

**ESTIMACIÓN AJUSTADA DEL COMPORTAMIENTO DE LA INFLACIÓN EN
COLOMBIA ENTRE 1994/95 – 2006/07: UNA APROXIMACIÓN USANDO CURVAS DE
ENGEL**

JEISON GERARDO RODRIGUEZ BURBANO



**UNIVERSIDAD SANTIAGO DE CALI
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
PROGRAMA DE ECONOMÍA
SANTIAGO DE CALI
2019**

**ESTIMACIÓN AJUSTADA DEL COMPORTAMIENTO DE LA INFLACIÓN EN
COLOMBIA ENTRE 1994/95 – 2006/07: UNA APROXIMACIÓN USANDO CURVAS DE
ENGEL**

JEISON GERARDO RODRIGUEZ BURBANO

JOHYNER OBREGON MORALES, MG
Director de trabajo de grado

MODALIDAD
Trabajo final de investigación

LÍNEA DE INVESTIGACIÓN
Desarrollo empresarial y competitividad

UNIVERSIDAD SANTIAGO DE CALI
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES
PROGRAMA DE ECONOMÍA
SANTIAGO DE CALI
2019

TABLA DE CONTENIDO

INTRODUCCIÓN	11
1. ANTECEDENTES	12
2. PROBLEMA DE INVESTIGACIÓN	16
2.1 PLANTEAMIENTO DEL PROBLEMA	16
2.2 FORMULACIÓN DEL PROBLEMA	18
2.3 SISTEMATIZACIÓN DEL PROBLEMA	18
3. OBJETIVOS	18
3.1 OBJETIVO GENERAL	18
3.2 OBJETIVOS ESPECÍFICOS	19
4. JUSTIFICACIÓN	19
5. MARCO DE REFERENCIA	20
5.1 MARCO CONTEXTUAL	20
5.2 MARCO TEÓRICO	26
5.3 MARCO CONCEPTUAL	31
5.4 MARCO LEGAL	32
6. DISEÑO METODOLÓGICO	32
6.1 TIPO DE ESTUDIO	33
6.2 MÉTODO	33
6.3 TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN	36
6.4 ACTIVIDADES DE LA INVESTIGACIÓN	36
7. RESULTADOS DE LA INVESTIGACIÓN	37
7.1 ESTADÍSTICAS DESCRIPTIVAS	37
7.2 RESULTADOS Y DISCUSIÓN	38
8. CONCLUSIONES Y RECOMENDACIONES	44
REFERENCIAS	46
ANEXOS	48

LISTA DE TABLAS

Tabla 1. Estadísticas descriptivas	37
Tabla 2. Prueba de diferencia de medias para gasto proporción de alimentos	38
Tabla 3. Índices para 2006/07	39
Tabla 4. Resultados	41
Tabla 5. Elasticidades alimentos	43

LISTA DE GRÁFICAS

Gráfico 1 Inflación interanual 1994-2007.....	21
Gráfico 2 Inflación interanual principales ciudades de Colombia 1994-2007	22
Gráfico 3 Inflación interanual por clasificación 1994-2007	23
Gráfico 4 Inflación interanual según grupo de gastos 1994-2007	24
Gráfico 5 Tasa de intervención de política monetaria 1999-2007	25
Gráfico 6 Crecimiento PIB 1994-2007	26
Gráfico 7 Población de Colombia 1994-2007	26
Gráfico 8 Inflación ajustada.....	43

LISTA DE ANEXOS

Anexo 1. Salidas Stata estadísticas descriptivas	48
Anexo 2. Colinealidad y heterocedasticidad del modelo.....	52
Anexo 3. Corrección colinealidad por exclusión de variable ln precio relativo de alimentos.....	55
Anexo 4. Regresión errores robustos de White	56
Anexo 5. Distribución de los errores	57
Anexo 6. Prueba de endogeneidad.....	57
Anexo 7. Salidas Stata elasticidades	58
Anexo 8. Do file Stata.....	58

RESUMEN

El presente trabajo estima el sesgo en la inflación medido a través del IPC. La metodología utilizada consistió en hallar curvas de Engel, es decir, se estimó una demanda de la proporción del gasto en alimentos en función del gasto real de los hogares, donde además, se incluyó un grupo de controles asociados a las características sociodemográficas de cada hogar en 23 áreas metropolitanas. Los datos utilizados corresponden a la Encuesta de Ingresos y Gastos de los años 1994/95 y 2006/07, donde se contó con 11260 observaciones. Los resultados arrojan una sobreestimación promedio anual de la inflación de 1.61 puntos porcentuales, congruente con resultados de autores anteriores. De igual forma, se incluyó cálculos de elasticidades de gasto (0,51) y precio (compensada (0,19) y no compensada (0,35)) de la demanda, encontrando que los alimentos son bienes normales-básicos e inelásticos, tal como lo plantea la teoría económica.

Palabras clave: IPC, inflación, sesgo, curva de Engel y alimentos.

ABSTRACT

This document estimates the bias in the inflation measured through the CPI. The methodology used was to find Engel curves, that is, a demand was estimated for the share of food expenditure based on the real expenditure of households, where, in addition, a group of controls associated with the sociodemographic characteristics of each household was included in 23 metropolitan areas. The data used correspond to the Survey of Income and Expenses for the years 1994/95 and 2006/07, where there were 11260 observations. The results show an average annual overestimation of inflation of 1.61 percentage points, consistent with the results of previous authors. Similarly, calculations of elasticities of expenditure (0,51) and price (compensated (0,19) and not compensated (0,35)) were included, finding that food are normal-basic and inelastic goods, as it is posed by the economic theory.

Key words: CPI, inflation, bias, Engel's curves and food.

INTRODUCCIÓN

Contar con indicadores que reflejen correctamente el comportamiento de las principales variables que repercuten en la actividad económica en general, resulta siempre de interés común dada la enorme incidencia que tienen sobre la toma de decisiones; algunos ejemplos son el producto interno bruto, el desempleo, las tasas de interés, la inflación, etc. Un caso particular corresponde al nivel de precios de cualquier economía. Un cambio en el nivel de precios se traduce en cambios en la planificación de consumo y producción, por parte de agentes públicos y privados, así como una enorme repercusión sobre el diseño y aplicación de la política pública.

La manera convencional como se miden los cambios en el nivel de precios de cualquier economía consiste en estimar índices de Laspeyres, donde la variación porcentual que refleje dicho índice entre dos periodos de tiempo corresponde a la inflación de dicho periodo. Este indicador cobra un papel protagónico en la economía dado que basado en su comportamiento se ajustan salarios, jubilaciones, prestaciones de seguridad social, pago de alquileres, pago de intereses, precio de bonos, y decisiones de política monetaria y fiscal.

Por lo anterior, resulta evidente que el valor de este indicador debe capturar la realidad sin incurrir en sesgos de estimación, que permeen hacia un detrimento del bienestar social. Sin embargo, la literatura internacional reconoce que el índice de Laspeyres no es una buena medida de los cambios en el costo de vida, ya que presenta varios sesgos que no permiten capturar correctamente dicho cambio (Caicedo, 2000). Es decir, dada la estructura metodológica de los índices de Laspeyres es inherente la existencia de un sesgo en su construcción que se transmite hacia un sesgo en el cálculo de su variación, esto es, la inflación.

De esta manera, en este trabajo se pretende encontrar el sesgo en el costo de vida medido a través del Índice de Precios al Consumidor (en adelante IPC) en Colombia a través de la estimación de curvas de Engel entre 1994/95 y 2006/07 usando las Encuestas de Ingresos y Gastos de dichos años.

1. ANTECEDENTES

El primer avance importante que se produjo respecto al estudio del sesgo en la medición de la inflación, lo realizó una comisión de expertos estadounidenses en 1991 bajo encargo de la Oficina Nacional de Investigación Económica conocida desde entonces como la Comisión de Stigler (Gordon, 2000). Dicha comisión externa, tenía por objetivo mejorar las estadísticas de precios en Estados Unidos para informar a la oficina de presupuestos. Gordon (2000) señala que la comisión no se enfocó en el estudio del sesgo en el IPC, de hecho, estudió tanto el índice de precios al productor (IPP), como también algunos índices de precios agrícolas de la época. La comisión no realizó ninguna estimación numérica del sesgo existente en ninguno de los índices, sin embargo, afirmó que se encuentra consenso entre la mayoría de los economistas acerca de la existencia de un sesgo hacia arriba por parte de los índices de precios (Gordon, 2000).

Más adelante, Boskin, Dulberger, Gordon, Griliches & Jorgenson (1996), bajo encomienda del comité financiero del senado de Estados Unidos, cuantificaron medidas numéricas del sesgo en el IPC de dicho país. La hipótesis de partida radicaba en encontrar un sesgo hacia arriba del IPC estadounidense. Utilizaron datos de la oficina de estadísticas laborales de Estados Unidos entre 1995 y 1996. La metodología que usaron se describe así: dividieron el IPC en 27 categorías de bienes diferentes y calcularon el sesgo en cada una de ellas; para la estimación de dichos sesgos utilizaron regresiones hedónicas en algunas categorías y extrapolación de series en otras. Uno de los aspectos más relevantes que destacaron los autores, fue la dificultad de medir correctamente la inflación, especialmente por la complejidad económica inherentemente ligada a los mercados; donde cada día se introducen nuevos bienes, se mejoran la calidad de otros, se sustituyen bienes por servicios, entre otros. La comisión encontró que durante el periodo 1995-1996, el IPC tuvo un sesgo de 1.1%, con un rango plausible entre 0.8% y 1.6%. Dentro de las principales recomendaciones que la comisión realizó, se encuentra el cambio de índices de canasta fija por índices que permitan absorber el sesgo de sustitución, actualizar más frecuentemente la ponderación de los bienes y servicios que hacen parte de una canasta base, y la más importante, establecer un índice del costo de vida como verdadera medida de los precios de los consumidores.

Basado en los recientes descubrimientos de Boskin et al. (1996), Shiratsuka (1999) estimó el sesgo en el IPC para Japón en el periodo 1991-1994. El autor argumenta que la simplicidad de un índice

de Laspeyres dificulta la correcta medición de cambios en el costo de vida, por tanto, plantea la hipótesis de encontrar un sesgo hacia arriba en el IPC. De esta manera, enfatiza que el sesgo total es explicado por variaciones de los precios relativos, la aparición de nuevos bienes y la obsolescencia de otros, y especialmente, por cambios en la calidad de los bienes propiciados por la innovación tecnológica. Así, el autor reconoce 4 fuentes diferentes por la que se sobreestima el IPC en Japón: efecto sustitución de los bienes, efecto por cambios en la calidad, efecto de nuevos bienes y por problemas técnicos en la construcción estadística del IPC. El autor utiliza datos de la encuesta familiar de ingresos y gastos de la oficina de estadísticas de Japón, donde, al dividir el índice en 88 bienes diferentes, estima regresiones hedónicas para calcular índices hedónicos y superlativos. El documento encuentra que el índice en Japón fue sobreestimado en 0.9% por año, durante el periodo correspondiente entre 1991 y 1994.

Más tarde, Hamilton (2001) estimó curvas de demanda por alimentos en Estados Unidos entre 1974 y 1991. La metodología de estimar curvas de Engel para medir el sesgo en el IPC fue novedosa en aquel momento, ya que, hasta la fecha solo se había considerado extrapolar series y calcular índices alternativos para cuantificar dicho sesgo. El autor plantea que, a medida que pasa el tiempo, cualquier cambio en la participación del gasto en alimentos medido por el ingreso real (deflactado por el IPC), se puede atribuir como el sesgo total en el IPC; esta proposición se basa en la ley de Engel, la cual establece que, a mayor ingreso, menor gasto en alimentos (Kaus, 2013). Hamilton (2001) utilizó dos bases de datos, la primera fue el panel de dinámica de ingresos de la Universidad de Michigan, y la segunda, las estadísticas de área metropolitanas estándar proveniente de la oficina de estadísticas laborales de Estados Unidos; lo anterior durante el periodo comprendido entre 1974 y 1991. Para la estimación, el autor modeló un sistema casi ideal de demanda expuesto por Deaton & Muellbauer (1980), y lo calculó a través de mínimos cuadrados ordinarios. Dentro de las variables incluidas se encuentra una dummy para cada año, edad del esposo y esposa, número de hijos, tasa de desempleo, horas anuales trabajadas del esposo y esposa, educación del esposo y esposa, ingreso real, precio relativo de alimentos y el porcentaje de la comida adquirida en restaurantes. Encontró que el sesgo total fue de 2.5% entre 1974 y 1981, y disminuye aproximadamente a menos de 1% en los años posteriores. Además, encuentra que la diferencia en los resultados con ambas bases de datos no cambia significativamente.

De igual forma, Costa (2001) intentó cuantificar cuál fue el verdadero ingreso de los ciudadanos estadounidenses durante el periodo comprendido entre 1888 y 1994 al denotar la existencia del sesgo en la medición del IPC. Para lograrlo utilizó la Gran Encuesta Nacional de Consumidores del Departamento del Trabajo de los años 1888-90, 1917-19, 1935-36, 1960-61, 1972-73 y 1980-94. Así pues, estimó el sesgo en el IPC entre los periodos 1888-1935 y 1960-1994; la metodología que usó fue la de estimar curvas de Engel, es decir, estimó un sistema casi ideal de demanda donde la variable dependiente era la participación del gasto en alimentos explicada por el gasto real, precio relativo de alimentos, proporción de alimentos consumidos fuera del hogar y variables dicótomas por cada año de la encuesta. La regresión la realizó por mínimos cuadrados ordinarios. Encontró que el sesgo en el IPC fue de -0.1% entre 1888 y 1919, 0.7% entre 1919 y 1935; en la década de 1960 fue de 0.4%, para posteriormente aumentar a 2.7% entre 1972 y 1982, y finalmente disminuir a 0.6% entre 1982 y 1994. En promedio, el sesgo comprendido entre 1972 y 1994 fue de 1.6%, ligeramente superior al encontrado por Boskin et al., (1996), aunque va en línea con los rangos que la comisión fijó en su momento.

Beatty & Roed (2004), basados en los trabajos de Hamilton (2001) y Costa (2001), cuantificaron el sesgo presente en el IPC para Canadá a través de la estimación de curvas de Engel, para el periodo 1978-2000. Utilizaron microdatos de la Encuesta de Gastos para los años 1978, 1982, 1984, 1986, 1990, 1992, 1996, así como también de la Encuesta de Gastos Domésticos de 1997, 1998, 1999 y 2000. Su hipótesis de partida consistía en medir el bienestar de dos hogares con características sociodemográficas similares en dos momentos distintos del tiempo, asumiendo que la proporción del gasto dedicada a alimentos, consistía en una buena medida de bienestar. El método de estimación fue el mismo utilizado por Hamilton (2001) y Costa (2001), con la diferencia que lo hicieron a través de un modelo no paramétrico; donde incluyeron la proporción del gasto en alimentos, el ingreso real, el precio relativo de alimentos y variables binarias por año de encuesta. Dentro de los principales hallazgos, encontraron que el IPC sobreestima el verdadero costo de vida en 1.81% y 0.785%, para hogares unipersonales y hogares con más de dos integrantes, respectivamente.

Carvalho & Chamon (2006) intentaron cuantificar el verdadero ingreso per cápita de la economía brasilera, bajo un contexto de estanflación económica que se produjo a finales de la década de 1980 y comienzos de la década de 1990. Los autores estimaron curvas de Engel por alimentos para

el periodo comprendido entre 1987-2002 basándose en la Búsqueda de Presupuestos Familiares del Instituto Brasileiro de Geografía y Estadística de los años 1987/88, 1995/96, y 2002/03. Dentro de las variables utilizadas se encuentran la proporción del gasto en alimentos, la proporción del gasto en el hogar, la proporción del gasto en restaurantes, el precio relativo de alimentos, el ingreso real, el gasto real, variables categóricas que incluían rangos de edades (0-4, 5-9, 10-14, 15-19 y mayor a 20 años), sexo del jefe de hogar y una variable dicótoma que indicaba si el hogar tenía casa propia; esto último con el fin de determinar el efecto riqueza sobre la proporción del gasto en alimentos. Para su estimación usaron dos modelos, en primer lugar, calcularon un sistema casi ideal de demanda utilizando mínimos cuadrados en dos etapas, en segundo lugar, usaron una regresión no paramétrica para su estimación. Los autores encontraron que el incremento del ingreso real per cápita en los hogares utilizando su metodología fue de 4.5%, muy por encima del 1.5% obtenido por las oficinas de estadística de dicho país.

