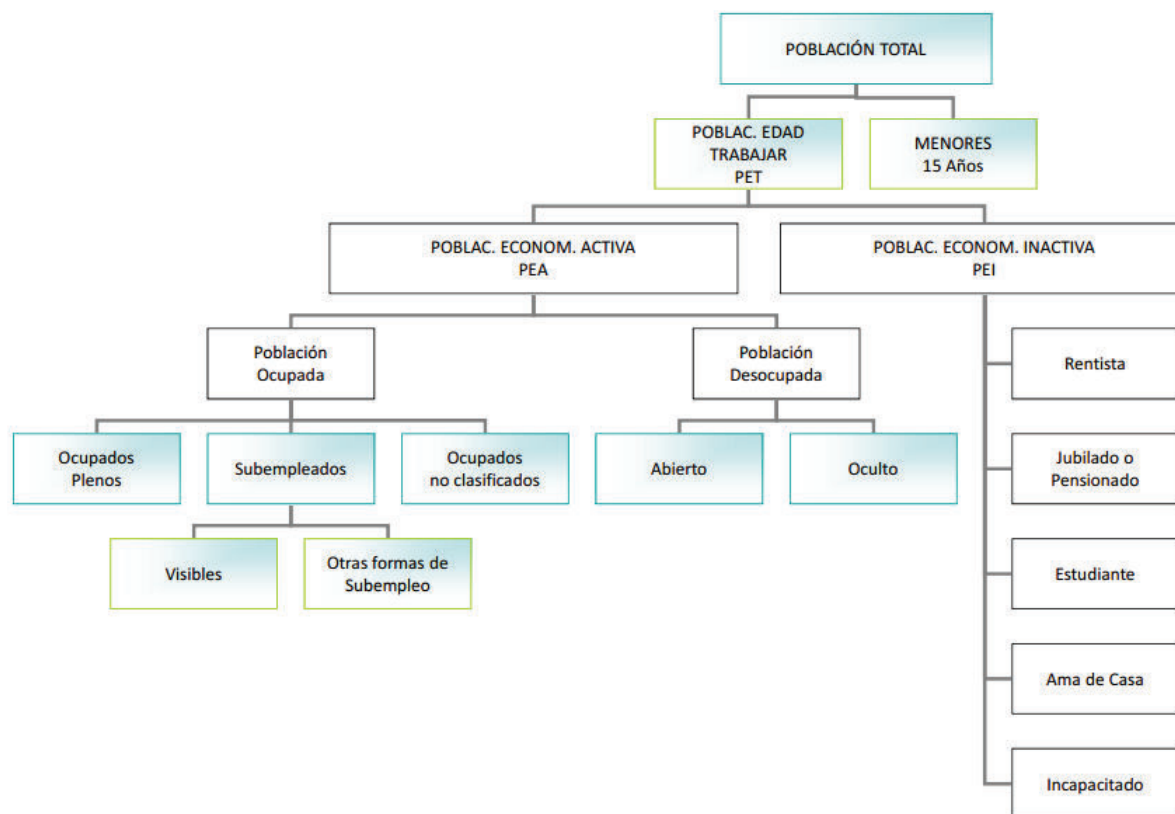
En esta sección, y con el objetivo de describir de una forma sucinta a la población de los diferentes grupos laborales en estudio se realizó un resumen de las características observables de dichas poblaciones.

En la tabla No. 1, en la primera sección, se resume para el año 2013, las características generales de toda la población del país. Los resultados demuestran que la población total del Ecuador fue de 15,9 millones de habitantes, de los cuales el 50,3% correspondió a mujeres y el 49,7% a hombres. Así mismo, la población que se consideró “No indígena” alcanzó un 93,7%, en tanto que el grupo que se auto identifica como “Indígena” fue de 6,3%.

Por otra parte, en la segunda sección de la tabla mencionada, realiza un análisis más particular de la población que se toma en consideración para el análisis. Esto quiere decir que a continuación se describe más particularmente a la población en edad de trabajar o entre 15 a 65 años. Como se puede observar la mayoría de hombres y mujeres viven en las zonas urbanas, son grupos relativamente jóvenes (ya que el 40% de la población se encuentra entre 15 y 30 años) y con un nivel de instrucción

predominantemente secundario. Para entender la condición de actividad de cada uno de los grupos a continuación, en el gráfico No. 1 se muestra la clasificación de la población por actividad conforme al INEC.

Gráfico 1. Clasificación de la Población por Actividad - INEC



Fuente y elaboración: Instituto Nacional de Estadística y Censos del Ecuador (INEC).

La mayoría de mujeres se encuentran en una condición de actividad considerada como inactiva (49,0%), o subempleada (29,1%), en tanto que los hombres también son altamente subempleados (37,8%) y ocupados plenos (40,0%). Así mismo, las mujeres en un alto porcentaje realizaron actividades como “Trabajadoras no calificadas” (24,8%), y como “Trabajadoras de los servicios y comerciantes” (32,1%); los hombres también son “Trabajadores no calificados” (25,5%), y “Oficiales, operarios y artesanos” (18,2%). Adicionalmente, se apreció que los hombres presentan un mayor salario por hora respecto a las mujeres. En este sentido se

muestra que los hombres ganan en promedio aproximadamente 360 dólares mensuales las mujeres tiene un ingreso mensual promedio de 288 dólares por lo que los diferenciales salariales entre estas dos poblaciones es de 62 dólares o los hombres ganan un 11% más que el grupo femenino.

Por otra parte, el grupo de personas ocupadas auto identificadas como “No Indígenas” viven en su mayoría en las zonas urbanas (72,4%), y el grupo de indígenas en zonas rurales (77,3%). Los dos grupos son relativamente jóvenes, pero existen diferencias significativas en los niveles de instrucción pues buena parte de la población no indígena alcanza una instrucción secundaria (47,9%), en tanto que aproximadamente el 50,9% los indígenas apenas lograron una instrucción primaria. Un alto porcentaje de “No indígenas” son inactivos (35,0%), o subempleados (32,0%), en tanto que los indígenas también son altamente subempleados (56,4%) e inactivos (25,4%).

Los “No indígenas” en un alto porcentaje son “Trabajadores no calificados” (24,4%), y “Trabajadores de los servicios y comerciantes” (22,9%); los “Indígenas” también son en su gran mayoría como “Trabajadores no calificados” (36,3%), y “Trabajadores calificados agropecuarios y pesqueros” (30,0%). Pero una de las características más diferenciadores entre estos dos grupos laborales son los niveles de ingresos mensuales promedio de los mismos. Mientras el grupo de trabajadores no indígenas ganan aproximadamente en promedio un monto de 320 dólares mensuales el grupo de indígenas percibe un ingreso promedio mensual de 192 dólares lo que significa un 40% menos.

Tabla 1. Características de la población del Ecuador por Género y Etnia en el año 2013
Porcentajes

Variables	Hombre	Mujer	No indígena	Indígena
i) Población total				
Millones de habitantes	7,9	8,0	13,6	0,9
Porcentajes	49,7	50,3	93,7	6,3
ii) Población de 15 a 65 años de edad				
Porcentajes	48,9	51,1	94,2	5,8
Condición de actividad				
Ocupados no clasificados	0,2	0,1	0,1	0,0
Ocupados plenos	40,0	18,9	30,0	17,1
Subempleo Visible	7,0	5,1	6,1	5,4
Otras formas de subempleo	30,8	24,0	25,9	51,0
Desempleo Abierto	2,1	2,0	2,1	0,6
Desempleo Oculto	0,8	0,8	0,8	0,4
Inactivo	19,1	49,0	35,0	25,4
Grupos de edad				
15-20	20,2	17,3	18,4	23,7
21-30	23,5	23,1	23,4	22,1
31-40	19,8	21,5	20,7	20,3
41-50	17,6	18,6	18,3	15,6
51-60	13,6	14,3	14,0	12,6
61-65	5,4	5,3	5,3	5,6
Nivel de instrucción				
Ninguno	0,0	0,0	0,0	0,1
Primaria	31,0	31,9	30,4	50,9
Secundaria	49,0	46,2	47,9	41,8
Superior	20,0	22,0	21,7	7,2
Grupo de ocupación				
Personal directivo y administrativo del sector público o privado	1,4	1,2	1,4	0,2
Profesionales científicos e intelectuales	7,0	12,7	9,6	3,7
Técnicos y profesionales de nivel medio	4,3	5,0	4,9	0,6
Empleados de oficina	3,9	7,2	5,4	1,6
Trabajadores de los servicios y comerciantes	15,8	32,1	22,9	12,4
Trabajadores calificados agropecuarios y pesqueros	11,8	8,0	8,9	30,0
Oficiales operarios y artesanos	18,2	7,5	14,2	10,9
Operadores de instalaciones, maquinaria y montadores	11,5	1,6	7,8	4,0
Trabajadores no calificados	25,5	24,8	24,4	36,3
Fuerzas Armadas	0,7	0,0	0,4	0,3
Área de residencia				
Urbana	68,5	70,5	72,4	22,7
Rural	31,5	29,5	27,6	77,3
Ingreso laboral en dólares por hora				
	2,0	1,8	2,0	1,2

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

3.3. DISCRIMINACIÓN SALARIAL POR GÉNERO Y ETNIA²¹

3.3.1. Estimación de ecuación de ingreso controlada por género y etnia.

Los resultados de las estimaciones de las ecuaciones de ingresos para toda la muestra de la población ocupada por género y etnia, corregidas por el método de (Heckman, 1979) (que incluye la muestra de la población desocupada) se observan en el Anexo B.

Como ya se mencionó en la sección anterior, si la decisión de participar en el mercado laboral es endógena entonces la estimación por MCO puede ser sesgada. Por lo que se requiere tratar el problema de sesgo de selección muestral a través del estimador propuesto por (Heckman, 1979). En la tabla No. 1 se muestra la estimación y significancia de la inversa de la ratio de Mills (y su error estándar estimado) para el contraste indirecto de la hipótesis nula de “independencia entre los errores de la ecuación de participación laboral y la ecuación de ingresos”. Los cálculos obtenidos muestran que el coeficiente de la inversa de la ratio de Mills es significativo. Esto significa implícitamente, que se rechaza la hipótesis nula de independencia a favor de la existencia de no independencia, por lo que la regresión minceriana corregida por el método de corrección de sesgo de selección muestral permiten obtener estimadores más eficientes. Adicionalmente, se puede decir que existe una correlación entre los errores de las ecuaciones de participación laboral y los errores de la ecuación de ingresos. Por lo que, fue conveniente la introducción de un control adicional a los determinantes de los salarios relacionado con la inversa del ratio de Mills, obtenido en una primera etapa a través de la estimación del modelo probit de participación laboral.

²¹ Para las estimaciones se tomó como población de estudio a las personas de 15 años o más, y a las personas clasificadas como ocupadas y no ocupadas (para la estimación del ratio de Mills). En algunas estimaciones se eliminaron de la muestra puntos atípicos, palancas o influyentes. Para más detalle de los ajustes realizados en las regresiones de salarios ver Anexo B.

En los siguientes párrafos se muestran los resultados obtenidos a través de la aplicación del método de (Heckman, 1979) en la estimación de la ecuación de ingresos para toda la población.

La estimación de ingresos mostró la existencia discriminación salarial en contra de las mujeres e indígenas pues las variables ficticias que controlan tales condiciones son negativas y significativas. En particular, se observó que manteniendo el resto de factores constantes, el grupo femenino recibe 13,3% menos en sus ingresos por el hecho de pertenecer a dicha población, en tanto que los indígenas 12,2% menos (respecto de los grupos de hombres y no indígenas, respectivamente).

Por otra parte, la significancia de la variable de “años de educación” ($p\text{-value} < 0.05$) permite decir los rendimientos promedio de la población es de aproximadamente 8% por cada año adicional de educación. La experiencia presenta un coeficiente positivo y significativo, por lo que personas con más años de experiencia laboral podrían tener mejores niveles de ingresos. La experiencia al cuadrado es negativa y significativa, lo que implica que dicha variable provoca un incremento en el perfil de los ingresos de los individuos de la población, pero a partir de cierto punto, los ingresos de personas con mayor experiencia podrían decrecer. Esto muestra una relación no necesariamente lineal entre los ingresos de los individuos y su experiencia.

Además personas con un estatus civil “casado o unido” reciben en promedio un pago adicional de aproximadamente 7% por el hecho de tener dicha situación civil; así mismo ser jefe(a) de familia podría favorecer para obtener mayores ingresos (el coeficiente de esta variable es de 5% y es significativo). Por otra parte, tener una condición migratoria puede ayudar a mejorar las condiciones salariales.

El hecho de vivir en el área rural tiene una relación negativa y significativa con el salario por hora, respecto de los trabajadores del área urbana que fue la categoría de referencia. Así mismo, vivir en la sierra sur respecto de otras áreas de residencia (como: Sierra Centro, Costa y Amazonía) tiene un efecto negativo en el salario por

hora. El efecto contrario se observó en la sierra norte, en la que posiblemente dicha ubicación favorecería a un incremento del perfil salarial respecto de las categorías de referencia mencionadas. Por otra parte, vivir en la ciudad de Quito, respecto de todas otras ciudades del Ecuador, tiene un efecto positivo sobre el salario por hora de alrededor de 9%.

