

## **Efectos de los ingresos no reportados en el nivel y tendencia de la pobreza laboral en México**

**Raymundo M. Campos-Vázquez\***

Fecha de recepción: 04/I/2013

Fecha de aceptación: 28/VI/2013

### **Resumen**

En México, en los últimos años, la proporción de trabajadores remunerados que no declara ingresos ha aumentado. Si la población que no declara ingresos es una muestra aleatoria o bien la composición de la misma no cambia, esto no afectaría la tendencia en el ingreso promedio de la población o en las mediciones de pobreza y desigualdad. Sin embargo, eso no se cumple para México, ya que la proporción con ingresos inválidos y educación superior era de 20% en 2005 y para 2012 era de 39%. En este artículo, se analiza la corrección de ingresos inválidos por medio de métodos de imputación. Se encuentra que el cálculo del ingreso es robusto para los diferentes métodos. Para 2012, se encuentra que el ingreso individual promedio con imputación es mayor en 4% que el reportado, y la desigualdad con imputación es 2.4% mayor que la observada. Asimismo, mientras que el ingreso familiar *per cápita* reportado en el periodo 2005-2012 ha caído cerca de 15% en términos reales, el ingreso familiar *per cápita* verdadero no muestra cambios sustanciales. La medición oficial de la pobreza laboral indica un crecimiento de la misma de 42% a 53%, en el periodo 2005-2012; pero la pobreza con ingresos imputados aumentó de 36% a 40%, un incremento mucho menor al observado. Los resultados sugieren la conveniencia de que las instituciones tomen en cuenta los ingresos no reportados, y que reporten estadísticas tanto con como sin corrección por esos ingresos.

**Palabras Clave:** Ingreso laboral; Ingreso no reportado; Imputación; Pobreza; Desigualdad.

**Clasificación JEL:** C10; D63; I30; I32; O54.

---

\* Centro de Estudios Económicos, El Colegio de México. Dirección: Camino al Ajusco 20, Col. Pedregal de Santa Teresa, C.P. 10740, México D. F. Tel: +52-55-54493000, ext. 4153. Correo electrónico: rmcampos@colmex.mx.

### **Abstract**

The proportion of wage workers with missing labor income has recently increased in Mexico. In this article, I correct the missing labor income through imputation methods. The official measurement of labor poverty indicates an increase from 42% to 53% in the 2005-2012 period, but poverty measured with imputed income increases only from 36% to 40%, a much lower increase than the one we observe. The results imply that policy makers should include in their statistics individuals with missing income and report statistics with and without correcting for missing income.

**Keywords:** Labor income; Missing income; Imputation; Poverty; Inequality.  
**JEL Classification:** C10; D63; I30; I32; O54.

### **Introducción**

La mayoría de los estudios en economía laboral, así como las mediciones de pobreza en México, ignoran a los individuos que reportan trabajar por remuneración pero que deciden no reportar ingresos. Los estudios de economía laboral generalmente eliminan de la muestra a esos individuos, mientras que en la medición del Índice de la Tendencia Laboral de la Pobreza (pobreza laboral), por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL, 2010), se incluyen pero como si no tuvieran un ingreso, es decir, con ingreso cero. El supuesto detrás de esas decisiones es que las personas que deciden no reportar ingresos son una muestra aleatoria de los trabajadores o bien que la composición de ese tipo de personas se mantiene constante en el tiempo.

En este artículo, se argumenta que esos supuestos dejaron de ser válidos para el caso de México. La proporción de individuos que decide no declarar ingresos aun cuando trabaja por remuneración, ha aumentado. Por ejemplo, en 2005, la proporción era de 10%, mientras que en 2012 era de 20%. Más aún, la proporción de ingresos inválidos ha aumentado principalmente entre los individuos con educación superior. Esta proporción era de 20% en 2005; pero, para 2012 era de 38%. Es decir, la proporción de trabajadores que no declara ingresos ha cambiado a través del tiempo, y no existe evidencia de que la decisión de no reportar ingreso sea aleatoria. Dado que, en promedio, los individuos con mayor educación tienen un ingreso más alto que individuos con menor educación, entonces, el incremento en la proporción de ingresos inválidos debido a individuos con educación superior pudiera tener impactos en el ingreso promedio de la población y, por tanto, en las mediciones de pobreza y desigualdad. En consecuencia, un tratamiento

correcto de los ingresos inválidos tiene implicaciones importantes en la medición de pobreza y en las políticas públicas destinadas a combatirlas.

En la literatura internacional, existe un amplio consenso sobre la conveniencia de corregir ingresos inválidos con métodos de imputación; por ejemplo, los tratamientos clásicos en Little y Rubin (2002) y Rubin (2004). Así que, se podría pensar que imputar por medio de la media observada es correcto, sin embargo, ese método produce una distribución de ingresos incorrecta: picos en partes de la distribución y una subestimación de la varianza. Para resolver el problema de ingresos inválidos, en esta investigación, se utilizan diferentes métodos de imputación que no poseen esas desventajas.<sup>1</sup> El propósito es calcular ingresos para aquellas personas que no declaran ingresos, de tal manera que se puedan analizar las tendencias de pobreza y desigualdad, si no existiera el problema de los ingresos inválidos. En este artículo se utilizan cuatro métodos para corregir por ingresos inválidos: pareamiento por puntajes de propensión, *hot-deck*, imputación en la mediana de un grupo, más un “ruido”, y el método de pareamiento por promedios predictivos.

La aplicación empírica utiliza la Encuesta Nacional de Ocupación y Empleo de cada trimestre de 2005 a 2012. Los resultados son robustos al método de imputación utilizado; es decir, cualquier método con las restricciones utilizadas en este artículo obtiene resultados similares. En términos de ingresos, encuentro que el ingreso promedio individual y familiar *per cápita* serían mayores si se corrigiera por el problema de ingresos inválidos. Después de 2008, este problema se agrava. Por ejemplo, en 2005, el ingreso promedio familiar *per cápita* era de \$1,700, mientras que para 2012 era de \$1,400. Con ingresos imputados se encuentra un ingreso promedio aproximado a \$1,900, tanto en 2005 como en 2012. Es decir, mientras que el ingreso familiar *per cápita* observado en ese mismo periodo ha caído cerca de 15% en términos reales, el ingreso con imputación (que llamaremos también ingreso verdadero) no muestra cambios sustanciales. En términos de desigualdad, medida con el coeficiente de Gini, incluyendo el ingreso imputado para los individuos que no reportan ingresos, se obtiene que la desigualdad es mayor. Finalmente, y más importante, la pobreza laboral

---

<sup>1</sup> En un procedimiento parcialmente similar al utilizado en este artículo, Rodríguez-Oreggia (2012) explica en una breve nota técnica la imputación por ingresos inválidos utilizando pareamiento por puntajes de propensión. Sin embargo, en esa nota se imputan ingresos utilizando múltiplos del salario mínimo. Como se explica en el texto, esto causa una subestimación de la varianza verdadera. Además en el presente trabajo abordamos diferentes métodos de imputación, y no únicamente el de pareamiento por puntajes de propensión.

como la calcula CONEVAL está sobreestimada<sup>2</sup> (CONEVAL, 2012). Cuando se construye el índice de la tendencia laboral de la pobreza, el CONEVAL encuentra que la pobreza ha aumentado un 25% en el periodo 2005-2012. Sin embargo, cuando se imputan los ingresos inválidos, se encuentra que la pobreza solo ha aumentado un 12%. Es decir, para el tercer trimestre del año 2012, existen 8.1 millones de personas que son declarados como pobres laborales, pero que de acuerdo con el método de imputación, no lo son. Por esta razón, se recomienda que las instituciones encargadas de medir pobreza y desigualdad tomen en cuenta los ingresos no reportados y reporten sus cálculos sin corrección y con ella, para esos ingresos.

El presente trabajo se encuentra organizado de la siguiente forma. En la sección 1, se discuten los datos y estadísticas descriptivas de la población con ingresos inválidos. En la sección 2, se explican los diferentes métodos de imputación así como sus ventajas y desventajas. En la sección 3, se presentan los resultados principales en términos de ingreso y desigualdad en el nivel individual. En la sección 4, se presentan resultados en términos de ingreso y pobreza en el nivel familiar. Por último, se ofrecen las conclusiones y algunas recomendaciones sobre política pública.

### **1. ¿Quiénes no reportan ingresos?**

En este estudio, se utiliza la Encuesta Nacional de Ocupación y Empleo (ENOE) para cada trimestre disponible de 2005 a 2012. La ENOE es una encuesta en hogares que es representativa en el nivel nacional y por entidad federativa. La encuesta se realiza aproximadamente a 120,000 hogares en cada trimestre, y recoge principalmente estadísticas relacionadas con el mercado laboral, tales como: participación laboral, desempleo, ingresos, horas trabajadas, ocupación, así como edad, sexo, educación, etc.