A nivel nacional son pocos los estudios que se han encargado de realizar estimaciones para el sesgo de medición en el IPC. El trabajo pionero corresponde a Caicedo (2000), quien cuantificó el sesgo por sustitución para Colombia en el periodo 1989-1998. Utilizó información suministrada por el DANE de 195 ítems diferentes, que correspondían al IPC desde diciembre de 1988 a diciembre de 1998. Para su estimación, y basado en trabajos de Diewert (1976), el autor calculó el IPC con diferentes fórmulas de números índices y computó el sesgo como la diferencia de cualquier índice numérico con un índice superlativo, que dotara, al menos en teoría, de racionalidad al consumidor ante cambios en los precios relativos de los bienes. Es decir, una vez que calculó de nuevo el IPC con los diferentes números índices (Laspeyres, Media Geometrica, Ideal de Fisher y Paasche), calculó la diferencia con un índice superlativo (Tornqvist) para obtener el sesgo por sustitución para el periodo en cuestión. El documento encontró que el sesgo de sustitución por año entre 1989 y 1998 fue de 0.7%.

A diferencia de Caicedo (2000), Langebaek y Caicedo (2007) estimaron el sesgo total del IPC en Colombia para el periodo comprendido entre 1984/85 y 1994/95 con base en la metodología propuesta por Hamilton (2001). El objetivo del estudio era cuantificar la evolución verdadera en el índice de costo de vida en Colombia. Es decir, calcularon curvas de demanda por alimentos en dos periodos distintos de tiempo, donde se supone que, si dos hogares con características demográficas e ingreso real similares tienen una proporción en alimentos diferente, dicha

diferencia es explicada por el sesgo en el IPC. Para esto, utilizaron las Encuestas de Ingresos y Gastos del DANE, para los periodos de referencia. Los autores encontraron dos resultados diferentes a partir de los supuestos utilizados en la metodología, y específicamente, en la relación existente entre el sesgo entre alimentos y el sesgo de los no alimentos. Así pues, utilizaron los resultados obtenidos por Caicedo (2000) del sesgo de sustitución como proxy del sesgo de alimentos y realizaron dos estimaciones diferentes a través de mínimos cuadrados ordinarios y mínimos cuadrados en dos etapas utilizando variables instrumentales. Dentro de las variables utilizadas se encuentra la proporción del gasto en alimentos, una variable binaria para el segundo año de la encuesta, el gasto real, el precio relativo de alimentos, el número de personas en el hogar, sexo jefe de hogar, edad jefe de hogar, escolaridad jefe de hogar, variables categóricas en 5 rangos de edades diferentes (0-4, 5-9, 10-15, 16-18, más de 18 años), número de trabajadores con remuneración, vivienda propia y varias variables binarias por región. El resultado, tomando como igual el sesgo entre alimentos y no alimentos, demuestra que la sobreestimación promedio anual del IPC estuvo entre 1.38 y 1.63 puntos porcentuales entre 1984/1985-1994/1995. Por el contrario, si se supone que el sesgo relativo de alimentos es proporcional al sesgo de sustitución, las estimaciones arrojaron una sobreestimación promedio anual de 1.44 a 1.69 puntos porcentuales para el mismo periodo.

Así pues, el aporte de la revisión bibliográfica para este trabajo estribó en la metodología, modelo econométrico y variables a utilizar. Basado en la literatura presentada en este literal se estimará el sesgo en el costo de vida medido a través del IPC usando curvas de Engel y modelando un sistema casi ideal de demanda. Las variables de interés serán la proporción del gasto en alimentos, el gasto real de los hogares, precio relativo de alimentos frente a los no alimentos, vivienda propia y variables asociadas al jefe de hogar (sexo, edad y educación). Finalmente, se espera encontrar que el IPC sobreestimo el costo de vida para Colombia entre los períodos 1995/95 – 2006-07.

2. PROBLEMA DE INVESTIGACIÓN

2.1 PLANTEAMIENTO DEL PROBLEMA

El IPC, utilizado como una herramienta para medir cambios en el nivel de precios tiene varias finalidades, de hecho, la OIT et al., (2006) establece que el IPC se utiliza ampliamente para indexar: salarios, jubilaciones, prestaciones de seguridad social, pago de alquileres, pago de intereses y precio de los bonos. De igual forma, se utiliza para elaborar otros índices que sirven para deflactar el gasto de consumo de los hogares, estimar paridades de poder adquisitivo entre países que permitan comparar sus niveles de consumo o para calcular tasas de interés real de países. También, al ser una medida representativa del nivel general de precios que permite calcular la inflación, el IPC tiene una incidencia directa en la toma de decisiones de los bancos centrales sobre temas de política monetaria, y especialmente, en aquellos que actúan bajo una política de inflación objetivo como lo es Colombia.

Por otra parte, Caicedo (2000) identifica tres implicaciones diferentes de sobreestimar el IPC: a) subvalorar el verdadero crecimiento de variables reales; b) incrementar la proporción de ajuste de ciertos indicadores; y c) distorsionar la toma de decisiones de los agentes. La primera hace referencia al proceso mediante el cual se deflactan algunas variables reales a través del IPC, en donde, sí el aumento en el denominador de un cociente es cada vez mayor, el resultado tendera a arrojar variables reales subvaloradas. El segundo, y como se mencionó anteriormente, hace referencia al uso del IPC para indexar diferentes variables. Por último, distorsiona las decisiones de consumo o producción de los agentes, que pueden conllevar a desequilibrios en el corto plazo de los mercados que a su vez significarían un detrimento del bienestar social y de la eficiencia económica.

Adicionalmente, Boskin & Jorgenson (1997) señalan las consecuencias presupuestarias históricas que tal sesgo puede generar en programas sociales, al generar sobrepagos a los beneficiarios adscritos a éste; como también en déficit fiscales, en donde, para Estados Unidos, se presume que una hipotética corrección de 0.5% en el índice del costo de vida, reduciría el déficit fiscal en aproximadamente 10% al final del periodo 1996-2000. Para aminorar este tipo de problemáticas, y debido a la imposibilidad de medir correctamente el costo de vida, Deaton (1998) propone indexar los pagos de seguridad social no sobre el IPC, sino sobre los incrementos en los salarios o en el consumo, además de utilizar encuestas de gastos con universos de estudio más amplios y de mayor frecuencia.

Así pues, y de acuerdo con Gordon (2000), se podrían resumir las implicaciones económicas de la siguiente manera: mala medición en estimaciones oficiales de producción, productividad, ingreso medio, déficits fiscales, salarios reales, rangos de pobreza, alteración de las tasas de crecimiento en ingresos fiscales, sobrecostos en programas sociales, decisiones equívocas por parte de bancos centrales, y todas las demás que se expusieron previamente en la presente sección.

De esta manera, resulta relevante intentar cuantificar este tipo de problemática bajo el contexto de un nuevo período de tiempo para determinar si el sesgo se ha mantenido en niveles cercanos a los encontrados por autores anteriores o si, por el contrario, se observa que dicho sesgo ha aumentado, así como también, todas las repercusiones que trae consigo.

2.2 FORMULACIÓN DEL PROBLEMA

¿Cuál fue el comportamiento ajustado de la inflación en Colombia entre 1994/95 – 2006/07 usando curvas de Engel?

2.3 SISTEMATIZACIÓN DEL PROBLEMA

- ¿Cómo ha sido el cambio en la estructura sociodemográfica de los hogares en Colombia entre 1994/95 – 2006/07?
- ¿Cuál ha sido el cambio en los precios del grupo de alimentos y no alimentos en Colombia entre 1994/95 – 2006/07?
- ¿Cuál fue el tamaño del sesgo del costo de vida medido por el IPC a través de la estimación de curvas de Engel para Colombia entre 1994/95 – 2006/07?

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Estimar el comportamiento ajustado de la inflación en Colombia entre 1994/95 – 2006/07 usando curvas de Engel.

3.2 OBJETIVOS ESPECÍFICOS

- Describir el cambio en la estructura sociodemográfica de los hogares en Colombia entre 1994/95 – 2006/07.
- Construir índices que permitan percibir los cambios en los precios del grupo de alimentos y no alimentos en Colombia entre 1994/95 – 2006/07.
- Calcular el tamaño del sesgo en el costo de vida medido por el IPC a través de la estimación de curvas de Engel para Colombia entre 1994/95-2006/07.

4. JUSTIFICACIÓN

Contar con una correcta medida del costo de vida medida a través del IPC resulta relevante para los agentes económicos dado que con éste se indexan salarios, jubilaciones, prestaciones de seguridad social, pago de alquileres, pago de intereses y precio de los bonos. Un posible sesgo en

dicho valor conduce a mala medición de estimaciones del PIB real, productividad, déficits fiscales, variables reales, sobrecostos en programas sociales, y especialmente, en decisiones de política pública como lo es la política fiscal y monetaria. Así pues, resulta pertinente resaltar la importancia de contar con información verídica que permita cuantificar cual ha sido el verdadero desarrollo del nivel general de precios en la economía colombiana.

De esta manera, la medición ajustada del costo de vida medida a través del IPC tiene una incidencia directa en toda la actividad económica y, por tanto, debe ser abordado sin dilación para evitar distorsiones en la medición del bienestar de la población en general, especialmente respecto a la ejecución de políticas públicas que se cimientan sobre este tipo de variables.

La investigación servirá para que instituciones, académicos, decisores de política pública y organismos internacionales conozcan la problemática que se puede generar cuando se intente medir variables reales y cuenten con una herramienta de ajuste. De igual forma, se pretende brindar información real a la oficina de estadísticas nacionales del comportamiento de dicho sesgo respecto a trabajos anteriores para el diseño eficiente de la política pública y su respectiva evaluación.

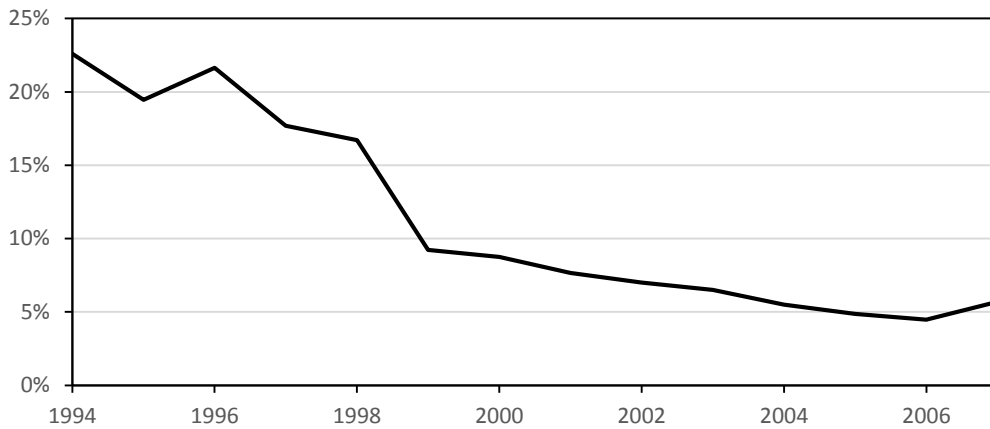
Respecto a la justificación empírico-teórica, se espera que el estudio aporte nueva información del sesgo existente en el IPC para determinar sí, desde trabajos anteriores hasta el resultado del presente trabajo, dicho sesgo ha aumentado, disminuido o se ha mantenido constante. De igual forma, se espera que sirva de referencia para determinar qué tan grande han sido los despropósitos que se han cometido, especialmente en términos de política públicas y política monetaria.

5. MARCO DE REFERENCIA

5.1 MARCO CONTEXTUAL

El marco donde se desarrolla la investigación corresponde a la Republica de Colombia. Más específicamente, se presta vital atención en los hogares de Colombia con ciertos patrones de consumo respecto a la demanda por bienes y servicios, y encontrar así, el verdadero aumento en el costo de vida en el periodo señalado.

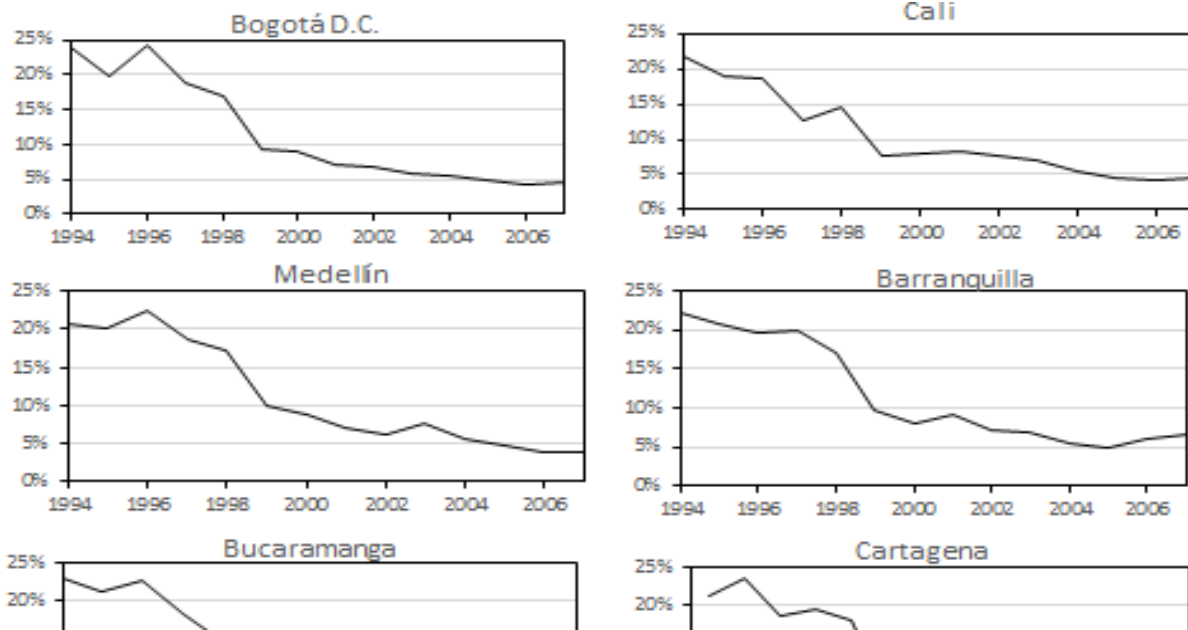
Históricamente Colombia no ha sufrido hiperinflaciones, sin embargo, si ha tenido que afrontar inflaciones altas, especialmente en la década de los setenta (Cárdenas, 2009). Respecto al periodo de estudio en cuestión, el gráfico 1 muestra que en 1994 la inflación interanual fue 22.59%, mientras que para el año 2007 fue de 5.69%. Esta constante disminución de la inflación tiene un precedente anterior al periodo comprendido en este estudio. Con la promulgación de la constitución política de Colombia de 1991, se dotó al Banco de la Republica de una mayor autonomía, y éste, en busca de preservar el poder adquisitivo, empezó un nuevo mecanismo de política monetaria conocido como inflación objetivo, el cual lo realiza a través de la tasa de intervención. Nótese que desde el siglo XXI, la inflación se redujo a un dígito.



Fuente: DANE

Gráfico 1 Inflación interanual 1994-2007

Por otra parte, resulta pertinente observar el comportamiento de la inflación de cada una de las principales ciudades del país.



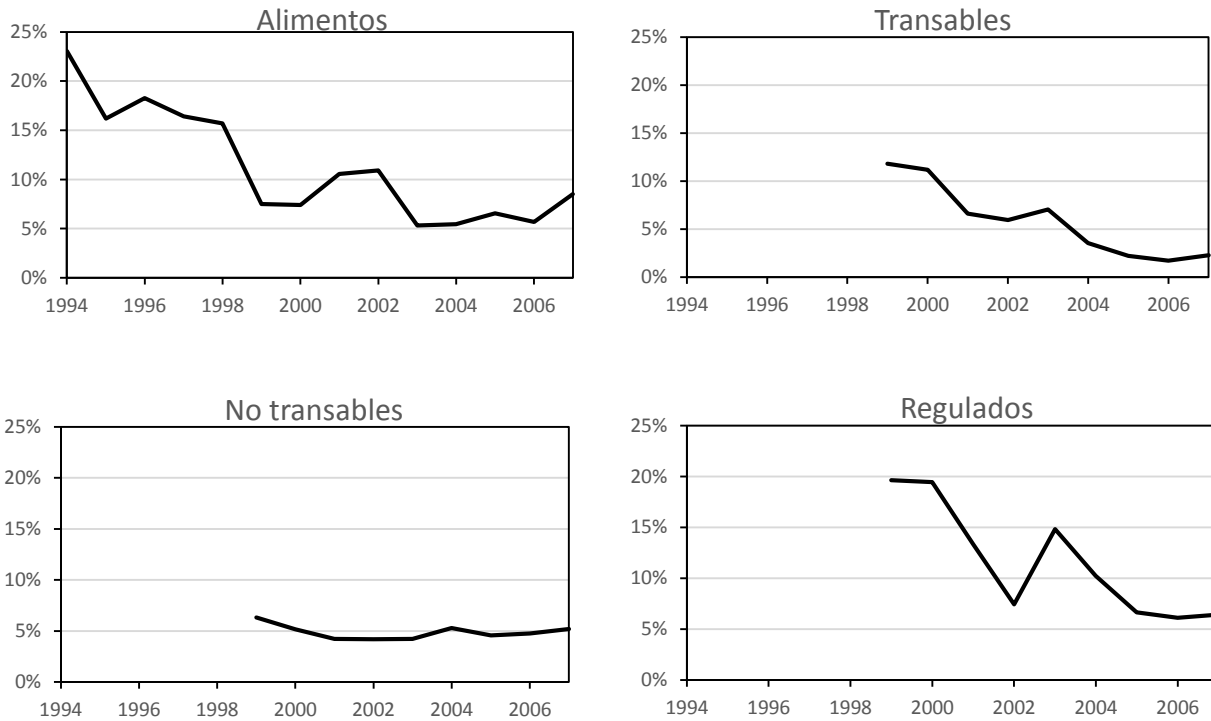
Fuente: DANE

Gráfico 2 Inflación interanual principales ciudades de Colombia 1994-2007

En todas las 10 ciudades que se muestran en el gráfico 2, la inflación cae constantemente a lo largo de la muestra. Las ciudades con la inflación más alta y más baja al comienzo de la muestra fueron Bogotá y Cúcuta con 24.08% y 19.3% respectivamente. Por otra parte, al final de la muestra las ciudades con la inflación más alta y más baja fueron Barranquilla con 6.62% y Medellín con 3.77%. Resulta relevante destacar la información anterior dado que muchas ciudades son el centro de áreas metropolitanas que abarcan otros municipios aledaños.

