



Universidad
Zaragoza

Trabajo Fin de Máster

*Factores determinantes de los salarios en
Colombia*

Autor

Edgar Silva Quintero

Director

José Ignacio Giménez Nadal

Profesor ayudante Doctor en análisis económico

Facultad de economía y empresa

2013

Índice

1. Introducción:	3
2. Estudios previos para Colombia.....	5
3. Los datos y descripción de las variables.....	6
4. Técnica de imputación de datos perdidos (missing), utilizando el metodo Hot-Deck.....	10
5. Técnicas econométricas. Metodología.....	13
6. Resultados	17
7. Conclusiones.....	23
Bibliografía	24

Factores determinantes de los salarios en Colombia

Resumen

El presente trabajo pretende identificar los factores asociados a los determinantes de los salarios en Colombia y su relación con las características personales de los trabajadores, el sector productivo en el que se encuentran y las ciudades en las cuales se efectúa la actividad económica. Igualmente, abordar la problemática que suscitan los datos faltantes en las encuestas de hogares, por medio de la utilización de técnicas de imputación como el Hot-Deck. Para la realización del estudio se utilizó la Gran Encuesta Integrada de Hogares de Colombia –GEIH- del año 2012, realizada por el Departamento Administrativo Nacional de Estadística, DANE. Dichas estimaciones se realizaron por separado para hombres y mujeres, analizando en primer lugar el comportamiento de las variables sin realizar la imputación y luego comparándola con los resultados obtenidos de las regresiones con variables imputadas. Se ratifica que el método Hot Deck permite reducir el sesgo de no respuesta, por tanto analiza de manera más completa la base de datos, su mayor ventaja es que imputa datos reales e incorpora información procedente de otras variables.

Palabras clave: Determinantes de los salarios, datos faltantes, Hot-Deck.

1. Introducción:

Es importante analizar por separado el comportamiento de los ingresos laborales entre hombres y mujeres ya que la diferencia salarial de ambos suscita un gran interés no solamente entre los investigadores del mercado de trabajo, sino también entre las otras áreas de las ciencias sociales ya que cada vez más ha aumentado la participación de la mujer en el mercado de trabajo; diferentes estudios tratan de explicar dicha conducta, destacando, los análisis sobre desigualdades de ingresos entre trabajadores con igual productividad e iguales funciones, solamente por el hecho de ser de distinto género, estos dos efectos fueron tratados por Blinder y Oaxaca (1973) y extendida por Oaxaca y Ransom (1994). Ellos identifican dos elementos principales: el primero se asocia a las diferencias en las características observables mientras que el segundo a la diferencia entre los coeficientes estimados y por tanto a la retribución marginal de cada una de estas características. Otros estudios han determinado diferencias regionales, (Rosen, 1986), en la cual encuentra, que el efecto de la variación de los salarios monetarios por regiones es el resultado de diferencias en el costo de la vida y en las condiciones no monetarias del trabajador.

A su vez otras investigaciones consideran el tamaño de la empresa y la temporalidad como factores importantes a la hora de medir una mayor retribución salarial, como son las diferencias compensatorias (Mellow, 1982, Faber, 1986), es decir las empresas con mayor número de trabajadores pagan salarios más elevados por contar con personal más cualificado o por compensar a sus trabajadores por ofrecer puestos de trabajo más desagradables con un ambiente laboral inadecuado como por ejemplo puesto de trabajo muy contaminantes así como también mayores costos de desplazamiento, entre otros.

Dickens y Katz (1987) y Genre et al., (2005), en las teorías no competitivas señalan que las diferencias salariales no sólo se ven afectadas por los atributos de los trabajadores, sino también por las características de la industria y de la empresa, que no necesariamente afectan a la utilidad de los empleados. Según estas teorías, las características específicas del sector, así como las

características de la empresa pueden contribuir a explicar los diferenciales salariales; Por ejemplo, Kruger y Summers (1986) encuentran que los salarios más altos tienden a ser pagados en industrias que están concentradas, que tienen utilidades altas y tienen una participación de mano de obra relativamente pequeña.

Este trabajo pretende determinar en una primera parte los factores asociados a los determinantes de los salarios en Colombia y su relación con las características personales de los trabajadores, el sector productivo en el que se encuentran y las regiones en las cuales se efectúa la actividad económica centrándonos en si esos determinantes entre hombres y mujeres pueden ocasionar brechas salariales. En una segunda parte, como novedad se aborda la problemática que suscita la presencia de datos faltantes en las encuestas de hogares, que repercuten en variables relevantes a la hora de explicar la diferencias salariales; evitando sesgos que pueden reducir el poder explicativo de los métodos estadísticos y que incluso conllevan a invalidar las conclusiones del estudio (Medina 2007), el metodo que utilizaremos es el procedimiento Hot-Deck, el cual se aplica para suplir información en grandes volúmenes de datos (censos y encuestas).

En general, el procedimiento Hot Deck es un proceso de duplicación: cuando un valor es faltante de una muestra, un valor registrado es duplicado para representar este valor faltante, esta metodología divide la base de datos en subgrupos utilizando variables correlacionadas; la razón principal para usar el procedimiento Hot Deck es que reduce el sesgo de no respuesta y su mayor ventaja es que imputa datos reales además puede incorporar información procedente de otras co variables, aunque también tiene una gran debilidad y es que requiere un buen emparejamiento de donante y receptores que reflejen la información de las co variables disponibles, (un buen emparejamiento se refiere a que será mejor si la muestra es grande que pequeña), otro inconveniente es que si hay escasez de donantes puede llevar al excesivo uso de un solo donante, por lo que muchas metodologías restringen el número de veces que un donante puede usarse para la imputación (Andridge y Little, 2010).

En esta investigación se utiliza la información suministrada por la Gran Encuesta Integrada de Hogares (GEIH) en el año 2012 para un trimestre (Octubre, Noviembre y Diciembre), la información que contiene esta base de datos y que utilizaremos para nuestro análisis es: características personales, en la cual encontramos variables sociodemográficas como la edad, el nivel educativo, estado civil y el genero, esta última nos permitirá realizar por separado las regresiones tanto para hombres como para mujeres. Para las variables del sector productivo tenemos: la experiencia que tiene el individuo dentro de la empresa, medida en meses, la profesión a la que se dedica y el número de personas que contiene la empresa , por último tenemos las Ciudades en las cuales desarrollan su labor.

Pretendo realizar una análisis sobre que factores determinan los salarios en Colombia realizando regresiones por separado entre Hombres y mujeres utilizando el metodo Mínimos Cuadrados Ordinarios MCO para corroborar si existe diferencia salarial; estos resultados finalmente se compararan con los obtenidos con regresiones de hombres y mujeres cuyas variables contengan datos perdidos y hallan sido imputadas a través de el metodo de imputación Hot Deck.

Los principales resultados encontrados arrojan que las variables analizadas influyen positiva o negativamente con respecto al ingreso laboral y actúan de diferente manera según sea hombre o mujer, confirmando así el peso que contienen las características personales dentro del análisis

como es el caso de la variable educación que en el caso de las mujeres es significativo al 90% para un nivel de educación superior o universidad, y para el nivel de educación secundario y medio es significativo al 95%. En el caso de los hombres aunque en la literatura se dice que el ingreso laboral crece a mayor nivel de educación, con las variables que intervienen en este estudio, no arrojan resultados significativos por tanto no podemos realizar una comparación de variables, esto se debe a que existe un alto grado de correlación entre la educación y la ocupación, de tal manera que la ocupación captura este efecto de la educación. Los retornos cuando se desagrega por niveles educativos indican que manteniendo constantes las demás variables de la regresión de mujeres por cada año adicional en el nivel educativo secundario-medio el ingreso laboral se incrementa en \$155000 pesos más que los individuos que solo han realizado ningún tipo de educación y en el caso del nivel educativo superior o universidad por cada año adicional aumenta en \$131200 pesos más que la variable base.

En cuanto al resto de variables se puede indicar que influyen positivamente la variable número de personas en la empresa corroborando así, que a mayor número de personas en la empresa el ingreso se incrementa tanto para hombres como para mujeres, observando en este sentido que en la mayoría de los casos es la mujer la que cobra más, en cuanto a las ciudades donde desarrollan su actividad, se analiza que para los hombres; Florencia, Montería, Cúcuta y Cali son las ciudades donde se obtienen menos ingresos y donde más se incrementa su ingreso son Bogotá, Neiva, Villavicencio y Bucaramanga. Para las mujeres como casos representativos a tener en cuenta las ciudades Ibagué y Cúcuta obtienen un menor retorno de ingresos y donde se incrementa, Bogotá y Medellín. Adicionalmente se ratifica que el método Hot Deck reduce el sesgo de no respuesta, permitiendo así analizar de manera más completa la base de datos, su mayor ventaja es que imputa datos reales e incorporar información procedente de otras variables.

El resto de trabajo de investigación se organiza de la siguiente manera:

La sección 2 reflexiona sobre los estudios que se han realizado en Colombia sobre este tema; en la sección 3 se comentara en detalle los datos utilizados en el trabajo, a su vez la sección 4 abordara la problemática que suscitan los valores missing por medio de la utilización de técnicas de imputación como el Hot-Deck, posteriormente se realizara un análisis de las estimaciones del modelo de ingresos laborales con datos perdidos (missing), con el modelo de ingresos laborales sin datos perdidos (sin missing), en cuanto a la sección 5 se dedicara a la descripción de las técnicas econométricas utilizadas (estrategia empírica), por último en las secciones 6 y 7 se presentan los resultados y las conclusiones más relevantes encontradas en esta investigación.