Tabla 2. Estadísticos descriptivos de la población ocupada de 15 a 65 años y estimación de ecuación de ingreso controlada por género y etnia 2013

i) Población total ocupada de 15 a 65 años de edad						
Variables	Obs.	Ponderadas	Media	Desv. Est.	Mín	Máx
Ingreso laboral en dólares por hora	27.675	5.640.172	2,0	2,2	,01	194,7
Logaritmo del salario por hora	27.675	5.640.172	0,67	0,80	-5,15	5,27
Indígena	31.595	6.276.767	0,07	0,25	0,00	1,00
Mujer	31.595	6.276.767	0,39	0,49	0,00	1,00
Edad	29.689	5.967.749	38,78	12,50	15,00	65,00
Años de escolaridad	30.405	6.096.942	10,04	4,45	0,00	20,00
Experiencia	29.689	5.967.749	22,80	13,86	1,00	58,50
Casado(a) /unido (a)	31.595	6.276.767	0,61	0,49	0,00	1,00
Jefe de hogar	31.595	6.276.767	0,49	0,50	0,00	1,00
Migración	31.595	6.276.767	0,17	0,38	0,00	1,00
Área urbana	31.595	6.276.767	0,68	0,47	0,00	1,00
Área rural	31.595	6.276.767	0,32	0,47	0,00	1,00
Sierra norte	31.595	6.276.767	0,23	0,42	0,00	1,00
Sierra centro	31.595	6.276.767	0,12	0,32	0,00	1,00
Sierra sur	31.595	6.276.767	0,11	0,31	0,00	1,00
Costa norte	31.595	6.276.767	0,06	0,23	0,00	1,00
Costa litoral	31.595	6.276.767	0,44	0,50	0,00	1,00
Amazonía	31.595	6.276.767	0,05	0,22	0,00	1,00
Quito	31.595	6.276.767	0,12	0,33	0,00	1,00
Guayaquil	31.595	6.276.767	0,16	0,37	0,00	1,00

ii) Estimación de ecuación de ingreso controlada por género y etnia 2013

Variables independientes	Coeficiente	Bootstrap Std. Err.	z	p> z	Intervalo de confianza al 95%	
					Inferior	Superior
Años de educación	0,081***	0,001	80,38	0,000	0,079	0,083
Experiencia	0,016***	0,001	12,82	0,000	0,014	0,019
Experiencia al cuadrado	-0,0002***	0,000	-11,70	0,000	-0,0003	-0,0002
Casado(a) / unido(a)	0,067***	0,009	7,15	0,000	0,049	0,085
Jefe(a) de hogar	0,045***	0,011	4,04	0,000	0,023	0,067
Migrante	0,089***	0,010	9,08	0,000	0,070	0,108
Quito	0,078**	0,031	2,51	0,012	0,017	0,139
Área rural	-0,113***	0,009	-13,26	0,000	-0,129	-0,096
Sierra sur	-0,032***	0,010	-3,15	0,002	-0,052	-0,012
Sierra norte	0,077***	0,011	7,13	0,000	0,056	0,098
Mujer	-0,133***	0,009	-14,87	0,000	-0,151	-0,116
Indígena	-0,122***	0,017	-7,01	0,000	-0,157	-0,088
Constante	-0,289***	0,036	-7,96	0,000	-0,360	-0,218
Ratio de Mills	-0,368***	0,128	-2,87	0,004	-0,619	-0,116
Número de observaciones	23053		Wald chi2(13)	10547		
Replicaciones para bootstrap	2000		p> chi2	0,000		
R-cuadrado	0,3359		R-cuadrado ajustado	0,3356		

Nota: *** p<0.01, ** p<0.05, * p<0.1

3.3.2. Descomposición de las diferencias de ingresos por género y etnia en el Ecuador

Luego de observar los principales resultados de la estimación de ingresos para toda la población en análisis, y su interpretación a continuación se muestran los resultados de la aplicación de los primeras dos metodologías para la descomposición salarial por grupos de trabajadores por género y etnia.

En la tabla No.3 se muestran los resultados de la descomposición de las diferencias de ingresos por género y etnia por método O-B y Neumark para los trabajadores asalariados y auto-empleados de 15 a 65 años de edad para el Ecuador. Es decir se muestra las estimaciones de los dos componentes de las diferencias salariales asumiendo: (i) la estructura salarial de los hombres $\hat{\beta}_h$ (o en el caso de la descomposición por etnia a la estructura de los no indígenas $\hat{\beta}_{NI}$) como no discriminatoria, la de las mujeres $\hat{\beta}_m$ (o la estructura de los indígenas $\hat{\beta}_I$ en el caso

de la descomposición por etnia).; y, (ii) una estructura no discriminatoria \hat{b} que resulta de la estimación conjunta de la ecuación de ingresos que implica la utilización de la metodología (Neumark, 1988).

Los resultados de las diferencias salariales por género muestran: (i) las diferencias salariales ajustadas (por las variables controladas en la ecuación de ingresos) entre hombres y mujeres son de 0,078 puntos logarítmicos (o USD 1,08 por hora)²²; (ii) en todas las estimaciones el componente atribuido a las características no observables o discriminación salarial es positivo y significativo, por lo que se puede concluir acerca de la existencia de discriminación salarial negativa en contra de las mujeres (en el caso en el que la estructura salarial no discriminatoria considerada sea la de los hombres $\hat{\beta}_h$), positiva en contra de los hombres (en el caso en el que la estructura salarial no discriminatoria sea la de las mujeres $\hat{\beta}_m$) y negativa o positiva en contra de las mujeres o hombres respectivamente; (iii) la parte de la brecha de ingresos atribuida a la discriminación es de alrededor de 0,17, puntos logarítmicos cuando se considera el método de (Neumark, 1988) (USD 1,19 por hora) que es un valor similar a lo obtenido por el método de O-B, cuando se considera como estructura salarial no discriminatoria a la de los hombres, pero dicho resultado es distinto cuando se toma en cuenta a la estructura salarial no discriminatoria a la de las mujeres pues el componente atribuido a la discriminación representa un mayor valor (0.20 puntos logarítmicos) en la diferencia salarial entre los grupos; y, (iv) el componente atribuido a las diferencias en las dotaciones (características grupales) es negativo y significativo en todas las estimaciones, esto sugiere que las características de las mujeres contribuyen a la disminución de las diferencias salariales por género.

Los resultados de las diferencias salariales por etnia muestran: (i) las diferencias de los ingresos ajustada entre la población no indígena e indígena es de

²² Los resultados mostrados en la tabla en mención son los que se obtuvieron de estimar y corregir las ecuaciones de ingreso de los grupos por el método de corrección de sesgo de selección muestral

aproximadamente 0,36 puntos logarítmicos una vez corregida por el sesgo de selección muestral; (ii) el componente atribuido al componente no observable de la diferencia salarial de este grupo de trabajadores o la discriminación en todas las estimaciones resulta ser positivo y significativo, lo que sugiere la existencia de discriminación negativa en contra de la población indígena (en el caso en el que la estructura salarial no discriminatoria considerada sea la de los no indígenas $\hat{\beta}_{NI}$), positiva en contra de la población no indígena (en el caso en el que la estructura salarial no discriminatoria sea la del grupo indígena $\hat{\beta}_I$) y negativa o positiva en contra de los indígenas y no indígenas, respectivamente; (iii) la parte de la brecha de ingresos atribuida a la discriminación es de aproximadamente 0,14 puntos logarítmicos (USD 1,15 por hora); en este caso es importante mencionar que cuando se utiliza la metodología de (Neumark, 1988), la discriminación tiene una mayor representatividad respecto de lo estimado por el método de O-B, esto sugiere que el proceso de valoración de las características observables es altamente sensible a la estructura salarial no discriminatoria y posiblemente el comportamiento discriminatorio de los empleadores sea de trascendental importancia al momento de la determinación del grado de discriminación salarial.

Tabla 3. Descomposición de salarios por los métodos O-B y Neumark por género y etnia 2013

	Estructura salarial no discriminatoria					
	Hombre y mujer	Hombre	Mujer	No indígenas e indígenas	No indígena	Indígena
Diferencia salarial		0,078***			0,364***	
Componente debido a la diferencia de dotaciones	-0,092*** (0,007)	-0,092*** (0,007)	-0,130*** (0,011)	0,206*** (0,010)	0,220*** (0,010)	0,232*** (0,032)
Componente atribuido a la discriminación	0,170*** (0,009)	0,169*** (0,010)	0,208*** (0,012)	0,158*** (0,017)	0,144*** (0,017)	0,131*** (0,033)
Número de observaciones	23.053	23.053	23.053	23.053	23.053	23.053

Nota: *** p<0.01, ** p<0.05, * p<0.1. Errores estándar entre paréntesis.

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

En resumen, las estimaciones de las brechas salariales y sus componentes se muestran un nivel de significancia elevado (en todos los casos son significativos al 99% de confianza) en los diferentes métodos de descomposición. Los resultados

anteriores sugieren la existencia de discriminación salarial en el mercado laboral ecuatoriano especialmente en contra de las mujeres; en tanto que por etnia, la discriminación salarial comparte una participación relativamente similar al componente asignado a las dotaciones.

3.4. DISCRIMINACIÓN SALARIAL POR GÉNERO Y ETNIA A TRAVÉS DE LA METODOLOGÍA MACHADO-MATA (M-M)

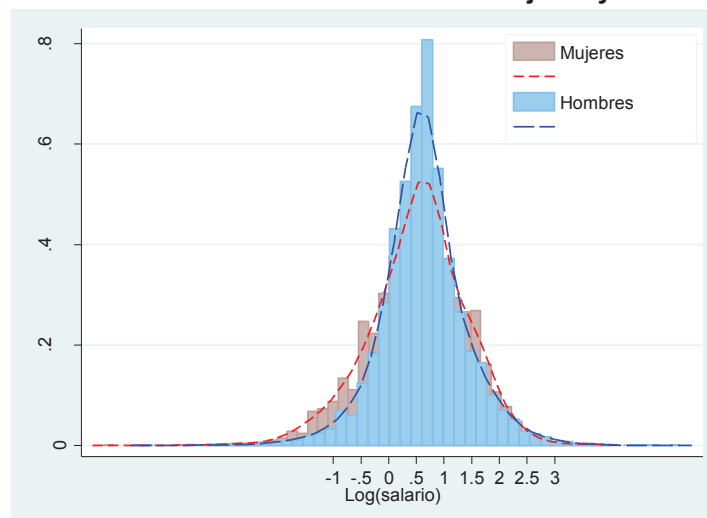
En este apartado, se muestran los resultados obtenidos a través de la utilización de la metodología (Machado & Mata, 2005). En este sentido, al igual que en el capítulo anterior se tomó la muestra de la población de ocupados entre 15 y 65 años de edad cuyas características en términos generales son las siguientes: la mayoría de hombres y mujeres viven en las zonas urbanas del país, son individuos relativamente jóvenes (pues aproximadamente el 40% de la población se encuentra entre 15 y 30 años); y, con un nivel de instrucción mayoritariamente secundario.

La condición de actividad de hombres y mujeres muestra que la mayoría de mujeres son inactivas (49,0%), o subempleadas (29,1%), en tanto que los hombres también son altamente subempleados (37,8%) y ocupados plenos (40,0%). En cuanto a las características de los puestos de trabajo o la actividad específica que desarrollan o desarrollaron los(as) trabajadores dentro del establecimiento, se observa que las mujeres en un alto porcentaje realizaron actividades como “Trabajadoras no calificadas” (24,8%), y como “Trabajadoras de los servicios y comerciantes” (32,1%); los hombres también fueron categorizados en su gran mayoría como “Trabajadores no calificados” (25,5%), y como “Oficiales, operarios y artesanos” (18,2%). Adicionalmente, se aprecia que los hombres presentan un mayor salario por hora respecto a las mujeres.

En el siguiente gráfico se muestra las distribuciones del logaritmo de los salarios por hora de mujeres y hombres. Se puede apreciar que los salarios de los hombres

parecen ser más altos que el de las mujeres; adicionalmente, se observa que ninguna de las dos muestras tiene una distribución normal.

Gráfico 2. Densidad de los salarios de mujeres y hombres



Fuente: ENEMDUR, diciembre de 2013.

Al realizar las pruebas de normalidad de los logaritmos de los salarios por hora de hombres y mujeres (representados por la letra w) se aprecia que estos no tienen una distribución normal. En las pruebas de normalidad se observa la existencia de asimetría y apuntalamiento en el caso de la distribución de las mujeres y apuntalamiento en la distribución de hombres, pero en conjunto, se podría decir que las variables analizadas siguen una ley diferente a la normal para las dos distribuciones (se rechaza la hipótesis nula de normalidad de las variables analizadas).