Respecto a la estructura del IPC, éste se puede desagregar en varios grupos dependiendo la necesidad de quien lo analice. Las más usuales corresponden a desagregarlo por clasificación y por grupos de gasto. La primera, la utilizan actores públicos para observar ciertos sectores en

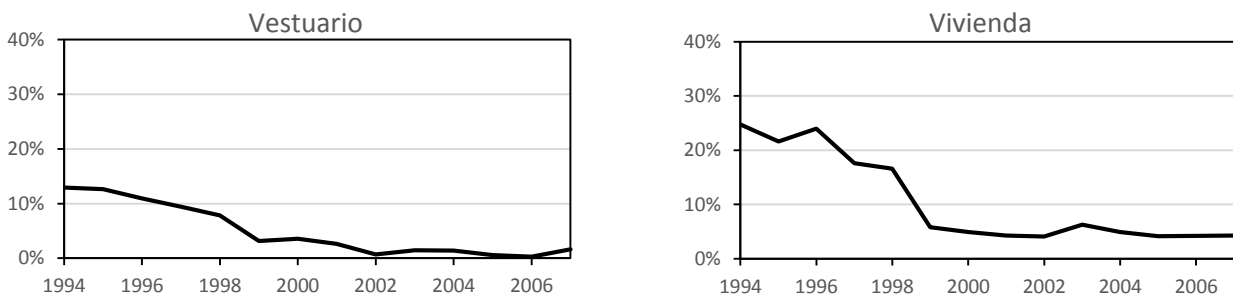
específico. Dentro de ella se encuentran el IPC de alimentos, de bienes transables, de bienes no transables y de regulados. Este tipo de clasificación lo utiliza, por ejemplo, el Banco de la República, donde presta especial atención a observar el comportamiento de la inflación básica, que corresponde a la inflación sin alimentos debido a su alta volatilidad.

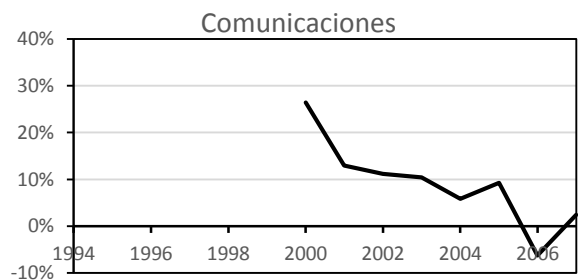
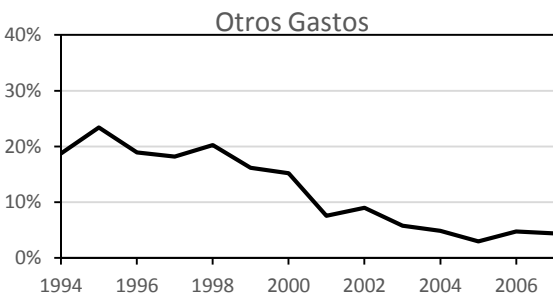
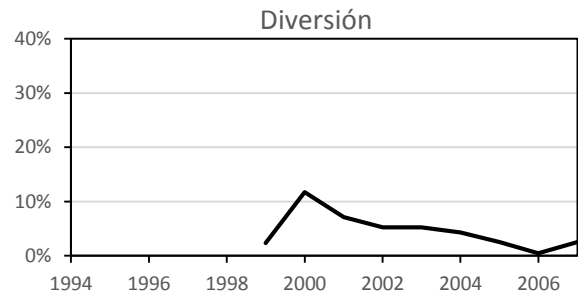
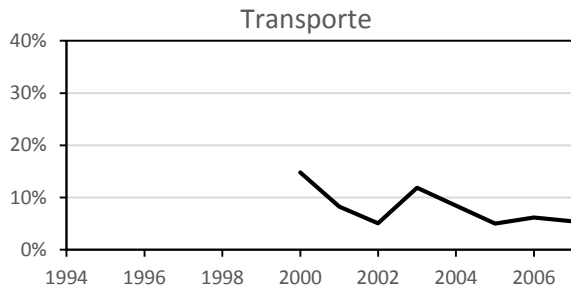
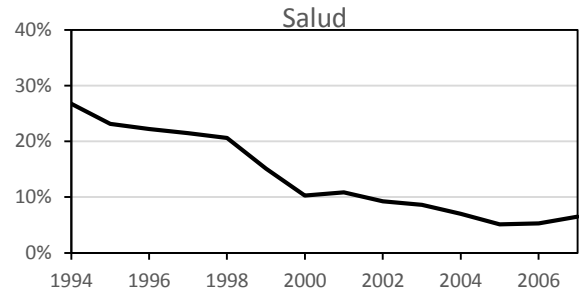
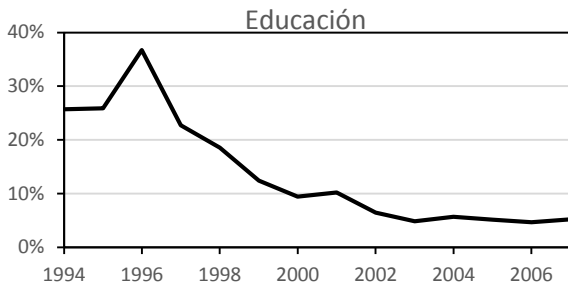


Fuente: DANE

Gráfico 3 Inflación interanual por clasificación 1994-2007

Las observaciones para los bienes transables, no transables y regulados solo se encuentran disponibles desde el año 1999. Así pues, se observa que la inflación en todas las clasificaciones disminuye a través del tiempo, a excepción de los no transables que parece permanecer más bien constante. Por otro lado, el IPC también se desagrega por grupos de gasto, tal como se muestra en el gráfico 4.





Fuente: DANE

Gráfico 4 Inflación interanual según grupo de gastos 1994-2007

Cabe resaltar que dentro del IPC por grupos de gasto también se encuentra el IPC de alimentos, que ya se presentó anteriormente. Respecto al gráfico 4, todos los grupos de gasto empiezan a disminuir tal como lo hace la inflación total. De igual forma, se observa que la inflación más alta la alcanzó el grupo de educación en el año 1996 con 36.74%; por otro lado, el grupo de comunicaciones es el único que presentó una deflación de 6.28% en el año 2006.

Como se mencionó previamente, el principal mecanismo de política monetaria que el Banco de la República utiliza para controlar la inflación es la tasa de intervención. El gráfico 5 muestra el promedio anual de la tasa de intervención.

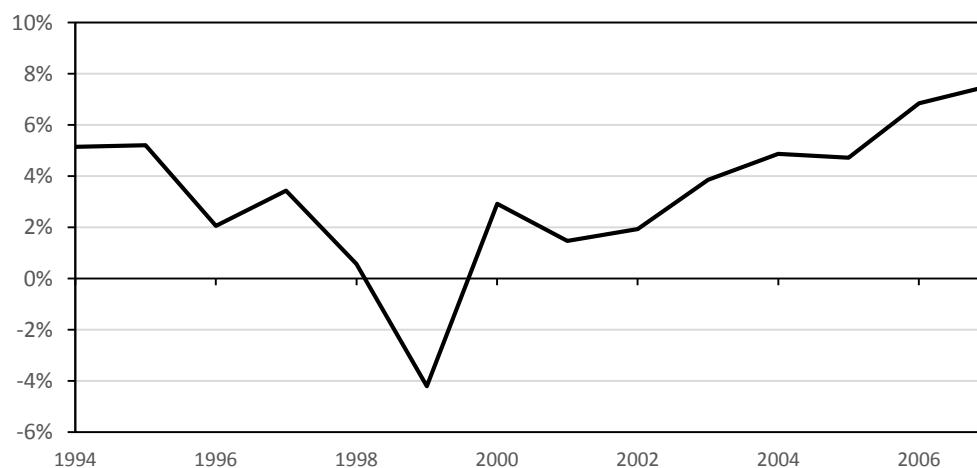


Fuente: Banco de la República

Gráfico 5 Tasa de intervención de política monetaria 1999-2007

Se observa que desde 1999 el banco empezó a disminuir su tasa de tasa de intervención acorde con el comportamiento de la inflación que ya se esbozó en gráficos anteriores. En el año 1999, la tasa de intervención promedio fue de 18.37%, en contraste con el 8.72% del año 2007.

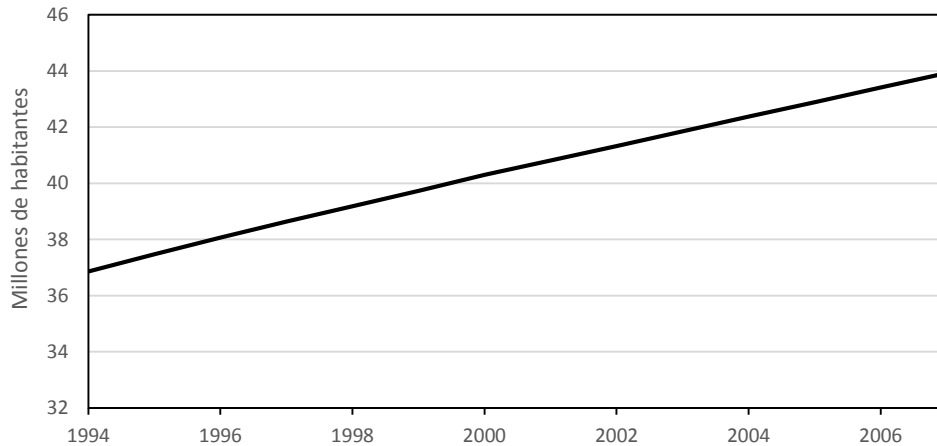
Respecto al comportamiento del PIB durante el periodo de estudio, el producto creció a tasas superiores al 5% en 1994 y 1995, posteriormente empezó a crecer a tasas más bajas, dando indicios de una desaceleración para finalmente contraerse 4.2% en el año 1999 a causa de la crisis financiera e hipotecaria. Después de este período el producto empieza a crecer positivamente, alcanzando un crecimiento de 7.58% en 2007. En el gráfico 6 se muestra el crecimiento del PIB entre 1994 y 2007.



Fuente: DANE

Gráfico 6 Crecimiento PIB 1994-2007

En cuanto a la población de Colombia entre 1994 y 2007, se debe decir que se contó con los censos nacionales de 1993 y 2005, por lo que las proyecciones deberían ser confiables. El gráfico 7 muestra el comportamiento de la población.



Fuente: DANE

Gráfico 7 Población de Colombia 1994-2007

La población de Colombia en el año 1994 fue de 36.853.905 habitantes, por otra parte, para el año 2007 fue de 43.926.929 habitantes; es decir, una variación porcentual de 19.19%.

5.2 MARCO TEÓRICO

Tal como De Gregorio y Huneus (1999) señalan, la primera aproximación en el estudio de índices de costo de vida lo realizó el economista ruso Alexander Alexandrovich Konüs en el siglo XX, y tiene un desarrollo basado en la teoría microeconómica del consumidor.

De esta manera, se tiene un agente que maximiza su utilidad en cada periodo t , y cuya forma funcional está determinada por preferencias regulares del tipo Cobb-Douglas, el agente resolvería:

$$\begin{aligned} \max_{(X,Y)} (X^\alpha Y^\beta) & \quad \forall (\alpha + \beta) = 1 \\ \text{s. a} & \end{aligned}$$

$$P_{x,t}X + P_{y,t}Y = M$$

El lagrangeano estaría dado por:

$$\mathcal{L} = \max_{(X,Y)} (X^\alpha Y^\beta) + \lambda(M - P_{x,t}X - P_{y,t}Y)$$

Se obtienen las condiciones de primer orden resultantes:

$$\frac{\partial \mathcal{L}}{\partial X} = 0$$

$$\frac{\partial \mathcal{L}}{\partial Y} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

Luego de resolver el sistema de ecuaciones que se forma se obtiene las demandas marshallianas del consumidor:

$$X^* = \alpha \frac{M}{P_{x,t}}$$

$$Y^* = \beta \frac{M}{P_{y,t}}$$

Como se desea medir el cambio en el costo de vida, se necesita una variable que proporcione el nivel de utilidad en función de los precios y la renta del agente. Dicha variable corresponde a la función de utilidad indirecta que se obtiene luego de reemplazar las demandas óptimas en la función de utilidad:

$$V(P_{x,t}, P_{y,t}, M) = \left(\frac{\alpha}{P_{x,t}} \right)^\alpha \left(\frac{\beta}{P_{y,t}} \right)^\beta M$$

Además, y siguiendo a Varian (1992), se utiliza la identidad subyacente de la dualidad entre la función de gasto mínimo y la función de utilidad indirecta, donde se obtiene:

$$e = \bar{v} \left(\frac{P_{x,t}}{\alpha} \right)^\alpha \left(\frac{P_{y,t}}{\beta} \right)^\beta$$

El verdadero aumento en el costo de vida de este consumidor entre el periodo t y el periodo $t-1$, corresponde a:

$$ACV_t = \frac{e(P_{x,t}, P_{y,t}, \bar{v})}{e(P_{x,t-1}, P_{y,t-1}, \bar{v})} = \frac{P_{x,t}^\alpha P_{y,t}^\beta}{P_{x,t-1}^\alpha P_{y,t-1}^\beta}$$

Y el índice verdadero del costo de vida de un periodo a otro estaría dado por:

$$ICV_t = \frac{P_{x,t}^\alpha P_{y,t}^\beta}{P_{x,0}^\alpha P_{y,0}^\beta}$$

Este índice del costo de vida corresponde al reajuste que se le debe hacer a un individuo para que tenga el mismo nivel de utilidad luego de un cambio en los precios.

El resultado anterior se puede agregar para un conjunto n de bienes:

$$ICV = \prod_{i=1}^N \left(\frac{P_{i,t}}{P_{i,0}} \right)^{\varphi_i}$$

Donde φ_i corresponde a la ponderación que el agente le da al bien i .

El mismo resultado se pudo haber obtenido luego de resolver el problema dual del consumidor y hallar las demandas compensadas.

$$\min \sum_{i=1}^N p_i x_i$$

s. a.

$$\bar{v}(x_i) = \prod_{i=1}^N x_i^{\alpha_i}$$

Como el nivel de utilidad es difícil de cuantificar en la práctica, se utiliza un índice de Laspeyres como la aproximación más común. El enfoque para medir el cambio en los precios de un periodo 0 a un periodo I consiste en especificar una canasta representativa de productos, es decir, un vector de cantidades que representa las compras realizadas entre el periodo 0 y el periodo I por un agente promedio (OIT et al., 2006). Para un índice de Laspeyres, este enfoque de canasta fija utiliza el vector de cantidades en el periodo 0 como la canasta de referencia, esto es:

$$P_L(p_0, p_1, q_0, q_1) = \frac{\sum_{i=1}^n p_{i,1} q_{i,0}}{\sum_{i=1}^n p_{i,0} q_{i,0}}$$

De acuerdo con OIT et al., (2006), se puede reformular la ecuación anterior de acuerdo a la aproximación convencional de las oficinas de estadísticas internacionales. Así pues, se define la participación del producto i en el gasto del período t :

$$s_{i,t} = \frac{p_{i,t} q_{i,t}}{\sum_{i=1}^n p_{i,t} q_{i,t}} \quad \forall i = 1, \dots, n$$

$$\quad \quad \quad \forall t = 0, 1$$

Ahora, se reorganiza el índice de Laspeyres multiplicando el numerador por $p_{i,0}/p_{i,0}$ se tendría:

$$P_L(p_0, p_1, q_0, q_1) = \frac{\sum_{i=1}^n (p_{i,1}/p_{i,0}) p_{i,0} q_{i,0}}{\sum_{i=1}^n p_{i,0} q_{i,0}}$$

$$P_L(p_0, p_1, q_0, q_1) = \sum_{i=1}^n \left(\frac{p_{i,1}}{p_{i,0}} \right) s_{i,0}$$

De esta forma, el índice de Laspeyres se obtiene del cociente de precios de ambos periodos, multiplicado por la participación del gasto en el periodo base. Lo que conlleva a la conclusión de que “el IPC de Laspeyres puede elaborarse en forma periódica sin necesidad de contar con información de las cantidades del periodo corriente” (OIT et al., 2006, p.310). La participación del gasto en el periodo base se obtiene luego de identificar un patrón agregado de consumo a través de encuestas nacionales. Para el caso colombiano, se denomina la Encuesta Nacional de Ingresos y Gastos realizada por el DANE cada 10 años aproximadamente.