2. Estudios previos para Colombia.

En Colombia, las desigualdades salariales entre sectores económicos no han recibido suficiente atención en la literatura. Entre las excepciones, se encuentran los documentos recientes de Urrutia y Ruiz (2010), Mesa et al. (2008) y Gracia et al. (2001). Urrutia y Ruiz (2010) analizan la evolución del salario real promedio por ramas de actividad económica, utilizando información trimestral para el período 1980-2006. Los autores muestran que a lo largo de este período, los sectores de servicios financieros y de electricidad, gas y agua registraron los niveles más altos de los salarios reales y fueron los únicos que tuvieron un crecimiento real de los salarios. Los otros cinco sectores presentan poca dinámica salarial en términos reales.

Según Ribero et al (2005), no toda diferencia salarial es discriminatoria, ni toda discriminación se interpreta en diferencias salariales, la discriminación se da cuando dos personas con las mismas características y habilidades realizan las mismas funciones laborales pero son tratadas de distinta manera por el empresario, los demás empleados o por el consumidor final, dicha discriminación además se puede manifestar también de otras formas, como beneficios laborales, ascensos, etc.

Rojas-Hayes, (2006), examina la presencia de brechas salariales por raza en Colombia, según la información de la Encuesta del Nivel de Vida 2003 del país, señala que las minorías afrocolombianas, específicamente los hombres en edad de trabajar ganan un 6.4% de los salarios medios por hora menos que los de raza blanca. Para explicar este caso se estimaron ecuaciones de Mincer estándar y se obtuvo evidencia de discriminación cuando se incluyeron en el modelo variables como niveles de educación, edad, experiencia, ubicación geográfica, la informalidad y el tamaño de la familia.

La literatura existente apunta que existen diferenciales salariales por raza u origen étnico. Arocha, et al. (2002). Confirma que el 25% de los afro-colombianos entrevistados en la capital, Bogotá informan que hay discriminación en el lugar de trabajo, por otra parte González-Rivas, (2012) encuentra evidencia cualitativa de diferenciales salariales en los barrios más pobres de Santiago de Cali y es que la raza afro-colombiana reporta un trato discriminatorio a la hora de solicitar empleo.

La desigualdad salarial en Colombia se analiza a través de la descomposición de la varianza del logaritmo de los ingresos laborales utilizando la metodología de Shorrocks, (1982) y de la cual se apoya Núñez y Sánchez (1998), concluye que el diferencial educativo es el factor que más afecta la desigualdad salarial.

Santamaría, (2004), investiga la evolución de la distribución de ingresos laborales en Colombia, encuentra que los cambios en la oferta relativa de trabajo calificado, combinado con un crecimiento constante del sesgo tecnológico con respecto a los trabajadores cualificados explican el incremento de los salarios de los mismos, su conclusión es que el incremento de la oferta relativa de los trabajadores con mayor cualificación ha sido más rápida que las necesidades del país.

Sin embargo, se puede argumentar que la metodología utilizada ha dejado por fuera del análisis características importantes de la mano de obra y los mercados regionales que pueden afectar a la igualdad en los salarios. Una de esas características es el nivel de educación, que fue estudiado por Galvis (2004) para el período 1984-2000. Este último realizó un análisis de las siete principales áreas metropolitanas para estudiar la integración del mercado de trabajo, incluyendo los trabajadores calificados y no calificados y diferenciando por el nivel de educación.

3. Los datos y descripción de las variables.

Para la realización del estudio se utilizará la Gran Encuesta Integrada de Hogares de Colombia GEIH- del año 2012, realizada por el Departamento Administrativo Nacional de Estadística, DANE. Esta encuesta, desde el año 2006, integra la Encuesta de Hogares de DANE, la de ingresos y Gastos y la de Calidad de Vida. Actualmente la encuesta se ha especializado en la medición de la estructura del mercado laboral y los ingresos de los hogares, tiene una muestra total anual de 248.028 hogares aproximadamente, lo que hace que sea la de mayor cobertura nacional. La encuesta cuenta con 24 ciudades, 13 áreas metropolitanas, anualmente se visitan aproximadamente 248.028 hogares,

concentrados en 22548 segmentos. La muestra mensual es de 20669 hogares, 18790 viviendas y 1879 segmentos que la hace muy fiable y completa a la hora de construir modelos y por tanto realizar investigación

Las variables socioeconómicas más importantes de la operación estadística son de vivienda (tipo de vivienda, características físicas), registro de personas (identificación), características generales (sexo, edad, estado civil), seguridad social en salud, educación, fuerza de trabajo (ocupados, desocupados, independientes, tipo de empleo) e ingresos. La cobertura geográfica es nacional y permite obtener resultados por zona urbana y rural, grandes regiones y total por departamento. Se incluyen en el estudio las trece principales ciudades y áreas metropolitanas, además se han incluido 11 ciudades más a partir del año 2006.

Para la realización de este estudio se toma la base de datos características generales dentro de la cual están las variables a analizar como edad, genero estado civil y nivel educativo adicionalmente tenemos la base de datos ocupados donde se analizara el sector y su actividad económica y cuyas variables son: el tipo de contrato de trabajo ya sea verbal o escrito, número de trabajadores en la empresa, la profesión a la que se dedica, las horas mensuales trabajadas, la experiencia laboral en meses y por último la ciudad donde desarrolla su trabajo.

Una cuestión a tener en cuenta es que para la elaboración de este trabajo se ha restringido la muestra debido a que solo se toma a las personas que están trabajando en el momento de realizar la encuesta, es decir no se toma en cuenta los desocupados, pensionistas y a los menores cuya edad es menor a 14 años, esto debido a que en Colombia la edad permitida para los menores con permisos especiales es de los 14 años en adelante. Por tal motivo se acota la edad de los encuestados; mayores o iguales a 14 años hasta los 55 años para las mujeres y 60 años para los hombres que es la edad para pensionarse. Teniendo en cuenta esta consideración la muestra final resultante es de 52095 observaciones de las cuales 27943 son hombres y las restantes 24152 son mujeres.

En la tabla 1 se describen las variables a analizar, resaltar que la educación juega un papel muy importante a la hora de interpretar el ingreso laboral en Colombia, es aquí donde se observa que el 53% de los encuestados hombres tienen como mayor nivel educativo la educación secundaria y media muy parecido al de las mujeres con un 50% pero las mujeres destacan aun más en la educación superior o universidad con un 32% con respecto los hombres con un 28%.

Observamos que para el caso de la variable estado civil son más los hombres soltero con un 42% con respecto a las mujeres con una 34% y se da una mayor porcentaje de mujeres separadas o divorciadas con un 16% mas del doble que para el caso de los hombres con un 7%. En cuanto a la profesión vemos el mayor porcentaje en hombres y mujeres que laboran en empresas particulares con 69% y 70% respectivamente. En las demás variables no se aprecian resultados que indiquen un cambio de tendencia entre salarios de hombres y mujeres en una primera instancia.

Tabla 1. Resumen estadístico: Media y desviación estándar de las variables

	Originales		Imputadas	
	(1)	(2)	(3)	(4)
	Hombre	Mujer	Hombre	Mujer
Ingreso laboral	965.972 (1216.81)	951.532 (1306.62)	—	—
Ingreso laboral imp.	—	—	819.478 (1115.69)	818.539(1400.40)
Edad	33.34 (13.24)	33.68 (12.80)	33.39(13.24)	33.63 (12.76)
Edad2	1287.06 (953.70)	1298.46 (910.33)	1290.71 (953.54)	1293.81 (906.22)
Experiencia	56.65 (85.19)	56.78 (86.32)	75.27 (103.08)	75.54 (103.92)
Experiencia2	10467 .66 (29073.9)	10675.6 (30617.2)	16292.94 (41359.2)	16505.25 (42138.6)
Educación				
Ninguna	0.02 (0.14)	0.02 (0.15)	0.02 (0.15)	0.02 (0.15)
Básica Preescolar	0.0003 (0.18)	0.0003 (0.017)	0.0004 (0.02)	0.0002 (0.015)
Básica Primaria 1º a 5º	0.16 (0.36)	0.16 (0.36)	0.16 (0.37)	0.16 (0.37)
Básica secund. y media 6º a 13º	0.53 (0.49)	0.50 (0.50)	0.53 (0.050)	0.50 (0.50)
Superior o Universidad	0.28 (0.45)	0.32 (0.46)	0.28 (0.45)	0.32 (0.47)
Otros estudios	0.0003 (0.018)	0.0002 (0.015)	0.0003 (0.018)	.00001 (0.011)
Estado civil				
No esta casado(a) y vive en pareja hace menos de dos años	0.04 (0.18)	0.03 (0.18)	0.03 (0.18)	0.03 (0.18)
No esta casado(a) y vive en pareja hace más de dos años	0.26 (0.43)	0.25 (0.43)	0.26 (0.44)	0.25 (0.43)
Casado	0.20 (0.39)	0.19 (0.39)	0.20 (0.40)	0.19 (0.39)
Esta separado(a) o divorciado(a)	0.07 (0.27)	0.16 (0.37)	0.08 (0.27)	0.16 (0.37)
Esta viudo (a)	0.005 (0.07)	0.03 (0.15)	0.005 (0.07)	0.03 (0.16)
Soltero	0.42 (0.49)	0.34 (0.47)	0.42 (0.49)	0.34 (0.47)
Tipo de contrato				
Contrato verbal	0.38 (0.48)	0.37 (0.48)	0.37 (0.48)	0.37 (0.48)
Contrato escrito	0.61 (0.48)	0.62 (0.48)	0.63 (0.48)	0.63 (0.48)
No sabe, no informa	0.0004 (0.02)	0.0009 (0.03)	0.0006 (0.25)	0.0009 (0.03)
Contrato verbal imp.	—	—	0.37 (0.48)	0.37 (0.48)
Contrato escrito imp.	—	—	0.63 (0.48)	0.63 (0.48)
No sabe, no informa imp.	—	—	0.0006 (0.02)	0.0008 (0.03)
Profesión				
Obrero o empleado de empresa particular	0.69 (0.46)	0.70 (0.45)	0.37 (0.48)	0.37 (0.48)
Obrero o empleado del Gobierno	0.11 (0.31)	0.10 (0.30)	0.06 (0.23)	0.06 (0.23)
Empleado doméstico	0.08 (0.26)	0.07 (0.26)	0.04 (0.20)	0.04 (0.19)
Trabajador por cuenta propia	0.11 (0.30)	0.11 (0.31)	0.45 (0.49)	0.45 (0.49)
Patrón o empleador	0.001 (0.04)	0.002 (0.04)	0.04 (0.19)	0.04 (0.20)
Trab. Familiar sin remuneración	—	—	0.04 (0.19)	0.04 (0.18)
Trab. Sin remun. en empresas	—	—	0.004 (0.06)	0.003 (0.05)
Jornalero o peón	0.007 (0.08)	0.008 (0.09)	0.003 (0.06)	0.004 (0.06)
Otro	0.002 (0.03)	0.001 (0.03)	0.001 (0.03)	0.0009 (0.03)

