Estimación de sendas salariales y aplicación al problema de la existencia de subsidios en el régimen de prima media

Luis Esteban Martínez Roldan

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

Estimación de sendas salariales y aplicación al problema de la existencia de subsidios en el régimen de prima media

Luis Esteban Martínez Roldan

Tesis de grado presentada como requisito parcial para optar al título de:
Magister en Ciencias-Estadísticas

Director:
Norman D. Giraldo Gomez, M.cs
Profesor Asociado Escuela de Estadística

Línea de Investigación:
Línea de Finanzas y Actuaria

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2016

A mi madre que está en el cielo e ilumina mi camino, a mi padre que ha sido mi fuerza y ejemplo.

"Venzan sus miedos y vencerán la muerte".

Alejandro Magno

Agradecimientos

Agradezco a mi asesor el profesor Norman Giraldo por su inmensa colaboración y dedicación para sacar este proyecto adelante. A mis amigos Sergio, Ricardo y Miguel por sus animos y apoyo durante la maestría. A Claudia Gisela, mi novia, por su amor y comprensión.

Resumen

Este trabajo muestra la metodología de cálculo de la senda salarial propuesta por Montenegro, Jiménez y Hurtado (2013), y además propone una serie de metodologías para la construción de la senda salarial por medio de regresiones no paramétricas, enfocándose principalmente en el Régimen de Prima Media con Prestación Definida (RPM-PD). Teniendo en cuenta esto, el objetivo de este trabajo es realizar un estudio comparativo entre estas dos metodologías utilizando los datos de la Alcaldía de Medellín para los periodos 2008, 2011 y 2014; luego se determina cuál metodología es la más apropiada para cuantificar la senda salarial. Así mismo, se utilizan las sendas salariales calculadas para estimar los aportes de un trabajador genérico durante su vida laboral, tales aportes más los rendimientos financieros conforman el capital acumulado que se invierte en la compra de una pensión mediante una renta vitalicia. Eventualmente, con base en esta estimación se establecerá, si las condiciones permiten que existan subsidios en el RPM-PD, y en caso afirmativo, cuantificar el subsidio que será transferido por el Estado.

Palabras clave: Esquemas pensionales, modelación econométrica, Senda Salarial, Subsidio.

Abstract

This paper shows the methodology for calculating the wage path proposed by Montenegro, Jiménez and Hurtado (2013), and also proposes a series of methodologies for the construction of the wage path through non-parametric regressions, focusing mainly on the Premium Regime Defined Benefit Media (RPM-PD). Taking into account this, the objective of this work is to conduct a comparative study between these two methodologies using the data of the Alcaldía of Medellín for the periods 2008, 2011 and 2014; then determine which methodology is most appropriate to quantify the path wage. Likewise, the salary paths calculated to estimate the contributions of a generic worker during his working life are used, such contributions plus the financial returns are the accumulated capital that is invested in the purchase of a pension through an annuity. Eventually, based on this estimate will be established, if the conditions allow for subsidies in the RPM-PD, and if so, quantify the subsidy that will be transferred by the State

Keywords: Pension schemes, econometric modeling, Wage Path, Subsidy.

Lista de Figuras

3-1. Mardia Test 12
3-2. Histogramas Salario y Edad 13
3-3. Normal Q-Q Salario-Edad 13
3-4. Boxplot 14
3-5. Violín plot 15
3-6. Gráfico de dispersión: Salario vs Edad 16
3-7. Vida media residual: (arriba)2008-2011 y (abajo) 2014 17
3-8. Contornos 18
3-9. Estimación escalas salariales N-W y Loess, 2008 29
3-10.Estimación escalas salariales N-W y Loess, 2011 30
3-11.Estimación escalas salariales N-W y Loess, 2014 30
3-12.Estimación de la escala salarial Loess, h=0.20 31
3-13.Estimación de la escala salarial Loess, h=2 32
3-14.Estimación de la escala salarial Loess, h=0.75 33
3-15.Estimación de los incrementos salariales 34
3-16.Estimación de la senda salarial Loess 34
3-17.Incrementos salariales Loess y N-W 35
3-18.Estimación de la senda salarial 36