Por otra parte, la literatura parece encontrar consenso cuando desagrega el sesgo en el IPC en cuatro sesgos diferentes: sesgo de sustitución, sesgo por cambios en la calidad de los bienes, sesgo por aparición de nuevos bienes y sesgo por aparición de nuevos expendios (Moulton (1996); Boskin et al. (1996); Shiratsuka (1999); Hausman (2003)).

El sesgo de sustitución surge cuando se experimentan cambios en los precios relativos de los bienes. De acuerdo con la teoría del consumidor, cuando se advierten cambios en los precios relativos, es de esperarse que el agente en cuestión modifique su comportamiento aumentando el consumo del bien que se ha abaratado y disminuyendo el consumo del bien que es relativamente

más costoso. Sin embargo, debido a la estructura metodológica del índice de Laspeyres lo anterior no es posible en la medida que utiliza ponderaciones fijas de la participación del gasto en cada uno de los bienes incluidos, que conlleva a sobreestimar el crecimiento del nivel de precios al valorar más los bienes caros que los baratos (Caicedo, 2000). Una de las formas más usuales para corregir este sesgo se presentó en la sección de antecedentes, y consiste, *grosso modo*, en estimar un índice superlativo que tenga en cuenta la racionalidad del agente ante cambios en los precios relativos, para posteriormente calcular la diferencia con los índices convencionales (Diewert, 1976).

El sesgo por cambios en la calidad de los bienes se presenta cuando algunas agregaciones de bienes son reemplazadas por otras de mayor calidad (Caicedo, 2000). De igual forma, el sesgo también se puede producir porque algunos de los bienes han salido del mercado, han perdido representatividad como resultado de algún fenómeno coyuntural o por el desarrollo de innovaciones tecnológicas (Shiratsuka, 1999). Lo anterior repercute directamente en la frecuencia en que los hogares demandan dichos bienes y tienden a sobreestimar el verdadero comportamiento del IPC. El método más usual para corregir este sesgo consiste en estimar regresiones hedónicas que permitan capturar todas las características del bien en cuestión.

El sesgo por la aparición de nuevos bienes se da en un contexto donde la innovación tecnológica, la competencia, la institucionalidad de un país o el cambio en las preferencias de los consumidores hacen que continuamente aparezcan nuevos bienes en los mercados. Si los nuevos bienes descritos anteriormente se convierten en una constante en el gasto de los hogares y no se incluyen en la elaboración de un índice de precios, se generaría entonces un sesgo hacia arriba al no ser tenidos en cuenta (Caicedo, 2000). Por lo general, este sesgo es el más fácil de corregir ya que solamente basta con actualizar la canasta de consumo que sirve como periodo base para calcular el IPC.

Por último, el sesgo por la aparición de nuevos expendios cobra relevancia cuando los almacenes deciden competir en el mercado con bienes de mejor calidad y menores precios; lo que conlleva a capturar una mayor proporción de la demanda al sacrificar algún margen de utilidad por un aumento de las ventas totales cobrando precios inferiores que otros establecimientos (Caicedo, 2000). Sí no se tienen en cuenta estos cambios en los lugares de compra de los consumidores por nuevos expendios, se generará un sesgo hacia arriba en estimaciones de inflación usando el IPC.

5.3 MARCO CONCEPTUAL

A continuación, se presentan algunas definiciones clave para comprender correctamente el desarrollo de ésta investigación.

Para empezar, un índice se define como “una cifra relativa que recoge las variaciones promedio en precios, cantidades o valores de una o más variables durante un período respecto a un período determinado” (Alonso, 2004, p. 3). Resulta relevante mencionar que los índices carecen de magnitudes dado que se interesa medir el comportamiento de alguna variable a través del tiempo.

De igual forma, el índice de costo de vida se define como el cociente de los gastos mínimos necesarios para obtener un nivel de utilidad específico en dos periodos de tiempo diferentes bajo dos niveles de precios diferentes (OIT et al., 2006). Por otra parte, el índice de precios al consumidor se define como un índice tipo Laspeyres que mide la variación porcentual promedio del nivel de precios de un periodo a otro (DANE, 2009).

Un índice tipo Laspeyres o de ponderaciones fijas se define como el aumento o disminución del valor de una canasta de un periodo a otro, manteniendo constante las cantidades del primer periodo. En contraste a lo anterior, un índice de Paasche o de ponderaciones variables mide el valor de compra de una canasta de bienes manteniendo los precios del periodo inicial. (DANE, 2009).

Un índice superlativo o ideal de Fisher se define como la media geométrica de los índices de Laspeyres y de Paasche (DANE, 2009). Este tipo de índices cobran relevancia dado que son una buena aproximación de cambios en el costo de vida, al permitir incorporar cambios en los patrones de consumo que el índice de Paasche si absorbe.

Finalmente, la Encuesta Nacional de Ingresos y Gastos, se define como una investigación que permite obtener información detallada sobre los ingresos que perciben los hogares, así como su distribución; es decir, identificar patrones de consumo. Con ésta encuesta se actualiza las ponderaciones del índice de precios al consumidor y se realiza aproximadamente cada 10 años (DANE, 2009).

5.4 MARCO LEGAL

De acuerdo al Decreto 3167 de 1968, se establece que el DANE debe producir los principales índices de precios del país a nivel de productor y consumidor.

Por otra parte, el artículo 371 de la constitución política de Colombia de 1991 establece que, el Banco de la República posee autonomía administrativa, permitiéndole enfocarse en establecer una meta de inflación para preservar el poder adquisitivo de los colombianos.

Respecto a algunas normas internacionales de estadística económica, destaca la Conferencia internacional de Estadísticos de 1925 convocada por la Organización Internacional del Trabajo, en ese momento se abrió el debate sobre la metodología y las pautas que se debían desarrollar para obtener índices de precios al consumidor que fueran comparables entre países. De esta manera, el primer conjunto de normas no hacía referencia al índice de precios al consumidor, sino al índice de costo de vida. No fue hasta el año 1962 cuando se adoptó formalmente el término de índice de precios del consumidor. Las normas fueron revisadas continuamente en 1947, 1962 y 1987 para afinar aún más las metodologías que se debían utilizar (OIT et al., 2006).

6. DISEÑO METODOLÓGICO

La metodología que se utilizará corresponde a la presentada por Hamilton (2001) y Costa (2001), y que previamente utilizaron Langebaek y Caicedo (2007) para el mismo propósito que el presente trabajo, aunque para el periodo comprendido entre 1984/85 a 1994/95. Cabe resaltar que el universo de estudio del presente trabajo es mucho más amplio ya que contempla 23 ciudades principales, a diferencia de Langebaek y Caicedo (2007) donde utilizan solamente 7 ciudades. La metodología supone que sí en dos periodos de tiempo distintos, dos hogares con el mismo nivel de ingreso real, medido a través del IPC, y con características sociodemográficas similares, presentan proporciones del gasto distinto en alimentos, dicha diferencia puede atribuirse al sesgo en el IPC. Lo anterior, se cimienta bajo el descubrimiento del economista Erns Engel, quien estudio el comportamiento de gasto en alimentos de 199 hogares en Bélgica en el siglo XX, y encontró que,

cuanto más aumentaban los ingresos, más disminuía la proporción del gasto en alimentos. A pesar de que la muestra era muy pequeña en aquella época, su consistencia a través del tiempo la han establecido como una de las leyes empíricas más regulares en la investigación económica (Kaus, 2013).

6.1 TIPO DE ESTUDIO

La presente investigación se enmarca dentro de un tipo de estudio explicativo, es decir, se busca encontrar, a través de estimaciones econométricas, la relación causa y efecto entre diferentes variables.

6.2 MÉTODO

Como se mencionó anteriormente, el modelo se cimienta bajo la Ley de Engel. Donde, de acuerdo con Houthakker (1957), de todas las regularidades empíricas observadas en la ciencia económica, la Ley de Engel es probablemente la mejor establecida. Dicha ley argumenta que existe una relación inversa entre la participación del gasto en alimentos y el ingreso real de los hogares. De esta manera, se podría inducir que, cambios en la participación del gasto en alimentos sería un buen indicador de cambios en el ingreso real de los hogares. Así pues, una buena aproximación del sesgo en el IPC utilizando curvas de Engel, consiste en estimar funciones de demanda por alimentos explicada por el ingreso real (entre otras variables). Además, si la función de demanda está bien especificada, las preferencias son estables y no hay errores sistemáticos en las variables, no debería existir cambios en las curvas de Engel de dos periodos diferentes (Hamilton, 2001).

Es decir, si se realizan dos estimaciones para diferentes periodos de tiempo utilizando el ingreso real como variable independiente explicando la participación del gasto en alimentos, las curvas de Engel tendrían que ser iguales, y el crecimiento real del ingreso estaría dado, principalmente, por el valor del IPC por el que se deflactaron los ingresos. Ahora bien, si la curva de Engel se desplaza hacia la izquierda (sobrestima) o hacia la derecha (subestima), la brecha que se genera entre ambas curvas puede atribuirse al sesgo en el IPC. A manera de ejemplo, si la curva se desplaza hacia la izquierda, esto diría que el ingreso real ha aumentado más que el ingreso real deflactado por el IPC, debido a que dicho índice ha sobrestimado la inflación (Hamilton, 2001). Podría

decirse que la metodología cuantifica indirectamente el sesgo en el IPC ya que no se utilizan series del comportamiento histórico del índice para obtenerlo.

Respecto al alcance del modelo, cabe destacar que la metodología tiene varias bondades. En primer lugar, y apuntando hacia el objetivo general, el estudio permitirá encontrar el sesgo en el IPC durante los dos periodos que se enuncian. Sin embargo, también permitirá encontrar características de los hogares, tales como la elasticidad del ingreso y la existencia de posibles economías a escala en el gasto de alimentos.

La estimación se hace a través del modelo casi ideal de demanda propuesto por Deaton & Muellbauer (1980):

$$\omega_{i,j,t} = \phi + \gamma \ln \left(\frac{P_{j,t}^f}{P_{j,t}^{nf}} \right) + \beta \ln \left(\frac{Y_{i,j,t}}{P_{j,t}} \right) + \sum_x \theta_x X_{i,j,t} + \mu_{i,j,t} \quad (1)$$

Dónde:

$\omega_{i,j,t}$ = proporción del gasto en alimentos del hogar i en la región j en el periodo t .

$\frac{P_{j,t}^f}{P_{j,t}^{nf}}$ = verdadero precio relativo de los alimentos frente a los no alimentos.

$\frac{Y_{i,j,t}}{P_{j,t}}$ = ingreso real del hogar i en la región j en el periodo t .

$X_{i,j,t}$ = vector de variables que captura características sociodemográficas.

$\mu_{i,j,t}$ = error aleatorio

Nótese que las variables referidas a los precios en la ecuación (1) hacen referencia a precios verdaderos pero inobservables de la economía. Ahora, si se considera:

$$\ln P_{j,t} = \ln P_{j,0} + \ln(1 + \Pi_{j,t}) + \ln(1 + E_t) \quad (2)$$

Dónde:

$\Pi_{j,t}$ = variación observada en los precios

E_t = sesgo no observado entre el periodo 0 y el periodo t .

Y a su vez, para simplificar el análisis se definen las expresiones de la ecuación (2) como $\pi_{j,t} = \ln(1 + \Pi_{j,t})$ y $\varepsilon_t = \ln(1 + E_t)$.

La ecuación (2) quedaría definida así:

$$\ln P_{j,t} = \ln P_{j,0} + \pi_{j,t} + \varepsilon_t \quad (3)$$

Reemplazando (3) en (1) se obtiene:

$$\begin{aligned} \omega_{i,j,t} = \phi + \gamma \left[\left(\ln P_{j,0}^f + \pi_{j,t}^f + \varepsilon_t^f \right) - \left(\ln P_{j,0}^{nf} + \pi_{j,t}^{nf} + \varepsilon_t^{nf} \right) \right] \\ + \beta \left[\ln Y_{i,j,t} - \left(\ln P_{j,0} + \pi_{j,t} + \varepsilon_t \right) \right] + \sum_x \theta_x X_{i,j,t} + \mu_{i,j,t} \end{aligned} \quad (4)$$

Reorganizando términos y definiendo $y_{i,j,t} = \ln Y_{i,j,t}$ y $p_{j,0} = \ln P_{j,0}$, se obtiene:

$$\omega_{i,j,t} = \phi + \gamma \left(\pi_{j,t}^f - \pi_{j,t}^{nf} \right) + \beta \left(y_{i,j,t} - \pi_{j,t} \right) + \sum_x \theta_x X_{i,j,t} + \sum_j \delta_j D_j + \delta_t D_t + \mu_{i,j,t} \quad (5)$$

Donde D_j y D_t son variables dicótomas de región y tiempo, cuyos parámetros se especifican a continuación:

$$\begin{aligned} \delta_j &= \gamma \left(p_{j,0}^f - p_{j,0}^{nf} \right) - \beta p_{j,0} \\ \delta_t &= \gamma \left(\varepsilon_t^f - \varepsilon_t^{nf} \right) - \beta \varepsilon_t \end{aligned} \quad (6)$$

Siguiendo a Costa (2001) y Carvalho & Chamon (2006) se asume que el sesgo entre alimentos y no alimentos es el mismo, por tanto, el sesgo acumulado corresponde a:

$$\begin{aligned} \varepsilon_t &= -\frac{\delta_t}{\beta} \\ E_t &= \exp \left\{ -\frac{\delta_t}{\beta} \right\} - 1 \end{aligned} \quad (7)$$

Así pues, la especificación econométrica viene dada por:

$$\omega_{i,j,t} = \phi + \gamma \left(\pi_{j,t}^f - \pi_{j,t}^{nf} \right) + \beta \left(y_{i,j,t} - \pi_{j,t} \right) + \sum_x \theta_x X_{i,j,t} + \sum_{j-1} \delta_{j-1} D_{j-1} + \delta_{2006} D_{2006} + \mu_{i,j,t} \quad (8)$$

Donde D_{j-1} y D_{2006} son variables binarias hace referencia a las $j-1$ ciudades diferentes a Bogotá y a los hogares que fueron encuestados en la Encuesta Nacional de Ingresos y Gastos 2006/07

respectivamente. Por otra parte, Deaton & Muellbauer (1980) modelaron los precios como un índice de Stone, sin embargo, Moschini (1995) demostró que el índice de Stone no es invariante a cambios en las unidades de medición de los precios, lo que puede conllevar a sesgos en la estimación de sistemas lineales. Para solucionar esto, Moschini (1995) recomendó usar índices de Laspeyres en lugar del índice de Stone.

Finalmente, dentro del vector de características sociodemográficas se encuentran: sexo del jefe de hogar, edad del jefe de hogar, educación del jefe de hogar, logaritmo natural del número de integrantes y una variable dummy que toma el valor de 1 si el hogar posee vivienda propia. La estimación se realizará a través de mínimos cuadrados ordinarios.

6.3 TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Para elaborar la presente investigación se utilizó fuentes de información secundaria como artículos científicos nacionales e internacionales. De igual forma, para el ejercicio econométrico se utilizará una fuente primaria, como lo es la Encuesta Nacional de Ingresos y Gastos de los años 1994/95 y 2006/07 que se encuentran en la página de microdatos del DANE.

6.4 ACTIVIDADES DE LA INVESTIGACIÓN

Para cuantificar el sesgo existente del IPC entre 1994/95 y 2006/07 se utilizó la Encuesta Nacional de Ingresos y Gastos de ambos años, que se encuentran disponibles para libre disposición en la página de microdatos del DANE.