El método de imputación del presente trabajo aplica únicamente para las personas que reportan trabajar por remuneración con horas de trabajo positivas, pero deciden no reportar ingresos. Es decir, dado que no hay un

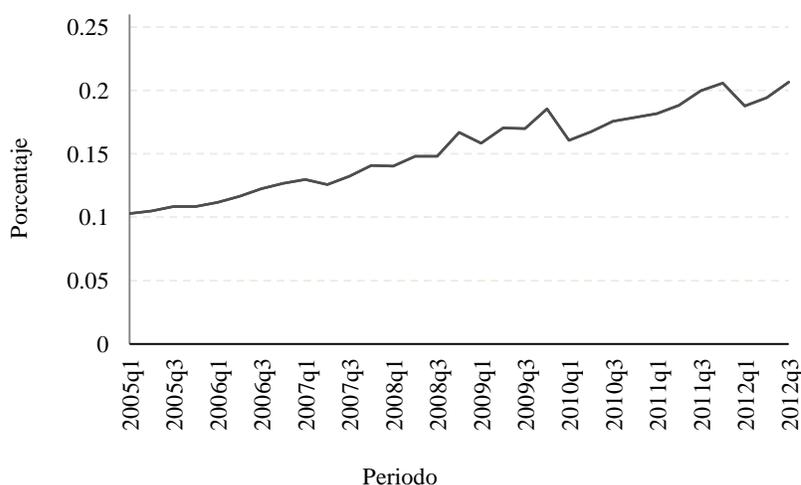
---

<sup>2</sup> La medición oficial de la pobreza en México utiliza la Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH). Sin embargo, recientemente el CONEVAL utiliza la Encuesta Nacional de Ocupación y Empleo (ENOE) para obtener la tendencia de la pobreza, utilizando únicamente ingresos laborales. Esto se debe a que los ingresos laborales representan el principal ingreso de los hogares, por lo que cambios en los mismos se ven reflejados en si el hogar es pobre o no. La ENOE incluye únicamente ingresos laborales, mientras que la ENIGH incluye además del ingreso laboral otro tipo de ingresos. La medición oficial del cálculo de ingresos para determinar si el hogar es pobre incluye no solo ingresos laborales, sino transferencias nacionales e internacionales, regalos, ingreso no monetario, entre otros. Los detalles de estas encuestas se encuentran en ENIGH (2010) y ENOE (2012).

salario inválido, no se les imputa ingreso a personas fuera de la fuerza laboral, desempleados, así como tampoco a trabajadores sin pago.<sup>3</sup> Si el trabajador remunerado reporta un ingreso positivo se considera un ingreso válido, si reporta un ingreso faltante se considera un ingreso inválido. Dado que CONEVAL y otras instituciones no restringen por años de edad los cálculos de pobreza, en este artículo se utiliza como muestra a toda la población en edad de trabajar en la ENOE, que es la población de edades mayores o iguales a 12 años. El ingreso laboral nominal se transforma a ingreso real con base en el segundo trimestre (mayo) de 2012 del Índice Nacional de Precios al Consumidor (INPC, 2012).

La figura 1 muestra la evolución del porcentaje de trabajadores que no declaran ingresos laborales. En el primer trimestre de 2005, el porcentaje era cercano al 10%; para el tercer trimestre de 2012, ese porcentaje superaba el 20%; es decir, en 8 años, el porcentaje de individuos que no declara ingresos se ha duplicado.

Figura 1  
**Porcentaje de trabajadores que no declaran ingresos**



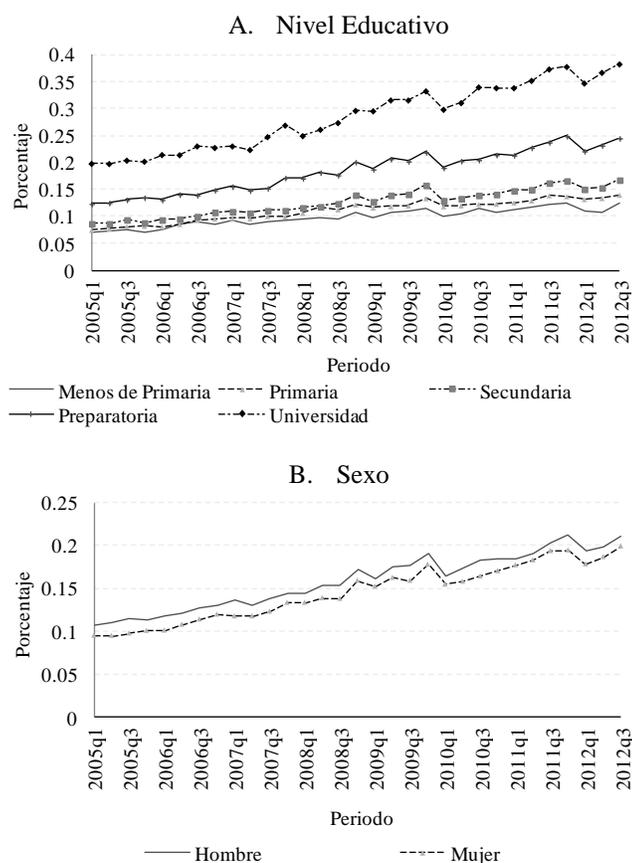
Nota: se graficó para los trimestres 1 y 3. La muestra utilizada se refiere a trabajadores con remuneración de edades mayores o iguales a 12 años de edad. No se incluyen trabajadores sin pago.

Fuente: cálculo realizado por el autor con datos de la ENOE.

<sup>3</sup> El método de imputación no incluye a los trabajadores sin pago y desempleados, ni incluso a los que reportan trabajar pero tienen cero horas trabajadas. Pero cuando se calcula pobreza e ingreso familiar, sí se incluyen (con ingreso cero).

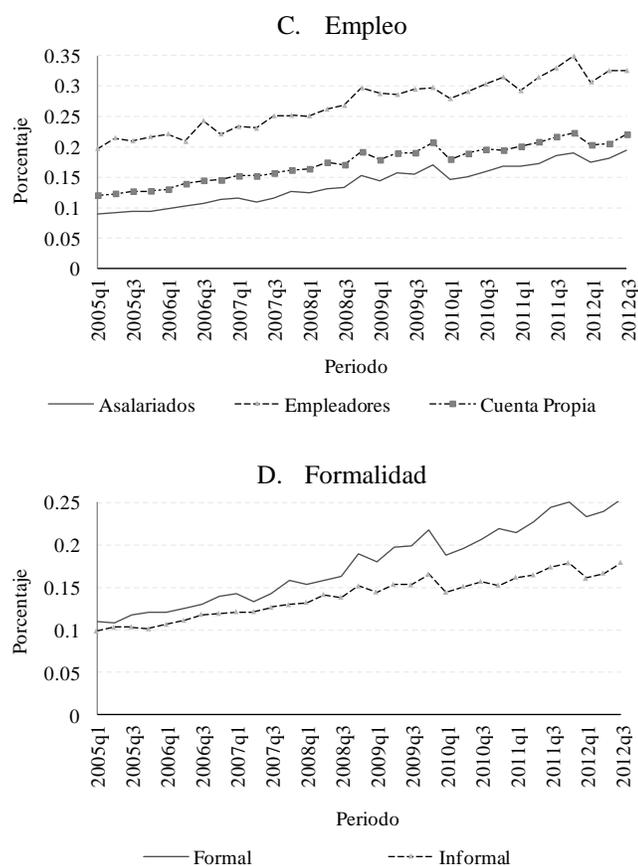
Si las personas que no declaran ingresos son una muestra aleatoria de la población de trabajadores, entonces no importaría en los cálculos de ingreso y pobreza. En la figura 2, se presentan diferentes paneles que muestran cómo ha cambiado entre diversos grupos sociodemográficos el porcentaje de trabajadores que no reporta ingresos. La figura incluye patrones clasificados según: nivel educativo, sexo, tipo de empleo y condición de formalidad (recibe seguro de salud en su trabajo). Los paneles muestran un patrón claro.

Figura 2  
**Porcentaje de trabajadores que no reportan ingresos, por grupos sociodemográficos.**



Continúa en la siguiente página.

Continuación... Figura 2  
**Porcentaje de trabajadores que no reportan ingresos, por grupos sociodemográficos.**



Nota: A, B, C y D se graficó para los trimestres 1 y 3. Cálculo obtenido por el autor con datos de la ENOE. La muestra utilizada se refiere a trabajadores con remuneración de edades mayores o iguales a 12 años de edad. No se incluyen trabajadores sin pago.

Fuente: elaboración del autor.

Si bien todos los grupos han aumentado su porcentaje de trabajadores con ingresos inválidos, el grupo que presenta un incremento notable es el de trabajadores con al menos educación universitaria (panel A). El grupo de trabajadores formales también ha tenido un incremento importante (panel D). En el caso de trabajadores con educación universitaria, cerca del 40% de los trabajadores ya no declara su ingreso laboral; mientras que, en 2005, era la mitad. En el caso de trabajadores formales, cerca del 25% ya no declara un

ingreso; mientras que en 2005, era únicamente el 10%. Por otro lado, la tendencia del porcentaje de trabajadores con ingresos inválidos no parece cambiar por sexo ni por tipo de ocupación en el empleo (paneles B y C).

Cuadro 1  
Diferencias de Medias en características observables: 2005 y 2012.

	2005			2012			Inválidos 2005-2012
	Válidos	Inválidos	Diferencia	Válidos	Inválidos	Diferencia	
Edad	36.86	40.25	-3.40 *	37.80	40.80	-2.99 *	-0.54 *
			(0.125)			(0.090)	(0.144)
Mujer	0.37	0.33	0.04 *	0.39	0.37	0.02 *	-0.04 *
			(0.004)			(0.003)	(0.004)
Casado	0.62	0.59	0.02 *	0.61	0.59	0.03 *	0.01
			(0.004)			(0.003)	(0.005)
Formal	0.41	0.41	0.00	0.39	0.46	-0.07 *	-0.06 *
			(0.004)			(0.003)	(0.005)
Rural	0.25	0.19	0.06 *	0.28	0.18	0.10 *	0.01
			(0.003)			(0.002)	(0.004)
Tiempo Completo	0.79	0.83	-0.04 *	0.76	0.83	-0.07 *	0.00
			(0.003)			(0.002)	(0.003)
Sin primaria completa	0.18	0.13	0.05 *	0.13	0.08	0.06 *	0.05 *
			(0.003)			(0.001)	(0.003)
Primaria completa	0.22	0.16	0.06 *	0.20	0.13	0.07 *	0.03 *
			(0.003)			(0.002)	(0.003)
Secundaria completa	0.27	0.22	0.05 *	0.31	0.23	0.08 *	-0.01 *
			(0.003)			(0.003)	(0.004)
Preparatoria completa	0.19	0.22	-0.03 *	0.21	0.25	-0.04 *	-0.03 *
			(0.003)			(0.003)	(0.004)
Universidad completa	0.14	0.27	-0.13 *	0.15	0.31	-0.16 *	-0.04 *
			(0.003)			(0.003)	(0.004)
N	132,849	13,977		120,491	29,344		

Nota: la muestra utilizada se refiere a trabajadores con remuneración de edades mayores o iguales a 12 años de edad. No se incluyen trabajadores sin pago. Nivel de significancia 5%.