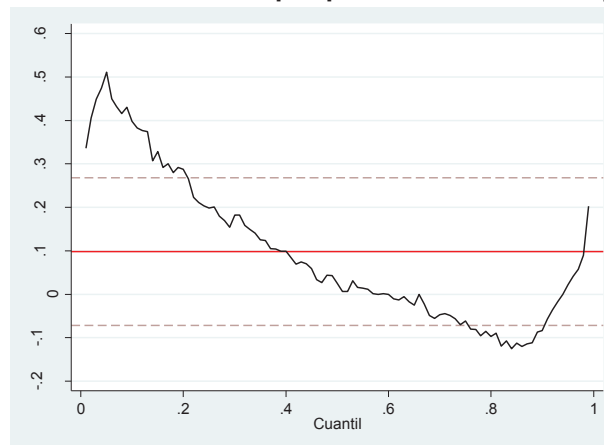
Tabla 4. Pruebas de normalidad de los logaritmos del salario por hora por género

Variable	Pr(Asimetría)	Pr(Curtosis)	Prob>chi2
w hombres	0.7323	0.0000	0.0000
w mujeres	0.0000	0.0000	0.0000

En el siguiente gráfico se aprecia la diferencia salarial por género para los distintos percentiles de la distribución de las brechas salariales. Se observa que la diferencia

salarial entre hombres y mujeres en el rango intercuantílico (20-70) está relativamente cercana a la media de dicha diferencia (0,0978177; USD 1,10) y a su desviación estándar (0,1699077; USD 1,19).

Gráfico 3. Diferencial salarial por percentil de hombres y mujeres



Fuente: ENEMDUR, diciembre de 2013.

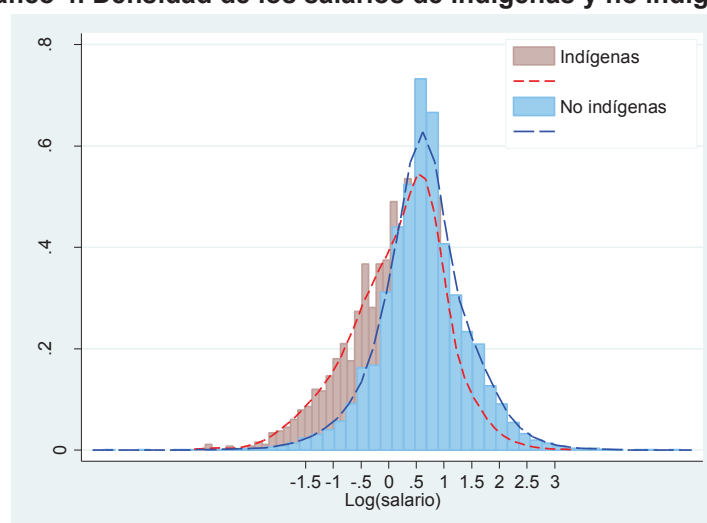
El análisis por etnia consideró a una muestra caracterizada por personas mayoritariamente auto identificadas como “No Indígenas” que viven más en las zonas urbanas (72,4%), en tanto que el grupo de indígenas se encuentran ubicados en su gran mayoría en las zonas rurales (77,3%). Los dos grupos son relativamente jóvenes pero su nivel de educación difiere, ya que la población no indígena en su gran mayoría alcanza una instrucción secundaria (47,9%), en tanto que aproximadamente el 50,9% de la población indígena apenas alcanza la instrucción primaria. La condición de actividad de los dos grupos muestra que un alto porcentaje los “No indígenas” son inactivos (35,0%), o subempleados (32,0%), en tanto que los indígenas también son altamente subempleados (56,4%) e inactivos (25,4%).

En cuanto a las características de los puestos de trabajo dentro del establecimiento donde laboran los grupos se observa que los “No indígenas” en un alto porcentaje son categorizados como “Trabajadores no calificados” (24,4%), y como “Trabajadores de los servicios y comerciantes” (22,9%); los “Indígenas” también son

categorizados en su gran mayoría como “Trabajadores no calificados” (36,3%), y como “Trabajadores calificados agropecuarios y pesqueros” (30,0%). Además, se aprecia una mayor diferencia de los ingresos promedio por hora entre los dos grupos que la diferencia de salarios por género.

A continuación se pueden observar las distribuciones del logaritmo de los salarios por hora de no indígenas e indígenas, donde los salarios de los no indígenas parecen ser más altos que los salarios de los indígenas; parecería además que ninguna de las dos variables presentan una distribución normal.

Gráfico 4. Densidad de los salarios de indígenas y no indígenas



Fuente: ENEMDUR, diciembre de 2013.

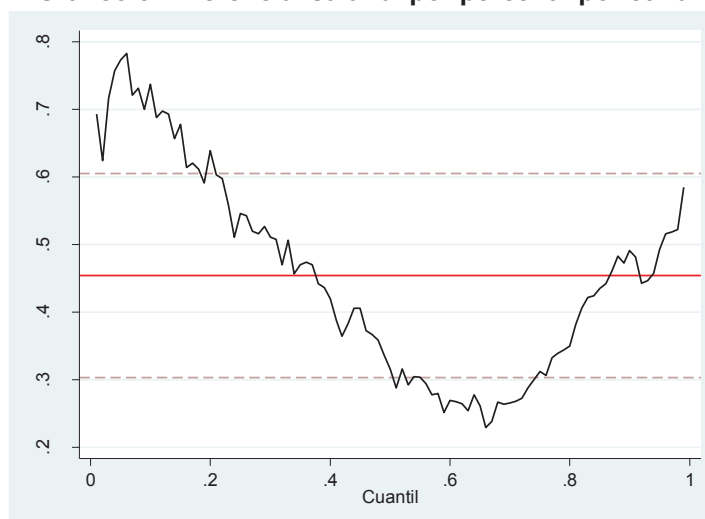
Al realizar las pruebas de normalidad de los logaritmos de los salarios por hora de no indígenas e indígenas (representados por la letra w) se observa que estos no tienen una distribución normal, pues se encontró la existencia de asimetría y apuntalamiento en el caso de las dos distribuciones, y por tanto en conjunto, se rechaza la hipótesis nula de normalidad.

Tabla 5. Pruebas de normalidad de los logaritmos del salario por hora por etnia

Variable	Pr(Asimetría)	Pr(Curtosis)	Prob>chi2
w no indígenas	0.0000	0.0000	0.0000
w indígenas	0.0000	0.0000	0.0000

Seguidamente, en el gráfico No. 5, se aprecia la diferencia salarial por etnia para los diferentes percentiles de la distribución de las brechas salariales; en dicho gráfico se observa que en el diferencial salarial por etnia particularmente en el rango intercuantílico (20-45 y 78-100) está relativamente cercana a la media de dicha diferencia (0,4542123; USD 1,57) y su desviación estándar (0,1510669; USD 1,16).

Gráfico 5. Diferencial salarial por percentil por etnia



Fuente: ENEMDUR, diciembre de 2013.

3.4.1. Regresiones cuantílicas de salarios.

Para poder estudiar las variaciones de las brechas salariales en cada uno de los “estratos” de ingreso, en la Tabla No. 6 se exponen los resultados de la aplicación de las regresiones cuantílicas mincerianas aplicadas a los cuantiles 1, 10, 25, 50, 75, 90 y 99 de la distribución del ingreso. Adicionalmente, se presenta los resultados obtenidos por el método de mínimos cuadrados ordinarios (MCO), el cual se muestra en la última columna de la tabla antes mencionada.

Los resultados muestran básicamente que:

- (i) la mayoría de los casos los coeficientes obtenidos varían en cada uno de los cuantiles; esto es, la respuesta de la variable dependiente a cambios de las covariantes es cambiante en los diferentes percentiles de la distribución de salarios;
- (ii) los denominados rendimientos de la educación parecen variar en cada uno de los cuantiles y se presentan particularmente más elevados para los percentiles más altos de la distribución (lo que se puede evidenciar por los valores del coeficiente asociado a la variable “Educación” y su nivel de significancia estadística relacionado);
- (iii) el coeficiente de la variable que identifica el género (mujer=1) muestra que aparentemente, desde una perspectiva estadística, las mujeres ganan menos ingresos promedio con respecto a los hombres en cada uno de los cuantiles. Sin embargo, esta diferencia es cada vez menor en la medida que se avanza hacia la parte más alta de la distribución de la renta, lo que desde una perspectiva empírica muestra la posible existencia de discriminación salarial negativa en contra de las mujeres, en casi toda las partes de la distribución del ingreso; en esta misma línea, es importante mencionar que los resultados obtenidos a través de MCO son relativamente parecidos o cercanos a los estimados en las regresiones cuantílicas de los percentiles 25 a 50;
- (iv) al igual que en caso anterior, desde una perspectiva estadística, los salarios de los indígenas parecen ser menores que los no indígenas en todos los percentiles de la distribución de la renta, aunque esta diferencia, parece ser menor en la parte más alta de la distribución, lo que nuevamente evidencia la existencia de discriminación salarial negativa en contra de los individuos autodenominados indígenas;
- (v) otro resultado importante es que la variable que controla la experiencia en la ecuación muestra que mientras mayor es ésta, posiblemente el ingreso sea

más alto en todos los puntos de la distribución y más aún, los rendimientos esperados por la experiencia parecen ser más elevados en los niveles más altos de la distribución del ingreso;

- (vi) por otra parte al parecer ser casado(a) respecto de otros estados o condiciones civiles, proporciona un rendimiento salarial más alto en todos los puntos de la distribución de la renta y estos últimos aumentan en la medida que se avanza hacia la parte más alta de la distribución;
- (vii) los resultados detallados en el punto anterior son similares a los obtenidos para la variable que controla el lugar de residencia de los individuos, particularmente los individuos que declararon vivir en la ciudad de Quito, parecen ganar mayores ingresos que los demás individuos residentes en las otras ciudades del país; sin embargo, es importante mencionar que aparentemente dicha ubicación no necesariamente es una ventaja desde la perspectiva salarial en las partes más altas de la distribución de la renta;
- (viii) finalmente, al parecer vivir en la zona de la Sierra norte genera posiblemente mayores rendimientos salariales, respecto de las demás zonas geográficas del país, y en casi todos los puntos de la distribución, a excepción del más bajo de la misma en la que el coeficiente se muestra negativo y no significativo desde una perspectiva estadística.

Tabla 6. Estimación de regresiones cuantílicas de ingresos a nivel poblacional, por género y etnia 2013

	Q01	Q10	Q25	Q50	Q75	Q90	Q99	MCO
Mujer	-0.401*** (-4.596)	-0.301*** (-14.945)	-0.198*** (-14.983)	-0.136*** (-13.316)	-0.093*** (-8.299)	-0.083*** (-4.688)	0.060 (0.910)	-0.170*** (-16.832)
Indígena	-0.603*** (-4.228)	-0.411*** (-8.658)	-0.267*** (-9.194)	-0.113*** (-6.208)	-0.082*** (-4.863)	-0.104*** (-3.741)	-0.067 (-0.456)	-0.185*** (-10.562)
Años de educación	0.060*** (5.423)	0.073*** (25.939)	0.074*** (44.963)	0.078*** (63.064)	0.087*** (79.022)	0.093*** (50.401)	0.098*** (14.702)	0.083*** (71.107)
Experiencia	0.003 (0.257)	0.013*** (5.376)	0.014*** (8.791)	0.019*** (15.167)	0.021*** (16.106)	0.019*** (8.957)	0.031*** (4.176)	0.016*** (13.487)
Experiencia al cuadrado	-0.000 (-0.835)	-0.000*** (-6.038)	-0.000*** (-9.146)	-0.000*** (-11.872)	-0.000*** (-10.617)	-0.000*** (-3.950)	-0.000* (-1.916)	-0.000*** (-10.684)
Casado(a) / unido(a)	0.006 (0.076)	0.061*** (3.375)	0.084*** (6.511)	0.079*** (7.930)	0.070*** (7.011)	0.103*** (6.460)	0.054 (0.866)	0.082*** (8.600)
Jefe(a) de hogar	-0.010 (-0.108)	0.089*** (4.547)	0.068*** (4.825)	0.056*** (5.191)	0.060*** (5.131)	0.075*** (4.259)	0.153** (2.026)	0.069*** (6.445)
Migrante	0.125 (1.338)	0.095*** (3.878)	0.090*** (6.056)	0.082*** (6.954)	0.071*** (6.210)	0.066*** (3.490)	0.057 (0.690)	0.082*** (7.043)
Quito	0.095 (0.318)	0.231*** (6.000)	0.100*** (4.470)	0.114*** (4.564)	0.067*** (3.126)	0.055* (1.631)	0.025 (0.155)	0.082*** (3.466)
Área rural	-0.217*** (-2.883)	-0.182*** (-9.276)	-0.122*** (-9.018)	-0.084*** (-8.950)	-0.067*** (-7.106)	-0.097*** (-5.732)	-0.196*** (-3.077)	-0.125*** (-12.933)
Sierra sur	-0.626*** (-4.834)	-0.259*** (-7.545)	-0.092*** (-4.608)	-0.025* (-1.765)	-0.008 (-0.660)	-0.016 (-0.793)	0.085 (0.973)	-0.088*** (-7.263)
Sierra norte	-0.041 (-0.356)	0.053** (2.117)	0.103*** (6.412)	0.085*** (6.719)	0.073*** (5.447)	0.063*** (3.162)	0.054 (0.486)	0.092*** (6.905)
Constante	-1.457*** (-7.769)	-0.890*** (-20.085)	-0.566*** (-20.547)	-0.345*** (-17.367)	-0.156*** (-7.957)	0.098*** (3.033)	0.694*** (5.660)	-0.377*** (-18.170)
Número de observaciones	26,322	26,322	26,322	26,322	26,322	26,322	26,322	26,322

Nota: *** p<0.01, ** p<0.05, * p<0.1. Valores del estadístico t entre paréntesis, para los errores estándar se usó la técnica de bootstrap con 2000 repeticiones.