Ambas encuestas cuentan con varios módulos donde se recoge toda la información que el DANE recopiló. Para empezar, respecto a la encuesta de 1994/95, se utilizaron los módulos: personas, vivi_hogar, ingresos y gasto hogar. En cuanto a la encuesta 2006/07 se utilizaron los módulos:

ig_ml_personas, ig_ml_hogar, ig_ml_ocuopado, ig_gsdu_caract_alimentos y ig_gsmf_compra. Respecto a los ingresos y gastos, se agregó por hogar ambas variables en ambos años. Siguiendo a Langebaek y Caicedo (2007) se eliminaron los hogares unipersonales y los hogares con el 5% de los ingresos más altos de cada ciudad. Para el análisis, las ciudades incluidas son: Medellín, Barranquilla, Bogotá, Cartagena, Tunja, Manizales, Florencia, Popayán, Valledupar, Montería, Quibdó, Neiva, Riohacha, Santa Marta, Villavicencio, Pasto, Cúcuta, Armenia, Pereira, Bucaramanga, Sincelejo, Ibagué y Cali. Se excluyó a San Andrés debido a que solo se tenía su información en la encuesta 2006/07.

Por otra parte, en busca de mejorar el análisis, se utilizó el gasto real en lugar del ingreso real para la estimación, y su explicación radica en que los hogares encuentran incentivos distorsionados para declarar correctamente sus ingresos (principalmente asociado a problemas de riesgo moral); en tanto que el gasto del hogar se presenta como una proxy adecuada para inferir el ingreso.

Para la variable de precios relativos, se crearon índices por ciudad con la información disponible del valor de compra de los hogares. Finalmente, y siguiendo a Costa (2001) se optó por conservar los hogares que tenían una participación del gasto en alimentos entre 5% y 80%, en busca de eliminar datos atípicos. Al final, se contó con 11260 observaciones para ambos años.

7. RESULTADOS DE LA INVESTIGACIÓN

7.1 ESTADÍSTICAS DESCRIPTIVAS

Inicialmente, se revisa el comportamiento de cada variable de interés presentando sus respectivas estadísticas descriptivas. Los resultados obtenidos para cada año se presentan en la siguiente tabla.

Tabla 1. Estadísticas descriptivas

Variable	1994/95					2006/07				
	Obs,	Media	Des. Est.	Mínimo	Máximo	Obs,	Media	Des. Est.	Mínimo	Máximo
Proporción gasto en alimentos	2436	0,64	0,14	0,05	0,79	8824	0,22	0,17	0,05	0,079
Gasto real	2436	116490,5	115536,7	2433,01	1527325	8824	961880,3	714226,7	4593,04	4556437
Integrantes	2436	4,18	1,69	2	13	8824	4,13	1,85	2	19

Precio relativo de alimentos	2436	1	0	1	1	8824	0,04	0,01	0,02	0,07
Sexo jefe hogar (1=Hombre)	2436	0,68	0,46	0	1	8824	0,63	0,48	0	1
Edad jefe hogar	2436	42,2	13,01	16	91	8824	47,94	15,06	15	96
Educación jefe hogar	2436	8,16	4,57	0	21	8824	11,44	5,6	0	26
Vivienda	2436	0,61	0,48	0	1	8824	0,53	0,49	0	1

Fuente: DANE. Cálculos propios.

Respecto al período 1994/95 el promedio de la proporción de gasto en alimentos de los hogares fue de 64%, gasto real de 116.490 COP y estaba compuesto de 4 personas aproximadamente. El hombre en su mayoría era el jefe de hogar y tenía 42 años de edad, 8 años de educación y más de la mitad contaba con vivienda propia. Respecto al período 2006/07 el promedio de la proporción de gasto en alimentos cambió a 22%, el gasto real aumentó a 714.226 COP, el promedio de los integrantes se mantuvo constante, la edad promedio del jefe de hogar aumentó a 48 años aproximadamente, al igual que su educación con 11 años de escolaridad.

Por otra parte, y teniendo en cuenta que descriptivamente se observa una disminución de la proporción del gasto en alimentos de un período a otro, se realizó una prueba de diferencia de medias para determinar dicho cambio de preferencias. La tabla 2 denota que la media de la proporción de alimentos en los dos años es la misma, por lo tanto, se puede proseguir con la estimación.

Tabla 2. Prueba de diferencia de medias para gasto proporción de alimentos

Año	Obs.	Media	Error Est.	Desv. Est.
1994/95	2436	0,6481	0,0028	0,14
2006/07	8824	0,2225	0,0018	0,1706
Diferencia		0,4255		
H ₀ : Media (1994/95)-Media(2006/07)=0			P-valor=1,0000	
H _a : Ho es falsa				

Fuente: DANE. Cálculos propios.

7.2 RESULTADOS Y DISCUSIÓN

Para obtener la variable “precio relativo de alimentos” por región, se optó por crear índices con los grupos de gasto (gasto en alimentos y gasto en no alimentos) como variables proxy. Así, utilizando como año base el gasto realizado por cada hogar en el período 1994/95, se calcularon los índices de gasto en alimentos, gasto en no alimentos y el cociente de ambos para el período 2006/07. La tabla 3 resume los índices calculados.

Se observa que, respecto al gasto en alimentos, la variación más grande se registró en Manizales y la más pequeña en Cali, con índices de 7,88 y 2,21 respectivamente. Respecto al cambio en el gasto en no alimentos, la variación más grande la tuvo Neiva, en contraste con Montería, donde se registró la menor variación, ambas ciudades con índices de 212,67 y 36,55 respectivamente.

Tabla 3. Índices para 2006/07

Departamento	Índice alimentos	Índice no alimentos	Precio relativo alimentos
Medellín	3,03542	71,07283	0,04271
Barranquilla	7,37177	105,14707	0,07011
Bogotá	4,24527	101,62483	0,04177
Cartagena	6,34591	179,95347	0,03526
Tunja	5,15004	128,74569	0,04000
Manizales	8,77848	197,79203	0,04438
Florencia	5,11198	116,09383	0,04403
Popayán	4,72951	206,99791	0,02285
Valledupar	7,01117	122,59624	0,05719
Montería	2,30236	36,55437	0,06298
Quibdó	3,93541	100,77856	0,03905
Neiva	5,73805	212,67219	0,02698
Riohacha	7,46981	182,76656	0,04087
Santa Marta	3,92614	73,08159	0,05372
Villavicencio	4,72238	114,13922	0,04137
Pasto	5,65711	152,47158	0,03710
Cúcuta	3,31557	70,82779	0,04681
Armenia	7,76846	194,57423	0,03993
Pereira	4,51842	84,36741	0,05356
Bucaramanga	2,87886	70,51060	0,04083
Sincelejo	4,64875	143,39791	0,03242
Ibagué	5,20380	169,40832	0,03072
Cali	2,21320	45,59498	0,04854

Fuente: DANE. Cálculos propios.

Estimando la función de demanda de la ecuación (8) se encontró que el modelo es heterocedástico y presenta colinealidad entre las variables “año” y “ln precio relativo de alimentos”.

El problema de colinealidad tiene su explicación en la estructura de la base de datos, especialmente debido a que al transformar la variable “precio relativo de alimentos” a logaritmos, todas las observaciones del año 1994/95 se convierten en cero, y teniendo en cuenta que la variable “año” toma el valor de cero para los hogares encuestados en 1994/95 es plausible la correlación existente entre las variables. Sin embargo, al realizar la correlación entre la variable “año” y los logaritmos de los precios de alimentos y no alimentos (como dos variables por separado), se encontró que la correlación de los precios con la variable “año” es prácticamente cero.

Además, se prueba excluyendo la variable “ln precio relativo de alimentos” como una posible solución para corregir esta problemática. Sin embargo, al realizar esto se estaría cometiendo un error de especificación aún mayor, ya que como demuestra Hamilton (2001), al excluir los precios de la regresión se estaría disminuyendo el valor del parámetro año, con lo que la sobreestimación sería aún mayor.

Así pues, y considerando que, la principal consecuencia de la colinealidad entre las variables explicativas radica en la pérdida de precisión de las varianzas de los parámetros, que a la vez conlleva a que los intervalos de confianza sean mucho más amplios y crezca la tendencia a que la inferencia estadística no sea válida (Gujarati y Porter, 2010); se optó por conservar el modelo teniendo en cuenta que la colinealidad solo hace que la variable año no sea significativa, sin embargo, al excluir los precios de la estimación, ésta se vuelve significativa en todos los niveles.

Por otra, para corregir la heterocedasticidad se calcularon los errores estándar robustos de White, esto es, re estimar las varianzas de los parámetros no en función de la varianza del error, sino del residuo al cuadrado¹ (Gujarati y Porter, 2010).

Finalmente, de acuerdo con la literatura sobre estimaciones de demanda, podría pensarse que los precios están correlacionados con el error aleatorio al existir características inobservables que causan shocks de demanda (Aguirregabiria, 2012). Sin embargo, al realizar una regresión del error con todas las variables explicativas, se encontró que ninguna de estas últimas es significativa.

¹ Todas las pruebas realizadas se encuentran en la sección de anexos.

Además que no se encontró ningún instrumento para los precios. La tabla 4 muestra los resultados encontrados.

Tabla 4. Resultados

Proporción gasto en alimentos	Coefficiente	Error Estándar Robusto
Año	-0,0168	0,0495
ln gasto real	-0,0786***	-0,0021
ln integrantes	0,0197***	0,0035
ln precio relativo alimentos	0,0794***	0,0155
Sexo jefe	0,0082***	0,0029
Edad jefe	0,0001	0,0001
Educación jefe	0,0039***	0,0002
Vivienda	0,0097***	0,0031
Efectos fijos por región	X	
Número de observaciones	11260	
Prob>F	0,0000	
R2	0,6112	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Fuente: DANE. Cálculos propios.

La variable año, al tener un coeficiente negativo indica que la Curva de Engel para el período 2006/07 es menor que la del año de referencia. Lo anterior, indica que efectivamente existe un sesgo en el IPC que se extiende hacia el cálculo del costo de vida. El coeficiente negativo del gasto real indica que se cumple la Ley de Engel, es decir, cuando aumenta el gasto (ingreso) de los hogares, la proporción destinada a alimentos disminuye. Estos dos resultados son consistentes con toda la literatura relacionada.

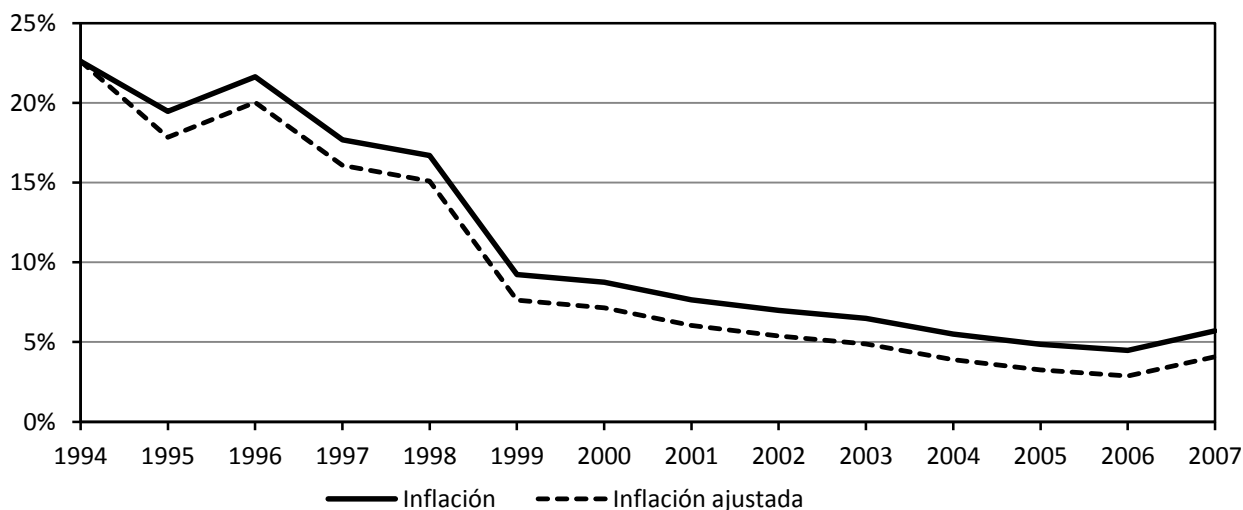
El signo positivo del precio relativo de alimentos coincide con lo encontrado por Costa (2001), pero es opuesto a los resultados de Langebaek y Caicedo (2007). Tal signo positivo resulta

plausible bajo la siguiente consideración: el precio relativo de alimentos, al ser un cociente del precio de alimentos y no alimentos, puede aumentar debido a un incremento del precio de alimentos o por una disminución del precio de no alimentos. Así, si se considera a los alimentos como un bien normal y ordinario, se espera que un aumento de los precios relativos se deba principalmente a una disminución del precio de los no alimentos, donde, el aumento en la participación del gasto en alimentos se deba más bien al ajuste entre los efectos renta y sustitución de los hogares que viene determinada por sus preferencias.

Por otra parte, las variables relacionadas con las características del jefe de hogar indican que un hogar donde el jefe es hombre, tiene una participación de alimentos mayor en 0.82%. De igual forma, la participación de gasto en alimentos aumenta 0.01% y 0.39% con la edad y la educación del jefe respectivamente. Así mismo, la variable vivienda indica que si el hogar posee vivienda propia la participación del gasto destinada a alimentos aumenta.

Aplicando la ecuación (7) a los parámetros de la regresión, se encontró que el sesgo acumulado en el IPC arrojó una sobreestimación del costo de vida de 19.33 puntos porcentuales. Lo anterior plantea que la sobreestimación promedio anual fue de 1.61 puntos porcentuales entre los periodos 1994/95 y 2006/07. Este resultado coincide con el rango encontrado por Langebaek y Caicedo (2007) de 1.63-1.69 puntos porcentuales.

De esta manera, ajustando el sesgo, se presenta en el gráfico 8 el comportamiento de la inflación y la inflación ajustada anual. La literatura sugiere que mayor nivel de inflación mayor sesgo, sin embargo, según los resultados encontrados se toma el sesgo como constante a lo largo de toda la muestra.



Fuente: DANE. Elaboración propia.

Gráfico 8 Inflación ajustada

Finalmente, y teniendo en cuenta que se estimó una función de demanda, se pueden hallar las elasticidades gasto y precio (marshalliana y hicksiana) de la demanda de alimentos. La obtención de las elasticidades es presentada en Green & Alston (1990)². La tabla 5 presenta los resultados de las elasticidades para el trabajo.

Tabla 5. Elasticidades alimentos

Elasticidad	Valor	Tipo
Gasto de la demanda	0,51	Normal - Básico
Precio de la demanda (no compensada)	0,35	Inelástica
Gasto de la demanda (compensada)	0,19	Inelástica

Fuente: DANE. Cálculos propios.

Estas elasticidades van en línea con la teoría económica al considerar a los alimentos como un bien(es) básico(s) e inelástico(s) respecto al precio de la demanda.

Los resultados encontrados en éste trabajo son congruentes con los hallazgos de Langebaek y Caicedo (2007) que encontraron un sesgo en el costo de vida en un rango de 1.63-1.69, sin embargo, se debe tener en cuenta que en el período de estudio de los autores la inflación fue mucho más alta, a diferencia de éste trabajo, donde existió una inflación alta al principio de la muestra

² Las fórmulas para el cálculo de las elasticidades corresponden a :

- Elasticidad gasto: $1 + \frac{\beta_i}{\omega_{i,t}}$
- Elasticidad precio Marshalliana: $-I_i + \frac{\gamma_{i,j}}{\omega_{i,t}} - \beta_i \frac{\omega_{j,t}}{\omega_{i,t}}$
- Elasticidad precio Hicksiana: $-I_i + \frac{\gamma_{i,j}}{\omega_{i,t}} - \omega_{j,t}$

para posteriormente disminuir. De igual forma, se demostró empíricamente que se cumple la ley de Engel para Colombia con nuevos datos.

Así pues, se demostró que el IPC sobreestimó el costo de vida. De hecho, se encontró que el sesgo acumulado en los precios fue de alrededor de 19.33 puntos porcentuales tomando todo el período de tiempo (desde 1994 hasta 2007). Lo anterior deja entrever que efectivamente se ha calculado erróneamente aquellas cifras deflactadas por el IPC donde se busca medir su variación real. De hecho, el IPC calculado por el DANE³ fue de 26.14 en 1994 y 92.87 en 2007, sin embargo, utilizando el sesgo encontrado en la regresión se encuentra que el verdadero valor del IPC en 2007 es de 87.81⁴.

Dentro de los resultados encontrados también se puede extraer algunas extensiones interesantes. Por ejemplo, la no existencia de economías a escala en el gasto en alimentos asociado al signo positivo del número de integrantes, lo anterior se explica bajo el argumento de que los alimentos son bienes excluyentes (Langebaek y Caicedo, 2007). De igual forma, el resultado de las elasticidades es congruente con la teoría económica. La elasticidad gasto de la demanda dio como resultado 0.51, es decir, los alimentos son considerados un bien básico o necesario para los hogares. Por otra parte, el resultado de las elasticidades precio de la demanda (marshalliana y hicksiana) concluyen que los alimentos son inelásticos ante variaciones de los precios.