Fuente: cálculos realizados por el autor con datos de la ENOE 2005 y 2012, tercer trimestre.

En el cuadro 1 se presenta la diferencia de medias en características observables para las personas que tienen ingresos válidos e inválidos. Se utilizan únicamente los años 2005 y 2012 con información del tercer trimestre. En general, se observa que las características de quienes no declaran ingresos son diferentes de los trabajadores que sí declaran ingresos. Por ejemplo, los que no declaran ingresos tienden a ser mayores en edad, menor proporción a ser mujer, menor propensión a ser casado y vivir en área rural, trabajar de tiempo completo (mayor a 35 horas a la semana) y, en general, con mayor educación que los trabajadores que sí declaran ingresos. Esto se repite para 2012, con la diferencia de que en 2012 existe una diferencia significativa en la tasa de formalidad. Si se analizan las diferencias entre las personas que no declaran ingresos entre 2005 y 2012 (última columna del cuadro 1), entonces se observará que también existen diferencias significativas; entre las más destacadas están un aumento en las personas que no son formales, y con mayor educación. En resumen, los trabajadores que no declaran ingresos son diferentes a los trabajadores que sí los declaran: además, los tipos de trabajadores que no declaran han ido cambiando en el tiempo. Por estas razones, es necesario e importante corregir el ingreso no reportado.

## **2. Métodos para imputar ingresos inválidos**

Existe una amplia literatura que describe diferentes métodos de imputación (Allison, 2001; Cameron y Trivedi, 2005; Little y Rubin, 2002). Por tanto, en este artículo únicamente explicamos de forma general el método de imputación, así como los métodos utilizados.<sup>4</sup> El problema de datos faltantes, en nuestro caso el ingreso laboral, puede llevar a inferencias incorrectas. Por ejemplo, el método más común ante la ausencia de datos es eliminar al individuo o al hogar de la muestra. Si las personas que no declaran ingresos son una muestra aleatoria de la población, en general, el único problema es de eficiencia (mayores errores estándar), sin problema de sesgo.<sup>5</sup> Sin embargo, si las personas con ingresos inválidos no son una

---

<sup>4</sup> En el presente artículo no se incluyen cálculos de imputación por máxima verosimilitud, ver por ejemplo Allison (2012) y Humphries (2012). La razón es que estos métodos requieren una forma funcional, y el método que se utiliza en este artículo es principalmente no paramétrico y por tanto más flexible y menos demandante en términos computacionales, además de ser más sencilla su aplicación para los hacedores de política.

<sup>5</sup> La notación de Rubin (1976, 2004) y de Cameron y Trivedi (2005), implica que los datos faltantes son faltantes de forma completamente aleatoria (*Missing Completely at Random, MCAR*). Para más detalle también ver Kennickell (1998) y Schafer (1997). En este artículo, no se utilizan métodos que dependan de supuestos sobre características no observables. Por ejemplo, que los individuos de ingresos altos no declaren ingresos por ciertas características no observables. Estos métodos incluyen el método de Heckman (1979), de corrección en dos etapas. Se decide no utilizar estos métodos por dos razones:

muestra aleatoria de la población, entonces los estadísticos de interés no son representativos de la población en su conjunto. Esto se debe a que la población con ingresos válidos e inválidos es diferente.

Dado lo analizado en la sección 1, no es posible afirmar que la proporción de personas que no declara ingresos, en México, se ha mantenido constante en el tiempo. Por el contrario, la proporción ha aumentado principalmente entre las personas con mayor educación, y en consecuencia, las de mayor salario. El descartar a estas personas o bien asignarles ingreso cero -como Coneval (2010)- podría llevar a inferencias incorrectas, por ejemplo: afirmar una caída mayor en salarios reales que la verdadera, o un mayor aumento en la pobreza que el verdadero valor.

Uno de los métodos comúnmente utilizados para resolver el problema de ingresos faltantes es imputar con la media del valor observado. O bien, otro método es imputar con el valor medio predicho de una regresión lineal (Haitovsky, 1968). Estos métodos son incorrectos (Rubin, 1976, 1996, 2004) por dos razones, principalmente. La primera es que la distribución de ingreso imputada tendría picos en ciertas partes de la distribución, algo que no se observaría con la verdadera distribución. La segunda es que se está subestimando sustancialmente la varianza del ingreso. Esto no solo tendrá implicaciones cuando se desee realizar inferencia estadística, sino que tiene implicaciones para el cálculo de desigualdad también, por citar un ejemplo. Por tanto, esos métodos no son recomendados para realizar imputaciones.

La mayoría de los estudios que utilizan imputación, usan el supuesto de: “Datos faltantes por razones aleatorias” (*Missing at Random, MAR*, Rubin 1976). En el caso del ingreso ( $Y$ ) faltante, este supuesto implica que las razones de su ausencia no dependen del ingreso verdadero sino de otras características observables (por ejemplo edad, educación, entre otras, y denotamos a esas características como  $X$ ). En notación matemática, esto implica que (Cameron y Trivedi, 2005; Rubin, 2004):

$$\Pr(Y_i \text{ faltante} | Y_i, X_i) = \Pr(Y_i \text{ faltante} | X_i) \quad (1)$$

De ahora en adelante, el trabajo asume que los ingresos faltantes cumplen el supuesto *MAR*. Sin embargo, en el trabajo vamos a utilizar diferentes métodos con el propósito de analizar la robustez y la sensibilidad de los datos faltantes, o al menos una de ellas.

---

(1) El objetivo es tener un método de corrección fácilmente aplicable para hacedores de política pública. (2) No queda claro que se pueda encontrar un instrumento que cumpla con la restricción de exclusión en el método de Heckman (1979).

Entre los métodos de imputación se pueden encontrar métodos paramétricos y no paramétricos. El primero asume una forma funcional de los datos faltantes y el segundo no la asume. Asimismo, entre las familias de métodos de imputación se pueden encontrar los de imputación única e imputación múltiple. La imputación única realiza una imputación solamente, mientras que la múltiple se realiza varias veces, y el estadístico final es el promedio del estadístico de interés en cada imputación (Little y Rubin, 2002). Rubin (1996, 2004) y Zhang (2003) han argumentado que la forma correcta de realizar imputaciones es por medio de la imputación múltiple. Esto se debe a que tiene que reconocerse que la imputación implica un proceso aleatorio, el cual puede cambiar en cada imputación. Sin embargo, Durrant (2005) encuentra que en la práctica la imputación única y la múltiple tienen pocas diferencias. En nuestro análisis, implementamos tanto métodos paramétricos como no paramétricos así como imputación única y múltiple.<sup>6</sup>

En el método paramétrico, dado que imputar con un valor promedio es incorrecto, la forma de resolver por el problema de poca varianza ha sido la de añadir cierto “ruido” a la imputación. Por ejemplo, se añade un error aleatorio que se asume con distribución normal estándar, multiplicado por la desviación estándar observada. Si es una imputación única se considera que la distribución de errores aleatorios es la “correcta”. En imputación múltiple se reconoce que ese error aleatorio puede cambiar de imputación a imputación, por lo que se toma el valor promedio del estadístico de interés (para cada muestra imputada se calcula el estadístico y al final se calcula el promedio de los estadísticos obtenidos). En general, se considera que el número de imputaciones óptima es relativamente bajo (de 5 a 10; Allison, 2001; Schaffer, 1997). Por otro lado, una desventaja de añadir “ruido” es que el valor imputado puede estar fuera del rango deseado, por ejemplo, el ingreso imputado podría tomar valores negativos.

A continuación se discuten los diversos métodos utilizados en este estudio. El primer método aplica la técnica de pareamiento por puntajes de propensión con base en las contribuciones de Di Nardo, Fortin y Lemieux (1996), y Rosenbaum y Rubin (1983). El segundo método es el conocido como Hot-Deck, el cual es ampliamente utilizado en la imputación de ingresos para el caso de encuestas en Estados Unidos. El tercer método es una combinación de los métodos anteriores, e implica: obtener el ingreso imputado con base en el valor medio condicional a ciertas características, más cierto “ruido”. Finalmente, el último método es el pareamiento por promedios predictivos.<sup>7</sup> En todos los métodos se discute si su aplicación es

---

<sup>6</sup> Aunque en la práctica tanto la imputación única como la múltiple obtienen resultados casi idénticos.

<sup>7</sup> Es importante mencionar que también se intentaron otros métodos como el de imputación múltiple por regresión, más un “ruido”. Sin embargo, la desventaja principal

relativamente sencilla, en términos de programación la computadora; esto con el fin de extender el uso de dichos métodos al público en general.

### *2.1. Pareamiento por puntajes de propensión.*

Este método fue desarrollado por Rosenbaum y Rubin (1983), Rubin y Thomas (1986), y explicado detalladamente en Caliendo y Kopeinig (2008). El método lleva dos pasos.<sup>8</sup> En el primer paso se estima un puntaje de propensión. Es decir, se estima un modelo de probabilidad donde se asumen formas funcionales logísticas o normales; lo cual significa que se estima la ecuación (1) asumiendo una forma funcional específica. Una vez estimado el modelo se calcula la probabilidad predicha o el puntaje de propensión. Las variables explicativas en el modelo tienen que incluir aquellas que determinan la probabilidad de ingreso faltante, tanto variables lineales como no lineales (Di Nardo et al., 1996).