Nº de pers. en la empresa				
Trabaja solo	0.10 (0.30)	0.10 (0.30)	0.37 (0.48)	0.37 (0.48)
2 a 3 personas	0.13 (0.33)	0.13 (0.34)	0.18 (0.39)	0.18 (0.39)
4 a 5 personas	0.08 (0.27)	0.08 (0.27)	0.07 (0.24)	0.07 (0.25)
6 a 10 personas	0.08 (0.28)	0.08 (0.27)	0.05(0.22)	0.05 (0.22)
11 a 19 personas	0.06 (0.22)	0.06 (0.23)	0.03 (0.17)	0.03 (0.18)
20 a 30 personas	0.05 (0.21)	0.05 (0.23)	0.03 (0.16)	0.03 (0.17)
31 a 50 personas	0.04 (0.20)	0.04 (0.19)	0.02 (0.14)	0.02 (0.14)
51 a 100 personas	0.04 (0.19)	0.04 (0.19)	0.02 (0.15)	0.02 (0.15)
101 o más personas	0.42 (0.49)	0.41 (0.49)	0.23 (0.42)	0.23 (0.41)
horas mensuales	193.8 (63.64)	193.9 (62.02)	185.8 (77.35)	185.4 (76.8)
Ciudades				
San Andrés	0.10 (0.29)	0.10 (0.30)	0.11 (0.31)	0.11 (0.32)
Medellín	0.08 (0.27)	0.078 (0.26)	0.06 (0.24)	0.06 (0.24)
Barranquilla	0.04 (0.20)	0.04 (0.20)	0.05 (0.22)	0.05 (0.22)
Bogota	0.07 (0.26)	0.08 (0.27)	0.06 (0.23)	0.06 (0.24)
Cartagena	0.02 (0.15)	0.02 (0.15)	0.04 (0.19)	0.04 (0.19)
Tunja	0.04 (0.20)	0.04 (0.20)	0.03 (0.17)	0.03 (0.17)
Manizales	0.05 (0.22)	0.05 (0.22)	0.04 (0.19)	0.04 (0.19)
Florencia	0.03 (0.17)	0.03 (0.16)	0.03 (0.16)	0.03 (0.16)
Popayán	0.03 (0.16)	0.03 (0.17)	0.03 (0.17)	0.03 (0.18)
Valledupar	0.03 (0.16)	0.03 (0.16)	0.04 (0.19)	0.04 (0.19)
Montería	0.04 (0.18)	0.03 (0.17)	0.04 (0.19)	0.04 (0.19)
Quibdo	0.03 (0.18)	0.03 (0.18)	0.03 (0.16)	0.028 (0.16)
Neiva	0.04 (0.20)	0.04 (0.20)	0.04 (0.20)	0.04 (0.20)
Riohacha	0.03 (0.18)	0.03 (0.18)	0.04 (0.19)	0.04 (0.19)
Santa Martha	0.03 (0.18)	0.03 (0.17)	0.04 (0.20)	0.04 (0.20)
Villavicencio	0.04 (0.19)	0.04 (0.19)	0.04 (0.19)	0.04 (0.19)
Pasto	0.03 (0.18)	0.03 (0.19)	0.03 (0.18)	0.03 (0.18)
Cúcuta	0.02 (0.15)	0.03 (0.15)	0.03 (0.18)	0.03 (0.18)
Armenia	0.03 (0.16)	0.03 (0.16)	0.04 (0.18)	0.03 (0.18)
Pereira	0.04 (0.19)	0.04 (0.20)	0.03 (0.18)	0.03 (0.18)
Bucaramanga	0.04 (0.20)	0.05 (0.21)	0.04 (0.20)	0.04 (0.20)
Sincelejo	0.03 (0.17)	0.04 (0.19)	0.04 (0.20)	0.04 (0.20)
Ibagué	0.04 (0.19)	0.037 (0.19)	0.04 (0.19)	0.04 (0.19)
Cali	0.04 (0.21)	0.04 (0.20)	0.04 (0.20)	0.04 (0.20)
Nº Observaciones	12.204	14.050	24.152	27.943

4. Técnica de imputación de datos perdidos (missing), utilizando el metodo Hot-Deck.

En las encuesta casi siempre se topa con dificultades para recopilar la información, el caso del ingreso laboral cuya importancia para realizar análisis económicos es esencial, no es la excepción.

Los datos faltantes (missing) originados por la falta de respuesta total o parcial en las encuestas de hogares puede generar dificultades en los procesos de análisis, ya que están afectadas por errores de muestreo, este inconveniente se debe en muchas ocasiones por haber encuestado sólo a una parte de las personas que habitan la vivienda, o también a la no respuesta de toda la familiar es decir cuando el hogar no desea participar en la encuesta o no puede ser localizada para tal fin y a las repuestas incoherentes entre otras; los diferentes análisis que se han realizado a lo largo de la historia de las encuestas muestran una muy buena calidad de la información pero que deja en muchas ocasiones diferencias considerables si la encuesta estuviera al cien por ciento desarrollada. Las estimaciones de una encuesta por muestreo varían debido al error muestral ya que se pueden estimar más o menos personas que habitan una vivienda, más o menos personas ocupadas o más o menos personas inactivas, etc. Dependiendo de la composición de los hogares y según el tamaño muestral de la encuesta en estudio estas variaciones tienen mayor o menor impacto en las estimaciones. En el caso de valores faltantes por la no respuesta puede deberse a:

- Rechazo a responder a una pregunta de la encuesta, bien sea por falta de comprensión de la pregunta, por desconocimiento de la misma o la renuncia a revelar la información (Barceló, 2008; Haziza, 2009).
- Ausencias de alguno de los miembros del hogar.
- Valores incoherentes que dan origen a valores Missing.
- Dificultad para el acceso a las viviendas.
- Desconocer la respuesta o no recordar la respuesta exacta.

Por tanto no siempre, la no respuesta se debe a la falta de cooperación de los encuestados sino a también a causas externas que hacen imposible recopilar toda la información. En la Gran encuesta de hogares la variable ingresos es generalmente la que presenta mayores dificultades a la hora de recopilar la información debido a que los personas con ingresos altos son reservados a la hora de dar este tipo de información y dicha variable queda sesgada. Esto ocasiona en la variable asimetría por la presencia de valores extremos, ya que muchos valores se concentran en la cola izquierda (valores bajos) y pocos valores extremos en la derecha. Pero ¿que genero la no respuesta? Para averiguarlo podemos identificar tres mecanismos por los cuales se generan estos datos faltantes.

a) Completamente aleatorio (Missing Completely At Random, MCAR).

Si la probabilidad de un elemento faltante es independiente de los datos observados como los no observados, en este caso es posible hacer inferencia correcta acerca de los parámetros poblacionales considerando solamente una submuestras de los que respondieron.

b) El aleatorio (Missing At Random, MAR).

Aquí y en la mayoría de los casos prácticos de métodos de imputación se usa un supuesto débil sobre el patrón de datos faltantes (Nicoletti y peracchi, 2006), el cual dice que la probabilidad de que un dato falte depende solamente de los datos observados.

c) El no aleatorio (Missing not At Random, MNAR).

Dice que la probabilidad de que un elemento sea faltante depende del valor no observado de los elementos faltantes. Como inconveniente los datos tienen poca información que nos permitan como investigadores saber si los datos faltantes son MCAR, MAR o MNAR y porque unos faltan y otros no. Solo es posible realizar pruebas para verificar el supuesto MCAR, aunque son deficientes, por otro lado los supuestos MAR y MNAR, sin información adicional no, ya que en el caso de MNAR depende de datos no observados.