8. CONCLUSIONES Y RECOMENDACIONES

En éste trabajo se calculó el sesgo en el IPC entre los periodos 1994/95 – 2006/07 para posteriormente calcular el sesgo en la inflación en dichos períodos. Lo anterior se realizó bajo la estimación de curvas de Engel, esto es, estimar funciones de demanda por alimentos en dos períodos de tiempo diferentes, tal como lo propusieron Hamilton (2001) y Costa (2001). La estimación incluyó un grupo de controles atribuidas a las características sociodemográficas del

³ Índices tomados del mes de diciembre de cada año.

⁴ Este cálculo se realiza mediante la fórmula de la variación porcentual de una variable: $\frac{X_t}{X_{t-1}} - 1$

hogar y efectos fijos por región. La información se obtuvo de las Encuestas de Ingresos y Gastos del DANE 1994/95 y 2006/07.

Dentro de los resultados más relevantes se encuentra que el IPC sobreestimó el costo de vida en 19.33 puntos porcentuales entre 1994/95 – 2006/07, que equivale a una sobreestimación promedio anual de 1.61 puntos porcentuales y que, es similar al encontrado por Langebaek y Caicedo (2007). Con los resultados encontrados se corrigió la inflación para observar cómo fue su evolución, lo anterior, bajo el supuesto de tomar como constante el sesgo durante todo el período de estudio.

De igual forma, se encontró que la participación del gasto en alimentos de cada hogar aumenta con la edad del jefe de hogar (0.01%), educación jefe de hogar (0.39%), número de integrantes del hogar (0.0197%), vivienda propia del hogar (0.97%) y si el sexo del jefe de hogar es hombre (0.82%). Por otra parte, el gasto en alimentos disminuye con aumentos del gasto real (-7.86%) algo que justamente se esperaba para el cumplimiento de la Ley de Engel (Kaus, 2013).

Finalmente, para ahondar en el análisis y dotar de mayor robustez a los resultados se estimaron elasticidades gasto y precio de la demanda. Se encontró que los alimentos son bienes necesarios e inelásticos ante variaciones en los precios con elasticidad gasto de la demanda igual a 0.51 y elasticidad precio de la demanda igual a 0.35 y 0.19 (no compensada y compensada).

Debido a la metodología de los índices de precios del consumidor, siempre existirá algún sesgo en su estimación. Aún si calcularan las ponderaciones más frecuentemente, el sesgo por sustitución siempre estará latente. Sin embargo, se celebra la nueva metodología que el DANE implementará desde principios del año 2019, donde se actualiza la canasta básica de acuerdo a la identificación de nuevos patrones de consumo siguiendo la Encuesta Nacional del Presupuesto de los Hogares 2018, ampliación de la cobertura al pasar de 24 ciudades principales a 38 ciudades y la inclusión de 12 grupos de gasto, en contraste con los 9 grupos de gasto con los que contaba anteriormente. Lo anterior permitirá una medida más acertada para medir el verdadero costo de vida. De igual forma, las oficinas de estadísticas de los países podrían, al mismo tiempo que calculan los índices, calcular el sesgo de dicho índice. Aunque esta opción acarrearía mayores costos pecuniarios y es algo que no muchos países estarían dispuestos a realizar.

Finalmente, vale la pena recordar que el aporte de este trabajo es meramente estimativo ya que se corrigió una medida sesgada como lo es la inflación, que surge por sesgos inherentes en la

estructura del IPC y, se espera, que estos resultados permitan continuar con el estudio de estimar variables más precisas para el análisis y aplicación desde la academia y la política pública.

REFERENCIAS

Constitución Política de Colombia. (1991). Bogotá, D.C: Imprenta Nacional de Colombia.

Aguirregabiria, V. (2012). Empirical Industrial Organization: Models, Methods and Applications. *MIMEO- University of Toronto.*

Alonso, J. C. (2004). Números Índices. *Apuntes de Economía No. 4, ICESI.*

Beatty, T., & Roed, E. (2004). Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index. *Food and Resource Economics, University of British Columbia. Working Paper.*

Boskin, M., & Jorgenson, D. (1997). Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress. *American Economic Association, Papers and Proceedings, Vol 87 No 2, 89-93.*

Boskin, M., Dulberger, E., Gordon, R., Griliches, Z., & Jorgenson, D. (1996). *Toward a More Accurate Measure of the Cost of Living.* Final Report to the Senate Finance Committee.

Caicedo, E. (2000). Problemas en la Medición del IPC, el caso colombiano. *Banco de la República, Borradores de Economía No 152.*

Cárdenas, M. (2009). *Introducción a la Economía Colombiana.* Bogotá D.C: Alfaomega Colombiana S.A.

Carvalho Filho, I., & Chamon, M. (2006). The Myth of Post-Reform Income Stagnation in Brazil. *International Monetary Found. Working Paper 06/275.*

Costa, D. (2001). Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves. *Journal of Political Economy. Volume 109. No 6, 1288-1310.*

DANE. (2009). *Metodología Encuesta Nacional de Ingresos y Gastos 2006-2007.* Bogotá, D.C: Colección de documentos No. 84.

DANE. (2009). *Metodología Índice de precios al Consumidor.* Bogotá, D.C: Colección de Documentos No. 62.

- De Gregorio, J., & Huneus, C. (1999). El Índice de Precios al Consumidor y el Costo de la Vida. *Centro de Economía Aplicada, Departamento de Ingeniería Industrial, Universidad de Chile*.
- Deaton, A. (1998). Getting Prices Right: What Should Be Done? *Journal of Economic Perspectives Volume 12, Number 1* , 37-46.
- Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *The Economic American Review, Vol 7, No 3*, 312-326.
- Diewert, W. E. (1976). Exact and Superlative Index Numbers. *Journal of Econometrics, Vol 4*, 115-145.
- Gordon, R. (2000). The Boskin Commission Report and its Aftermath. *National Bureau of Economic Research*.
- Green, R., & Alston, J. M. (1990). Elasticities in AIDS Models. *American Journal of Agricultural Economics, Vol. 72, No. 2*, 442-445.
- Gujarati, D. N., & Porter, D. C. (2010). *Econometría*. Ciudad de México: McGraw Hill.
- Hamilton, B. (2001). Using Engel's Law to Estimate CPI Bias. *The American Economic Review, Volume 91 No 3*, 619-630.
- Hausman, J. (2003). Sources of Bias and Solutions to Bias in the Consumer Price Index. *The Journal of Economic Perspectives, Vol. 17, No. 1*, 23-44.
- Houthakker, H. S. (1957). An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law. *Econometrica, Vol 25*. pp. 532-521.
- Kaus, W. (2013). Beyond Engel's law - A cross-country analysis. *The Journal of Socio-Economics, Vol 47*, , 118-134.
- Langebaek, A., & Caicedo, E. (2007). Sesgo de Medición del IPC: nueva evidencia para Colombia. *Banco de la República, Borradores de Economía No 435*.
- Moschini, G. (1995). Units of Measurement and the Stone Index in Demand System Estimation. *American Journal of Agricultural Economics, Vol. 77, No. 1*, 63-68.
- Moulton, B. (1996). Bias in the Consumer Price Index: What is the Evidence? *The Journal of Economic Perspectives. Vol. 10, No. 4*, 159-177.
- OIT, FMI, OCDE, Eurostat, ONU y Banco Mundial. (2006). *Manual del índice de precios al consumidor. Teoría y práctica*. Washington: OIT / FMI / OCDE / Oficina Estadística de las Comunidades Europeas / Organización de las Naciones Unidas / Banco Mundial.

Presidencia de la República de Colombia. (26 de Diciembre de 1968). DECRETO 3167 DE 1968.
Por el cual se reorganiza el Departamento Administrativo Nacional de Estadística.

Shiratsuka, S. (1999). Measurement Errors in Japanese Consumer Price Index. *Bank of Japan and Federal Reserve Bank of Chicago. Working Paper.*

Varian, H. (1992). *Análisis Microeconómico*. Barcelona: Antoni Bosch.

ANEXOS

Anexo 1. Salidas Stata estadísticas descriptivas

```
. sum w ln_gasto_real ln_integrantes ln_precio_relativo_alimentos sex edad_jefe educacion_jefe vivi
```

Variable	Obs	Mean	Std. Dev.	Min	Max
w	11,260	.3146111	.2403797	.050046	.7997968
ln_gasto_real	11,260	13.02518	1.212264	7.796883	15.33205
ln_integrantes	11,260	1.336455	.4108372	.6931472	2.944439
ln_precio_relativo_alimentos	11,260	-2.503461	1.333922	-3.778887	0
sex	11,260	.6487567	.47738	0	1
edad_jefe	11,260	46.70009	14.83406	15	96
educacion_jefe	11,260	10.7357	5.564529	0	26
vivi	11,260	.5508881	.4974257	0	1

```
. sum w gasto_real integrantes precio_relativo_alimentos sex edad_jefe educacion_jefe vivi
```

Variable	Obs	Mean	Std. Dev.	Min	Max
w	11,260	.3146111	.2403797	.050046	.7997968
gasto_real	11,260	778987.8	723749.7	2433.005	4556437
integrantes	11,260	4.14627	1.819116	2	19
precio_relativo_alimentos	11,260	.2494651	.3944738	.0228481	1
sex	11,260	.6487567	.47738	0	1
edad_jefe	11,260	46.70009	14.83406	15	96
educacion_jefe	11,260	10.7357	5.564529	0	26
vivi	11,260	.5508881	.4974257	0	1

. by año: sum w gasto_real integrantes precio_rel~s sex edad_jefe educacion_jefe vivi

-> año = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
w	2,436	.6481114	.1400685	.0527313	.7997968
gasto_real	2,436	116490.5	115536.7	2433.005	1527325
integrantes	2,436	4.186371	1.697127	2	13
precio_rel~s	2,436	1	0	1	1
sex	2,436	.6843186	.4648821	0	1
edad_jefe	2,436	42.20402	13.01144	16	91
educacion_~e	2,436	8.166256	4.572378	0	21
vivi	2,436	.6190476	.4857206	0	1

-> año = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
w	8,824	.2225433	.170692	.050046	.7991111
gasto_real	8,824	961880.3	714226.7	4593.046	4556437
integrantes	8,824	4.135199	1.851316	2	19
precio_rel~s	8,824	.0422685	.0105677	.0228481	.0701091
sex	8,824	.6389393	.4803353	0	1
edad_jefe	8,824	47.9413	15.06516	15	96
educacion_~e	8,824	11.44504	5.605099	0	26
vivi	8,824	.5320716	.4989986	0	1

. ttest w, by (año)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	2,436	.6481114	.0028379	.1400685	.6425463	.6536764
1	8,824	.2225433	.0018171	.170692	.2189814	.2261053
combined	11,260	.3146111	.0022653	.2403797	.3101707	.3190516
diff		.425568	.0037662		.4181857	.4329504

diff = mean(0) - mean(1)

t = 112.9972

Ho: diff = 0

degrees of freedom = 11258

Ha: diff < 0

Ha: diff != 0

Ha: diff > 0

Pr(T < t) = 1.0000

Pr(|T| > |t|) = 0.0000

Pr(T > t) = 0.0000

. tab departamento

departament o	Freq.	Percent	Cum.
1	700	6.22	6.22
2	535	4.75	10.97
3	606	5.38	16.35
4	454	4.03	20.38
5	627	5.57	25.95
6	754	6.70	32.65
7	396	3.52	36.16
8	475	4.22	40.38
9	334	2.97	43.35
10	331	2.94	46.29
11	216	1.92	48.21
12	499	4.43	52.64
13	253	2.25	54.88
14	282	2.50	57.39
15	431	3.83	61.22
16	820	7.28	68.50
17	375	3.33	71.83
18	699	6.21	78.04
19	659	5.85	83.89
20	460	4.09	87.98
21	206	1.83	89.80
22	614	5.45	95.26
23	534	4.74	100.00
Total	11,260	100.00	

Fuente: DANE. Cálculos propios.

Anexo 2. Colinealidad y heterocedasticidad del modelo

Source	SS	df	MS	Number of obs	=	11,260
Model	397.602577	30	13.2534192	F(30, 11229)	=	588.30
Residual	252.969204	11,229	.022528204	Prob > F	=	0.0000
				R-squared	=	0.6112
				Adj R-squared	=	0.6101
Total	650.571781	11,259	.057782377	Root MSE	=	.15009

	w	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	año	-.0168926	.0484772	-0.35	0.727	-.1119163	.0781312
	ln_gasto_real	-.0786079	.0017513	-44.88	0.000	-.0820408	-.075175
	ln_integrantes	.0197357	.0036095	5.47	0.000	.0126605	.0268109
	ln_precio_relativo_alimentos	.0794827	.015335	5.18	0.000	.0494234	.1095421
	sex	.008278	.0030245	2.74	0.006	.0023495	.0142065
	edad_jefe	.0001325	.0001134	1.17	0.243	-.0000898	.0003549
	educacion_jefe	.0039352	.0002894	13.60	0.000	.0033679	.0045025
	vivi	.0097781	.0031323	3.12	0.002	.0036382	.0159179
	Medellin	-.0431529	.0083513	-5.17	0.000	-.0595229	-.0267828
	Barranquilla	-.0354251	.0108647	-3.26	0.001	-.0567218	-.0141283
	Cartagena	.0105022	.0096009	1.09	0.274	-.0083173	.0293216
	Tunja	-.026484	.0085895	-3.08	0.002	-.043321	-.009647
	Manizales	-.0110784	.0082432	-1.34	0.179	-.0272366	.0050798
	Florencia	-.0476123	.0097728	-4.87	0.000	-.0667686	-.028456
	Popayan	-.0218718	.0124854	-1.75	0.080	-.0463454	.0026019
	Valledupar	-.0665751	.0110109	-6.05	0.000	-.0881584	-.0449918
	Monteria	-.0719335	.0111513	-6.45	0.000	-.0937919	-.0500751
	Quibdo	-.0169983	.011985	-1.42	0.156	-.0404909	.0064944
	Neiva	-.0133646	.0108515	-1.23	0.218	-.0346355	.0079063
	Riohacha	-.0079412	.0112769	-0.70	0.481	-.030046	.0141635
	Santa_marta	-.0338934	.0112102	-3.02	0.003	-.0558673	-.0119195
	Villavicencio	-.041704	.0094791	-4.40	0.000	-.0602847	-.0231233
	Pasto	-.028666	.0082266	-3.48	0.000	-.0447916	-.0125405
	Cucuta	-.0492678	.0099652	-4.94	0.000	-.0688014	-.0297342
	Armenia	.000013	.0083759	0.00	0.999	-.0164053	.0164313
	Pereira	-.0064105	.0088779	-0.72	0.470	-.0238128	.0109917
	Bucaramanga	-.0522599	.0093028	-5.62	0.000	-.0704949	-.0340248
	Sincelejo	-.0101789	.0126414	-0.81	0.421	-.0349582	.0146003
	Ibague	-.0159644	.0094681	-1.69	0.092	-.0345235	.0025948
	Cali	-.0411885	.0090392	-4.56	0.000	-.0589069	-.0234701
	_cons	1.49078	.0217413	68.57	0.000	1.448163	1.533397

. estat hettes

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of w

chi2(1) = 220.91

Prob > chi2 = 0.0000

. corr año ln_precio_relativo_alimentos
(obs=11,260)

	año ln_pre~s	
año	1.0000	
ln_precio_~s	-0.9861	1.0000

. vif

Variable	VIF	1/VIF
ln_precio_~s	209.12	0.004782
año	199.14	0.005022
Popayan	3.15	0.317647
Barranquilla	2.67	0.374522
Neiva	2.49	0.401173
Ibague	2.31	0.432896
Pasto	2.28	0.437836
ln_gasto_r~l	2.25	0.443917
Pereira	2.17	0.460693
Manizales	2.12	0.471259
Armenia	2.04	0.489796
Medellin	2.03	0.492030
Tunja	1.94	0.515706
Cali	1.84	0.542032
Cartagena	1.78	0.560946
Monteria	1.77	0.563908
Valledupar	1.74	0.573339
Bucaramanga	1.69	0.590010
Villavicenc~o	1.65	0.604872
Florencia	1.62	0.617369
Cucuta	1.60	0.625792
Santa_marta	1.53	0.652032
Sincelejo	1.43	0.697093
edad_jefe	1.42	0.706509
Riohacha	1.40	0.716298
Quibdo	1.35	0.740306
educacion_~e	1.30	0.771404
vivi	1.21	0.824227
ln_integra~s	1.10	0.909912
sex	1.04	0.959841
Mean VIF	15.31	

```
. corr año ln_precio_no_alimentos ln_precio_alimentos
(obs=11,260)
```

	año	ln_precio_no_alimentos	ln_precio_alimentos
año	1.0000		
ln_precio_no_alimentos	0.0499	1.0000	
ln_precio_alimentos	-0.0078	0.9212	1.0000

Fuente: DANE. Cálculos propios.