El segundo paso implica utilizar un procedimiento de pareamiento con base en el puntaje de propensión. En este paso, para cada puntaje de propensión de personas con ingreso no válido, se busca el puntaje más similar posible de otra persona con ingreso válido. Después de esto, a la persona con ingreso inválido se le asigna el ingreso válido de la persona escogida. Esto implica que las personas escogidas con ingresos válidos e inválidos, tienen características similares en promedio. Dado que se planean utilizar otros métodos de imputación, en este trabajo se ha implementado el método de pareamiento del vecino más cercano. La ventaja de este método de pareamiento sobre otros es que se utiliza como imputación un valor observado, a diferencia de otros métodos como el de más vecinos o el de tomar el promedio dentro de un radio en el puntaje de propensión.

El método de pareamiento por puntajes de propensión presenta ventajas y desventajas. Entre las ventajas podemos distinguir la fácil interpretación y entendimiento del método. Además, el ingreso imputado es un valor observado en la muestra (de utilizarse el método del vecino más cercano). Sin embargo, el método presenta varias desventajas. Primero, el método es intensivo, computacionalmente. Segundo, si bien existe una rutina en el

---

de este método es que se tuvieron demasiados ingresos negativos, los cuales fueron declarados inválidos, y por tanto no se pudo disminuir sustancialmente el porcentaje de individuos con ingresos inválidos.

<sup>8</sup> Existe una amplia literatura sobre este procedimiento. El método tiene dos supuestos principales: (1) selección por características observables, que es básicamente el supuesto MAR, descrito anteriormente; (2) soporte común, el cual indica que los individuos tienen que ser comparables entre grupos de ingreso válidos e inválidos. Para más detalles, véase Caliendo y Kopeinig (2008).

software Stata para realizar el pareamiento, no existe una rutina para realizar la imputación.<sup>9</sup> Tercero, no es posible su extensión a imputación múltiple.

## 2.2. *Hot-deck.*

Este método es comúnmente utilizado para corregir por datos faltantes. Una revisión y explicación más profunda sobre el método puede ser encontrada en Andridge y Little (2010), Rubin (1996) y West, Butani y Witt (1990). Este método consiste en reemplazar datos faltantes con valores observados de forma aleatoria.

Supongamos que tenemos  $g$  grupos. Por ejemplo, si definimos género y formal tenemos cuatro grupos posibles (hombre formal, hombre informal, mujer formal, mujer informal). Dentro de cada grupo se divide por los ingresos: válidos e inválidos. El procedimiento *hot-deck* es el siguiente, para cada grupo  $g$  (Allison, 2001; Rubin y Schenker, 1986):

1. Tomar una muestra aleatoria con reemplazo (bootstrap) de los ingresos válidos (la muestra se mantiene del mismo tamaño  $n_0$ ). Por ejemplo, para cada grupo de género-formal se toma una muestra con reemplazo de las personas con ingreso válido. Se mantiene el mismo tamaño de muestra que el original.
2. De esta muestra  $n_0$ , tomar una muestra aleatoria con reemplazo de tamaño  $n_1$ . Este tamaño  $n_1$  es el número de observaciones con ingreso inválido dentro del grupo  $g$ . Entonces, por ejemplo, si tenemos tres observaciones con ingreso inválido dentro del grupo  $g$  (en nuestro ejemplo anterior puede ser cualquier grupo de género-formal), entonces, se toma una muestra aleatoria con reemplazo de tamaño tres (3) proveniente del grupo  $n_0$ . Por tanto, después de este paso, tenemos el mismo número de observaciones con ingresos válidos e inválidos dentro de cada grupo, ambas ordenadas aleatoriamente.
3. Sustituir el ingreso de la muestra  $n_1$  en el paso dos (2), como ingreso imputado, para aquellos con ingreso inválido. En el ejemplo anterior, del paso dos (2) tenemos tres observaciones (tanto para el grupo de ingresos válidos como inválidos) dentro de un grupo, ordenadas aleatoriamente, por lo que se sustituye el valor del ingreso reportado válido en los valores con ingreso inválido.
4. Repetir los pasos uno a tres (1-3) para todos los grupos  $g$ .

---

<sup>9</sup> Se utiliza el procedimiento *psmatch2* de Leuven y Sianesi (2012) para estimar el vecino más cercano.

Este procedimiento es no paramétrico porque no se asume una forma funcional de los datos. Asimismo, pudiera parecer que el paso uno (1) es innecesario. Pero Allison (2001) menciona que eliminar este paso implicaría reducir variabilidad en la estimación y, por tanto, con errores estándar inválidos.

Este procedimiento presenta varias ventajas. Primero, los valores imputados toman valores observados. Segundo, es fácilmente aplicado a imputación única o múltiple. Tercero, en términos de tiempo computacional, es relativamente eficiente. Cuarto, el procedimiento es utilizado para imputar ingresos en otros países, como en el Censo y otras encuestas en Estados Unidos (Allison, 2001; Andridge y Little, 2010). Quinto, y finalmente, el procedimiento tiene una rutina en el software *Stata* (véase por ejemplo, el comando *hotdeck* de Mander y Clayton, 2007), lo que evita errores de programación además de hacerlo de fácil aplicación para los tomadores de políticas y para el público en general.

### 2.3. *Por grupos con aleatoriedad.*

Este método es una combinación de los anteriores. En este método, en cada grupo  $g$  se calcula la mediana (no se utiliza la media para evitar problemas de sensibilidad a valores exageradamente grandes). Para evitar el problema de subestimar la verdadera varianza del ingreso, se le suma a esa mediana la desviación estándar observada de ese grupo, multiplicada por una variable que se distribuye normal estándar.

Este método tiene similitud con el método de pareamiento por puntajes de propensión, porque se limita a imputar valores “similares” dentro de cada grupo. Mientras que en el método de puntajes de propensión se asume que en promedio tienen los mismos valores, con este método existe la certeza de que están tomando valores similares en el ingreso, dentro de cada grupo. El método presenta ventajas y desventajas. La ventaja principal es su sencilla aplicación en programación y su rápida ejecución. Otra ventaja es su extensión en imputación única o múltiple. Sin embargo, tiene desventajas importantes. Primero, se asume siempre que la mediana y la desviación estándar observada dentro de cada grupo son las correctas. Es decir, se asume que esos estadísticos no cambian con los datos faltantes, y no se añade incertidumbre en ese paso. Por ejemplo, no se propone tomar muestras con reemplazo para calcular diferentes medianas o desviaciones estándar. Segundo, si la distribución dentro de cada grupo no sigue una normal, se estaría cambiando la distribución verdadera de los datos. En la aplicación empírica comparamos este método con otros, para valorar sus ventajas y desventajas.

#### 2.4. Pareamiento por promedios predictivos.

Una desventaja del método de *hot-deck* y de la mediana por grupos es que cada grupo tiene que ser definido categóricamente (Allison, 2001). Es decir, si se tienen variables continuas se tendrían que categorizar en grupos (por ejemplo, rangos de edad o rangos de horas trabajadas). Si al hacer esto, se pierde información valiosa, se podrían tener consecuencias en la medición del ingreso imputado. Para evitar este problema, Little (1988) propuso el método de pareamiento por promedios predictivos, el cual es relativamente similar a *hot-deck* y el método de pareamiento por puntajes de propensión.

El método propuesto es el siguiente (Allison, 2001):

1. Se estima una regresión de los ingresos observados y características observables,  $Y_i = \beta X_i + e_i$ . De esta regresión se obtienen los parámetros de los coeficientes y de la varianza:  $(\hat{\beta}, \hat{\sigma}^2)$ .
2. Obtener un parámetro de varianza de forma aleatoria. Esto se logra por medio de calcular  $(n_0 - k) \frac{\hat{\sigma}^2}{\chi^2}$  donde  $\chi^2$  es un número aleatorio proveniente de la distribución  $\chi^2$  con  $(n_0 - k)$  grados de libertad. Denotemos como  $s_{[1]}^2$  ese cálculo.
3. Obtener los coeficientes de forma aleatoria. Los parámetros de mínimos cuadrados se comportan de acuerdo con una distribución normal multivariada, por lo que se obtienen de forma aleatoria mediante la siguiente distribución:  $(\hat{\beta}, s_{[1]}^2 (X'X)^{-1})$ . Se obtiene  $b_{[1]}$ .
4. Con el nuevo parámetro  $b_{[1]}$  se toman predicciones para las personas con ingreso válidos e inválidos.
5. Se realiza un pareamiento con el o los vecinos más cercanos según la distancia en el ingreso. Es decir, a cada persona con ingreso inválido se le busca el vecino más cercano según la distancia en el ingreso predicho. Una vez encontrado ese vecino, el ingreso imputado es el ingreso observado de ese vecino más cercano, bajo esa métrica.

Este método tiene como ventaja su robustez estadística, valores imputados de valores observados, su sencilla expansión a imputación múltiple (repetiendo, el número de veces deseado, los pasos 1-5), así como su fácil implementación en el software Stata. Sin embargo, en aplicaciones empíricas, el método es computacionalmente intensivo. Se discutirá este aspecto en los resultados encontrados.

### 3. Resultados

En esta sección, se presentan los resultados de la imputación de ingresos de acuerdo con los métodos descritos. Se implementa cada método de la forma más similar posible. Además, para algunos individuos con ingreso inválido se tiene información de su ingreso en términos de los múltiplos de salario mínimo. Para esos individuos, siempre se busca un ingreso dentro de ese intervalo. Estos individuos representan cerca del 50% del total de trabajadores con ingreso inválido. Es decir, en los métodos utilizados se hace primero el pareamiento dentro de cada grupo de múltiplos de salario mínimo (además de las características observables que se mencionan abajo) y después, para los individuos restantes, el pareamiento general.

Los cálculos de ingresos individuales se realizan utilizando a los trabajadores por remuneración. En este caso únicamente se imputan ingresos para las personas que trabajan por remuneración, pero que deciden no reportar ingresos. Por otro lado, cuando se calculan ingresos familiares (suma individual de ingresos por familia) se incluye a los trabajadores sin pago, desempleados y fuera de la fuerza laboral. A los trabajadores sin pago no se les imputa ingreso, ya que su ingreso verdadero es cero, de igual forma para los otros miembros del hogar sin trabajo.