Durante las últimas décadas se han propuesto distintas metodologías para sustituir datos faltantes; la mayoría de estos métodos pueden acarrear varias variantes que pueden ser de tipo aleatorio, no aleatorio, etc. Sin embargo, es frecuente que estos procedimientos se apliquen sin tener en cuenta sus fundamentos teóricos y sus limitaciones prácticas, entre los cuales están:

- **Eliminación de casos.**

La técnica más empleada que se utiliza en diferentes estudios es la eliminación de casos, la cual es la más sencilla para solucionar este tipo de problema, pero esta técnica puede provocar errores de estimación ya que podemos eliminar información de encuestados que han respondido diferente de encuestados que sí lo han hecho, esta técnica excluye un caso completo cuando al menos una variable tiene un dato faltante, su ventaja es que es de fácil aplicación, pero hay que tener cuidado ya que la eliminación de casos incompletos puede reducir el tamaño de la matriz de datos de forma drástica si el número de datos es elevado ya que ocasionaría que la muestra aumentará los errores estándar, reduciendo así el nivel de significancia de los estadísticos calculados, sesgando los resultados (Acock, 2005; Olin ski et al., 2003).

- **la imputación por la media.**

Consiste en utilizar la media muestral de los valores disponibles como donante en cada uno de los valores perdidos, es el método más simple pero el menos atractivo de los métodos de imputación, su ventaja es que proporciona estimaciones insesgadas para la media poblacional pero su desventaja es que distorsiona la distribución de los datos debido a la concentración de valores en torno a la media, (valor constante), casualmente los datos que están en los extremos de la distribución son los que no responden, es decir a los de ingresos altos, subestimando así, el valor de la varianza (Olin ski et al., 2003).

- **El método de máxima verosimilitud.**

La idea fundamental de este método es tomar como estimación del parámetro m estudiado el valor que haga máxima la probabilidad de obtener la muestra observada. En este tipo de método se supone que los datos completos siguen un determinado modelo multivariante. Por tanto es

importante elegir un modelo que sea suficientemente flexible para reflejar las características de los datos estudiados.

- **Imputación múltiple.**

Método propuesto por Rubin en 1987, consiste en asignar a cada valor faltante varios valores (m), generando m conjuntos de datos completos.

En cada conjunto de datos completo se estiman los parámetros de interés y posteriormente se combinan los resultados obtenidos, el objetivo de la imputación múltiple es hacer un uso eficiente de los datos que se han recogido, obtener estimadores no sesgados y reflejar adecuadamente la incertidumbre que la no respuesta parcial introduce en la estimación de los parámetros.

- **Imputación Hot Deck.**

En general, el procedimiento Hot Deck es un proceso de duplicación: cuando un valor es faltante de una muestra, un valor registrado es duplicado para representar este valor faltante, esta metodología divide la base de datos en subgrupos utilizando variables correlacionadas. Los valores faltantes se sustituyen con la información de un registro con información similar en las co variables. Los donantes y receptores de los datos pertenecen al mismo subgrupo.

La razón principal para usar el procedimiento Hot Deck es que reduce el sesgo de no respuesta. Para reducir este sesgo, el procedimiento Hot Deck por lo general tiene un proceso de clasificación asociada a ella. Todas las unidades de la muestra están clasificadas en grupos disjuntos así que las unidades son tan homogéneas como sea posible dentro de cada grupo. Para cada valor faltante, un valor registrado es imputado el cual está en el mismo grupo de clasificación. Así la suposición se basa en que dentro de cada grupo de clasificación las unidades que no responden siguen la misma distribución como aquellos que responden (Ávila, 2002).

Una de las mayores ventajas de este método es que imputa datos reales y por tanto, realistas y además puede incorporar información procedente de otras co variables, aunque también tiene una gran debilidad y es que requiere un buen emparejamiento de donante y receptores que reflejen la información de las co variables disponibles, (un buen emparejamiento se refiere a que será mejor si la muestra es grande que pequeña), otro inconveniente es que si hay escasez de donantes puede llevar al excesivo uso de un solo donante, por lo que muchas metodologías restringen el número de veces que un donante puede usarse para la imputación (Andridge y Little, 2010).

Aunque algunos autores señalan que algunos métodos de imputación presentan deficiencia como por ejemplo la imputación de la media y la eliminación de casos, no son adecuados ya que no preservan la distribución de los datos (distribución de datos observados y la distribución de datos faltantes) reduciendo el tamaño de la muestra, (Sande (1982), (Barceló, 2008) , a su vez el conjunto de datos es susceptible a posibles sesgos si las personas que no respondieron son sistemáticamente diferentes de las que sí lo hicieron, (Horton y Lipsitz, 2001) y Olin ski et al, 2003).

El desarrollo de métodos estadísticos para resolver este problema es un área de investigación en constante movimiento en los últimos años (Horton y Lipsitz, 2001).

Con la imputación de datos faltantes surge una alternativa para no perder esta información, preservando las características de la distribución de los datos y la relación entre las variables.

En este trabajo se evaluarán los datos con missing y posteriormente sin missing, asumiendo que los datos siguen el patrón MNAR y que los individuos con mayores ingresos son contrarios a relevar esta información.

5. Técnicas econométricas. Metodología.

La metodología empírica más utilizada en la mayoría de trabajos de investigación se realiza por medio de una ecuación minceriana de ingresos (Mincer 1974) a través de la cual se estima el impacto de un año adicional de estudios en los ingresos laborales. En un modelo de regresión lineal simple, se analiza la influencia de una variable explicativa χ en los valores que toma otra variable denominada dependiente (ω). En la regresión lineal múltiple vamos a utilizar más de una variable explicativa; esto nos va a ofrecer la ventaja de utilizar más información en la construcción del modelo y consecuentemente, realizar estimaciones más precisas.

Modelo de regresión lineal múltiple:

$$\omega_i = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3 + \beta_4\chi_4 + \beta_5\chi_5 + \dots + \beta_k\chi_k + \mu$$

Donde

ω_i Representa el salario mensual de un trabajador i .

β_0 Es el intercepto o parte fija del salario del trabajador.

$\chi_1, \chi_2, \chi_3, \chi_4, \chi_5, \dots, \chi_k$ Son la combinación lineal de los valores de variables explicativas.

μ Es el efecto aleatorio de los salarios asociado a las características individuales del trabajador.

Con la intención de explicar las diferencias salariales en Colombia se plantea un método en dos fases. La primera fase comienza eligiendo las variables que considero van a interactuar e influir en los ingresos laborales y que recopilé de la base de la Gran Encuesta Integrada de Hogares, GEIH 2012 para un trimestre (Octubre, Noviembre y Diciembre), en dicha encuesta las bases de datos que tomo son características generales, dentro de esta base se encuentran las variables edad, género, nivel educativo, estado civil y la base de datos ocupados donde se encuentran las variables contrato de trabajo, número de trabajadores en la empresa, profesión, horas mensuales trabajadas y la experiencia laboral en meses y por último la ciudad donde desarrolla su trabajo; que uno a través del programa econométrico SPSS, una vez obtenida la base de datos, realizamos una selección de los mismos, reducimos la base de datos original de 80484 a 52095 encuestados, de los cuales 27943 son hombres y las restantes 24152 son mujeres, por medio de la variable edad que contiene datos desde 0 hasta 101 años, para nuestro trabajo tomamos la edad de encuestados desde 14 años, que en Colombia es la edad mínima para trabajar con un permiso especial de los padres hasta los 55 años para las mujeres y 60 años para los hombres que es la edad de jubilación, a partir de este programa generamos un archivo con extensión dta para poder trabajar esta nueva base de datos en el programa econométrico STATA, posteriormente de las variables elegidas se

comprueba cuales tienen datos perdidos (missing) concluyendo que ingreso laboral y contrato de trabajo contienen datos perdidos, es importante tener en cuenta estas variables ya que serán utilizadas más adelante para realizar la imputación Hot Deck.

El primer paso de este método consiste en estimar ecuaciones de salarios para hombres y mujeres por separado a través de mínimos cuadrados ordinarios (MCO), además tenemos varios supuestos que acompañan al presente modelo:

1. Las tasas de retornos son semejantes para todos los individuos; es decir, los beneficios que obtenga un individuo de un año adicional de educación o experiencia no se verá afectada por las cualidades de este.

2. Que los ingresos aumentan con la experiencia laboral, pero a una tasa decreciente.

Con el objeto de conocer cómo retribuye el mercado a cada una de las características mencionadas anteriormente cuando las controlamos por la variable dependiente ingresos laborales. Las ecuaciones son las siguientes:

Ecuación 1 (Regresión para hombres)

$$\begin{aligned} \text{Inglabomiles}_{Ht} = & \beta_{H0} + \beta_{H1} * \text{edad}_{Ht} + \beta_{H1} * \text{edad}^2_{Ht} + \beta_{H1} * \text{genero}_{Ht} + \beta_{H1} * \\ & \text{educación}_{Ht} + \beta_{H1} * \text{estadocivil}_{Ht} + \beta_{H1} * \text{tipocontrato}_{Ht} + \beta_{H1} * N_{\text{pers}_{\text{empresa}}}_{Ht} \\ & + \beta_{H1} * \text{profesión}_{Ht} + \beta_{H1} * \text{horas}_{\text{mensuales}}_{Ht} + \beta_{H1} * \text{experiencia}_{\text{meses}}_{Ht} \\ & + \beta_{H1} * \text{experiencia}^2_{Ht} + \beta_{H1} * \text{ciudad}_t + \text{Fex}_{c_{2011}} + \mu_{Ht} \end{aligned}$$

Ecuación 2 (Regresión para mujeres)

$$\begin{aligned} \text{Inglabomiles}_{Mt} = & \beta_{M0} + \beta_{M1} * \text{edad}_{Mt} + \beta_{M1} * \text{edad}^2_{Mt} + \beta_{M1} * \text{genero}_{Mt} + \beta_{M1} * \\ & \text{educación}_{Mt} + \beta_{M1} * \text{estadocivil}_{Mt} + \beta_{M1} * \text{tipocontrato}_{Mt} + \beta_{M1} * N_{\text{pers}_{\text{empresa}}}_{Mt} \\ & + \beta_{M1} * \text{profesión}_{Mt} + \beta_{M1} * \text{horas}_{\text{mensuales}}_{Mt} + \beta_{M1} * \text{experiencia}_{\text{meses}}_{Mt} \\ & + \beta_{M1} * \text{experiencia}^2_{Mt} + \beta_{M1} * \text{ciudad}_t + \text{Fex}_{c_{2011}} + \mu_{Mt} \end{aligned}$$

Descripción de las variables.