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

3.4.2. Descomposición de salarios de (Machado & Mata, 2005)

Utilizando el procedimiento de (Machado & Mata, 2005) descrito anteriormente, en la tabla 7, se observa la descomposición salarial por género y etnia en cada uno de los cuantiles de la distribución.

Tabla 7. Descomposición de ingresos de Machado y Mata por género y etnia 2013

Cuantil	Género			Etnia		
	Diferencial salarial	Componente debido a diferencias en dotaciones	Componente atribuido a discriminación	Diferencial salarial	Componente debido a diferencias en dotaciones	Componente atribuido a discriminación
1	0.4336 (0.0513)	-0.0108 (0.0320)	0.4443 (0.0597)	0,6761 (0.0951)	0,1691 (0.0317)	0,5071 (0.1075)
10	0.3724 (0.0219)	-0.0379 (0.0110)	0.4103 (0.0227)	0,6286 (0.0393)	0,1865 (0.0122)	0,4421 (0.0385)
25	0.1980 (0.0153)	-0.0444 (0.0073)	0.2423 (0.0156)	0,4536 (0.0293)	0,1796 (0.0096)	0,2740 (0.0275)
50	0.0265 (0.0109)	-0.0586 (0.0067)	0.0851 (0.0101)	0,3095 (0.0204)	0,2014 (0.0090)	0,1082 (0.0186)
75	-0.0655 (0.0118)	-0.0742 (0.0082)	0.0087 (0.0104)	0,3148 (0.0188)	0,2489 (0.0103)	0,0659 (0.0164)
90	-0.0551 (0.0162)	-0.0766 (0.0106)	0.0215 (0.0158)	0,3910 (0.0295)	0,2834 (0.0141)	0,1076 (0.0248)
99	0.0845 (0.0539)	-0.0585 (0.0257)	0.1430 (0.0571)	0,4926 (0.0858)	0,3263 (0.0311)	0,1664 (0.0879)

Nota: Errores estándar entre paréntesis.

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

Elaboración: Autores.

En términos generales, lo que se puede observar es que la diferencia salarial, desde la perspectiva de género es relativamente más elevada en los percentiles más bajos y hasta aproximadamente la tercera parte de la distribución de la renta. En este sentido como se muestra en la tabla anterior, los diferenciales salariales ajustados o controlados por las diferentes características observables, varía desde 0.4336 puntos logarítmicos, hasta 0.1980 puntos en el cuantil 1 y 25 respectivamente. Asimismo, los diferenciales salariales a partir del cuantil 50 y 99 son de 0.0265 y 0.0845 puntos logarítmicos, respectivamente. Así mismo, otro resultado interesante muestra que el mayor porcentaje de dichos diferenciales es debido básicamente al componente atribuido a la discriminación salarial; en todos los casos, dicho componente representa más del 100% de dicho diferencial lo que significa una clara evidencia desde la perspectiva estadística de la existencia de una discriminación salarial en contra de la mujer. Sin embargo de lo anterior, existen dos cuantiles que no responden a lo determinado en los párrafos mencionados. En este sentido, los cuantiles 75 y 90 se observa primeramente que las brechas salariales son negativas

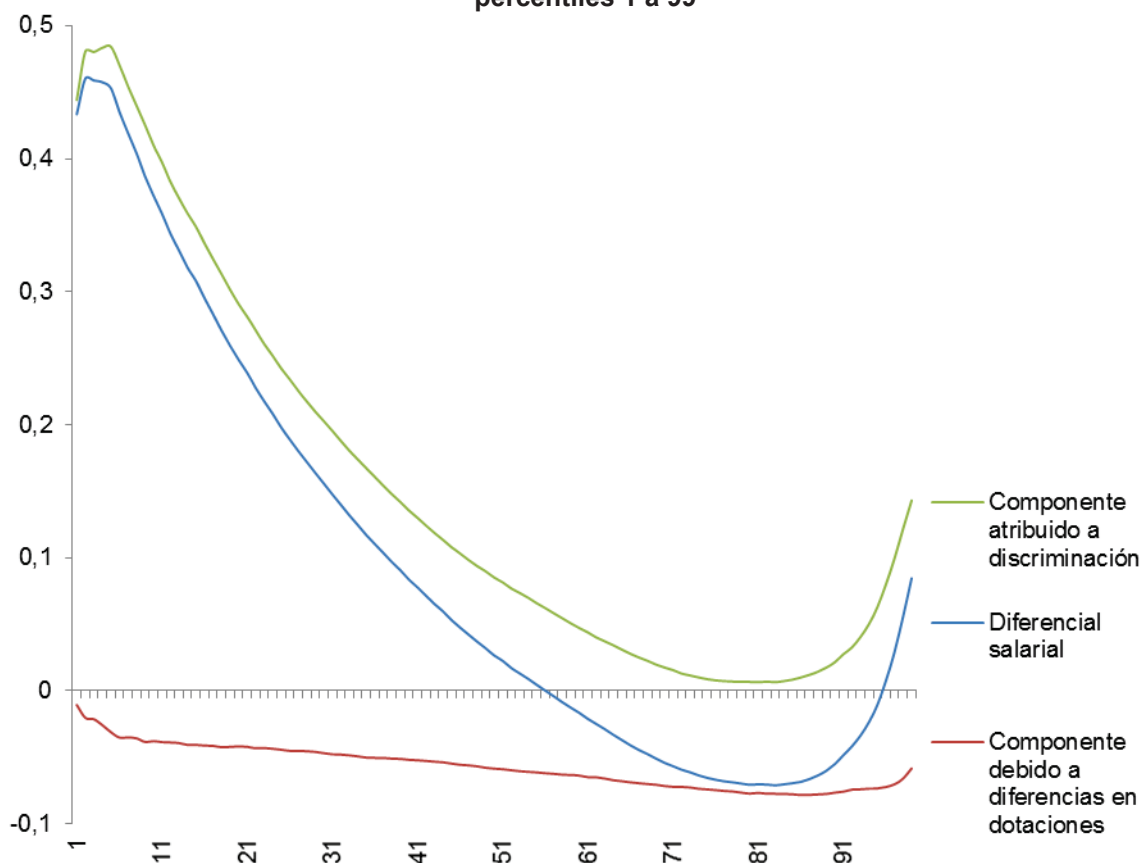
(-0.0655 y -0.0551 puntos logarítmicos). Seguidamente en los mismos cuantiles, se aprecia que el componente atribuido las dotaciones es negativo y marca la tendencia de la brecha salarial (dicho componente representa más del 100% de la diferencia salarial), lo que significa que en estos caso las diferencias de ingresos entre los hombres y mujeres están mucho más relacionadas con las diferencias en las características observables. Por lo que posiblemente, en dichos cuantiles de la distribución las mujeres parecen tener mayor dotación individual lo que explica el diferencial salarial respecto de los hombres.

En base a los resultados anteriormente descritos, en el gráfico 6 se representa el diferencial salarial por género de manera agregada. Se observa una tendencia decreciente en la curva de diferencia salarial; en otras palabras, existe una mayor diferencia salarial por género en los primeros cuantiles, mientras que en los cuantiles superiores la diferencia es mínima. Es decir, los diferenciales salariales por género están determinados básicamente por el comportamiento de las brechas en la parte inferior de la distribución del ingreso, reduciéndose paulatinamente en la medida que se avanza hacia la derecha de la distribución. De hecho, como se observa en el gráfico mencionado, los percentiles hacia el centro de la distribución muestran brechas cercanas a cero, especialmente a partir del percentil 55. En esta línea, y en concordancia con lo establecido en el párrafo anterior, a partir del percentil 70 y aproximadamente hasta el cuantil 90 el efecto discriminación es cercano a cero. En tanto que, la brecha salarial y el efecto dotación son negativos por lo que las diferencias salariales están más asociadas con las brechas de características observables a favor de las mujeres.

También se encuentra que el componente de los diferenciales atribuido a las características observables o efecto dotación es negativo en todos los puntos de la distribución de la renta lo que es coherente con los resultados obtenidos anteriormente. Finalmente, por el lado de la curva de componentes atribuidos a la discriminación tiene la misma tendencia de la curva de diferencial salarial, lo que

significa que dicho componente es el que más explica las brechas salariales, pues en todos los percentiles éste compensa el efecto negativo de las dotaciones o características observables, sobrepasando en toda la distribución a la brecha total (excepto en una parte de la distribución de la renta relacionado con los cuantiles 70 y 90 aproximadamente como se mencionó anteriormente), lo que implica que en todas las partes de la distribución del ingreso el diferencial sea o esté a favor de los hombres y en contra de las mujeres.

Gráfico 6. Diferencia del logaritmo del salario por hora entre hombres y mujeres entre los percentiles 1 a 99



Fuente: ENEMDUR, diciembre de 2013.

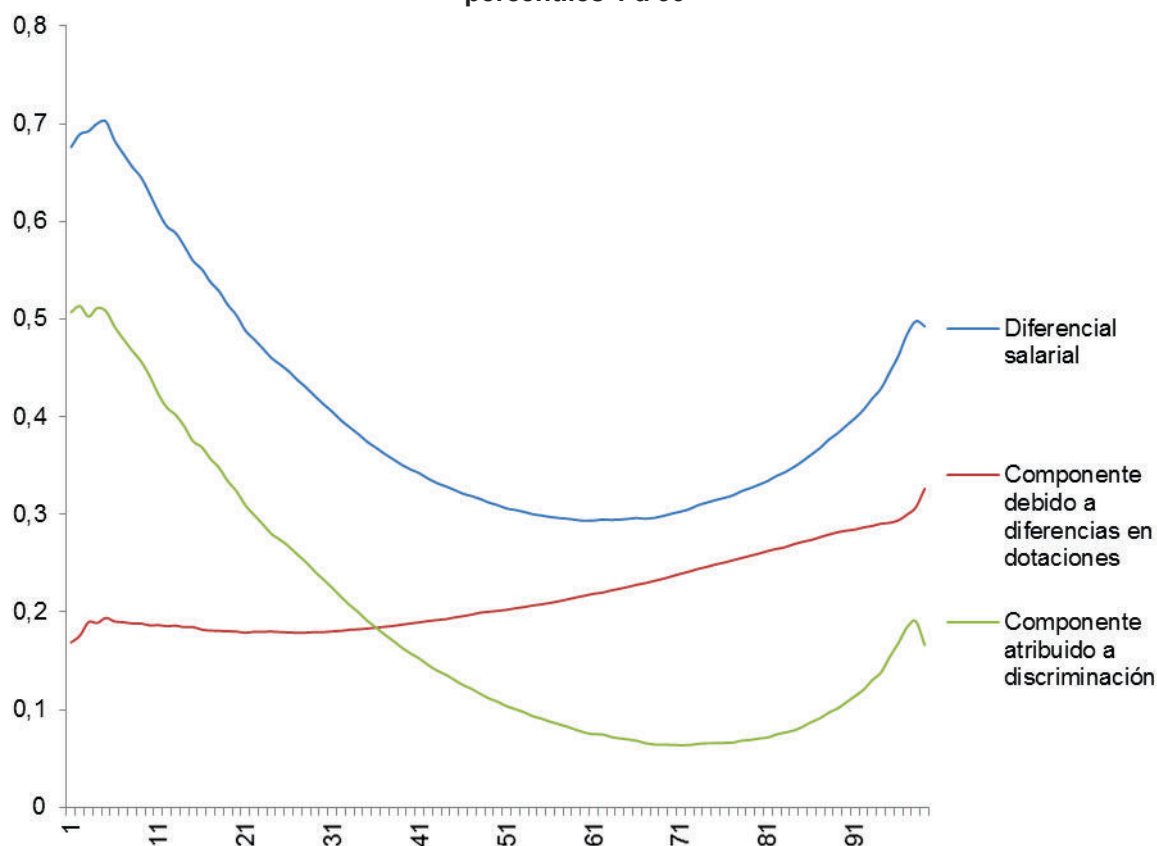
Descomposición de los diferenciales salariales por etnia por método (Machado & Mata, 2005): en este caso los resultados se muestran relativamente distintos. Como se puede observar en la tabla 7, desde la perspectiva étnica, las brechas salariales parecen ser positivas en todos los puntos de la distribución pero estas

tiene una forma de U en la medida que se desplaza hacia la derecha de la distribución, lo que significa que hacia la mitad de la misma las diferencias son positivas pero menores respecto de la cola baja y alta de la distribución del ingreso.