El método de puntajes de propensión utilizó como características observables un polinomio cúbico de edad, variables dicotómicas de educación (primaria incompleta, primaria completa, secundaria completa, preparatoria completa y universidad completa), estado civil casado, mujer, trabajador formal, vive en sector rural, trabajador de tiempo completo (más de 35 horas a la semana), geográficas (una por cada área metropolitana definida en ENOE) e interacciones entre algunas variables. Aunque se estima un único puntaje de propensión, el pareamiento se realiza diferenciado por género.

El método *hot-deck* utiliza variables dicotómicas de edad (12 grupos: menores de 20, 20-24 y así sucesivamente, hasta 70 años y más de 70), área metropolitana, nivel educativo, trabajador formal, vive en sector rural, trabajador de tiempo completo. Como este método es muy demandante en términos de que tiene que haber individuos con ingresos válidos e inválidos en cada combinación posible, se realiza una segunda imputación para aquellos trabajadores donde no se pudo hacer el pareamiento. En este segundo proceso, se eliminan las variables de trabajo formal y tiempo completo. La gran diferencia entre este método y el anterior en términos de características observables es que el método de *hot-deck* busca un pareamiento por área metropolitana.

El método de la mediana más “ruido” utiliza las mismas variables categóricas que el método hot-deck, así como el mismo procedimiento. Finalmente, el método de pareamiento por promedios predictivos utiliza las mismas variables que el método de pareamiento por puntajes de propensión, así como los grupos de pareamiento (por género).<sup>10</sup>

Todos los métodos, con excepción del método de pareamiento por puntajes de propensión, utilizan imputaciones múltiples. El procedimiento se repite diez veces; en cada repetición se calcula el estadístico de interés (media, desigualdad), y después se obtiene el promedio entre los estadísticos obtenidos. No se incluyen resultados de imputación única, dado que los resultados son muy similares.

### *3.1 Ingresos individuales*

La figura 3, panel A, muestra los resultados del promedio de ingreso laboral mensual por método de imputación y se incluye, además, el valor reportado.<sup>11</sup> El panel B muestra los resultados en términos relativos cuando se compara con el periodo base (2005:1). El panel A muestra que el ingreso promedio observado en 2005:1 era de \$5,500, mientras que para 2012:3 fue de cerca de \$4,800. Una disminución de 12.7%, en términos reales. El panel también muestra los resultados para diferentes métodos de imputación. En general, los resultados son muy similares entre sí. El método *hot-deck* implica que el ingreso laboral fue de \$5,600 en 2005:1 y \$5,100 en 2012:3, es decir, una disminución del ingreso real de 9%.

El panel B muestra las tendencias del ingreso con respecto al periodo base 2005:1. La gráfica es clara con respecto a que, antes de 2008, la tendencia de los ingresos (reportados o imputados) es la misma. Sin embargo, para 2008 en adelante se tiene una diferente tendencia; lo cual es consistente con los cambios de tendencia mostrados en la figura 2, con respecto al porcentaje de trabajadores que declaran ingreso inválido en términos educativos y de formalidad. Es decir, el panel muestra que los ingresos no debieron de haber caído tanto como lo observado.

El resultado principal de la figura 3 es que los ingresos promedios observados, sobreestiman la caída en el ingreso real verdadero (individuos

---

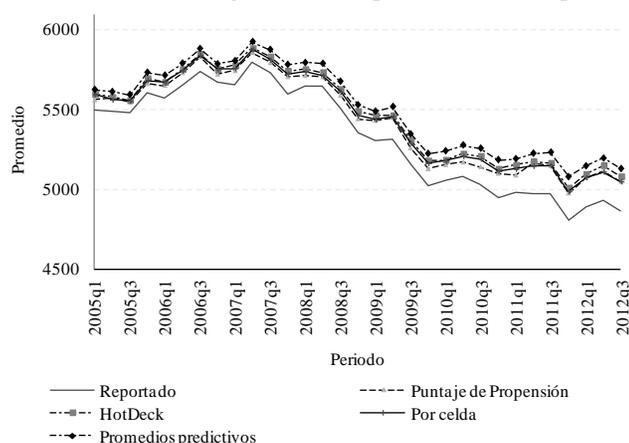
<sup>10</sup> En los métodos del pareamiento del puntaje de propensión y de la mediana no fue posible imputar todos los valores de ingreso inválido. Sin embargo, menos del 1% de los trabajadores sigue teniendo un ingreso inválido después del método de imputación. Dado que es un porcentaje pequeño no debería de afectar en las estimaciones. En los resultados mostrados abajo, todos los métodos obtienen el mismo resultado en promedio.

<sup>11</sup> Esta muestra incluye únicamente a trabajadores por remuneración. Los trabajadores sin pago no son incluidos en esta muestra.

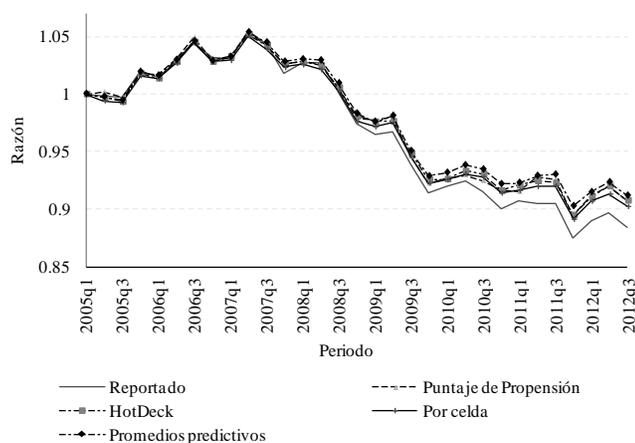
con ingreso observado e individuos con ingreso imputado). El ingreso verdadero es cercano a 4% mayor que el ingreso observado en 2012. Entonces, omitir ese ingreso podría tener consecuencias importantes para mediciones de desigualdad y de pobreza.

Figura 3  
**Promedio de ingreso laboral individual: 2005-2012.**

A. Promedio de ingreso laboral por método de imputación



B. Razón de ingreso laboral con respecto al periodo 2005:1



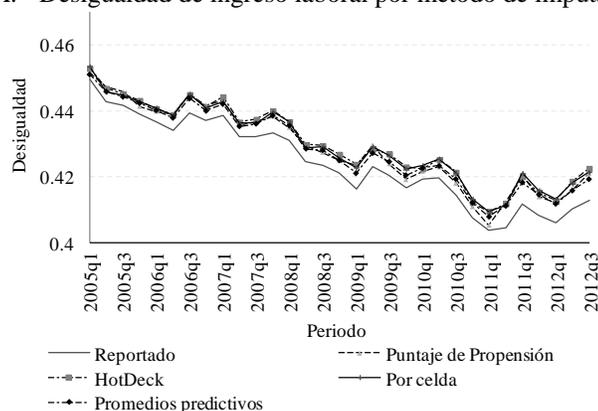
Nota: A y B se graficó para los trimestres 1 y 3. Los métodos de imputación aparecen explicados en el texto. Ingresos en pesos constantes del segundo trimestre de 2012. La muestra incluye únicamente trabajadores por remuneración, por tanto, no incluye trabajadores sin pago. Fuente: cálculos realizados por el autor con datos de la ENOE.

### 3.2 Desigualdad individual

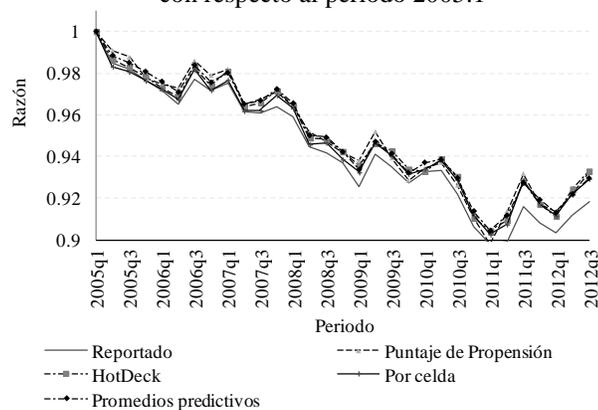
La figura 4 muestra ahora los resultados, en términos de la desigualdad observada y verdadera, por método de imputación en el ingreso laboral individual. La desigualdad se calcula con el coeficiente de Gini. El panel A muestra que la desigualdad ha caído durante el periodo 2005-2012.

Figura 4  
**Desigualdad del ingreso laboral individual: 2005-2012.**

A. Desigualdad de ingreso laboral por método de imputación



B. Razón de desigualdad de ingreso laboral con respecto al periodo 2005:1



Nota: A y B se graficó para los trimestres 1 y 3. Los métodos de imputación aparecen explicados en el texto. La muestra incluye únicamente trabajadores por remuneración; por tanto, no incluye trabajadores sin pago.

Fuente: cálculos realizados por el autor con datos de la ENOE.

La diferencia entre desigualdad reportada y desigualdad verdadera es mayor en 2012, que en periodos anteriores. Esto también puede observarse en el panel B. La desigualdad reportada cae cerca de 8% en el periodo; mientras que la verdadera, únicamente 7%. Si comparamos la desigualdad reportada y la imputada (Hot-Deck) en 2012:3, la desigualdad con el método de imputación es 2.4% mayor que la observada.

Los resultados en las figuras 3 y 4 llevan a la conclusión de que incluir a los trabajadores que no declaran ingresos brinda una fotografía del mercado laboral más precisa. Por ejemplo, los ingresos son mayores que los observados, no se presenta una caída tan precipitosa como la observada, así como la desigualdad no ha caído tanto como se aprecia. Sin embargo, los cálculos anteriores se han obtenido en el nivel individual. Dado que CONEVAL y otros estudios utilizan el ingreso familiar (la suma de ingresos individuales en el nivel hogar), el problema de omitir a los trabajadores sin ingreso o bien imputarles un valor de cero puede llevar a conclusiones imprecisas. En la siguiente sección se analizan las tendencias del ingreso familiar, y de pobreza en el nivel hogar, cuando se imputa el ingreso individual.