Inglabomiles, H_t , M_t = Hace referencia al ingreso laboral del hombre y la mujer t en miles de pesos colombianos.

χ_t Representa un vector de variables como características generales que considero relevantes para explicar las diferencias salariales (edad, edad al cuadrado, género, educación, estado civil, tipo de contrato, número de personas en la empresa, profesión, horas mensuales, experiencia, experiencia al cuadrado y ciudad.

La variable "experiencia" es muy difícil de observar directamente, por lo que puede elegirse una de las siguientes alternativas: tomar el tiempo que ha transcurrido desde que el individuo dejó de estudiar, o tomar el tiempo que el individuo lleva en el trabajo actual, en este caso tomaremos la segunda. Se incluye el cuadrado de la experiencia para introducir la idea de que el perfil de ingresos del individuo tiene forma cóncava; es decir, pasado un cierto límite, un año adicional de edad

Disminuye los ingresos en vez de aumentarlos. El efecto total se halla mediante el cálculo que incluye los coeficientes de la experiencia y de la experiencia al cuadrado.

β_{Ht}, β_{Mt} Son las tasas de retorno de dichas características tanto de hombre t como de la mujer t .

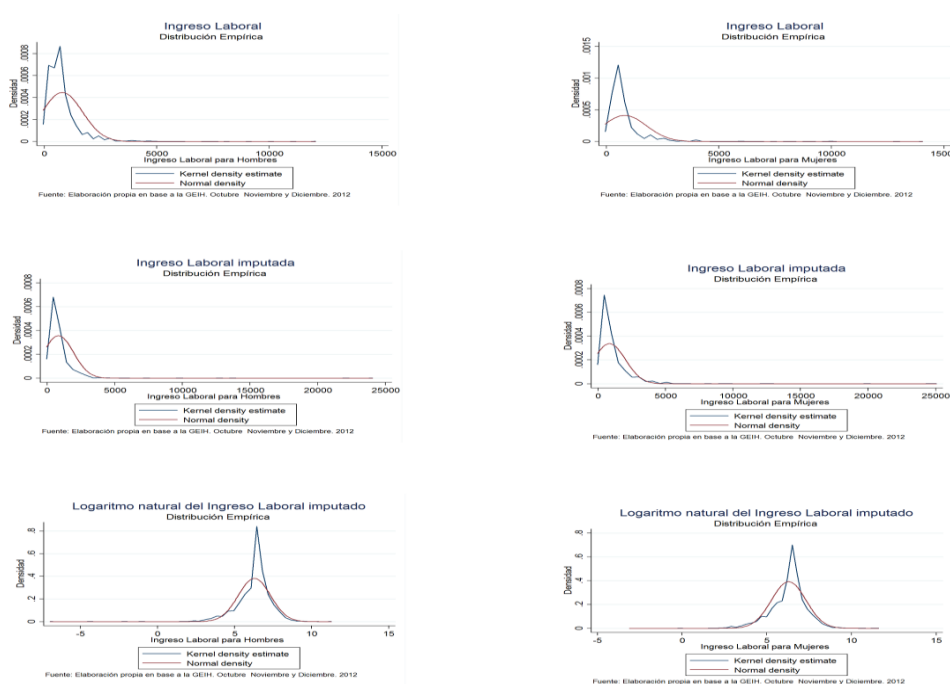
μ_{Ht}, μ_{Mt} Es el correspondiente término error tanto de hombres t como de mujeres t .

Fex_c_2011 Es el factor expansión, son los pesos originales incluidos en la encuesta, estos efectos fijos muestran que dos trabajadores con las mismas dotaciones de capital humano perciben salarios diferentes bien porque se encuadran en diferente sector dentro de la misma ciudad, bien porque desarrollan su actividad en el mismo sector pero en ciudades diferentes o por el efecto conjunto de factores sectoriales/ciudades.

En la segunda fase se realiza una regresión, de nuevo imputando la variable ingreso laboral $Inglabomiles$ y la variable dicotómica $tipocontrato$ a través del metodo Hot Deck, con esta regresión se trata de mejorar las estimaciones de los factores que determinan las diferencias salariales.

En la grafica 1 realizamos la función de densidad de la variable dependiente ingreso laboral para hombres y mujeres por separado arrojando como resultado que el ingreso laboral no sigue una distribución normal agregando además mucha kurtosis, tanto para la variable dependiente original como para la imputada, por tanto aplicamos logaritmo natural a la variable para intentar solucionar esta kurtosis y al no poder normalizar la variable optamos por realizar el análisis aplicando la opción “robust” en las estimaciones.

Grafica 1. Función de densidad de la variable ingreso laboral.



Adicionalmente realizamos un contraste tipo t de las variables originales e imputadas. Ver tabla 2 ; tipo de contrato y ingreso laboral, para comparar las medias de dichas variables por separado tanto

para hombres como para mujeres, su resultado nos indica que las medias de ambas variables son iguales ya que $H_a:diff$ es mayor de 0.05 , por tanto nos esta indicando que su distribución no cambia.

Tabla 2

Contraste tipo t de las variables originales y imputadas				
Variables	Obs.	Media	Desv. estándar	$H_a:diff \neq 0$
Inglabomiles				
Hombre				
X	24152	819.478	1.115.693	
Y	21831	818.375	1.123.335	
$H_a:diff \neq 0$				0.916
Mujer				
X	27943	818.539	1.400.395	
Y	25235	819.608	1.435.819	
$H_a:diff \neq 0$				0.9308
Contrato verbal				
Hombre				
X	24152	0.37301	0.48361	
Y	12204	0.38257	0.48603	
$H_a:diff \neq 0$				0.0754
Contrato escrito				
X	24152	0.62636	0.48377	
Y	12204	0.61701	0.48613	
$H_a:diff \neq 0$				0.0821
No sabe, no contesta				
X	24152	0.000621	0.024914	
Y	12204	0.000409	0.020237	
$H_a:diff \neq 0$				0.4169
Contrato verbal				
Mujer				
X	27943	0.37272	0.48353	
Y	14050	0.37466	0.48405	
$H_a:diff \neq 0$				0.6983
Contrato escrito				
X	27943	0.62641	0.48376	
Y	14050	0.62448	0.48427	
$H_a:diff \neq 0$				0.6992
No sabe, no contesta				
X	27943	0.000858	0.29294	
Y	14050	0.000854	0.029213	
$H_a:diff \neq 0$				0.9873

6. Resultados

Las ecuaciones han sido estimadas utilizando información de corte transversal proveniente de la Gran Encuesta Integrada de Hogares (GEIH), aplicada para Colombia para el trimestre Octubre, Noviembre y Diciembre de 2012.

Posteriormente se determinó si el modelo es robusto y consistente; el cual a su vez nos permitió determinar que variables en estudio presentan mayores retornos respecto a la dispersión salarial.

Criterio Económico

Observamos que los resultados de la estimación cumplen con los criterios expuestos por la teoría económica neoclásica, del capital humano. Los coeficientes tienen el signo esperado y están en los rangos de otras estimaciones.

Criterio Estadístico

De las estimaciones anteriores apreciamos que las variables exógenas que mejor explica el comportamiento de los ingresos son la experiencia y la educación con respecto a las mujeres tanto en el modelo sin imputar como en el modelo imputado, presenta un R cuadrado de 0.1032 que explican conjuntamente el modelo sin imputar y un R cuadrado de 0.083 del modelo imputado. En cuanto a el modelo para los hombres, la experiencia y la profesión explican satisfactoriamente el modelo e incluso mejorando el R cuadrado pasando de 0.138 a un 0.15, lo que demuestra una buena capacidad explicativa conjunta de las variables introducidas en el modelo.

Antes de la imputación:

Analizando las estimaciones entre hombres y mujeres, comenzando por los rendimientos de las variables relativas a las características individuales, destacaremos los siguientes resultados, recordando que en todas las regresiones la variable dependiente es el ingreso laboral en miles de pesos colombianos.

La tabla 3 muestran los resultados de la estimación de la ecuación 1 y 2 para hombres y mujeres respectivamente con respecto a las variables edad, edad², experiencia, experiencia² y horas mensuales trabajadas; a su vez presentamos seis grupos de variables ficticias Dummies para el análisis econométrico, tenemos: educación, estado civil, número de personas en la empresa, profesión a la que se dedican, la ciudad en que labora y tipo de contrato.

En primer lugar se analizara la variable educación, antes hay que recordar que el grupo base de la educación es la variable educación₁ (Ninguno), de tal manera que las dummies de este grupo mide la diferencia proporcional del salario con relación a los individuos que no tienen ninguna clase de educación. Dicho esto se observa que el parámetro de la educación secundaria y media para la mujer es significativo individualmente al 95% para esta variable ya que el p-valor es menor que 0.05, mientras que la educación superior o universidad es significativa al 90%. Los resultados básicos de estas estimaciones señalan un aumento de desigualdad de salarios entre trabajadores con distinto nivel educativo, los cuales presentan una asociación positiva, es decir a mayor

educación mayor nivel de ingresos que luego al llegar al nivel educativo superior o universidad aunque es positiva comienza a decrecer.