5-1. Senda Salariales 2014. Metodología de Montenegro et al (2013) 49

Lista de Tablas

2-1. Salarios Medios Nacionales 4

3-1. Estadísticas Descriptivas 10
3-2. Estadísticas Descriptivas 11
3-3. Test de Normalidad Multivariada de Mardia 11
3-4. Test de Normalidad: Kolmogorov-Smirnov 14
3-5. Estimación de la Correlación Salario vs Edad 16
3-6. Resultados de la validación cruzada 28

4-1. Parámetros Gompertz-Makeham por género 40
4-2. Parámetros de la Renta Vitalicia 48
4-3. Estimaciones para afiliado hombre, edad = 62 48
4-4. Estimaciones para afiliado mujer, edad = 57 48

5-1. Estimaciones para afiliado hombre, edad = 62 53
5-2. Estimaciones para afiliado mujer, edad = 57 54

A-1. Transformación Box-Cox Robusta 57
A-2. Modelo de Regresión No-lineal, modelo 1= $Y^{-1,65}$ 58
A-3. Modelo de Regresión No-lineal, modelo 2 = $Y^{-1,65}$ 58
A-4. Modelo de Regresión No-lineal, modelo 3 = $Y^{-1,65}$ 59
A-5. Modelo de Regresión No-lineal, modelo 4= Y^2 59
A-6. Modelo de Regresión No-lineal, modelo 5 = Y^2 60
A-7. Modelo de Regresión No-lineal, modelo 6 = Y^2 60
A-8. Modelo de Regresión lineal múltiple, modelo 1= $Y^{-1,65}$ 61
A-9. Modelo de Regresión lineal múltiple, modelo 2 = $Y^{-1,65}$ 61
A-10. Modelo de Regresión lineal múltiple, modelo 3 = Y^2 61
A-11. Modelo de Regresión lineal múltiple, modelo 4 = Y^2 62

Contenido

Agradecimientos vii

Resumen ix

List of Figures x

List of Tables xi

1. Introducción 1

2. Revisión de la definición de senda salarial 3
2.1. Definición en la legislación pensional 3
2.2. Definición en Carriere y Shand (1998) y Modelos estadísticos para la función salarial S_x 5
2.3. Definición en Montenegro, Jiménez y Hurtado (2013) 7
2.4. Metodología de la tesis 8

3. Estimación de los Modelos 10
3.1. Análisis preliminar de los datos de los salarios 10
3.2. Observaciones en la implementación de métodos paramétricos 19
3.3. Estimación mediante regresión no paramétrica 20
3.3.1. Supuestos de la regresión no paramétrica 20
3.3.2. Estimación del modelo de regresión no paramétrica 21
3.3.3. Selección del núcleo, K 23
3.3.4. Selección del polinomio, p 24
3.3.5. Métodos de selección de banda, h 24
3.3.6. Efecto de borde en N-W 26
3.3.7. Relación Loess y regresión lineal local Nadaraya-Watson 26
3.3.8. Estimaciones 27
3.4. Conclusiones acerca de las estimaciones 36

4. Aplicación: la cuestión de los subsidios en el RPM-PD 38
4.1. Cálculo de la mesada pensional 38
4.2. Cálculo de las rentas vitalicias 39
4.3. Cálculo de los subsidios 43

4.4. Ecuación de capitalización de aportes 44
4.5. Ejemplos numéricos 47

5. Análisis de la metodología en Montenegro, Jiménez y Hurtado (2013) 49
5.1. Análisis de la metodología para sendas salariales 49
5.1.1. Cálculo del capital acumulado 50
5.1.2. Fórmulas de renta vitalicia 52
5.1.3. Fórmulas del Ingreso Base de Liquidación (IBL) 52
5.1.4. Estimaciones 53
5.2. Conclusiones 54

6. Conclusiones 55

A. Aplicaciones con transformaciones Box-Cox y logarítmica 57
A.1. Resultados regresión no lineal con la tranformación B-C 57
A.2. Resultados regresión no lineal con la tranformación logarítmica 59
A.3. Resultados regresión lineal múltiple con la tranformación B-C 60
A.4. Resultados regresión lineal múltiple con la tranformación logarítmica . 61