En este caso, el componente atribuido a la discriminación salarial para ser más relevante en la diferencia salarial especialmente en la parte más baja de la distribución de la renta; sin embargo, este componente disminuye en la medida que se desplaza a lo largo de la distribución, aunque vuelve a tomar relevancia en la parte más alta de la distribución. En esta misma línea y en concordancia con lo anterior, el componente atribuido a las dotaciones o a las características observables, parece describir el comportamiento de las brechas salariales especialmente en la parte más alta de la distribución de la renta mientras en la más baja no necesariamente.

Lo mencionado anteriormente se puede observar en el gráfico 7, donde se aprecia la dinámica del comportamiento de las brechas salariales por etnia y sus componentes, concluyendo que, a diferencia de lo observado en el punto anterior, existe una relevancia compartida entre los efectos dotación y discriminación en las brechas salariales por etnia, lo que puede significar que en los diferentes puntos de la renta, es necesario un análisis mucho más minucioso respecto de la discriminación salarial por etnia, la cual parece estar más presente en la parte más baja de la renta.

Gráfico 7. Diferencia del logaritmo del salario por hora entre no indígenas e indígenas entre los percentiles 1 a 99



Fuente: ENEMDUR, diciembre de 2013.

3.5. DESCOMPOSICIÓN DE DIFERENCIAS SALARIALES EN EL TIEMPO

3.5.1. Enfoque de (Juhn, Murphy, & Pierce, 1991)

En este punto, se implementó la descomposición (Juhn, Murphy, & Pierce, 1991) usando en la especificación de las regresiones de salarios variables de capital humano (educación, experiencia y experiencia al cuadrado), estado civil, jefe de hogar y variables de localización geográfica; sin embargo no se introdujo variables de control para las categorías de ocupación y rama de actividad económica para no subestimar los efectos de la discriminación salarial, tal como se mencionó en los capítulos iniciales.

Los datos examinan las relaciones en las estructuras salariales por género y etnia de la Encuesta Nacional de Empleo y Desempleo (ENEMDU) de diciembre de 2007 y diciembre 2013. Para los dos años analizados, se utilizó la muestra de ocupados de 15 a 65 años de edad. Los salarios se expresan en términos reales utilizando el Índice de Precios al Consumidor de 2004. Los promedios de las variables y los resultados de las ecuaciones de salarios de cada año se presentan en los Anexos C y D. Para la estimación de las ecuaciones de salarios del Anexo D se excluyeron los puntos atípicos, influyentes y palanca, y para el cálculo de los errores estándar se utilizó la técnica de re muestreo bootstrap con 2000 repeticiones.

En la tabla 8 se observa que parece hay una leve disminución de la brecha salarial entre hombres y mujeres e indígenas y no indígenas entre los años 2007 y 2013. Además, los salarios de las mujeres e indígenas crecen a mayor velocidad que los salarios de hombres y no indígenas. En los dos casos los errores estándar de los salarios son relativamente bajos con lo cual tiene una mínima variabilidad

Tabla 8. Evolución del salario real (base del IPC 2004), 2007-2013

	2007	2013		2007	2013
Log salario por hora hombre	0,0963	0,3351	Log salario por hora no indígena	0,0827	0,3205
	(0,8959)	(0,7576)		(0,9065)	(0,7878)
Log salario por hora mujer	-0,0232	0,2214	Log salario por hora indígena	-0,4163	-0,1714
	(0,9467)	(0,8650)		(0,9461)	(0,8704)
Diferencia salarial	0,1195	0,1137	Diferencia salarial	0,4990	0,4918

Nota: errores estándar entre paréntesis.

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

En el siguiente gráfico se aprecia la diferencia salarial por género y etnia para los diferentes percentiles de la distribución de las brechas salariales en los años 2007 y 2013. En el Anexo E se tiene la brecha salarial por género y etnia por percentiles de la distribución para los años 2007 y 2013.

En la parte izquierda del gráfico siguiente, se aprecia las diferencias salariales por género, en los años 2007 y 2013. Se puede observar que la brecha salarial en el

año 2013 es menor a la del año 2007 en el rango intercuantílico (40-90), mientras que por etnia (que se observa en la parte derecha del gráfico) se observa que la diferencia salarial entre no indígenas e indígenas es menor en el año 2013 respecto del año 2007 en el rango intercuantílico (50-99); adicionalmente, se encuentra que la brecha salarial por género y etnia ha disminuido entre 2013 y 2007 en la mayor parte de la distribución, especialmente en los individuos de ingreso medio y medio alto, al contrario de lo que sucede con las brecha por género en el rango intercuantílico (1-20) y por etnia en el rango intercuantílico (5-40), donde la brecha entre los años de análisis ha aumentado.

Gráfico 8. Diferencial salarial por percentil, 2007-2013
Género Etnia



Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

Para la descomposición de la brecha salarial entre los años 2007 y 2013 (Tabla 9), en las ecuaciones de salarios se utilizaron variables explicativas de control en las ecuaciones de salarios de capital humano como los años de escolaridad, los años de experiencia laboral y experiencia laboral al cuadrado, también se utilizaron otras como estado civil, jefe de hogar, migración y variables de localización geográfica (si la persona vive en Quito, en el área rural, en la sierra sur o sierra norte). Los resultados corroboran lo presentado en la tabla 8, parece haber una disminución de la brecha salarial entre hombres y mujeres e indígenas y no indígenas entre los años 2007 y 2013.

La brecha salarial entre 2007 y 2013 disminuyó en -0,0357 puntos logarítmicos (USD 0,96 por hora) por género y en -0,0046 puntos logarítmicos (USD 1,00 por hora) entre no indígenas e indígenas. El componente de la brecha salarial asociado a discriminación disminuyó entre hombres y mujeres (-0.0309 puntos logarítmicos), mientras que aumentó para la población no indígena e indígena (0.0551).

Entre hombres y mujeres la parte de la brecha que se explica por la diferencia de dotaciones en las variables observadas X representó el 13,4% (-0,0048/-0,0357), mientras que el componente atribuido a la discriminación el 86,6% (-0,0309/-0,0357). Por tanto, la mayor parte de la brecha salarial se explicaría por el componente asociado a discriminación. Por otro lado el lado del componente de la brecha explicado por la diferencia de dotaciones se observó que de manera conjunta (es decir si se analiza el efecto de las cantidades X y los coeficientes o “precios” β), las variables que más contribuyen a la disminución de la brecha salarial son experiencia laboral y jefe de hogar, mientras que las que más aumentan la brecha son las variables de localización geográfica: sierra sur, sector rural y Quito.

Entre hombres y mujeres la parte de la brecha que se explica por la diferencia de dotaciones en las variables observadas X fue 13,4% (0,0048/0,0357), mientras que el componente atribuido a la discriminación fue 86,6%. Por género la mayor parte de la brecha salarial se explica por el efecto de discriminación. Por el lado del componente de la brecha debido a la diferencia de dotaciones se observa que de manera conjunta al analizar el efecto de las cantidades X y los coeficientes o “precios” β , las variables que más contribuyen a la disminución de la brecha salarial la experiencia laboral y el hecho de ser jefe de hogar, mientras que las que aumentan la brecha principalmente son las variables de localización sierra sur, sector rural y Quito.

Por otro lado, entre no indígenas e indígenas el componente observado representó el 1288,8% y el atribuido a discriminación -1188,8%. De forma global se observa que la

brecha salarial entre los años 2013 y 2007 entre no indígenas e indígenas tiene un componente fuerte tanto de dotaciones como de discriminación.

La brecha salarial como se señaló se redujo en -0.0046 puntos logarítmicos (-0.0597+0.0551), lo que se explica porque el componente debido a la diferencia de dotaciones contribuyó a su reducción con -0.0597 puntos logarítmicos, mientras que el componente atribuido a la discriminación contribuyó a su aumento con 0.0551 puntos logarítmicos. Al analizar el componente de la brecha debido a dotaciones, tanto el efecto de las cantidades X como el de los coeficientes o “precios” β , se obtiene que las variables que más contribuyen a la disminución de la brecha salarial fueron años de educación, experiencia laboral, el hecho de estar casados o unidos, mientras que las que aumentan la brecha principalmente son las variables de migración y sector rural.

Entre hombres y mujeres la parte de la brecha que se explica por la diferencia de dotaciones en las variables observadas X representó el 13,4% (-0,0048/-0,0357), mientras que el componente atribuido a la discriminación el 86,6% (-0,0309/-0,0357). Por tanto, la mayor parte de la brecha salarial se explicaría por el componente asociado a discriminación. Por otro lado el lado del componente de la brecha explicado por la diferencia de dotaciones se observó que de manera conjunta (es decir si se analiza el efecto de las cantidades X y los coeficientes o “precios” β), las variables que más contribuyen a la disminución de la brecha salarial son experiencia laboral y jefe de hogar, mientras que las que más aumentan la brecha son las variables de localización geográfica: sierra sur, sector rural y Quito.

Por otro lado, entre no indígenas e indígenas el componente observado representó el 1288,8% (-0.0597 puntos logarítmicos) y el atribuido a discriminación -1188,8% (0.0551 puntos logarítmicos). Se observa que la brecha salarial entre los años 2013 y 2007 entre no indígenas e indígenas tuvo un componente fuerte tanto de dotaciones como de discriminación.

La brecha salarial como se señaló para no indígenas e indígenas se redujo en -0.0046 puntos logarítmicos $(-0.0597+0.0551)$, lo que se explica porque el componente debido a la diferencia de dotaciones contribuyó a su reducción con -0.0597 puntos logarítmicos, mientras que el componente atribuido a la discriminación contribuyó a su aumento con 0.0551 puntos logarítmicos. Al analizar el componente de la brecha debido a dotaciones, tanto el efecto de las cantidades X como el de los coeficientes o “precios” β , se encontró que las variables que más contribuyen a la disminución de la brecha salarial fueron años de educación, experiencia laboral, y el hecho de estar casados o unidos, mientras que las que aumentan la brecha principalmente fueron las variables de migración y sector rural.

Tabla 9. Descomposición de los cambios en la brecha salarial por género y etnia, 2007-2013

	Hombres y mujeres	No indígenas e indígenas
Brechas salariales		
2007	0,0847	0,3254
2013	0,0490	0,3208
Diferencia brechas 2013 – 2007	-0,0357	-0,0046
Componente debido a la diferencia de dotaciones	-0,0048	-0,0597
Cantidades X	-0,0112	-0,0079
Mujer		-0,0049
Indígena	-0,0005	
Años de educación	-0,0129	-0,0041
Experiencia	-0,0051	-0,0006
Experiencia al cuadrado	0,0046	-0,0125
Casado(a) / unido(a)	0,0024	-0,0084
Jefe(a) de hogar	-0,0044	-0,0038
Migrante	0,0004	0,0090
Quito	0,0018	0,0009
Rural	0,0008	0,0137
Sierra sur	0,0005	-0,0010
Sierra norte	0,0013	0,0038
Precios β	0,0064	-0,0518
Mujer		0,0024
Indígena	-0,0007	
Años de educación	0,0109	-0,0210
Experiencia	-0,0107	-0,0062
Experiencia al cuadrado	0,0054	0,0018
Casado(a) / unido(a)	0,0026	-0,0010
Jefe(a) de hogar	-0,0114	-0,0002
Migrante	0,0000	0,0008
Quito	0,0030	0,0003
Rural	0,0041	-0,0112
Sierra sur	0,0047	-0,0105
Sierra norte	-0,0014	-0,0069
Componente atribuido a la discriminación	-0,0309	0,0551

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

4.1. CONCLUSIONES

Los resultados obtenidos por el método de (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) - (Blinder, 1973) y Neumark (1988) muestran que existen diferenciales salariales en el año 2013 en los grupos laborales por género y etnia; además el principal componente de estas brechas parece ser la discriminación salarial en el caso del grupo por género en tanto que desde la perspectiva de la etnia la parte por características observables es un componente más predominante.

Para el año 2013, a través de la utilización de la metodología M-M la diferencia salarial, desde la perspectiva de género es relativamente más elevada en los percentiles más bajos y hasta aproximadamente la tercera parte de la distribución de la renta. Así mismo, otro resultado interesante muestra que el mayor porcentaje de dichos diferenciales es debido básicamente al componente atribuido a la discriminación salarial; en todos los casos, dicho componente representa más del 100% de dicho diferencial lo que significa una clara evidencia desde la perspectiva estadística de la existencia de una discriminación salarial en contra de la mujer. Desde la perspectiva étnica, para el año 2013, las brechas salariales parecen ser positivas en todos los puntos de la distribución, pero estas tienen una forma de U en la medida en que se desplaza hacia la derecha de la distribución, lo que significa que hacia la mitad de la misma las diferencias son positivas, pero menores respecto de la cola baja y alta de la distribución del ingreso.