En el caso de los hombres aunque en la literatura se dice que el ingreso laboral crece a mayor nivel de educación, con las variables que intervienen en este estudio, no arrojan resultados significativos por tanto no podemos realizar una comparación de variables.

Los retornos cuando se desagrega por niveles educativos indican que manteniendo constantes las demás variables de la regresión, para el caso de las mujeres por cada año adicional en el nivel educativo secundario y medio el ingreso laboral se incrementa en \$155000 pesos más que los individuos que no tienen ninguna educación y en el caso del nivel educativo superior o universidad por cada año adicional aumenta en \$131200 pesos con respecto a la variable base.

Para interpretar los coeficientes de la variable estado civil que también se presenta en grupo de variables ficticias dummies, hay que recordar que el grupo base es hombres solteros por tanto las demás miden la diferencia proporcional entre el ingreso laboral con relación a los hombres y mujeres solteros en sus respectivas regresiones, observando los resultados de la tabla sin imputar podremos decir que manteniendo constantes las demás variables; los hombres que no están casado pero que viven con su pareja hace menos de dos años (estadocivil1) ven disminuido su ingreso laboral en \$148300 pesos con respecto a los hombres soltero.

Para el caso de las mujeres viudas por cada año adicional con respecto al las mujeres solteras, su ingreso se reduce en \$135.300 pesos.

Con respecto a la cualificación requerida para el puesto (variable experiencia) se observa que tanto para hombres como para mujeres la cualificación aumenta el salario considerablemente siendo mayor en la mujer, por tanto por cada mes adicional de experiencia en el trabajo por parte de los hombres el salario se incrementa \$3365 pesos mientras que por cada mes adicional de experiencia por parte de la mujer el salario se incrementa \$4364 pesos.

Se observa que la experiencia al cuadrado el coeficiente es negativo tenemos que la función del ingreso laboral es cóncava con respecto a la experiencia en ambos casos (hombres y mujeres), es decir los salario crecen llegan a un máximo y luego decrecen.

En cuanto a la variable profesión, en este grupo tomamos como referencia la variable empleado domestico, encontramos que todos los coeficientes son estadísticamente significativos en los hombres salvo la profesión5 (patrón o empleador) que no es significativa y resaltando el caso de la profesión obrero o empleado del gobierno (profesión2) para las mujeres ya que su retorno es de \$447700 pesos en comparación con los hombres con \$380200 pesos, esto podría estar demostrando un cambio de tendencia en Colombia en cuanto a la discriminación por genero ya que se supone que con los mismo estudios las mujeres generarían un ingreso menor con respecto a los hombres.

En lo relativo al tamaño de la empresa, se observa que los rendimientos son crecientes en el tamaño tanto para hombres como para mujeres, siendo los salarios notablemente mayores en las empresas de mayor tamaño confirmando la teoría acerca de esta variable de que mayor número de empleados mejora considerablemente el ingreso con la particularidad de que son superiores en las mujeres.

En el caso de los hombres empleados en empresas de entre 101 o más trabajadores el salario es superior al de grupo de referencia N_pers_empresa1, trabaja solo con \$525200 pesos inferior al de las mujeres con \$657800 pesos.

Finalmente, se contrasta que a mayor tamaño de la empresa mayor es la remuneración por efecto de economías de escala.

En cuanto a la variable horas mensuales se observa que es significativa tanto para hombres como para mujeres, en la cual por cada hora adicional de trabajo por parte de los hombres el ingreso se incrementa en \$804 pesos mientras que en la mujer es superior con \$1057 pesos.

Finalmente la variable ciudad que representa a las 24 ciudades de Colombia, observamos que en cuanto a los hombres hay 13 que son significativas, todas positivamente con respecto al ingreso laboral, siendo relevante señalar que la ciudades con menor ingreso son Montería y Cúcuta con \$99430 y \$85930 pesos respectivamente, mientras que las de mayores ingresos son Bogota y Villavicencio con \$318400 y \$344700 pesos respectivamente.

Para el caso de las mujeres hay 10 ciudades que son significativas todas positivamente con respecto al ingreso laboral, como casos representativos a tener en cuenta Ibagué y Cúcuta que obtienen un menor retorno con \$129100 y \$109700 pesos respectivamente y con mayor retorno de ingresos Bogota y Medellín con \$371600 y \$291400 pesos respectivamente.

Como caso particular se observa que la ciudad de Popayán arroja unas estimaciones negativas con respecto a las mujeres, ya que trabajar en esta ciudad disminuye el salario en \$82300 con respecto a la variable de control.

Señalar que la ciudad de Cúcuta presentan los ingresos más bajos tanto para hombres como para mujeres y confirmando que las grandes ciudades como la capital Bogota presenta retornos más altos para los géneros.

Después de la imputación

Conociendo que en las encuestas a nivel general se encuentran con el problema de datos perdidos, esta trabajo analiza las variables que dentro de nuestra investigación contienen este tipo de datos faltantes, en este sentido la variable dependiente Ingreso Laboral (Inglabomiles) cuenta con 52095 datos de los cuales 5029 son datos perdidos y la variable tipo de contrato contiene en total 52095 datos de los cuales 24616 son datos perdidos; al realizarles imputación Hot Deck llegan a mejorar el modelo, generando más observaciones, y lo más importante aumentando el nivel de significancia de las variables e incluso variables con no son significativas al aumentar el número de datos se convierten en significativas y nos permiten poder analizarlas.

Si comparamos los resultados en la tabla2 de las regresiones 1 y 2 para hombres y mujeres sin imputar con las regresiones 3 y 4 ya imputadas de las variables Ingreso laboral y tipo de contrato a través del metodo hotdeck, observamos que al obtener más observaciones, el nivel educativo secundario - medio y el nivel educativo superior o universidad continúan siendo significativas incluso mejorando el nivel de significancia de la variable educación superior o universidad y manteniendo el de educación secundaria y media para el caso de las mujeres ; por otra parte continua siendo no significativas las de los hombres.

En cuanto al estado civil al obtener los resultados imputados continua siendo significativa la variable estadocivil1 para los hombres y la variable estadocivil5 (viuda) en el caso de las mujeres pasa de no ser significativa a serlo, es decir que manteniendo constantes las demás variables; los hombres que no están casado pero que viven con su pareja hace menos de dos años ven disminuir su ingreso laboral en \$71960 pesos con respecto a los hombres soltero.

Al realizar la imputación la variable experiencia casi no presenta cambio con respecto al escenario anterior ya que sigue siendo la mujer la que cobra más por su experiencia en el trabajo, \$3805 mientras que los hombres \$2460 pesos.

Para la variable profesión al realizar la imputación la profesión5 (patrón o empleador) pasa de no ser significativa a serlo en el caso de los hombres, confirmando la aleatoriedad del metodo hot deck y observando que el retorno a puestos de trabajo continua siendo mayor en las mujeres con respecto a los hombres, profesión obrero o empleado del gobierno (profesión2) ya que su retorno es de \$425300 pesos en comparación con los hombres con \$351300 pesos.

Finalmente la variable ciudad que representa a las 24 ciudades de Colombia, observamos que en cuanto a los hombres pasamos de 12 significativas, todas positivamente con respecto al ingreso laboral a 19, siendo relevantes señalar que la ciudades con menor ingreso son Barranquilla y Santa Martha con \$45920 y \$55500 pesos respectivamente, mientras que las de mayores ingresos siguen siendo Bogota y Villavicencio con \$294500 y \$347900 pesos respectivamente aunque ahora se le suma Bucaramanga con \$301000 pesos .

Finalmente al realizar la imputación con respecto a las ciudades y en el caso de mujeres pasamos de 10 significativas, todas positivamente con respecto al ingreso laboral a 17 ciudades significativas para el modelo, como casos representativos a tener en cuenta las ciudades Ibagué y Cúcuta siguen obteniendo un menor retorno con \$ y \$113000 y \$ 149200 pesos respectivamente pero ahora se suman con ingresos menores Pereira, Armenia y Riohacha con \$89710, \$80020 y \$100800 respectivamente y con mayor retorno de ingresos continúan siendo Bogota y Medellín con \$439900 y \$293500 pesos respectivamente.