#### **4. Implicaciones en la medición de pobreza**

Dado que los distintos métodos obtienen el mismo resultado, esta sección se limita a utilizar únicamente el método de *hot-deck* para calcular el ingreso laboral familiar. Este ingreso está definido como la suma de ingresos laborales individuales. A diferencia de la sección anterior, en esta sección se incluye a todas las personas en la ENOE: trabajadores (por remuneración y sin pago), desempleados y los que están fuera de la fuerza laboral. La imputación de ingresos laborales se realiza únicamente para los trabajadores por remuneración, pero que deciden no reportar ingresos. Los trabajadores sin pago, los desempleados y las personas fuera de la fuerza laboral tienen un ingreso cero. Se utiliza como escala de equivalencia el tamaño de hogar para calcular en términos *per cápita*, dado que es como CONEVAL calcula la pobreza laboral, pero si se utilizan otras escalas de equivalencia se obtienen resultados similares. Asimismo, se utilizó imputación múltiple con diez (10) repeticiones.<sup>12</sup>

---

<sup>12</sup> Realizamos el ejercicio de imputación única y los resultados fueron prácticamente idénticos a los de imputación múltiple.

#### 4.1 Ingreso familiar per cápita

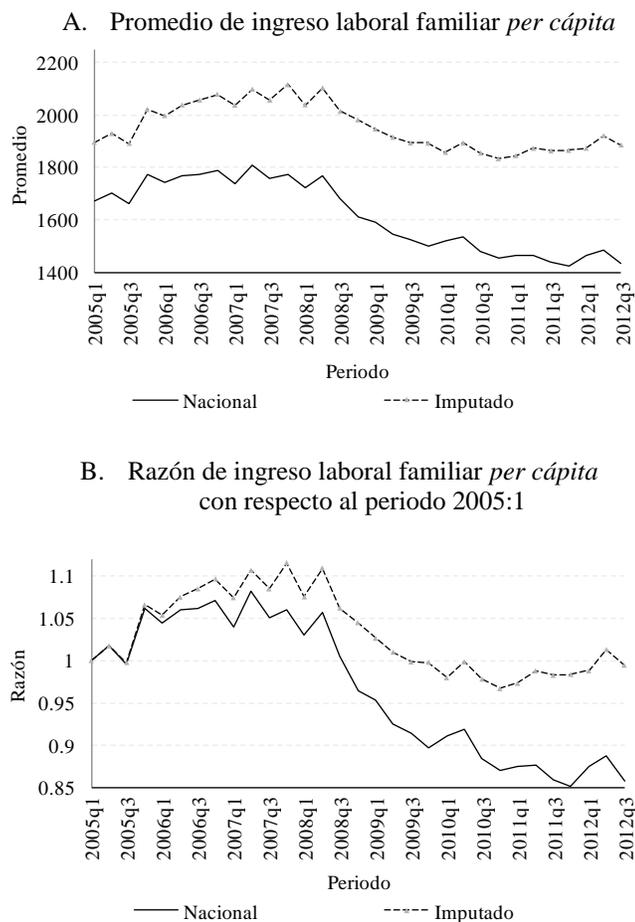
La figura 5 muestra los resultados para el ingreso laboral familiar *per cápita*. Los resultados de imputación son contundentes. El ingreso familiar *per cápita* verdadero es sustancialmente mayor que el observado. Por ejemplo, en el tercer trimestre de 2012, el ingreso familiar *per cápita* reportado fue de cerca de \$1,400, mientras que el ingreso verdadero fue de cerca de \$1,900. Esta diferencia implica que el ingreso verdadero es 36% mayor que el observado. El panel B muestra que el ingreso familiar *per cápita* verdadero (en pesos constantes) tiene el mismo valor aproximadamente en 2005, primer trimestre, y 2012, tercer trimestre. Mientras que el ingreso reportado en ese mismo periodo ha caído cerca de 15%.

La diferencia entre el ingreso familiar reportado y verdadero (36% en 2012) es mayor que en el caso de ingreso individual (4%). Existen varias razones para explicar esa diferencia.<sup>13</sup> Primero, el ingreso familiar incluye a todas las familias en la ENOE y no únicamente a los trabajadores por remuneración, como es el caso del ingreso individual. Esto implica que el porcentaje de zeros se disminuye considerablemente. Por ejemplo, el porcentaje de hogares con ingreso cero, debido a que nadie en el hogar tiene un trabajo remunerado (ya sea porque está formado por trabajadores sin pago o fuera de la fuerza laboral), es de aproximadamente de 11% en todo el periodo. En cambio, el porcentaje de hogares conformado por individuos que únicamente tienen un trabajo remunerado, pero que deciden no declarar ingresos, cambia de 6% en 2005:1 a 13% en 2012:3.

---

<sup>13</sup> El ingreso familiar promedio puede ser descompuesto como  $\bar{W} = \theta_{reporta}\bar{W}_{reporta} + \theta_{inválido}\bar{W}_{imputado} + \theta_{sinpago}\bar{W}_{sinpago}$ , donde cada  $\theta$  es la proporción de hogares con ingreso reportado, imputado o sin pago tal que  $\theta_{reporta} + \theta_{inválido} + \theta_{sinpago} = 1$ . Cada  $\bar{W}$  se refiere al ingreso promedio reportado, imputado o de los trabajadores sin pago. Dado que el ingreso de los trabajadores sin pago es cero, el ingreso familiar se descompone únicamente en los que reportan y en los inválidos. Si a los trabajadores con ingreso inválido se les imputa un valor de cero cuando en realidad tiene un valor positivo, entonces, por definición el ingreso familiar disminuye. Por tanto, el ingreso familiar aumentará dependiendo de la proporción de hogares con ingreso inválido, así como, según la magnitud del ingreso imputado.

Figura 5  
**Promedio de ingreso laboral familiar *per cápita*: 2005-2012.**



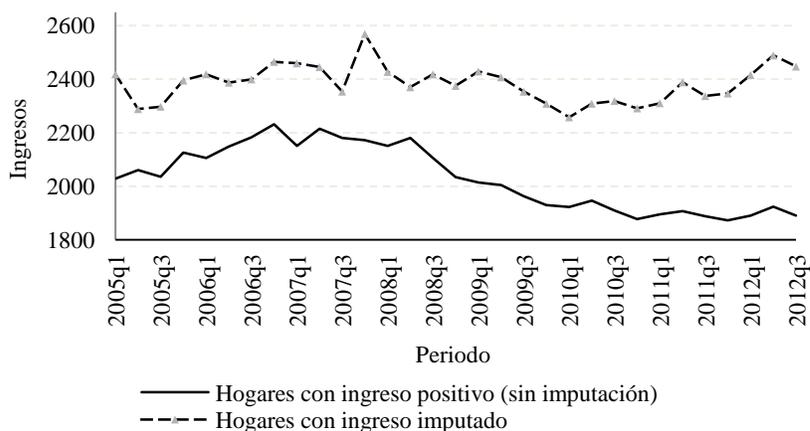
Nota: A y B se graficó para los trimestres 1 y 3. Se utiliza el método de imputación Hot-Deck, imputación múltiple 10 repeticiones. Ingreso en pesos constantes del segundo trimestre de 2012. La muestra incluye a todas las familias en la ENOE, incluyendo las familias con ingreso cero. Aunque la línea indicada como “Nacional” es calculada por el autor, se utilizan los programas provistos por CONEVAL y, por tanto, se obtienen los mismos números que CONEVAL (2012).

Fuente: cálculos realizados por el autor con datos de la ENOE.

Segundo, la tendencia del promedio de ingresos para los hogares con ingreso inválido y válido ha cambiado. La figura 6 muestra el promedio del ingreso familiar *per cápita* para los hogares con ingreso imputado únicamente y para

los hogares con ingreso positivo excluyendo a los hogares con ingreso imputado. La figura 5 muestra que el ingreso real reportado ha caído 15%, pero ese estadístico incluye a los hogares con ingreso cero o ingreso faltante. Una vez que se excluyen esos hogares, la figura 6 muestra que el ingreso real promedio ha caído 7% de 2005 a 2012. No ha sucedido lo mismo para los hogares con ingreso imputado. Además, la brecha en el promedio de ingreso ha aumentado entre los hogares que sí reportan ingresos y los que deciden no reportar ingresos.

Figura 6  
**Ingresos Laborales Familiares *per cápita* (Excluye hogares sin ingreso):  
 2005-2012**



Nota: se graficó para los trimestres 1 y 3. Se utiliza el método de imputación Hot-Deck, imputación múltiple 10 repeticiones. La muestra de hogares con ingreso positivo pero sin imputación incluye la muestra total de hogares en la ENOE, pero excluye a los que CONEVAL declara con ingreso cero (hogares con trabajadores sin pago y hogares con trabajo por remuneración, pero que no declara ingresos). La muestra de hogares con ingreso imputado incluye únicamente los hogares que tuvieron imputación de ingreso. El porcentaje de hogares con ingreso cero que están formados por trabajadores sin pago o fuera de la fuerza laboral es de aproximadamente 11% en todo el periodo. El porcentaje de hogares con ingreso imputado es de 6% en 2005:1 y sube a 13% en 2012:3.

Fuente: cálculos realizados por el autor con datos de la ENOE.