B. Estimación de modelos no paramétricos 67

C. Estimación de las sendas salariales bajo la metodología de Montenegro et al. (2013) 72

D. Cálculos de la renta vitalicia 74

Bibliografía 78

1. Introducción

Esta tesis es una elaboración crítica del cálculo realizado por Montenegro, Jiménez y Hurtado (2013), en la que desarrollaron: primero, una metodología estadística para estimar las denominadas sendas ó carreras salariales; y segundo, con dichas sendas, efectuaron la cuantificación de un subsidio pensional en el Régimen de Prima Media con Prestación Definida (RPM-PD), administrado por Colpensiones y otras entidades de fondos pensionales. Su punto de partida fue calcular los aportes a pensiones de un trabajador genérico durante su vida laboral, tal que, al aplicar las fórmulas para el capital acumulado, se genera una diferencia o déficit entre el valor de la mesada y el costo de la renta vitalicia, que se supone son las transferencias fiscales o subsidios por parte del Estado al sistema pensional de prima media administrado por Colpensiones y otros fondos pensionales.

Para llevar a cabo esta crítica, se toma como fundamento la teoría actuarial y el modelo de senda salarial propuesto por Carriere y Shand (1998), y se tiene como referente la definición de la función salarial según la legislación colombiana. Luego se realiza un análisis de los datos y, mediante múltiples ensayos y estimaciones, se busca un modelo que permita calcular la senda salarial capturando la mayor información de los datos. En este caso los modelos adecuados son los no paramétricos Loess y Nadaraya-Watson, dado que poseen propiedades que combinan lo mejor de la regresión lineal con la flexibilidad de la regresión no lineal por medio del ajuste de un modelo sobre un subconjunto de datos locales, y posibilitan el análisis para poblaciones con datos extremos. Después de estimar la senda salarial se realizan los cálculos actuariales para determinar la mesada pensional y la renta vitalicia con el objetivo de investigar la existencia y condiciones de subsidios en el PRM-PD. Por último, se revisa la metodología de construcción de la senda salarial en Montenegro et al. (2013) para realizar una aplicación con los datos de estudio.

Teniendo en cuenta lo anterior, es indispensable abordar el supuesto de la existencia de tales subsidios desde la teoría actuarial de los regímenes de prima media, entendidos como un sistema de cotización basado en un porcentaje del salario del afiliado, más una prestación definida, establecida como una fórmula que cuantifica el monto de la mesada pensional con base en los salarios de los últimos años previos a la jubilación. Por consiguiente, en este estudio no se parte del supuesto de la existencia de tales sub-

sidios; pues esto equivale a afirmar que existe un desbalance entre lo acumulado por el trabajador y el costo de una pensión en la forma de una renta vitalicia, en la cual la mesada pensional se estable al momento de la jubilación (es la prestación definida). En este punto, el estudio se fundamenta en la teoría actuarial de sistemas de pensiones con prestación definida donde se establecen condiciones para que el costo de la renta vitalicia, dado el valor de la mesada, esté en equilibrio con el capital acumulado. Por último, se advierte que este trabajo, a pesar de que menciona el déficit de las pensiones, no tiene como propósito realizar una cuantificación del mismo.

La estructura de la tesis está dividida en seis apartados, que comienzan con esta introducción. En el segundo se realiza una revisión de las definiciones de la senda salarial, específicamente las definiciones de Carriere y Shand (1998) y la de Montenegro et al (2013). En el tercero se presenta una descripción de los datos y se estiman los modelos no paramétricos utilizados. En el cuarto se muestra la aplicación al tema de los subsidios en el Régimen de Prima Media con Prestación Definida (RPM-PD), focalizándose en la construcción teórica y los resultados por medio de las sendas salariales construidas. En el quinto se analiza la metodología y se realizan las críticas al trabajo de Montenegro et al (2013). Y finalmente, en el sexto apartado se presentan las conclusiones.

2. Revisión de la definición de senda salarial

Un problema clásico en econometría y economía laboral consiste en determinar un modelo para explicar el nivel de los salarios a partir de ciertas características del trabajador como nivel educativo, experiencia, tipo de empresa, edad, género, etc. Una solución recurrente para afrontar este problema es el modelo de Mincer (1974) que se define a continuación. Si AS_j es el salario del empleado j-ésimo, D_j son los años de escolaridad y U_j es la experiencia laboral en años, entonces el modelo plantea una relación cuadrática entre el logaritmo de AS_j y U_j, y lineal con D_j y un error aleatorio ϵ_j, de la siguiente manera.

$$log(AS_j) = \beta_0 + \beta_1 U_j + \beta_2 U_j^2 + \beta_3 D_j + \epsilon_j.$$

Un modelo alterno se basa en la escala ó función salarial S_x, una función de la edad x del trabajador, tal que si AS_y es el salario a una edad $y > x$, se tiene que

$$ASy = AS_x \frac{S_y}{S_x}. \tag{2-1}$$

En otras palabras, la función salarial permite proyectar los salarios futuros de un empleado a partir del salario en una edad inicial. Los salarios así proyectados se denominan 'senda salarial' ó 'carrera salarial'. Se supone que S_x es un modelo que incorpora el incremento del salario mínimo anual (que no siempre coincide con la variación de la inflación) y el incremento por mejora profesional, debido a cualificación por estudios, experiencia, etc.