Tabla 3. Regresiones MCO sobre el Ingreso Laboral

Variables	Variable originales				Variables Imputadas			
	(1) Hombre		(2) Mujer		(3) Hombre		(4) Mujer	
Edad	-3.019	(7.108)	21.38	(13.61)	-2.299	(4.677)	21.56*	(12.51)
Edad2	0.0116	(0.0942)	-0.284	(0.175)	0.0109	(0.0614)	-0.237*	(0.132)
Experiencia	3.365***	(0.696)	4.364***	(0.947)	2.460***	(0.319)	3.805***	(0.744)
experiencia2	-0.00341**	(0.00167)	-0.00490*	(0.00251)	-0.00389***	(0.000626)	-0.00604***	(0.00128)
Educación Preescolar	-104.1	(216.0)	-88.04	(100.4)	119.5	(153.8)	-188.8**	(86.60)
Educación Primaria	-33.45	(180.2)	108.6	(74.63)	70.33	(93.27)	86.62	(56.79)
Educación Secundaria _ media	-112.8	(171.1)	155.0**	(62.98)	3.590	(88.60)	135.6**	(60.15)
Educación Superior o universidad	-89.66	(174.7)	131.2*	(74.30)	36.04	(91.24)	164.5*	(85.16)
Educación No sabe, no informa	-134.0	(319.7)	-203.9	(238.3)	-232.1	(174.0)	-181.7	(198.2)
Estado civil. No esta casado (a) y vive en pareja hace menos de dos años	-148.3***	(54.69)	473.8	(428.9)	-71.96*	(41.34)	210.0	(234.8)
Estado civil. (No esta casado (a) y vive en pareja hace más de dos años	47.43	(50.25)	-26.33	(58.45)	52.07	(33.33)	-68.86	(97.94)
Estado civil casado (a)	30.28	(87.79)	-64.31	(73.70)	18.89	(51.05)	-166.8	(127.8)
Estado civil(separado (a) o divorciado (a)	3.267	(61.11)	13.67	(81.73)	8.226	(43.87)	-92.00	(115.1)
Estado civil. (viudo (a)	-8.517	(132.9)	-135.3*	(80.66)	-50.61	(88.58)	-235.1*	(135.7)
Obrero o empleado de empresa particular	-131.3**	(51.54)	-139.3**	(66.80)	-194.6***	(39.66)	-216.1***	(56.79)
Obrero o empleado de Gobierno	380.2***	(89.66)	447.7***	(169.5)	351.3***	(74.84)	425.3***	(160.8)
Trabajador por cuenta propia	142.6**	(62.51)	193.9***	(62.90)	-58.17**	(24.48)	-16.86	(33.49)
Empleador	816.6	(509.0)	1,446**	(654.1)	789.9***	(99.73)	1,270***	(393.2)
Trabajador familiar sin remuneración	—	—	—	—	154.2**	(63.45)	271.3**	(120.5)
Trabajador sin remuneración en la empresa	—	—	—	—	195.7	(161.8)	110.7	(150.5)
Jornalero o peón	-119.1*	(62.58)	-135.1*	(75.49)	-253.6***	(54.91)	-240.4***	(73.12)
Otro	-308.8***	(67.98)	-269.0***	(97.56)	-383.9***	(59.07)	-328.2***	(73.82)
Tipo contrato (escrito imp.)	—	—	—	—	78.05***	(19.55)	9.350	(82.20)
Tipo contrato (No sabe, no contesta imp.)	—	—	—	—	-86.11	(86.54)	-43.71	(167.5)
Nº personas en la empresa (2 a 3)	163.4***	(53.09)	183.1***	(61.42)	93.05***	(29.29)	76.89	(70.22)
Nº personas en la empresa (4 a 5)	216.8***	(59.17)	201.3***	(66.72)	275.7***	(44.80)	172.7**	(84.75)
Nº personas en la empresa (6 a 10)	218.2***	(57.31)	261.0***	(97.07)	396.4***	(60.93)	684.0**	(315.4)
Nº personas en la empresa (11 a 19)	236.2***	(64.03)	312.4***	(70.11)	490.0***	(72.82)	494.0***	(68.89)
Nº personas en la empresa (20 a 30)	313.1***	(98.91)	340.7***	(81.19)	506.4***	(96.44)	534.0***	(86.38)
Nº personas en la empresa (31 a 50)	366.7***	(89.54)	410.1***	(83.81)	557.6***	(81.77)	582.7***	(87.14)

Nº personas en la empresa (51 a 100)	289.8*** (68.88)	387.8*** (78.95)	505.8*** (54.57)	609.5*** (85.65)
Nº personas en la empresa (101 o más)	525.2*** (69.13)	657.8*** (89.59)	750.7*** (54.86)	873.4*** (98.58)
Horas meses	0.804*** (0.185)	1.057** (0.494)	0.901*** (0.114)	1.410*** (0.394)
Medellín	265.6*** (45.24)	291.4*** (81.48)	294.0*** (36.96)	293.5*** (55.59)
Barranquilla	21.80 (35.88)	-36.21 (40.28)	45.92* (27.05)	17.59 (28.48)
Bogota	318.4*** (50.04)	371.6*** (70.93)	294.5*** (35.38)	439.9*** (79.33)
Cartagena	57.66 (84.88)	66.59 (81.41)	91.87** (37.04)	105.3** (45.22)
Tunja	238.3*** (65.36)	56.28 (44.14)	266.5*** (48.95)	125.4** (49.12)
Manizales	45.31 (34.65)	29.82 (44.03)	82.96*** (30.69)	106.7*** (36.55)
Florencia	129.9* (71.10)	62.64 (52.86)	179.4*** (48.35)	226.6*** (74.59)
Popayán	-10.85 (44.76)	-82.30* (49.48)	33.41 (31.05)	-9.815 (34.23)
Valledupar	312.9* (190.1)	163.4*** (63.03)	232.5*** (73.50)	157.4*** (35.97)
Montería	99.43** (48.68)	46.15 (39.12)	91.36* (53.10)	52.86 (56.98)
Quibdo	-39.01 (46.84)	-51.35 (48.15)	42.83 (36.99)	17.33 (33.49)
Neiva	269.4*** (57.10)	195.8*** (58.32)	257.6*** (36.87)	189.4*** (38.27)
Riohacha	215.8*** (55.23)	238.5*** (61.88)	126.1*** (33.27)	100.8*** (35.56)
Santa Martha	-23.33 (34.53)	-35.81 (47.42)	55.50** (22.47)	54.83* (31.99)
Villavicencio	344.7*** (79.01)	206.7*** (53.72)	347.9*** (62.29)	241.2*** (57.28)
Pasto	-29.94 (37.24)	-17.80 (43.04)	-6.913 (28.84)	4.369 (37.56)
Cúcuta	85.93** (42.58)	109.7*** (40.97)	137.4*** (27.45)	149.2*** (30.48)
Armenía	88.46 (58.85)	40.25 (60.82)	100.1** (39.85)	80.02** (36.23)
Pereira	156.0 (152.0)	32.88 (49.18)	220.8** (96.51)	89.71** (40.69)
Bucaramanga	277.2*** (54.72)	275.9*** (52.66)	301.0*** (39.60)	279.0*** (41.78)
Sincelejo)	47.86 (78.64)	-63.76 (41.29)	61.51 (38.36)	-13.91 (26.57)
Ibagué	182.9*** (54.34)	129.1** (54.16)	138.9*** (35.03)	113.0*** (40.54)
Cali	128.9*** (39.30)	183.9** (76.91)	141.7*** (45.57)	137.9*** (48.66)
Tipo contrato (escrito)	271.9*** (31.83)	215.8*** (37.09)	—	—
Tipo contrato (No sabe, no contesta)	264.6 (337.5)	419.9*** (149.2)	—	—
Constant	223.9 (195.3)	-538.8* (288.0)	132.5 (112.6)	-559.6* (311.7)
Observaciones	12,204	14,050	24,152	27,943
R2	0.138	0.103	0.150	0.083

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

7. Conclusiones.

Las características de los hombres en edad de trabajar expresada en educación (años de estudio) y la experiencia laboral, son estadísticamente significativas y ayudan a explicar de manera consistente la dispersión salarial de los trabajadores, es decir, ganan más los que tienen mayores logros educativos y los que tienen mayor experiencia laboral; De allí la importancia de invertir en educación, para mejorar las condiciones de accesibilidad al mercado laboral.

En cuanto al sector productivo donde laboran, el número de personas y la profesión a la que se dedica son igualmente significativas corroborando así la teoría económica adicionalmente se verifica que la mayoría de las ciudades donde desarrolla su trabajo obtienen tanto hombres como mujeres resultados distintos que demuestra el grado de dispersión que contiene el ingreso laboral en Colombia

Que el R cuadrado sea bajo, puede deberse a la no incorporación de ciertas variables en los modelos, como por ejemplo las habilidades y destreza de cada persona, la calidad educativa, la segmentación del mercado, entre otras, que en última instancia repercute en el aprendizaje y posteriormente en los ingresos laborales.

En cuanto a los modelos imputados tanto para hombres como para mujeres brinda un horizonte mejor a hora de intentar explicar el comportamiento de las variables explicativas que utilizamos con respecto al ingreso laboral. Por tanto la imputación surge como una opción a la hora de evitar perder información valiosa para el análisis, preservando la distribución de los datos y las relaciones entre las variables explicativas y la variable dependiente.

Como recomendación debemos seguir investigando si el cambio de tendencia en Colombia sobre el ingreso laboral entre hombres y mujeres continua, debido a que en la actualidad la mujer presenta unos índices de educación superiores a época pasadas y por tanto pueden desempeñar profesiones que requieren alto grado de cualificación.

Bibliografía

Acock, A. (2005). "Working with missing values", *Journal of Marriage and Family*, 67:1012-1028.

Akerlof, A. y Yellen, J.L. (1986): *Efficiency Wage Models of the Labor Market*. Cambridge University Press, Cambridge.

Andrigge, R. y Little, R. (2010). "A review of hot deck imputation for survey non-response", *International Statistical Review*, 78:40-64.

Ávila, G. Una implicación del procedimiento Hot Deck como método de imputación. Universidad Mayor de San Marcos, 2002.

Barceló, C. (2008). "The impact of alternative imputation methods on the measurement of income and wealth: Evidence from the Spanish survey of household finances", *Documentos de Trabajo (0829)*:9-64.

Barro, R (1991), "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics*, 106(2), 407-43.

Becker, G. (1975). *Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Education*, Columbia University Press, New York.

Bonet, J. & Meisel, A. (1999). "La convergencia regional en Colombia: Una visión de largo plazo, 1926- 1995" *Coyuntura económica*, 29(1):69-106.

Bonet, J., & Meisel, A. (2006). "Polarización del ingreso per cápita departamental en Colombia, 1975 - 2000", *Documentos de Trabajo sobre Economía Regional*, Banco de la República-Cartagena.