Por tanto, la diferencia entre el ingreso familiar *per cápita* reportado (incluyendo ingresos ceros de trabajadores sin pago y de los que no deciden reportar ingresos) y el ingreso familiar *per cápita* verdadero (incluye la imputación) se debe a dos razones. Principalmente, se debe a un aumento en el porcentaje de hogares que ha decidido no declarar ingresos, aun cuando ese hogar tiene trabajadores por remuneración. Esto ocasiona, por construcción, que el ingreso promedio nacional disminuya. Y segundo, los

hogares que deciden no reportar ingresos son diferentes a los que sí reportan algún ingreso: tienen más educación y por tanto un mayor ingreso en promedio. Por consiguiente, si se le asigna un ingreso cero a los hogares que deciden no declarar ingreso se está subestimando seriamente el ingreso verdadero de los hogares. En la siguiente subsección analizamos estas implicaciones en el cálculo del índice de la tendencia laboral de la pobreza elaborado por CONEVAL (2010).

#### 4.2 *Índice de la Tendencia Laboral de la Pobreza*

La figura 7 incluye la tendencia laboral de la pobreza: tanto la medición oficial por CONEVAL (2010), como la calculada con el ingreso imputado. La pobreza laboral se determina si la suma de ingresos laborales de los miembros del hogar dividida por el tamaño de hogar es menor que la línea de pobreza correspondiente. La línea de pobreza la determina CONEVAL, y se refiere al ingreso mínimo requerido para satisfacer las necesidades de alimentación de acuerdo con la canasta básica.<sup>14</sup> Es importante mencionar también que la definición de pobreza laboral no ha cambiado en el periodo de estudio. Mientras la medición de CONEVAL indica un aumento de la pobreza laboral de 42% a 53% en el periodo 2005-2012, la pobreza laboral con ingresos imputados aumentó únicamente de 36% a 40%, un aumento mucho menor al observado. Cuando se construye el índice de la tendencia laboral de la pobreza, el CONEVAL encuentra que la pobreza ha aumentado un 25% en el periodo 2005-2012 (Coneval, 2012, p. 4). Sin embargo, cuando se imputan los ingresos inválidos se encuentra que la pobreza solo ha aumentado un 12%, es decir menos de la mitad que la cifra oficial.

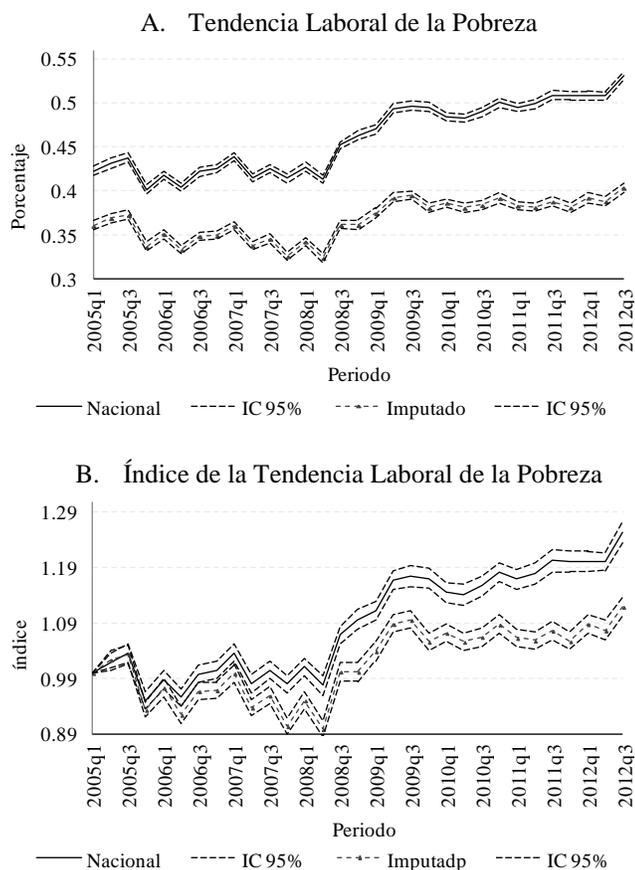
Para construir intervalos de confianza de estos cálculos, se utiliza la metodología de remuestreo (*bootstrap*, para más detalles de esta metodología véase Cameron y Trivedi, 2005, y Chernick, 2008) con 150 replicaciones.<sup>15</sup> Los intervalos de confianza al 95% son pequeños, de tal forma que los valores del índice reportado y verdadero son estadísticamente diferentes.

---

<sup>14</sup> CONEVAL (2010) menciona: “La línea de bienestar mínimo permite identificar a la población que, aun haciendo uso de todo su ingreso en la compra de alimentos, no podría adquirir los productos de una canasta básica alimentaria” (p. 13).

<sup>15</sup> El procedimiento de remuestreo es sumamente tardado. El procedimiento tardó aproximadamente 80 horas en terminar. Se aplica el remuestreo no paramétrico con estratificación en el sector rural y en los hogares con trabajadores sin pago.

Figura 7  
**Índice de la Tendencia Laboral de la Pobreza: 2005-2012**



Nota: A y B se graficó para los trimestres 1 y 3. Se utiliza el método de imputación Hot-Deck, imputación múltiple 10 repeticiones. Intervalos de confianza calculados con el método de remuestreo no paramétrico con 150 replicaciones. La muestra incluye a todas las familias en la ENOE, incluyendo las familias con ingreso cero. Aunque la línea indicada como “Nacional” es calculada por el autor, se utilizan los programas provistos por CONEVAL, y por tanto se obtienen los mismos números que CONEVAL (2012).

Fuente: cálculos realizados por el autor con datos de la ENOE

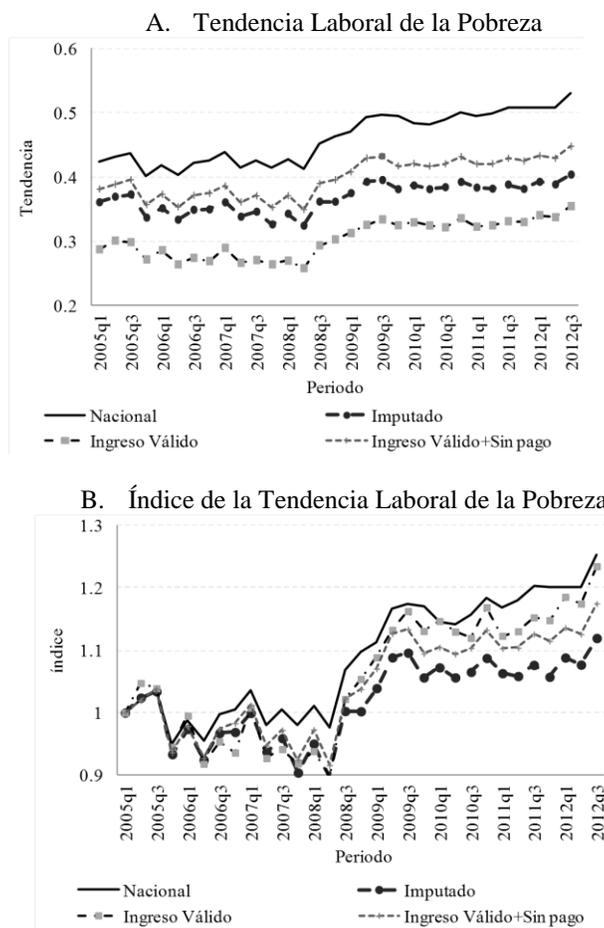
El resultado de la figura 7 panel A puede interpretarse en números absolutos, en lugar de porcentajes. Utilizando la ENOE del primer trimestre del año 2005, INEGI reporta que el país tenía 103,597,366 habitantes, y para el tercer trimestre de 2012, se reporta que el país tenía 115,296,767 habitantes.

Por tanto, utilizando los porcentajes de pobreza de CONEVAL, en 2005 se puede decir que el país tenía 42.8 millones de pobres de ingreso laboral. Sin embargo, utilizando el ingreso imputado se encuentran 37.3 millones de personas en pobreza de ingresos laborales para ese mismo año. Para el tercer trimestre de 2012, las diferencias son mayores: los porcentajes publicados por CONEVAL implican que existen 61.1 millones de pobres por ingresos laborales, pero usando el ingreso imputado se encuentra que existen solo 46.5 millones de pobres por ingresos laborales. Dado que el número de individuos que no reporta ingresos aumentó principalmente entre las personas con mayor educación, y dado que CONEVAL toma a esas personas como ingreso cero, se está sobreestimando tanto el número de pobres como el crecimiento en la pobreza. Los porcentajes de CONEVAL implican un aumento de la pobreza laboral de 17 millones de personas en el periodo 2005 a 2012, pero utilizando el ingreso imputado se tiene un incremento de 9.2 millones de personas en ese periodo. Asimismo, para el tercer trimestre de 2012, existen 8.1 millones de personas que son declarados como pobres; pero, de acuerdo con método de imputación, no lo son.

Las figuras 5, 6 y 7 resaltan la importancia de imputar los ingresos inválidos. La proporción de trabajadores que no declara ingresos ha ido en aumento, y su composición está cambiando de tal manera que cada vez más trabajadores con educación superior evitan dar información de ingreso. Esta información es importante, tanto que no tomarla en cuenta sesga sustancialmente las cifras oficiales de pobreza y del ingreso familiar *per cápita*. Los resultados anteriores implican que: las instituciones oficiales en México deben considerar imputar los ingresos inválidos siguiendo la misma metodología que en este artículo, o alguna cercana.

Por otro lado, la figura 8 muestra que el ejercicio antes presentado es robusto, cuando se elimina a hogares con individuos con ingresos inválidos. Es decir, aquí se elimina todo hogar que tenga al menos un individuo que no reporta ingresos (inválidos y sin pago). Recordemos que CONEVAL asigna un ingreso cero a los individuos que no reportan ingresos. También se añade a la figura qué pasa cuando se incluye en el cálculo a individuos sin pago, así como el método de imputación mencionado. El incremento en la pobreza laboral con el cálculo de imputación es el menor de todos los casos. Esto implica que es muy importante corregir en la muestra por los ingresos inválidos.