Existe una relevancia compartida entre los efectos dotación y discriminación en las brechas salariales por etnia en cada punto de la distribución de los ingresos, lo que puede significar que es necesario un análisis mucho más minucioso respecto de la

discriminación salarial por etnia, la cual parece estar más presente en la parte más baja de la renta.

Finalmente, la brecha salarial entre 2007 y 2013, calculada a través de JMP disminuyó entre hombres y mujeres y entre no indígenas e indígenas, pero el componente de la brecha asociado a discriminación disminuyó por género, mientras que aumentó por etnia. Entre hombres y mujeres la mayor parte de la brecha se explicó por el componente no observado atribuido a la discriminación; mientras que entre no indígenas e indígenas el componente observado y el atribuido a discriminación fue similar.

4.2. RECOMENDACIONES

Promover la utilización de metodologías cuantitativas para los análisis de brechas salariales considerando el método de (Oaxaca, Male-Female Wage Differentials in Urban Labor Markets, 1973) - (Blinder, 1973), Newmark (1988), los diferentes puntos de la distribución de la renta y sus cambios en el tiempo con el objetivo de generar políticas salariales no discriminatorias en favor de las mujeres e indígenas las cuales deben estar orientadas y particularizadas conforme la naturaleza, cuantificación y grado de discriminación salarial, ya que la evidencia empírica para el Ecuador muestra que la discriminación salarial no necesariamente es la misma en el comportamiento de las brechas salariales por género y etnia en los diferentes partes de la distribución del ingreso y a través del tiempo.

Considerando que la discriminación salarial marca el comportamiento de las brechas salariales por género, es necesario identificar los individuos y los sectores económicos que se encuentran inmersos en dicho fenómeno para generar políticas antidiscriminatorias y una cultura empresarial más igualitaria, particularmente en los estratos más bajos de la renta.

Por otra parte, debido a que la discriminación salarial por etnia y el efecto dotación son relativamente similares en casi toda la distribución, es necesario generar políticas para el mejoramiento de las características observables de los individuos auto considerados indígenas, particularmente las políticas educacionales y de acceso al mercado laboral pueden resultar las de mayor relevancia. Sin embargo, tomando en consideración que, especialmente en el sector público ya existe este tipo de políticas, quizá lo que convenga es una evaluación de las mismas para definir sus efectos.

Aun cuando, en este proyecto de tesis no se ha analizado el efecto de los contratos colectivos en el mercado laboral, se requiere estudiar el impacto cuantitativo y cualitativo de los mismos, pues es importante para evaluar si estos son los posibles causantes de la discriminación salarial y explicativos de las brechas salariales en los diferentes puntos de la renta.

BIBLIOGRAFÍA

- Albrecht, J., Van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, 16(4), 383-396.
- Arrow, K. (1971). *The Theory of Discrimination*. . Industrial Relations Section Working Paper, no. 30A, Princeton University.
- Atonji, J., & Blank, R. (1999). Race and gender in the labor market. *Handbook of labor economics*, 3, 3143-3259.
- Becker, G. (1985). *Human Capital, Effort, and the Sexual Division of Labor*. Journal of Labor Economics; vol. 3, no. 1.
- Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4), 436-455.
- Breusch, T., & Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287-1294.
- Buchinsky, M. (1998). Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of human resources*, 88-126.
- Cahuc, P., & Zylberberg, A. (2004). *Labor Economics*. MIT, Press.
- Cain, G. (1986). The economic analysis of labor market discrimination: A survey. *University of Wisconsin--Madison*, 693-785.
- Castro, A. (2010). Regresión lineal. *Monografías de Matemática y Estadística*, 1-192.
- Cook, D. (1977). Detection of influential observation in linear regression. *Technometrics*, 15-18.
- Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition methods in economics. *Handbook of labor economics*, 4, 1-102.
- Galindo, E. (2006). *Estadística métodos y aplicaciones para administración y economía*. Quito: ProCiencia Editores.
- Galvis, L. (2010). Diferenciales salariales por género y región en Colombia: Una aproximación con regresión por cuantiles. *Revista de Economía del Rosario*, 13(2), 1-10.
- García, J., Hernández, P., & Lopez-Nicolas, A. (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical economics*, 26(1), 149-167.
- Gini, C. (febrero de 1913). Variabilità e Mutabilità. *Journal of the Royal Statistical Society*, 76(3), 326-327.
- Guan, W. (2003). From the help desk: bootstrapped standard errors. *The Stata Journal*, 3(1), 71-80.
- Heckman, J. J. (enero de 1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161.
- Hernández, P. (1995). Análisis empírico de la discriminación salarial de la mujer en España. *Investigaciones económicas*, 19(2), 195-215.

- Jann, B. (2008). A Stata implementation of the Blinder-Oaxaca decomposition. *Stata Journal*, 8(4), 453-479.
- Juhn, C., Murphy, K. M., & Pierce, B. (1991). Accounting for the Slowdown in Black-White Wage. *M. H. Koster*, 107-143.
- Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.
- Machado, J. A., & Mata, J. (junio de 2005). Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics*, 20(4).
- Madden, D. (2000). Towards a broader explanation of male-female wage differences. *Applied Economics Letters*, 7(12), 765-770.
- Mincer, J. (agosto de 1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66(4), 281-302.
- Mincer, J., & Polachek, S. (1974). *Family Investments in Human Capital: Earnings of Women*. *Journal of Political Economy*; vol. 82, no. 2.
- Moffitt, R. (1999). New developments in econometric methods for labor market analysis. *Handbook of labor economics*, 1367-1397.
- Neumark, D. (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. *The Journal of Human Resources*, 23(3), 279-295.
- Oaxaca, R. (octubre de 1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3), 693-709.
- Oaxaca, R., & Ransom, M. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1), 5-21.
- Phelps, E. (1972). *The Statistical Theory of Racism and Sexism*. *The American Economic Review*, vol. 62, no. 4.
- Sen, A. (2001). *The Many Faces of Gender Inequality*. *New Republic*, vol. 225, no.12.
- STATA. (2012). STATA Base Reference Manual. 1731-1733.
- Torresano Melo, D. (2009). Análisis empírico de la discriminación salarial por género y etnia en el Ecuador en el año 2008. *Universitat Autònoma de Barcelona*, 1-90.
- Willis, R. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of labor economics*, 1, 525-602.

ANEXOS

ANEXO- A Validación del modelo

Para encontrar la regresión de salarios que se ajuste de la mejor manera a los datos, se realizaron varias pruebas y procedimientos. Para la estimación se utilizó los datos de personas ocupadas y desocupadas del mercado laboral ecuatoriano de 15 a 65 años.

La regresión de salarios corregida por el método de (Heckman, 1979) y el gráfico de los residuos y valores predichos por el modelo se muestran a continuación:

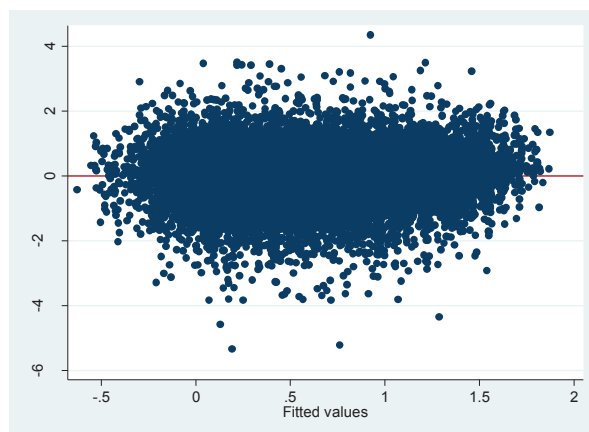
Source	SS	df	MS	
Model	4343.79173	13	334.137826	Number of obs = 26322
Residual	12442.5434	26308	.472956644	F(13, 26308) = 706.49
				Prob > F = 0.0000
				R-squared = 0.2588
				Adj R-squared = 0.2584
Total	16786.3351	26321	.637754459	Root MSE = .68772

w	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mujer	-.1620838	.0103671	-15.63	0.000	-.1824039 -.1417637
indigena_1	-.1934499	.0177199	-10.92	0.000	-.2281819 -.1587178
año_escola~d	.0824462	.0011863	69.50	0.000	.080121 .0847713
experiencia	.014194	.0014072	10.09	0.000	.0114358 .0169523
experienci~2	-.0002094	.0000228	-9.18	0.000	-.0002542 -.0001647
casados_un~s	.0667288	.0107345	6.22	0.000	.0456885 .087769
jefe_hogar	.0484631	.0125138	3.87	0.000	.0239354 .0729909
migra	.0810341	.0115918	6.99	0.000	.0583135 .1037547
quito	.0826501	.023706	3.49	0.000	.0361851 .129115
area_2	-.1364226	.0103428	-13.19	0.000	-.1566951 -.1161501
sierrasur	-.0894652	.0121618	-7.36	0.000	-.1133029 -.0656275
sierranorte	.0897981	.0133214	6.74	0.000	.0636874 .1159087
lambda1	-.3878231	.1272294	-3.05	0.002	-.6371997 -.1384465
_cons	-.2793939	.0382686	-7.30	0.000	-.3544025 -.2043853

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	26322	-31428.96	-27487.98	14	55003.96	55118.45

Gráfico de residuos en función del pronóstico



Se observa una nube de puntos que sugiere que no hay una relación entre los residuos y los valores predichos por el modelo, parece que no hay problemas de heteroscedasticidad, ya que las observaciones están distribuidas aleatoriamente en una franja horizontal simétrica; sin embargo, al existir puntos fuera del intervalo $[-3, 3]$, se procedió a corregir por puntos atípicos, influyentes y palanca.

Para identificar a los puntos palanca se obtiene los elementos de la diagonal de la matriz de proyección denominada “matriz sombrero”. Para (Castro, 2010) si estos valores son mayores al límite $27/n$, entonces el punto correspondiente es un punto palanca. Así también, para reconocer a los puntos influyentes se obtiene la distancia de Cook, cuyos valores mayores a $4/n$ son considerados puntos influyentes²³. El paquete estadístico (STATA Corp, 2012) obtiene la distancia de Cook así:

$$D_j = \frac{\hat{\epsilon}_{sj}^2 (s_{pj} s_{rj})^2}{k} = \frac{h_j \hat{\epsilon}_j^2}{k s^2 (1 - h_j)^2}$$

Donde k es el número de regresores más la constante, s_{pj} es el error estándar de la predicción, s_{rj} el error estándar de los residuos y $\hat{\epsilon}_{sj} = \hat{\epsilon}_j / s_{rj}$ son los residuos estandarizados.

²³ Al respecto véase (STATA, 2012), pp. 1731-1733.

Por otra parte, para los puntos atípicos se calculan los residuos estudentizados; así el paquete estadístico (STATA, 2012) los calcula mediante la siguiente fórmula:

$$r_j = \frac{\hat{e}_j}{s_{(j)}\sqrt{1-h_j}}$$

Sea \hat{e}_j los residuos, $s_{(j)}$ representa la raíz del error cuadrático medio eliminando la j_{th} observación y h_j es el elemento de la diagonal de la matriz de proyección “sombbrero”. Por tanto, según (Castro, 2010), si r_j es mayor a 3 se considera punto atípico.