Anexo 3. Corrección colinealidad por exclusión de variable ln precio relativo de alimentos

Linear regression	Number of obs	=	11,260
	F(29, 11230)	=	655.36
	Prob > F	=	0.0000
	R-squared	=	0.6102
	Root MSE	=	.15027

w	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
año	-.2666468	.0062593	-42.60	0.000	-.278916	-.2543775
ln_gasto_real	-.0786706	.0021744	-36.18	0.000	-.0829329	-.0744083
ln_integrantes	.0198526	.0035854	5.54	0.000	.0128245	.0268806
sex	.0080055	.0029887	2.68	0.007	.0021471	.0138638
edad_jefe	.000144	.0001125	1.28	0.201	-.0000765	.0003645
educacion_jefe	.0039515	.0002834	13.94	0.000	.0033959	.0045071
vivi	.0096899	.003167	3.06	0.002	.003482	.0158978
Medellin	-.0418012	.0082618	-5.06	0.000	-.0579957	-.0256067
Barranquilla	-.0033769	.009727	-0.35	0.728	-.0224435	.0156898
Cartagena	-.0009436	.0098627	-0.10	0.924	-.0202761	.0183889
Tunja	-.0295319	.0079915	-3.70	0.000	-.0451965	-.0138672
Manizales	-.0073768	.0080693	-0.91	0.361	-.0231942	.0084405
Florencia	-.0443363	.0092878	-4.77	0.000	-.062542	-.0261307
Popayan	-.0653665	.008592	-7.61	0.000	-.0822083	-.0485247
Valledupar	-.0460191	.0103415	-4.45	0.000	-.0662904	-.0257479
Monteria	-.049549	.0105501	-4.70	0.000	-.0702291	-.0288689
Quibdo	-.0216351	.0118327	-1.83	0.068	-.0448293	.0015591
Neiva	-.0439609	.0086471	-5.08	0.000	-.0609108	-.027011
Riohacha	-.0099675	.0117732	-0.85	0.397	-.033045	.01311
Santa_marta	-.0192134	.0118098	-1.63	0.104	-.0423627	.0039358
Villavicencio	-.042638	.0093215	-4.57	0.000	-.0609097	-.0243662
Pasto	-.0371536	.0077423	-4.80	0.000	-.0523299	-.0219773
Cucuta	-.0431328	.0099284	-4.34	0.000	-.0625943	-.0236714
Armenia	-.0033558	.0082311	-0.41	0.684	-.0194903	.0127787
Pereira	.0075899	.0080439	0.94	0.345	-.0081775	.0233573
Bucaramanga	-.0536441	.0090449	-5.93	0.000	-.0713737	-.0359146
Sincelejo	-.028285	.0126093	-2.24	0.025	-.0530014	-.0035687
Ibague	-.0364269	.0083149	-4.38	0.000	-.0527256	-.0201282
Cali	-.0340064	.0088976	-3.82	0.000	-.0514472	-.0165656

Fuente: DANE. Cálculos propios.

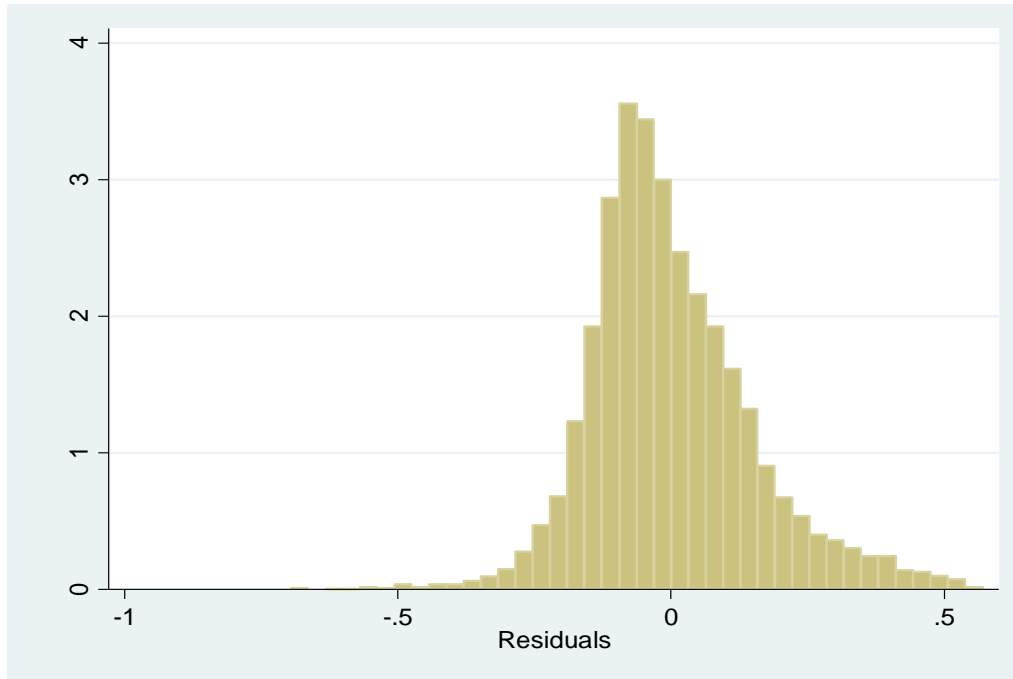
Anexo 4. Regresión errores robustos de White

```
Linear regression            Number of obs   =   11,260
                             F(30, 11229)    =   642.10
                             Prob > F      =   0.0000
                             R-squared     =   0.6112
                             Root MSE   =   .15009
```

w	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
año	-.0168926	.0495982	-0.34	0.733	-.1141137	.0803285
ln_gasto_real	-.0786079	.0021714	-36.20	0.000	-.0828642	-.0743517
ln_integrantes	.0197357	.0035815	5.51	0.000	.0127154	.0267561
ln_precio_relativo_alimentos	.0794827	.0155756	5.10	0.000	.0489517	.1100137
sex	.008278	.0029875	2.77	0.006	.0024221	.014134
edad_jefe	.0001325	.0001125	1.18	0.239	-.0000879	.000353
educacion_jefe	.0039352	.0002832	13.89	0.000	.00338	.0044903
vivi	.0097781	.003166	3.09	0.002	.0035722	.015984
Medellin	-.0431529	.0082623	-5.22	0.000	-.0593484	-.0269573
Barranquilla	-.0354251	.0113359	-3.13	0.002	-.0576455	-.0132046
Cartagena	.0105022	.0101008	1.04	0.298	-.0092972	.0303015
Tunja	-.026484	.0080113	-3.31	0.001	-.0421875	-.0107805
Manizales	-.0110784	.0080967	-1.37	0.171	-.0269494	.0047927
Florencia	-.0476123	.0093073	-5.12	0.000	-.0658562	-.0293684
Popayan	-.0218718	.0122612	-1.78	0.074	-.0459058	.0021623
Valledupar	-.0665751	.0110768	-6.01	0.000	-.0882875	-.0448627
Monteria	-.0719335	.0114299	-6.29	0.000	-.094338	-.049529
Quibdo	-.0169983	.0118717	-1.43	0.152	-.0402689	.0062724
Neiva	-.0133646	.0104169	-1.28	0.200	-.0337836	.0070544
Riohacha	-.0079412	.0117924	-0.67	0.501	-.0310565	.015174
Santa_marta	-.0338934	.0121009	-2.80	0.005	-.0576133	-.0101736
Villavicencio	-.041704	.0093162	-4.48	0.000	-.0599653	-.0234427
Pasto	-.028666	.0079395	-3.61	0.000	-.0442288	-.0131033
Cucuta	-.0492678	.0100004	-4.93	0.000	-.0688704	-.0296653
Armenia	.000013	.008254	0.00	0.999	-.0161663	.0161923
Pereira	-.0064105	.0084245	-0.76	0.447	-.022924	.0101029
Bucaramanga	-.0522599	.0090359	-5.78	0.000	-.0699717	-.034548
Sincelejo	-.0101789	.0131135	-0.78	0.438	-.0358837	.0155259
Ibague	-.0159644	.0092623	-1.72	0.085	-.03412	.0021913
Cali	-.0411885	.0089864	-4.58	0.000	-.0588033	-.0235737
_cons	1.49078	.0265145	56.23	0.000	1.438807	1.542753

Fuente: DANE. Cálculos propios.

Anexo 5. Distribución de los errores



Fuente: DANE. Cálculos propios.

Anexo 6. Prueba de endogeneidad

```
. reg residuo año ln_gasto_real ln_integrantes ln_precio_relativo_alimentos sex edad_jefe ed
> ucacion_jefe vivi
```

Source	SS	df	MS	Number of obs =	11,260
Model	0	8	0	F(8, 11251) =	0.00
Residual	252.969204	11,251	.022484153	Prob > F =	1.0000
Total	252.969204	11,259	.022468177	R-squared =	0.0000
				Adj R-squared =	-0.0007
				Root MSE =	.14995

residuo	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
año	8.33e-10	.0212247	0.00	1.000	-.0416041 .0416041
ln_gasto_real	-2.83e-11	.001743	-0.00	1.000	-.0034167 .0034167
ln_integrantes	1.18e-10	.0035461	0.00	1.000	-.0069509 .0069509
ln_precio_relativo_alim~s	2.29e-10	.0063921	0.00	1.000	-.0125296 .0125296
sex	-5.43e-11	.0030106	-0.00	1.000	-.0059012 .0059012
edad_jefe	3.44e-12	.0001122	0.00	1.000	-.0002198 .0002198
educacion_jefe	1.67e-11	.0002866	0.00	1.000	-.0005618 .0005618
vivi	3.73e-11	.00311	0.00	1.000	-.0060962 .0060962
_cons	-1.48e-10	.0206408	-0.00	1.000	-.0404596 .0404596

Fuente: DANE. Cálculos propios.

Anexo 7. Salidas Stata elasticidades

```
. sum elast_gasto
```

Variable	Obs	Mean	Std. Dev.	Min	Max
elast_gasto	11,260	.5112205	.3916315	-.5707131	.9017152

```
. sum elast_mars elast_hick
```

Variable	Obs	Mean	Std. Dev.	Min	Max
elast_mars	11,260	-.3540505	.5168453	-.8767869	1.180874
elast_hick	11,260	-.1911699	.39299	-.662619	.9623713

Fuente: DANE. Cálculos propios.

Anexo 8. Do file Stata

Base de datos "personas" de la página de microdatos del DANE ENIG 1995/95

```
gen sex= sexo
```

```
recode sex (2=1) (else=0)
```

```
gen jefe= parentesco
```

```
recode jefe (1=1) (else=0)
```

```
gen edad1=edad
```

```
recode edad1 (0/4=1) (5/9=2) (10/15=3) (16/18=4) (18/98=5)
```

```
quietly tabulate edad, generate(new_)
```

```
sort num_orden
```

```
by num_orden: egen number_new_1= sum (new_1)
```

```
by num_orden: egen number_new_2= sum (new_2)
```

```
by num_orden: egen number_new_3= sum (new_3)
```

```
by num_orden: egen number_new_4= sum (new_4)
```

```
by num_orden: egen number_new_5= sum (new_5)
```

```
gen integrantes = number_new_1+ number_new_2+ number_new_3+ number_new_4+ number_new_5
```

```
gen nivel_edu= nivel_educativo
```

```
recode nivel_edu (6/9=0)
```

```
generate edu= ultimo_ano_aprob
```

```
recode nivel_edu (0=99)
```

```
recode nivel_edu (5=16)
```

```
recode nivel_edu (4=11)
recode nivel_edu (3=5)
recode nivel_edu (2=0)
gen educacion= nivel_edu+ edu
recode educacion (99=0)
rename edad edad_jefe
rename nivel_edu nivel_edu_jefe
rename number_new_1 edad_1_a_4
rename number_new_2 edad_5_a_9
rename number_new_3 edad_10_a_15
rename number_new_4 edad_16_a_18
rename number_new_5 edad_mayor_18
rename educacion edu_jefe
drop nivel_edu_jefe
drop new_1 new_2 new_3 new_4 new_5
quietly tabulate departamento, generate(new_)
rename new_1 Medellin
label variable Medellin ""
rename new_2 Barranquilla
label variable Barranquilla ""
rename new_3 Bogota
label variable Bogota ""
rename new_4 Cartagena
label variable Cartagena ""
rename new_5 Tunja
label variable Tunja ""
rename new_6 Manizales
label variable Manizales ""
rename new_7 Florencia
label variable Florencia ""
rename new_8 Popayan
label variable Popayan ""
rename new_9 Valledupar
label variable Valledupar ""
rename new_10 Monteria
label variable Monteria ""
rename new_11 Quibdo
label variable Quibdo ""
rename new_12 Neiva
label variable Neiva ""
rename new_13 Riohacha
label variable Riohacha ""
rename new_14 Santa_marta
label variable Santa_marta ""
```



```

rename new_15 Villavicencio
label variable Villavicencio ""
rename new_16 Pasto
label variable Pasto ""
rename new_17 Cucuta
label variable Cucuta ""
rename new_18 Armenia
label variable Armenia ""
rename new_19 Pereira
label variable Pereira ""
rename new_20 Bucaramanga
label variable Bucaramanga ""
rename new_21 Sincelejo
label variable Sincelejo ""
rename new_22 Ibague
label variable Ibague ""
rename new_23 Cali
label variable Cali ""
keep if jefe==1
***mÃ³dulo "vivi_hogar"***
gen vivienda_propia= cond_viven
recode vivienda_propia (2/3=1) (else=0)
merge 1:1 num_orden using vivi_hogar1
***archivo personas 1.1***
***mÃ³dulo ingresos***
sort num_orden
by num_orden: egen total_income= sum ( tot_ing_mes_dis)
keep if orden_per==1
***se cambia a la base de datos personas, y se cambia de directorio***
drop _merge
merge 1:1 num_orden using ingresos1
drop _merge
***mÃ³dulo gasto hogar***
*se elimina las observaciones donde el cÃ³digo del artÃ­culo es 0 o donde no se tiene informaciÃ³n sobre
el tipo de artÃ­culo que sea (cÃ³digo 9999),*
*en total se eliminaron 12776 observaciones del cÃ³digo 0, y 223 obs del cÃ³digo 9999*
drop if cod_art==9999
drop if cod_art==0
recode cod_art (1920/8615=0) (else=1)
quietly tabulate cod_art, generate(new_)
rename new_2 Alimentos
label variable Alimentos ""
rename new_1 No_alimentos
label variable No_alimentos ""

```

```

sort num_orden
by num_orden: egen gasto_total=sum ( valor_tot)
by num_orden: egen gasto_alimentos=sum ( valor_tot) if cod_art==1
by num_orden: egen gasto_noalimentos=sum ( valor_tot) if cod_art==0
recode gasto_alimentos (.=0)
recode gasto_noalimentos (.=0)
duplicates drop num_orden gasto_alimentos gasto_noalimentos, force
by num_orden: egen gasto_alimentos1=sum(gasto_alimentos)
by num_orden: egen gasto_noalimentos1=sum(gasto_noalimentos)
duplicates drop num_orden gasto_alimentos1 gasto_noalimentos1, force
drop gasto_alimentos gasto_noalimentos
rename gasto_alimentos1 gasto_alimentos
rename gasto_noalimentos1 gasto_noalimentos
*abro el mÃ³dulo donde ya depure las variables anteriores y las incluyo*
merge 1:1 num_orden using gas_hogar
drop _merge

```

me quedo con las observaciones donde el hogar reportÃ³ algun tipo de gasto en alimentos y en no alimentos, que en total son 14804

```

drop if gasto_alimentos==0
drop if gasto_noalimentos==0

```

se eliminan hogares unipersonales 523 observaciones eliminadas

```

drop if npers_hog==1

```

ahora se elimina el 5% de los hogares con mayores ingresos en cada ciudad porque presentan gran dispersiÃ³n en los hÃ¡bitos de consumo

```

sort Medellin total_income
by Medellin: sum total_income, detail
browse Medellin total_income
drop in 14220/14281
*62 obs eliminadas para Medellin*

```

```

sort Barranquilla total_income
by Barranquilla: sum total_income, detail
browse Barranquilla total_income
drop in 14187/14219
*33 obs eliminadas para Barranquilla*

```

```

sort Bogota total_income
by Bogota: sum total_income, detail
br Bogota total_income
drop in 14149/14186
*38 obs eliminadas para BogotÃ¡*