Figura 8  
**Robustez de la imputación al eliminar hogares con individuos con ingresos inválidos**



Nota: A y B se graficó para los trimestres 1 y 3. Se utiliza el método de imputación Hot-Deck, imputación múltiple 10 repeticiones. La línea con ingreso válido se refiere al cálculo que excluye de la muestra a aquellos hogares con al menos 1 integrante con ingresos inválidos (es decir, no reporta ingresos ese individuo pero reporta tener un trabajo remunerado) y aquellos hogares con únicamente trabajadores sin pago. La línea con ingreso válido y sin pago se refiere a que se incluyen los hogares con trabajadores sin pago, pero se siguen excluyendo hogares con algún miembro con ingreso inválido. Aunque la línea indicada como “Nacional” es calculada por el autor, se utilizan los programas provistos por CONEVAL, y por tanto se obtienen los mismos números que CONEVAL (2012).

Fuente: cálculos realizados por el autor con datos de la ENOE.

### Conclusiones y recomendaciones

En este artículo, se argumenta que asignar ingresos cero para las personas que deciden no declarar ingresos en México ha dejado de ser válido. La Encuesta Nacional de Ocupación y Empleo muestra que, de 2005 a 2012, la proporción que no declara ingresos pero trabaja remuneradamente con horas positivas de trabajo, aumentó de 10% a 20%. La mayoría de los trabajos en economía laboral y los cálculos de pobreza oficiales en México no corrigen por ese posible sesgo. En este artículo, se discuten y analizan diversos métodos de imputación con su aplicación en el cálculo de pobreza y desigualdad.

La metodología compara cuatro métodos: pareamiento por puntajes de propensión, método de hot-deck, método de asignación de la mediana más un ruido, y finalmente el método de pareamiento por puntaje predictivo. Los cálculos de ingreso, pobreza y desigualdad son robustos al uso de diferentes métodos. Por tanto, se recomienda aquel de más fácil aplicación. En este artículo, se recomienda el uso del método de *hot-deck* por su facilidad de aplicación, y tiempo estimado de cómputo y ejecución.

Las implicaciones del método de corrección de ingresos son las siguientes. Primero, para el año 2012 el ingreso individual está subestimado en aproximadamente 4%. Segundo, la caída en la desigualdad de ingresos individuales está sobreestimada. De considerarse toda la muestra de trabajadores remunerados, la verdadera desigualdad sería 2.4% mayor que lo que es ahora. Tercero, el ingreso familiar está subestimado en 36%. En particular, se encuentra que el ingreso familiar *per cápita* verdadero, tiene el mismo valor real en el año 2005 primer trimestre y el 2012 tercer trimestre. Mientras que el ingreso reportado en ese mismo periodo ha caído cerca de 15% en términos reales. Cuarto, y más importante, la pobreza laboral como la calcula CONEVAL está sobreestimada. Cuando se construye el índice de la tendencia laboral de la pobreza, el CONEVAL (2012) encuentra que la pobreza ha aumentado un 25% en el periodo 2005-2012. Sin embargo, cuando se imputan los ingresos inválidos se encuentra que la pobreza laboral solo ha aumentado un 12%. Tanto la subestimación de ingresos familiares como la sobreestimación de la pobreza laboral, se deben principalmente a que se toman como ingresos cero los ingresos de personas que sí tienen un ingreso positivo (pero deciden no reportarlo). Esto causa una seria sobreestimación en la pobreza del país, y en la tendencia observada. Los resultados de este artículo sugieren que las instituciones encargadas de medir pobreza y asignar recursos con base en ella, como por ejemplo CONEVAL y SEDESOL, deben incluir en sus cálculos un método de corrección para los trabajadores que deciden no declarar ingresos.

Por otro lado, el presente trabajo presenta una limitante importante. El trabajo se refiere únicamente a los cálculos estadísticos basados en la Encuesta Nacional de Ocupación y Empleo. Sin embargo, la medición oficial de la pobreza se realiza con la Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH). Y si bien la tendencia en la pobreza entre ambas encuestas es muy similar, para poder hacer inferencias con la medición de pobreza oficial habría que analizar el problema de ingresos inválidos en la ENIGH. Sin embargo, ese procedimiento representa retos importantes dado que los ingresos en la medición oficial incluyen no solo ingresos laborales sino otros ingresos también (por ejemplo: remesas, ingreso no monetario, entre otros). Por tanto, un procedimiento correcto sería imputar todas las fuentes de ingreso y no únicamente el ingreso laboral. Definir esos criterios, así como una comparación de métodos de imputación para diferentes encuestas, es una labor para futuras investigaciones.

### Referencias

- Allison, P. D. (2001). *Missing Data*. Beverly Hills, California: Sage Publications.
- Allison, P. D. (2012). *Modern Methods for Missing Data*. Disponible en internet <http://www.amstat.org/sections/srms/webinarfiles/ModernMethodWebinarMay2012.pdf>, último acceso 28 de junio de 2013.
- Andridge, R. R. y Little, R. J. A. (2010), "A Review of Hot Deck Imputation for Survey Non-response". *National Institutes of Health*, 78(1), 40-64.
- Caliendo, M., y Kopeinig, S. (2008). "Some Practical Guidance for the Implementation of Propensity Score Matching". *Journal of Economic Surveys*, 22(1), 31-72.
- Cameron, A. C. y Trivedi, P. K. (2005). *Microeconometrics Methods and Applications*. New York: Cambridge University Press.
- Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and researchers*. Segunda edición, Hoboken, New Jersey: Wiley-Interscience.
- CONEVAL (2010). "Tendencias Económicas y sociales de corto plazo y el Índice de la tendencia laboral de la pobreza (ITLP)". *Consejo Nacional de Evaluación de la Política de Desarrollo Social*. México: Gobierno Federal. Disponible en: [http://www.coneval.gob.mx/cmsconeval/rw/resource/coneval/med\\_pobreza/TendencialaboralpobrezaCONEVAL.pdf?view=true](http://www.coneval.gob.mx/cmsconeval/rw/resource/coneval/med_pobreza/TendencialaboralpobrezaCONEVAL.pdf?view=true), último acceso diciembre 12, 2012.
- CONEVAL (2012). "Tendencias Económicas y sociales de corto plazo. Resultados Nacionales. Noviembre 2012". *Consejo Nacional de Evaluación de la Política de Desarrollo Social*. México: Gobierno Federal. Disponible en: <http://web.coneval.gob.mx/Informes/ITLP/TERCER%20TRIMESTRE%202012/ITLP%20NACIONAL%20noviembre%202012.pdf>, último acceso diciembre 12, 2012.
- Di Nardo, J., Fortin, N. M. y Lemieux, T. (1996). "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach". *Econometrica*, 64(5), 1001-1044.
- Durrant, G. B. (2005). "Imputation Methods for Handling Item-Nonresponse in the Social Sciences: A Methodological Review". *National Centre for Research Methods Working Paper Series*.
- ENIGH (2010). Encuesta Nacional de Ingreso-Gasto de los Hogares en México. *Instituto Nacional de Estadística, Geografía e Informática (INEGI)*. México.
- ENOE (2012). Encuesta Nacional de Ocupación y Empleo. *Instituto Nacional de Estadística, Geografía e Informática (INEGI)*. México.

- Haitovsky, Y. (1968). "Missing data in regression analysis". *Journal of the Royal Statistical Society, Series B*, 30, 67-82.
- Heckman, J. J. (1979). "Sample selection bias as a specification error". *Econometrica*, 47(1), 153-161.
- Humphries, M. (2012). *Missing Data & how to deal: an overview of missing data*. Disponible en: [http://www.utexas.edu/cola/centers/prc/\\_files/cs/Missing-Data.pdf](http://www.utexas.edu/cola/centers/prc/_files/cs/Missing-Data.pdf), último acceso 28 de junio de 2013.
- INPC (2012). Índice Nacional de Precios al Consumidor. *Banco de México*. México.
- Kennickell, A. B. (1998). "Multiple Imputation in the Survey of Consumer Finances". Working Paper, Board of Governors of the Federal Reserve System. Washington, D.C.
- Leuven, E. y Sianesi, B. (2012). *PSMATCH2: Stata Module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Disponible en: <http://ideas.repec.org/c/boc/bocode/s432001.html>, último acceso en diciembre 12, 2012.
- Little, R. J. A. (1988). "Missing-data adjustments in large surveys". *Journal of Business and Economics and Statistics*, 6(3), 287-296.
- Little, R. J. A. y Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. 2a edición. Wiley-Interscience,
- Mander, A. y Clayton, D. (2007). *HOTDECK: Stata Module to impute missing values using the hotdeck method*. Disponible en: <http://ideas.repec.org/c/boc/bocode/s366901.html>, último acceso en diciembre 12, 2012.
- Rodríguez-Oreggia, E. (2012). "Índice de la tendencia laboral de la pobreza. Consideraciones sobre adiciones al grupo de trabajadores con ingreso reportado". Mimeo.
- Rosenbaum, P. R., y Rubin, D. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". *Biometrika*, 70(1), 41-55.
- Rubin, D. B. (1976). "Inference and missing data (with discussion)". *Biometrika*, 63, 581-592.
- Rubin, D. B. (1996). "Multiple Imputation After 18+ Years". *Journal of the American Statistical Association*, 91(434), 473-489.
- Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. Wiley-Interscience.

- Rubin, D. B., y Schenker, N. (1986). "Multiple imputation for interval estimation from simple random samples with ignorable non-response". *Journal of the American Statistical Association*, 81(394), 366-374.
- Rubin, D. B. y Thomas, N. (1986). "Matching Using Estimated Propensity Score: Relating Theory to Practice". *Biometrika*, 52(1), 249-264.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London, Chapman and Hall.
- West, S., Butani, S. y Witt, M. (1990) "Alternative Imputation Methods for Wage Data". *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 254-259.
- Zhang, P. (2003). "Multiple Imputation: Theory and Method". *International Statistics Review*, 71(3), 582-592.