Al corregir la regresión por puntos atípicos, influyentes y palanca, se obtuvo los siguientes resultados:

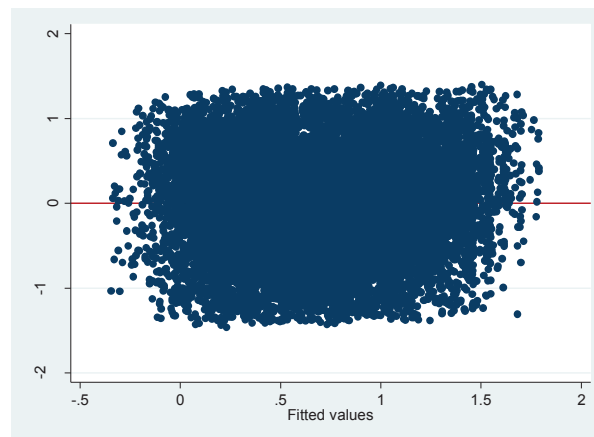
Source	SS	df	MS	
Model	3218.79253	13	247.599425	Number of obs = 23053
Residual	6362.53336	23039	.276163608	F(13, 23039) = 896.57
Total	9581.32589	23052	.415639679	Prob > F = 0.0000
				R-squared = 0.3359
				Adj R-squared = 0.3356
				Root MSE = .52551

w	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mujer	-.1333633	.0086892	-15.35	0.000	-.1503947 -.116332
indigena_1	-.1224683	.0175051	-7.00	0.000	-.1567795 -.0881571
año_escola~d	.0809563	.0009859	82.11	0.000	.0790239 .0828888
experiencia	.0161281	.0012533	12.87	0.000	.0136717 .0185846
experienci~2	-.0002427	.0000202	-11.99	0.000	-.0002824 -.000203
casados_un~s	.0670015	.0092318	7.26	0.000	.0489065 .0850964
jefe_hogar	.045064	.0109023	4.13	0.000	.0236947 .0664332
migra	.0892222	.0096968	9.20	0.000	.0702159 .1082285
quito	.0778645	.0389394	2.00	0.046	.0015407 .1541884
area_2	-.1127526	.008604	-13.10	0.000	-.129617 -.0958882
sierrasur	-.0319031	.0099602	-3.20	0.001	-.0514258 -.0123803
sierranorte	.0771126	.0109336	7.05	0.000	.0556819 .0985432
lambda1	-.3677919	.1238613	-2.97	0.003	-.6105684 -.1250155
_cons	-.28921	.0358774	-8.06	0.000	-.3595321 -.2188879

. estat ic

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	23053	-22590.75	-17871.93	14	35771.86	35884.49

Gráfico de residuos en función del pronóstico



Prueba de heteroscedasticidad

Según (Castro, 2010), los primeros detectores de heteroscedasticidad son los gráficos de dispersión y los gráficos de residuos; por consiguiente, anteriormente ya se analizó el gráfico de residuos en función de las predicciones y se concluyó que no existen indicios de heteroscedasticidad; es decir, al parecer la varianza es constante; sin embargo, para confirmar esta afirmación se realiza la prueba de (Breusch & Pagan, 1979).

La hipótesis nula de la prueba de (Breusch & Pagan, 1979) hace referencia a que el modelo es homoscedástico; mientras que la hipótesis alternativa se refiere a la heteroscedasticidad del modelo; analíticamente se tiene:

$$H_0: \sigma_i^2 = \sigma^2$$

$$H_1: \sigma_i^2 = \sigma^2 h(\alpha_2 z_{i2} + \dots + \alpha_n z_{in})$$

Los resultados del modelo son:

Prueba de Heteroscedasticidad de (Breusch & Pagan, 1979)

Ho: Varianza
constante

chi2(1)	9.90
Prob> chi2	0.0017

Según esta prueba, se confirma que el modelo presenta problemas de heteroscedasticidad, ya que se rechaza la hipótesis nula de homogeneidad de los errores.

Para corregir este problema se procede a estimar la matriz de covarianzas mediante el comando *robust* del programa STATA, obteniéndose los siguientes resultados:

```
Linear regression                               Number of obs = 23053
                                                F( 13, 23039) = 810.37
                                                Prob > F       = 0.0000
                                                R-squared     = 0.3359
                                                Root MSE     = .52551
```

w	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
mujer	-.1333633	.0089002	-14.98	0.000	-.1508083	-.1159184
indigena_1	-.1224683	.0171855	-7.13	0.000	-.156153	-.0887835
año_escola~d	.0809563	.0010144	79.80	0.000	.078968	.0829447
experiencia	.0161281	.0012612	12.79	0.000	.0136561	.0186002
experienci~2	-.0002427	.0000204	-11.90	0.000	-.0002827	-.0002027
casados_un~s	.0670015	.009461	7.08	0.000	.0484572	.0855457
jefe_hogar	.045064	.0111935	4.03	0.000	.0231239	.067004
migra	.0892222	.0096947	9.20	0.000	.0702199	.1082245
quito	.0778645	.0309595	2.52	0.012	.0171818	.1385473
area_2	-.1127526	.0085807	-13.14	0.000	-.1295714	-.0959339
sierrasur	-.0319031	.0102251	-3.12	0.002	-.0519449	-.0118613
sierranorte	.0771126	.0106488	7.24	0.000	.0562402	.0979849
lambda1	-.3677919	.1284732	-2.86	0.004	-.6196081	-.1159758
_cons	-.28921	.0367752	-7.86	0.000	-.3612918	-.2171281