2.1. Definición en la legislación pensional

La función salarial es definida en la legislación pensional de Colombia bajo el concepto de Salario Medio Nacional (SMN) que se determina de acuerdo a la edad del individuo, como un elemento para el cálculo de los bonos pensionales. En el decreto 2779 de 1994 se establece que "para el efecto del cálculo de Bonos y Títulos Pensionales se adopta la tabla de Salarios Medios Nacionales a la cual se hace referencia en

el artículo 117 de la Ley 100 de 1993 y consagrado en la Resolución 1588 de 1994 expedida por el Departamento Nacional de Estadística, DANE. La tabla de SMN es de utilización obligatoria para todo cálculo de Bono de Título Pensional". Es necesario anotar que el cálculo de los subsidios está relacionado con el cálculo del valor de los bonos pensionales tipo A y B, porque éstos representan también el valor acumulado de los aportes del trabajador durante su vida laboral.

Si bien, en el Decreto 2779 de 1994 la función salarial se denomina S_x, para evitar confusiones en este trabajo, se denominará por IS_x. Entonces, si IS_x indica el SMN a edad x, el salario a la edad $x > e$, con e como la edad inicial del trabajador, se estima por

$$AS_x = AS_e\Big(1 + i_\Delta\Big)^{x-e}\frac{IS_x}{IS_e}, \tag{2-2}$$

donde i_Δ es una tasa promedio del incremento del salario mínimo anual. En el Decreto 2779 de 1994 se encuentran los valores IS_x para $x = 13, ..., 71$. La siguiente tabla contiene los valores de las edades x y los factores IS_x

Tabla 2-1.: Salarios Medios Nacionales

13	1.066649	14	1.135599	15	1.206678	16	1.279752	17	1.354687
18	1.431251	19	1.509273	20	1.588504	21	1.668693	22	1.749612
23	1.830933	24	1.912388	25	1.993653	26	2.074416	27	2.154318
28	2.233031	29	2.310209	30	2.385489	31	2.458524	32	2.528971
33	2.596463	34	2.660676	35	2.721264	36	2.777921	37	2.830357
38	2.878287	39	2.921441	40	2.959579	41	2.992482	42	3.020004
43	3.041946	44	3.058210	45	3.068682	46	3.073323	47	3.072096
48	3.065019	49	3.052111	50	3.033468	51	3.009187	52	2.979401
53	2.944284	54	2.904026	55	2.858858	56	2.809030	57	2.754790
58	2.696446	59	2.634304	60	2.568691	61	2.499933	62	2.428355
63	2.354341	64	2.278237	65	2.200368	66	2.121119	67	2.040814
68	1.959800	69	1.878421	70	1.796985	71	1.715798		

En esta tesis, se utilizará la siguiente aproximación para los valores de IS_x.

$$IS_x = ax + bx^2 + cx^3, \quad 13 \leq x \leq 71 \tag{2-3}$$

donde $a = 0.0744989069, \quad b = 0.0008268515, \quad c = -0.0000221148.$

2.2. Definición en Carriere y Shand (1998) y Modelos estadísticos para la función salarial S_x

En esta sección se describen los métodos utilizados en la estimación de la senda salarial en Carriere y Shand (1998), los cuales son el fundamento teórico del presente trabajo. Se inicia denotando la función salarial S_x, donde x es la edad de un empleado que tiene un plan de pensión. Esta función se supone no decreciente con la edad, y refleja incrementos del salario debido a la inflación y al mérito (experiencia). Carriere y Shand (1998, pág. 18), definen la senda salarial así

Definición 2.1. La función de escala salarial S_x, con $x = e, ..., r$, se define como una función de la edad x, a partir de una edad e, la edad de ingreso al mercado laboral, y hasta r, la edad de jubilación, dada por

$$S_x = exp\left(\int_e^x \xi_s ds + \int_e^x \psi_s ds\right) \tag{2-4}$$

donde $\int_e^x \xi_s ds$ representa el ajuste acumulado hasta la edad x por incremento del salario mínimo anual y $\int_e^x \psi_s ds$ representa el incremento por mejoramiento profesional hasta la edad x. Las funciones ξ_s y ψ_s se denominan fuerzas de incremento y se asume que son funciones positivas e integrables en cada subintervalo del eje positivo.