```

sort Cartagena total_income
by Cartagena : sum total_income, detail
browse Cartagena total_income
drop in 14119/14148
30 obs eliminadas para Cartagena

sort Tunja total_income
by Tunja : sum total_income, detail
browse Tunja total_income
drop in 14088/14118
31 obs eliminadas para Tunja

sort Manizales total_income
by Manizales : sum total_income, detail
browse Manizales total_income
drop in 14055/14087
33 obs eliminadas para Manizales

sort Florencia total_income
by Florencia : sum total_income, detail
browse Florencia total_income
drop in 14033/14054
22 obs eliminadas para Florencia

sort Popayan total_income
by Popayan: sum total_income, detail
browse Popayan total_income
drop in 14009/14032
24 obs eliminadas para popayan

sort Valledupar total_income
by Valledupar : sum total_income, detail
browse Valledupar total_income
drop in 13992/14008
17 obs eliminadas para Valledupar

sort Monteria total_income
by Monteria : sum total_income, detail
browse Monteria total_income
drop in 13956/13991
36 obs eliminadas para Monteria

sort Quibdo total_income

by Quibdo : sum total_income, detail
browse Quibdo total_income
drop in 13938/13955
18 obs eliminadas para QuibdÃ³

sort Neiva total_income
by Neiva : sum total_income, detail
browse Neiva total_income
drop in 13915/13937
23 obs eliminadas para Neiva

sort Riohacha total_income
by Riohacha : sum total_income, detail
browse Riohacha total_income
drop in 13904/13914
11 obs eliminadas para Riohacha

sort Santa_marta total_income
by Santa_marta : sum total_income, detail
browse Santa_marta total_income
drop in 13881/13903
23 obs eliminadas para santa marta

sort Villavicencio total_income
by Villavicencio : sum total_income, detail
browse Villavicencio total_income
drop in 13859/13880
22 obs eliminadas para villavicencio

sort Pasto total_income
by Pasto : sum total_income, detail
browse Pasto total_income
drop in 13822/13858
37 obs elimimnadas para Pasto

sort Cucuta total_income
by Cucuta : sum total_income, detail
browse Cucuta total_income
drop in 13787/13821
35 obs eliminadas para cucuta

sort Armenia total_income
by Armenia : sum total_income, detail
browse Armenia total_income

drop in 13756/13786

31 obs eliminadas para Armenia

sort Pereira total_income

by Pereira : sum total_income, detail

browse Pereira total_income

drop in 13713/13755

43 obs eliminadas para Pereria

sort Bucaramanga total_income

by Bucaramanga : sum total_income, detail

browse Bucaramanga total_income

drop in 13674/13712

39 obs eliminadas para Bucaramanga

sort Sincelejo total_income

by Sincelejo : sum total_income, detail

browse Sincelejo total_income

drop in 13660/13673

14 obs eliminadas para Sincelejo

sort Ibague total_income

by Ibague : sum total_income, detail

browse Ibague total_income

drop in 13628/13659

32 obs eliminadas para ibague

sort Cali total_income

by Cali : sum total_income, detail

browse Cali total_income

drop in 13557/13627

71 obs eliminadas para Cali

ahora se transforman las variables de gasto e ingreso al deflactarlas por el IPC base 2008, para esta base de datos se utilizÃ³ el dato 1994-12 de la pÃ¡gina de banrep

gen ipc=0.3123709 (ipc 1995/12)

gen ingreso_real= total_income/ ipc

gen gasto_real= gasto_total/ ipc

drop if gasto_total==.

gen w= gasto_alimentos/ gasto_total

se elimina los valores donde el gasto en alimentos es mayor al ingreso, es decir, aquellos valores de w1 maypres a 1

gen w1= gasto_alimentos/total_income

```
sort w1
br w1
drop in 13361/13474
*se eliminaron 114 obs*
```

```
sort departamento
by departamento: egen gasto_alimentosxdepart= sum( gasto_alimentos)
browse departamento gasto_alimentosxdepart
by departamento: egen gasto_Noalimenxdepart= sum( gasto_noalimentos)
browse departamento gasto_alimentosxdepart gasto_Noalimenxdepart
*pendiente obtener el gasto por departamento de la segunda encuesta para crear los índices*
```

```
gen ln_ingreso_real=log( ingreso_real)
gen ln_gasto_real=log( gasto_real)
gen ln_integrantes=log( integrantes)
```

Se empieza a depurar los módulos de la ENIG 2006/017

Módulo ig_ml_persona

```
gen sex= p6020
recode sex (2=1) (else=0)
gen jefe= p200
recode jefe (1=1) (else=0)
gen edad_jefe= p6040
gen edad= p6040
recode edad (0/4=1) (5/9=2) (10/15=3) (16/18=4) (18/100=5)
quietly tabulate edad, generate(new_)
sort vivienda
by vivienda : egen number_new_1= sum (new_1)
by vivienda : egen number_new_2= sum (new_2)
by vivienda : egen number_new_3= sum (new_3)
by vivienda : egen number_new_4= sum (new_4)
by vivienda : egen number_new_5= sum (new_5)
gen integrantes= number_new_1+ number_new_2+ number_new_3+ number_new_4+ number_new_5
gen nivel_educativo= p6210
recode nivel_educativo (1=0) (9=0)
recode nivel_educativo (6=11)
recode nivel_educativo (4/5=5)
recode nivel_educativo (2/3=0)
gen anios_edu= p6210s1
gen educacion= nivel_educativo + anios_edu
rename educacion educacion_jefe
recode educacion_jefe (99=0) (.=0)
```

```

rename number_new_1 edad_1_a_4
rename number_new_2 edad_5_a_9
rename number_new_3 edad_10_a_15
rename number_new_4 edad_16_a18
rename number_new_5 edad_mayor_18
drop new_1 new_2 new_3 new_4 new_5 nivel_educativo anios_edu edad
keep if jefe==1
  duplicates drop vivienda, force
*se guarda como 1 en escritorio*
*se abre el módulo ig_ml_hogar*
gen vivi=p5090
recode vivi (1/2=1) (else=0)
  duplicates drop vivienda, force
*Se unen las bases de datos*
* Cree una variable id en cada base de datos para que stata lo reconozca y las pueda unir*
merge 1:1 id using vivienda1
drop _merge

```

```

*se abre el módulo de ig_ml_ocupado*
gen sueldo=p7080/12
replace p6510s1=0 if p6510s2==1
gen horas_extra=p6510s1
gen alimento_como_sueldo_especie=p6590s1
gen vivi_como_salario_especie=p6600s1
gen transporte_como_salario_especie=p6610s1
replace p6585s1a1=0 if p6585s1a2==1
gen subsidio_aliment= p6585s1a1
replace p6585s2a1=0 if p6585s2a2==1
gen subsidio_transporte= p6585s2a1
replace p6585s3a1=0 if p6585s3a2==1
gen subsidio_familiar= p6585s3a1
gen subsidio_educativo= p6585s4a1
gen ingresos_gastos_representacion= p6530s1
replace p6540s1=0 if p6540s2==1
gen prima_tecn= p6540s1
replace p6550s1=0 if p6550s2==1
gen prima_antig= p6550s1
replace p6560s1=0 if p6560s2==1
gen prima_clima= p6560s1
replace p6570s1=0 if p6570s2==1
gen primas_otras= p6570s1
replace p6580s1=0 if p6580s2==1
gen bonif_mens= p6580s1
gen prima_servicios=p6630s1a1/12

```

```

gen prima_navidad=p6630s2a1/12
gen prima_vacaciones=p6630s3a1/12
gen viaticos_perm=p6630s4a1/12
gen bonif_anuales=p6630s5a1/12
gen honorarios= p6750/p6760
gen cosechas=p550/12
gen sueldo_2_empleo=p7070
gen todos_trabajos=p7080/12
gen aportes_fondo_solid=p6625s1
gen aportes_pens=p6950
gen ingreso_disponible= sueldo+ horas_extra+ alimento_como_sueldo_especie+
vivi_como_salario_especie+ transporte_como_salario_especie+ subsidio_aliment+ subsidio_transporte+
subsidio_familiar+ subsidio_educativo+ ingresos_gastos_representacion+ prima_tecn+ prima_antig+
prima_clima+ primas_otras+ bonif_mens+ prima_servicios+ prima_navidad+ prima_vacaciones+
viaticos_perm+ bonif_anuales+ honorarios+ cosechas+ sueldo_2_empleo+ todos_trabajos+
aportes_fondo_solid+ aportes_pens
sort vivienda
by vivienda: egen ingreso_total= sum ( ingreso_disponible )
duplicates drop vivienda ingreso_total, force
drop if ingreso_total==0
*se guarda como ingresos 06*
*abre la base de datos 06-07.Personas1*

merge 1:1 vivienda using ingresos1
keep if _merge==3
drop _merge

***se abre el módulo de ig_gsdu_caract_alim***
sort vivienda
by vivienda: egen gasto_alimentos= sum ( gdu_valor_ultima_cmpra_mes)
duplicates drop vivienda gasto_alimentos,force
sum gasto_alimentos

*se anexa a la base de datos 06-07.Personas1*
merge 1:1 vivienda using alimentos1
keep if _merge==3
drop _merge

***se abre el módulo ig_gsmf_compra***

sort vivienda
by vivienda: egen gasto_noalimentos=sum( gmf_cmpra_vlr_pago_mes)
duplicates drop vivienda gasto_noalimentos,force

```


se combinan las BD

```
merge 1:1 vivienda using ENIG2007.8.dta
keep if _merge==3
drop _merge
drop if gasto_noalimentos==0
drop if gasto_alimentos==0
gen gasto_total= gasto_alimentos+ gasto_noalimentos
gen w= gasto_alimentos/ gasto_total
gen w1= gasto_alimentos/ ingreso_total
```

se eliminan las obs donde el gasto en alimentos es mayor a 1 (tomando el ingreso)

```
sort w1
br w1
drop in 20644/21380
```

*para identificar el departamento al que pertenece el hogar se utiliza el módulo `ig_gs_vivienda`, para tener consistencia con la BD 94-95 se eliminan los hogares de otras cabecera y el dominio resto de la variable "dominio"

de igual forma, solo se trabajarán con las mismas ciudades que la BD 94-95, por tanto, solo se eliminan a San andrés y providencia

```
merge 1:1 vivienda using departamento1
keep if _merge==3
drop _merge
sort departamento
encode dominio, gen (a)
drop if a==14
drop if departamento==88
```

se eliminan hogares unipersonales

```
drop if integrantes==1
```

se elimina el 5% de los hogares con mayores ingresos por ciudad

```
quietly tabulate departamento, generate(new_)
```

```
rename new_1 Medellin
```

```
rename new_2 Barranquilla
```

```
rename new_3 Bogota
```

```
rename new_4 Cartagena
```

```
rename new_5 Tunja
```

```
rename new_6 Manizales
```

```
rename new_7 Florencia
```

```
rename new_8 Popayan
```

```
rename new_9 Valledupar
```

```
rename new_10 Monteria
```

```
rename new_11 Quibdo
```

rename new_12 Neiva
rename new_13 Riohacha
rename new_14 Santa_marta
rename new_15 Villavicencio
rename new_16 Pasto
rename new_17 Cucuta
rename new_18 Armenia
rename new_19 Pereira
rename new_20 Bucaramanga
rename new_21 Sincelejo
rename new_22 Ibague
rename new_23 Cali

sort Medellin ingreso_total
by Medellin: sum ingreso_total, detail
browse Medellin ingreso_total
drop in 16695/16737
43 obs eliminadas para medellin

sort Barranquilla ingreso_total
by Barranquilla : sum ingreso_total, detail
browse Barranquilla ingreso_total
drop in 16653/16694
42 obs eliminadas para barranquilla

sort Bogota ingreso_total
by Bogota : sum ingreso_total, detail
browse Bogota ingreso_total
drop in 16618/16652
35 obs eliminadas para bogotÃ¡

sort Cartagena ingreso_total
by Cartagena : sum ingreso_total, detail
browse Cartagena ingreso_total
drop in 16586/16617
32 obs eliminadas para cartagena

sort Tunja ingreso_total
by Tunja : sum ingreso_total, detail
browse Tunja ingreso_total
drop in 16544/16585
42 obs eliminadas para tunja

sort Manizales ingreso_total
by Manizales : sum ingreso_total, detail
browse Manizales ingreso_total
drop in 16498/16543
46 obs eliminadas para manizales

sort Florencia ingreso_total
by Florencia : sum ingreso_total, detail
browse Florencia ingreso_total
drop in 16463/16497
35 obs eliminadas para florencia

sort Popayan ingreso_total
by Popayan : sum ingreso_total, detail
browse Popayan ingreso_total
drop in 16418/16462
45 obs eliminadas para popayan

sort Valledupar ingreso_total
by Valledupar : sum ingreso_total, detail
browse Valledupar ingreso_total
drop in 16392/16417
26 obs eliminadas para valledupar

sort Monteria ingreso_total
by Monteria : sum ingreso_total, detail
browse Monteria ingreso_total
drop in 16363/16391
29 obs eliminadas para monteria

sort Quibdo ingreso_total
by Quibdo : sum ingreso_total, detail
browse Quibdo ingreso_total
drop in 16347/16362
16 obs eliminadas para quibdo

sort Neiva ingreso_total
by Neiva : sum ingreso_total, detail
browse Neiva ingreso_total
drop in 16303/16346
44 obs eliminadas para neiva

sort Riohacha ingreso_total
by Riohacha : sum ingreso_total, detail

browse Riohacha ingreso_total
drop in 16279/16302
24 obs eliminadas para riohacha

sort Santa_marta ingreso_total
by Santa_marta : sum ingreso_total, detail
browse Santa_marta ingreso_total
drop in 16255/16278
24 obs eliminadas para santa marta

sort Villavicencio ingreso_total
by Villavicencio : sum ingreso_total, detail
browse Villavicencio ingreso_total
drop in 16218/16254
37 obs eliminadas villavicencio

sort Pasto ingreso_total
by Pasto : sum ingreso_total, detail
browse Pasto ingreso_total
drop in 16163/16217
55 obs eliminadas para Pasto

sort Cucuta ingreso_total
by Cucuta : sum ingreso_total, detail
browse Cucuta ingreso_total
drop in 16125/16162
38 obs eliminadas para cÃ³cuta

sort Armenia ingreso_total
by Armenia : sum ingreso_total, detail
browse Armenia ingreso_total
drop in 16090/16124
37 obs eliminadas para armenia

sort Pereira ingreso_total
by Pereira : sum ingreso_total, detail
browse Pereira ingreso_total
drop in 16018/16052
35 obs eliminadas para pereira

sort Bucaramanga ingreso_total
by Bucaramanga : sum ingreso_total, detail
browse Bucaramanga ingreso_total
drop in 15971/16017

47 obs eliminadas para bucaramanga

```
sort Sincelejo ingreso_total
by Sincelejo : sum ingreso_total, detail
browse Sincelejo ingreso_total
drop in 15948/15970
*23 obs eliminadas para sincelejo*
```

```
sort Ibague ingreso_total
by Ibague : sum ingreso_total, detail
browse Ibague ingreso_total
drop in 15897/15947
*51 obs eliminadas para Ibague*
```

```
sort Cali ingreso_total
by Cali : sum ingreso_total, detail
browse Cali ingreso_total
drop in 15863/15896
*34 obs eliminadas para Cali*
```

se deflactan las series de ingreso y gasto por el ipc 2007-12 que fue de 0.9287228

```
gen ipc=0.9287228
gen ingreso_real= ingreso_total/ ipc
gen ln_ingreso_real=log( ingreso_real)
gen ln_gasto_real=log( gasto_real)
gen ln_integrantes=log( integrantes)
```

```
recode sex (1=0) (0=1)
```

```
sort departamento
by departamento: egen gasto_alimentosxdepart= sum( gasto_alimentos)
browse departamento gasto_alimentosxdepart
by departamento: egen gasto_Noalimenxdepart= sum( gasto_noalimentos)
```

***se sacan dos bases de datos**

base de datos w=> para el año 94-95 el gasto en no alimentos es = ingreso total - gasto total

base de datos w1=> para el año 94-95 el gasto en no alimentos es el mismo que el reportado en la base de datos

Base de datos W

se eliminan las observaciones donde la participacion del gasto es menor que 0.05 o mayor a 0.8

```
gen departamento=0
```

replace departamento=1 if Medellin==1
replace departamento=2 if Barranquilla ==1
replace departamento=3 if Bogota ==1
replace departamento=4 if Cartagena ==1
replace departamento=5 if Tunja ==1
replace departamento=6 if Manizales ==1
replace departamento=7 if Florencia ==1
replace departamento=8 if Popayan ==1
replace departamento=9 if Valledupar ==1
replace departamento=10 if Monteria ==1
replace departamento=11 if Quibdo ==1
replace departamento=12 if Neiva ==1
replace departamento=13 if Riohacha ==1
replace departamento=14 if Santa_marta ==1
replace departamento=15 if Villavicencio ==1
replace departamento=16 if Pasto ==1
replace departamento=17 if Cucuta ==1
replace departamento=18 if Armenia ==1
replace departamento=19 if Pereira ==1
replace departamento=20 if Bucaramanga ==1
replace departamento=21 if Sincelejo ==1
replace departamento=22 if Ibague ==1
replace departamento=23 if Cali ==1

gen B=-0.0786079

gen elast_gasto=1+(B/w)

sum elast_gasto