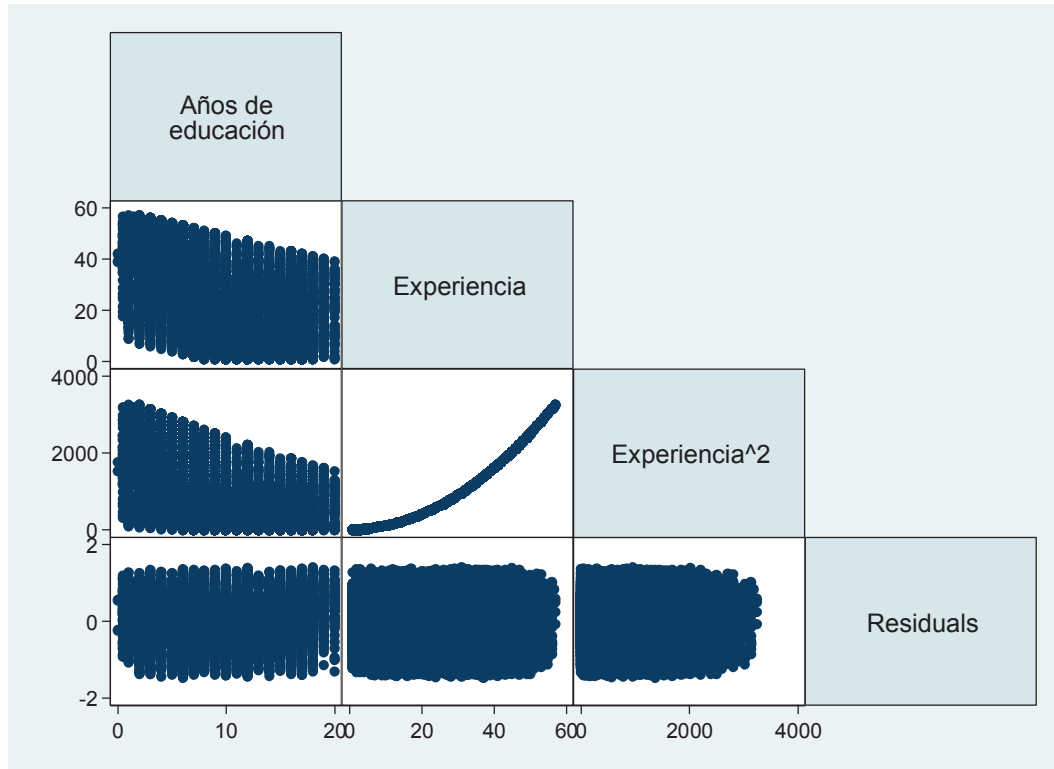
```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	23053	-22590.75	-17871.93	14	35771.86	35884.49

Linealidad del modelo

Los gráficos de años de educación, experiencia y experiencia al cuadrado respecto de los residuos no muestran puntos atípicos; ni problemas con la varianza y la hipótesis de linealidad. Por consiguiente, se podría validar la regresión de salarios.

Gráfico de residuos en función de años de educación, experiencia y experiencia al cuadrado



Prueba de multicolinealidad

Para detectar problemas de Multicolinealidad del modelo, el paquete estadístico Stata (2012) calcula el Factor Inflación de la Varianza centrado (VIF_c^{24}) a través de la siguiente fórmula:

$$VIF_c(X_j) = \frac{1}{1 - \hat{R}_j^2}$$

Donde \hat{R}_j^2 es el cuadrado del coeficiente de correlación múltiple centrado en la regresión de X_j en función de los restantes $k-2$ regresores.

A continuación se muestra el VIF de cada regresor:

²⁴ Por sus siglas en inglés.

Variable	VIF	1/VIF
experiencia	24.05	0.041578
experienci~2	16.90	0.059185
lambda1	7.02	0.142458
jefe_hogar	2.46	0.406163
casados_un~s	1.66	0.600620
año_escola~d	1.60	0.626348
area_2	1.50	0.667413
mujer	1.41	0.710127
sierranorte	1.10	0.905887
indigena_1	1.09	0.915905
quito	1.08	0.923209
migra	1.06	0.944944
sierrasur	1.04	0.963439
Mean VIF	4.77	

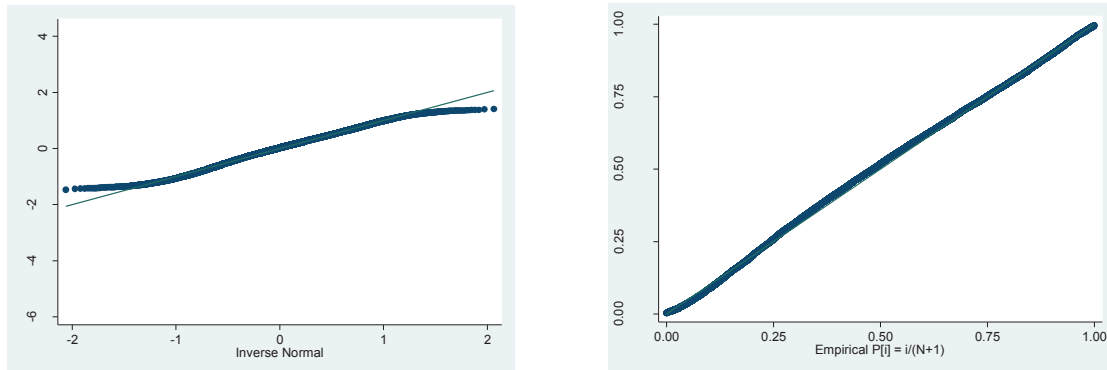
El Factor Inflación de la Varianza (VIF) de las variables experiencia y experiencia al cuadrado es mayor a 10, por tanto, existe problema de multicolinealidad²⁵ en el modelo; no obstante, las variables que evidencian este problema son significativas y el regresor experiencia al cuadrado se construye a partir de la variable experiencia; por ende, no es necesario corregir la multicolinealidad. Además que la teoría económica, en especial las ecuaciones de ingresos propuestas por (Mincer, 1958) requieren de la presencia de dicha variable; pues una de las razones es que se verifica la concavidad negativa de la variable experiencia al cuadrado ya que como se mencionó los ingresos de los individuos crecen hasta un punto determinado de la experiencia y luego empieza su decrecimiento.

Normalidad de los residuos

El Factor Inflación de la Varianza (VIF) de las variables experiencia y experiencia al cuadrado es mayor a 10, por tanto, existe problema de multicolinealidad en el modelo; no obstante, las variables que evidencian este problema son significativas y

²⁵ Para mayor información del tema véase (Castro, 2010) (Cook, 1977) (Galindo, 2006), pp. 71-92.

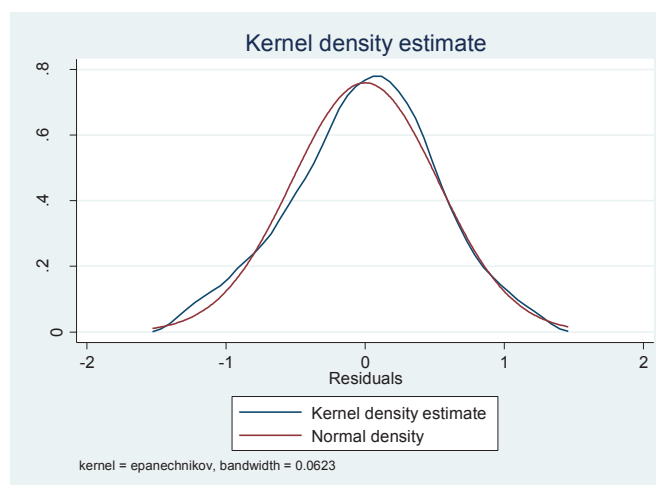
Gráfico de cuantiles y percentiles de los residuos estandarizados frente a los cuantiles de la distribución normal



El gráfico de cuantiles indica la violación de la hipótesis de normalidad de los residuos ya que se observa tanto en el extremo derecho como en el izquierdo puntos que se alejan de la recta. También, el gráfico de percentiles muestra evidencia de la violación de la hipótesis de normalidad ya que no todos los puntos se ajustan a la línea recta.

Lo que se confirma también con el gráfico de la distribución de los residuo sobre una distribución normal.

Distribución de residuos estandarizados



Aparentemente la distribución muestra que los residuos se aproximan a una ley normal; no obstante, se observa apuntamiento y asimetría.

Prueba de apuntamiento y asimetría de los residuos estandarizados

Variable	Pr(Asimetría)	Pr(Curtosis)	Prob>chi2
Rstandard	0.0000	0.0000	0.0000

En las pruebas de normalidad se observa la existencia de asimetría y apuntamiento, en conjunto, se podría decir que los errores siguen una ley diferente a la normal (se rechaza hipótesis nula de normalidad de los residuos).

Según (Galindo, 2006), la suposición de que los residuos estén normalmente distribuidos no es necesaria para la estimación de los parámetros de regresión pero si es indispensable para la construcción de las pruebas de hipótesis e intervalos de confianza de los parámetros. Para solucionar el problema de normalidad de los residuos se procedió a utilizar el método bootstrap con 2000 repeticiones para obtener mejores errores estándar e intervalos de confianza de la regresión de salarios, lo que se muestra a continuación:

Regresión de salarios e intervalos de confianza mediante el método bootstrap con 2000 repeticiones

Linear regression	Number of obs	=	23053
	Replications	=	2000
	Wald chi2(13)	=	10547.96
	Prob > chi2	=	0.0000
	R-squared	=	0.3359
	Adj R-squared	=	0.3356
	Root MSE	=	0.5255

w	Observed Coef.	Bootstrap Std. Err.	z	P> z	Normal-based [95% Conf. Interval]	
mujer	-.1333633	.0089693	-14.87	0.000	-.1509428	-.1157839
indigena_1	-.1224683	.0174601	-7.01	0.000	-.1566895	-.0882471
año_escola~d	.0809563	.0010072	80.38	0.000	.0789823	.0829304
experiencia	.0161281	.0012577	12.82	0.000	.0136631	.0185932
experienci~2	-.0002427	.0000207	-11.70	0.000	-.0002834	-.000202
casados_un~s	.0670015	.0093763	7.15	0.000	.0486242	.0853787
jefe_hogar	.045064	.0111481	4.04	0.000	.0232142	.0669137
migra	.0892222	.0098212	9.08	0.000	.0699729	.1084714
quito	.0778645	.0310388	2.51	0.012	.0170295	.1386995
area_2	-.1127526	.0085038	-13.26	0.000	-.1294197	-.0960855
sierrasur	-.0319031	.0101199	-3.15	0.002	-.0517376	-.0120685
sierranorte	.0771126	.0108224	7.13	0.000	.055901	.0983241
lambda1	-.3677919	.1282158	-2.87	0.004	-.6190903	-.1164936
_cons	-.28921	.036339	-7.96	0.000	-.3604331	-.2179868

ANEXO- B Estimaciones de ecuaciones de ingreso a nivel poblacional, por género y etnia 2013

	Aditivo	Multiplicativo género	Multiplicativo etnia	Hombre	Mujer	No indígena	Indígena
Años de educación	0,081*** (0,001)	0,072*** (0,001)	0,081*** (0,001)	0,072*** (0,002)	0,091*** (0,002)	0,081*** (0,001)	0,083*** (0,005)
Experiencia	0,016*** (0,001)	0,015*** (0,001)	0,015*** (0,001)	0,014*** (0,002)	0,007** (0,003)	0,015*** (0,002)	0,011** (0,005)
Experiencia al cuadrado	-0,0002*** (0,000)	-0,0002*** (0,000)	-0,0002*** (0,000)	-0,0002*** (0,000)	-0,0001*** (0,000)	-0,0002*** (0,000)	-0,0002*** (0,000)
Casado(a) / unido(a)	0,067*** (0,009)	0,032*** (0,011)	0,059*** (0,009)	0,076*** (0,013)	0,038** (0,016)	0,069*** (0,009)	0,004 (0,051)
Jefe(a) de hogar	0,045*** (0,011)	0,061*** (0,013)	0,029*** (0,011)	0,093*** (0,015)	0,014 (0,018)	0,047*** (0,011)	-0,032 (0,045)
Migrante	0,089*** (0,010)	0,083*** (0,009)	0,079*** (0,009)	0,102*** (0,014)	0,045** (0,017)	0,087*** (0,010)	0,042 (0,053)
Quito	0,078** (0,032)	0,080*** (0,020)	0,117*** (0,017)	0,085*** (0,029)	0,134*** (0,029)	0,112*** (0,025)	-0,062 (0,089)
Área rural	-0,113*** (0,008)	-0,120*** (0,009)	-0,118*** (0,009)	-0,154*** (0,016)	-0,119*** (0,016)	-0,107*** (0,009)	-0,147*** (0,039)
Sierra sur	-0,032*** (0,010)	-0,027*** (0,010)	-0,047*** (0,010)	-0,047*** (0,017)	-0,112*** (0,019)	-0,037*** (0,010)	0,126** (0,052)
Sierra norte	0,077*** (0,011)	0,084*** (0,011)	0,068*** (0,011)	0,085*** (0,016)	0,062*** (0,020)	0,053*** (0,011)	0,359*** (0,037)
Mujer	-0,133*** (0,009)	-0,347*** (0,026)	-0,127*** (0,008)			-0,122*** (0,009)	-0,247*** (0,040)
Indígena	-0,122*** (0,017)	-0,132*** (0,016)	-0,109*** (0,027)	-0,197*** (0,024)	-0,223*** (0,037)		
Mujer por años de escolaridad		0,025*** (0,002)					
Mujer por jefe de hogar		-0,078*** (0,020)					
Mujer por área rural		-0,005 (0,017)					
Indígena por años de escolaridad			0,004 (0,003)				
Indígena por migrante			-0,078*** (0,028)				
Indígena por sierra centro			-0,113*** (0,024)				
Constante	-0,289*** (0,037)	-0,142*** (0,040)	-0,218*** (0,038)	-0,172*** (0,064)	-0,278*** (0,080)	-0,274*** (0,043)	-0,300*** (0,112)
Ratio de Mills	-0,368*** (0,127)	-0,655*** (0,140)	-0,654*** (0,135)	-0,695*** (0,267)	-1,007*** (0,180)	-0,430*** (0,139)	-1,253*** (0,476)
Número de observaciones	23.053	23.526	24.409	17.059	8.284	21.950	1.589

Nota: *** p<0.01, ** p<0.05, * p<0.1. Errores estándar entre paréntesis.

ANEXO- C Promedio de las variables explicativas para el modelo JMP

Variable	2007		2013		2007		2013	
	Hombres	Mujeres	Hombres	Mujeres	No indígenas	Indígenas	No indígenas	Indígenas
Indígena	0,06	0,06	0,06	0,06				
Mujer					0,51	0,51	0,51	0,51
Años de educación	4,27	4,24	4,07	4,21	4,25	3,52	4,13	3,72
Experiencia	20,80	21,36	21,35	21,94	21,16	19,88	21,70	20,75
Experiencia al cuadrado	743,71	740,92	748,80	748,02	741,30	758,41	747,83	759,20
Casado(a) / unido(a)	0,50	0,50	0,50	0,50	0,50	0,48	0,50	0,49
Jefe(a) de hogar	0,55	0,14	0,53	0,16	0,34	0,35	0,34	0,35
Migrante	0,13	0,14	0,15	0,16	0,14	0,09	0,15	0,10
Quito	0,12	0,13	0,12	0,12	0,13	0,09	0,13	0,04
Área rural	0,32	0,30	0,32	0,29	0,28	0,71	0,28	0,77
Sierra sur	0,09	0,10	0,09	0,10	0,10	0,05	0,10	0,06
Sierra centro	0,31	0,32	0,30	0,31	0,29	0,49	0,28	0,50
Sierra norte	0,25	0,25	0,23	0,23	0,25	0,31	0,23	0,26
Costa norte	0,04	0,03	0,06	0,06	0,04	0,01	0,06	0,02
Costa litoral	0,50	0,50	0,50	0,50	0,50	0,24	0,50	0,23
Amazonía	0,05	0,04	0,05	0,05	0,03	0,19	0,04	0,18

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural, diciembre 2007 y 2013.

ANEXO- D Regresiones de Mínimos Cuadrados Ordinarios de salarios

Variable dependiente: logaritmo del salario por hora

	2007		2013		2007		2007	
	Hombres	Mujeres	Hombres	Mujeres	No indígenas	Indígenas	No indígenas	Indígenas
Indígena	-0,054*** (-2,951)	-0,043 (-1,476)	-0,094*** (-4,900)	-0,174*** (-4,942)				
Mujer					-0,185*** (-18,820)	-0,236*** (-6,373)	-0,134*** (-15,460)	-0,290*** (-6,020)
Años de educación	0,082*** (60,896)	0,103*** (59,023)	0,073*** (60,912)	0,095*** (56,137)	0,091*** (82,152)	0,082*** (17,743)	0,081*** (79,824)	0,086*** (15,992)
Experiencia	0,026*** (19,553)	0,028*** (15,037)	0,018*** (14,342)	0,019*** (10,505)	0,027*** (23,912)	0,032*** (7,357)	0,019*** (18,405)	0,007 (1,027)
Experiencia al cuadrado	-0,000*** (-14,594)	-0,000*** (-8,900)	-0,000*** (-12,036)	-0,000*** (-7,190)	-0,000*** (-15,869)	-0,000*** (-6,723)	-0,000*** (-14,467)	-0,000 (-0,659)
Casado(a) / unido(a)	0,042*** (3,099)	0,064*** (4,009)	0,068*** (5,853)	0,058*** (3,783)	0,069*** (7,499)	-0,007 (-0,194)	0,082*** (10,520)	0,075 (1,506)
Jefe(a) de hogar	0,096*** (6,445)	0,031 (1,492)	0,092*** (7,288)	0,020 (1,144)	0,059*** (5,731)	0,058 (1,383)	0,066*** (7,345)	0,001 (0,010)
Migrante	0,089*** (6,453)	0,051** (2,559)	0,103*** (8,975)	0,063*** (3,581)	0,073*** (6,416)	0,098** (2,206)	0,088*** (9,022)	0,172** (1,967)
Quito	0,273*** (11,376)	0,233*** (7,790)	0,067* (1,761)	0,098* (1,783)	0,266*** (14,083)	0,165** (2,499)	0,091*** (2,860)	
Rural	-0,153*** (-14,283)	-0,122*** (-7,245)	-0,106*** (-11,596)	-0,101*** (-6,730)	-0,137*** (-14,593)	-0,134*** (-3,972)	-0,101*** (-12,840)	-0,133*** (-2,998)
Sierra sur	0,119*** (8,659)	0,100*** (5,462)	-0,001 (-0,075)	-0,071*** (-4,112)	0,111*** (9,784)	0,017 (0,301)	-0,033*** (-3,200)	0,060 (0,685)
Sierra norte	0,019 (1,380)	0,053*** (2,694)	0,084*** (6,557)	0,071*** (3,694)	0,019 (1,566)	0,120*** (3,571)	0,061*** (5,577)	0,294*** (7,724)
Constante	-1,093*** (-51,434)	-1,526*** (-50,547)	-0,680*** (-34,726)	-1,028*** (-32,247)	-1,187*** (-66,941)	-1,149*** (-17,172)	-0,763*** (-44,662)	-0,710*** (-7,830)
R ²	0,35	0,39	0,31	0,39	0,36	0,31	0,33	0,32
Número de observaciones	14.658	7.813	15.284	7.769	21.215	1.256	22.023	1.030

Nota: *** p<0.01, ** p<0.05, * p<0.1. Valores del estadístico t entre paréntesis, para los errores estándar se usó la técnica de bootstrap con 2000 repeticiones.

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2013.

ANEXO- E Brecha salarial por género y etnia por percentiles de la distribución para los años 2007 y 2013. Variable dependiente: logaritmo del salario por hora

<i>Género</i>							
Percentil	2007			2013			2013 - 2007
	Hombres	Mujeres	Diferencia	Hombres	Mujeres	Diferencia	
10	-0,956	-1,331	0,375	-0,581	-0,979	0,398	0,024
20	-0,616	-0,861	0,244	-0,224	-0,512	0,288	0,043
30	-0,391	-0,573	0,182	-0,042	-0,224	0,182	0,000
40	-0,207	-0,350	0,143	0,113	0,014	0,099	-0,044
50	-0,050	-0,155	0,105	0,246	0,221	0,025	-0,080
60	0,114	0,069	0,045	0,364	0,364	0,000	-0,045
70	0,343	0,354	-0,011	0,515	0,562	-0,047	-0,036
80	0,624	0,694	-0,070	0,757	0,855	-0,098	-0,027
90	1,085	1,085	0,000	1,162	1,246	-0,084	-0,084
90-10	2,041	2,416	-0,375	1,743	2,225	-0,482	-0,108

<i>Etnia</i>							
Percentil	2007			2013			2013 - 2007
	No indígenas	Indígenas	Diferencia	No indígenas	Indígenas	Diferencia	
10	-1,043	-1,671	0,629	-0,666	-1,402	0,737	0,108
20	-0,637	-1,123	0,486	-0,278	-0,917	0,639	0,153
30	-0,425	-0,846	0,421	-0,070	-0,581	0,511	0,090
40	-0,232	-0,637	0,405	0,113	-0,307	0,420	0,014
50	-0,050	-0,398	0,349	0,246	-0,070	0,316	-0,033
60	0,128	-0,167	0,295	0,382	0,113	0,270	-0,026
70	0,371	0,015	0,356	0,558	0,292	0,266	-0,090
80	0,684	0,238	0,446	0,815	0,465	0,350	-0,096
90	1,106	0,546	0,559	1,223	0,732	0,491	-0,068
90-10	2,148	2,218	-0,069	1,889	2,134	-0,245	-0,176

Fuente: Instituto Nacional de Estadística y Censos. Encuesta Nacional de Empleo, y Subempleo Urbana y Rural 2007-2013.