La función salarial S_x permite estimar el salario futuro del trabajador de edad x, mediante la función AS_{x+t}, para t años adelante, mediante la ecuación recursiva siguiente

$$AS_{x+t} = \frac{S_{x+t}}{S_x} AS_x, \tag{2-5}$$

Definición 2.2. La senda salarial AS_x para un empleado con edad inicial e y edad de jubilación r se define como la sucesión AS_x calculada a partir de (2-5) para $x = e, e+1, ..., r-1$.

Además de estas definiciones, los autores analizan dos metodologías de estimación tradicional no paramétrico, CAM (Current Average Method) e IRM (Increase Ratio Method), y adicionalmente introducen un modelo estadístico paramétrico. En los métodos tradicionales se tienen las siguiente denotaciones: AS_x es el salario actual de la persona x en el tiempo $t = 0$ y $AS_{[x]+1}$ es el salario actual de la misma persona, un año mas tarde en tiempo $t = 1$. Se define N_x como el número de personas de edad x en el tiempo $t = 0$ y $AS^k_{[x]+t}$, con $k = 1, 2, ..., N_x$, como el salario actual del empleado k con edad $x + t$.

La estimación del CAM se puede realizar con grupos de edades quinquenales o anuales. En particular, los autores denotaron s_k como la cantidad de servicio que el empleado tiene en $t = 0$ y $AS_{[x]-s_k}$ como el salario actual contractuado. Empleando la técnica CAM, utilizando las anteriores denotaciones y por medio de una serie de transformaciones como la regla del punto medio, se obtiene la siguiente ecuación, para mayores detalles, ver Carriere y Shand (1998, pág. 20)

$$\overline{AS}_{[x]+t} = \frac{1}{N_x}\sum_{k=1}^{N_x}\left[AS^k_{[x]-s_k}.exp\left\{\int_{-s_k}^{t}\xi_z d_z + \int_{x-s_k}^{x+t}\psi_z d_z\right\}\right] \tag{2-6}$$

se observa que este método no permite extraer la función de salario definida en (2-4), sin embargo, el método IRM si. Para entender por qué, se inicia desde

$$\overline{AS}_{[x]+t+1} \approx \frac{1}{N_x}\sum_{k=1}^{N_x}\left[AS^k_{[x]+t}.exp\left\{\int_{t}^{t+1}\xi_z d_z + \int_{x+t}^{x+t+1}\psi_z d_z\right\}\right] \tag{2-7}$$

Teniendo en cuenta las denotaciones anteriores y utilizando el método IRM, el siguiente paso es calcular el promedio de las tasas y se obtiene

$$\overline{R}^*_x = \frac{1}{N_x}\sum_{k=1}^{N_x} R^k_x, \tag{2-8}$$

donde, $R^k_x = \frac{AS^k_{[x]+1}}{AS^k_{[x]}}$ denota las tasas individuales observadas.

A continuación, los autores definen

$$R^*_x \approx exp\left\{\xi_{1/2} + \int_{x}^{x+1}\psi_z d_z\right\} \tag{2-9}$$

Multiplicando estas tasas sucesivamente generan el estimador:

$$S^*_x = \prod_{a=0}^{x-1} R^*_a \approx exp\left\{\xi_{1/2} + \int_{x}^{x+1}\psi_z d_z\right\} \tag{2-10}$$

el cual es exactamente igual a (2-4), la definición de la función de salario.

Para sintetizar, el resultado permite explicar la función salarial por medio de dos componentes: La inflación y el mérito, lo cual es aproximadamente la productoria de la variación de las tasas de salarios. En otras palabras, la senda salarial se puede expresar como un ecuación, donde se permite definir una variable dependiente (la senda salarial) y unas variables independientes (inflación y mérito). Esta aproximación será el punto de partida de la metodología del trabajo de tesis.