Cárdenas, M., Pontón, A. & Trujillo, J. (1993). "Convergencia, y Migraciones Inter departamentales: Colombia 1950-1989", *Coyuntura Económica*, Abril.

Carline, D. (1985): "Trade Unions and Wages", en Carline et al. (eds.). *Surveys in Economics: Labour Economics*, Longman, Londres.

Castillo, J. 2009. "¿Cuál es la brecha salarial justa? Opinión pública y legitimación de la desigualdad económica en Chile" *Estudios públicos* 113:237-266.

CEPAL (2002), *Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida (MECOVI)*, Tercer Taller Regional, Santiago de Chile.

Dickens, W., y Katz, L.F. (1987). Inter-industry wage differences and industry characteristics. In: Lang, K., & Leonard, J. (Eds.). *Unemployment and the structure of labour markets*, p. 48-89. Londres: Basil Blackwell.

Farber, H. S. (1986): "El análisis de la conducta sindical, en O. Ashenfelter, y R. Layard (Comp.): *Manual de Economía del Trabajo*, cap.18, pp. 1.341-1.406, Ministerio de Trabajo y Seguridad Social, Madrid 1991.

Galvis, L. (2004), "Integración Regional de los Mercados Laborales en Colombia, 1984-2000", in: Adolfo Meisel (2004) (ed.). *Macroeconomía y Regiones en Colombia*, Bogotá: Banco de la República.

Galvis, L. & Meisel, A. (2001). "El crecimiento económico de las ciudades colombianas y sus determinantes, 1973-1998", in: Adolfo Meisel (ed.). *Regiones, ciudades y crecimiento económico en Colombia*, Bogotá: Banco de la República.

Galvis, L. A., 2010. *Diferencias salariales por género y región en Colombia: Una aproximación con regresión por cuantiles*, Banco de la República, Documentos de Trabajo sobre Economía Regional, No. 131.

Gracia; O.; Hernández, G., y Ramírez; J.M. (2001). *Diferenciales salariales y mercados laborales en la industria colombiana*. *Desarrollo y Sociedad* No. 48.

González-Rivas, N. (2012). *Discriminación salarial: un análisis entre mujeres afrocolombianas en el área metropolitana de Cali*. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 10(1), pp. 563-578.

HAZIZA, D. (2009). "Imputation and inference in the presence of missing data", *Sample Surveys: Design, Methods and Applications*, 29A:+ 215-246.

Heckman J. (1979). "Sample Selection Bias as a Specification Error", *Econometría*, 47(1), 153-162.

Heckman J. (1980). "Sample Selection Bias as a Specification Error with an Application to the Estimation of Labor Supply Functions", in: J. Smith (ed.) (1980), *Female Labor Supply: Theory and Estimation*, Princeton University Press.

Hewings G.J.D. (1977). *Regional Industrial Analysis and Development*, Great Britain: Methuen & Co Ltd.

Hoyos, A, Ñopo, H., y Peña, X. (2010). *The Persistent Gender Earnings Gap in Colombia, 1994-2006*. Universidad de los Andes, Facultad de Economía, Documento CEDE, No. 32.

Izquierdo, M. y A. Lacuesta (2006): "Wage inequality in Spain: Recent developments", *Documento de Trabajo del Banco de España* 0615.

Jimeno, J. F. (1987): *La flexibilidad de los costes laborales nominales en la industria española (1978-1982)*, *Investigaciones Económicas*, Vol. XI, n.3.

Jimeno, J.F. y Bentolila, S. (1998): *Regional Unemployment Persistence (Spain, 1976-1994)*, *Labour Economics*, 5, pp. 25-51.

Jimeno, J.F., Izquierdo, M. y V. Hernanz, V. (2001): "La desigualdad salarial en España: Descomposición y variación por niveles de salarios", *Papeles de Economía Española* 88, pp. 113-125.

Jaramillo, C.F., Romero, C.A., y Nupia, O. (2000). *Integración en el mercado laboral colombiano: 1945-1998*. Banco de la República, *Borradores de Economía*, No. 148, Bogotá, Colombia.

- Katz, L.F. y D.H. Autor (1999): "Changes in the wage structure and earnings inequality", en O. Ashenfelter y D. Card (eds.), *Handbook of Labor Economics*, North-Holland, Amsterdam, pp. 1463-1555.
- Krueger, A.B., y Summers, L.H. (1987). Reflections on the Inter-industry wage structure. In: Lang, K., & Leonard, J. (Eds.). *Unemployment and the structure of labour markets*, p.17-47. Londres: Basil Blackwell.
- Krueger, A.B., y Summers, L.H. (1988). Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica*, 56 (2), 259-293.
- Manero, M. (1999): "La discriminación salarial en el mercado de trabajo español", Tesina CEMFI no. 9906.
- Medina, F. Galván M. Imputación de datos: teoría y práctica. CEPAL, 2007.
- Mellow, W. (1982): "Employer Size and Wages". *Review of Economics and Statistics*.
- Mesa, D.C., García, A.F., y Roa, M. (2008). Estructura salarial y segmentación en el mercado laboral de Colombia: Un análisis de las siete principales ciudades, 2001-2005. Universidad del Rosario, Facultad de Economía, Serie Documentos de Trabajo, No. 52, Bogotá, Colombia.
- Mesa, D., García, A. (2008). "Estructura salarial y segmentación en el mercado laboral de Colombia: un análisis de las siete principales ciudades, 2001-2005", Documentos de Trabajo # 52, Departamento de Economía, Universidad del Rosario.
- Mincer, J. (1974). Schooling, Experience, and Earnings. *Human Behavior & Social Institutions* No. 2, New York: National Bureau of Economic Research.
- Moncayo, E. (2002). "Nuevos enfoques de política regional en América Latina: El caso de Colombia en perspectiva histórica. Enfoques teóricos y evidencias empíricas sobre el desarrollo regional en Colombia", Archivos de Macroeconomía, Departamento Nacional de Planeación.
- Mortensen, (1986): Búsqueda de empleo y análisis del mercado laboral, en O.Ashenfelter, y R. Layard (Comp.): *Manual de Economía del Trabajo*, Vol. 1, cap.15, pp. 1.111-1.198, Ministerio de Trabajo y Seguridad Social, Madrid 1991.
- Nicoletti, C. y Peracchi, F. (2006). "The effects of income imputation on microanalyses: Evidence from the European Community Household Panel", *Journal of the Royal Statistical Society*, 169(3):625-646.
- Nupia, Oskar (1997). "Integración espacial de los mercados laborales: evidencia para las regiones colombianas", *Desarrollo y sociedad*, No.40, Septiembre. Neumar, D. (1988): "Employers Discriminatory behaviour and the estimation of wage discrimination". *Journal of Human Resources* 23, pp. 279-295.
- Oaxaca, R. (1973): "Male-female wage differentials in urban labor markets." *International Economic Review*, 14(3), pp. 693-709.

Oaxaca, R., RANSOM, M. (1994): "On Discrimination and the Decomposition of Wage Differentials", *Journal of Econometrics*, 61, pp. 5-21.

Olinsky, A., Chen, S. y Harlow, L. (2003). "The comparative efficacy of imputation methods for missing data in structural equation modeling", *European Journal of Operational Research*, 15:53-79.

Ortiz, C.H, Uribe, J.I., y García, G.A. (2007). La segmentación del mercado laboral colombiano en la década de los noventa. *Revista Economía Institucional*, 9(16), 189-221.

Ortiz, C., Uribe, J. & Badillo, E. (2009). "Segmentación Inter e intrarregional en el mercado laboral urbano de Colombia, 2001–2006", *Ensayos Sobre Política Económica*, 27(58).

HORTON, N. J. y LIPSITZ, S. R. (2001). "Múltiple imputation in practice:

Palacio, J.I. y H. Simón (2004): "Dispersión salarial entre establecimientos y desigualdad salarial en España", *Revista de Economía Aplicada* Vol. XII, pp. 47-81.

Persson, T. & G.Tabellini. (1994). "Is Inequality Harmful for Growth?," *American Economic Review* 84, 600- 621.

Posso, C. (2008). Desigualdad salarial en Colombia 1984-2005: Cambios en la composición del mercado laboral y retornos a la educación post-secundaria. Banco de la República, Borradores de Economía, No 529, Bogotá, Colombia.

Rocha, R. & Vivas, A. (1998). "Crecimiento Regional en Colombia: Persiste la Desigualdad", *Revista de Economía del Rosario*, 1(1), 67-108.

Rojas-Hayes, (2006). Race determinants of wage gaps in Colombia.

Rosen, S. (1986): La teoría de las diferencias igualadoras, en O. Ashenfelter, y R. Layard (Comp.): *Manual de Economía del Trabajo*, Vol. 1, cap.12, pp. 813-880, Ministerio de Trabajo y Seguridad Social, Madrid 1991.

Rubin, D.B.(1987). *Múltiple Imputation for Non response in Surveys*. Wiley, New York.

Santamaría, M. (2004).. "Income inequality, skills and trade: Evidence from Colombia Turing the 80s and 90s" (Documentos CEDE 2). Universidad de los Andes.

Slichter, S.H. (1950). Notes on the structure of wages. *Review of Economics and Statistics*, 32 (1), 80-91.

Urrutia, M., y Ruiz, M. (2010). Ciento sesenta años de salarios reales en Colombia. *Ensayos Sobre Política Económica*, No. 63. Por aparecer.

Yun, M.S. (2006): "Earnings inequality in the USA, 1961-1999: Comparing inequality using earnings equations", *Review of Income and Wealth* 52, pp. 127-144.