Adicionalmente, los autores demostraron que entre los métodos no paramétricos CAM e IRM, este último es el más indicado para analizar la función salarial; sin embargo, como alternativa propusieron una estimación paramétrica con un modelo de regresión no lineal para la función salarial, que es muy usada en otras investigaciones, pero no se considerará en esta tesis.

2.3. Definición en Montenegro, Jiménez y Hurtado (2013)

En esta sección se retoma la metodología utilizada en Montenegro, Jiménez y Hurtado (2013) para calcular las 'carreras salariales'. Aunque matemáticamente son diferentes las expresiones 'carreras salariales', empleada en Montenegro et al (2013), y 'senda salarial ', expresada en Carriere y Shand (1998), este trabajo supone que ambos conceptos son equivalentes con el propósito de mantener una coherencia conceptual y una posible aplicabilidad de un marco comparativo. Aclarado esto, se modifica ligeramente la notación original en Montenegro et al. (2013).

Considere un conjunto de salarios y edades, por género (h=hombre, m=mujer), en una fecha determinada para un colectivo de empleados. En el caso de Montenegro et al. (2013) se tomó el total de afiliados al Régimen de Ahorro Individual con Solidaridad (RAIS) durante los años 2007, 2008, 2009. Dado que los salarios se denotaron anteriormente como AS_j, entonces se expresarán como $AS^{(g)}_{x,j}$, donde x recorre las edades del conjunto, j es un contador que recorre todos los individuos de edad x, y $g \in \{h, m\}$ denota el género del empleado.

Defina ${}_xT_g$ el total de individuos de edad x y género g. Defina ${}_xK_g = [0.1{}_xT_g]$, y los siguientes intervalos inter-decílicos por edad y género:

$$
\begin{aligned}
{}_xI_{g,1} &= \Big\{ j : 1 \leq j \leq {}_xK_g \Big\}, \\
{}_xI_{g,2} &= \Big\{ j : {}_xK_g + 1 \leq j \leq 2{}_xK_g \Big\}, \\
{}_xI_{g,3} &= \Big\{ j : 2{}_xK_g + 1 \leq j \leq 3{}_xK_g \Big\}, \\
&. \\
&. \\
&. \\
{}_xI_{g,10} &= \Big\{ j : 9{}_xK_g + 1 \leq j \leq 10{}_xK_g \Big\}
\end{aligned}
$$

Definición 2.3. La senda salarial para el intervalo inter-decílico para cada edad x y género g, se define como el promedio

$$ {}_xS_{g,i} = \frac{1}{{}_xK_g} \sum_{j \in {}_xI_{g,i}} AS_{x,j}^{(g)} \tag{2-11} $$

Esta definición es básica para realizar la comparación con el método adoptado en esta tesis y permite construir las sendas salariales que se muestran en el capítulo 5.

2.4. Metodología de la tesis

La población de estudio son 3.551 empleados del Municipio de Medellín comprendidos en 10.653 datos para los años 2008, 2011 y 2014, [1] con los cuales se realiza la construcción de las sendas salariales.

Para analizar estos datos, el punto de partida es la notación AS_{x+t} dada por Carrire y Shand (1998), donde los cálculos entre años se hacen por referencia a una persona de edad x, en un periodo determinado t. En este trabajo se utiliza esa notación para identificar por cédula a una persona de edad x que recibe ingresos salariales en la forma AS_{x_k}, $AS_{x_{k+3}}$ y $AS_{x_{k+6}}$, que corresponden a los años 2008, 2011 y 2014 respectivamente. Además, N_x es el número de trabajadores con una edad x, para los cuales se formarán los promedios de salarios para cada cambio de periodo, donde $t = 0$ representa la varición salarial entre $2011 - 2008$, y $t = 1$ la variación para $2014 - 2011$; en este sentido Y_{x_k} denotará la variación salarial de una persona k a la edad x para el tiempo $t = 0, 1$. Así mismo, e y r son las edades de ingreso y retiro del sistema pensional, para este caso $e = 20$, $r = 62$.

Siguiendo con la propuesta metodológica se realizan varios pasos: primero, suavizar $AS_{x_{k+1}}$ y AS_{x_k}, mediante los diferentes tipos de regresiones no-paramétricas, para producir incrementos salariales por edad, de la forma

$$ \overline{Y_x} = \frac{\widetilde{AS}_{x+1}}{\widetilde{AS}_x} \quad x = 24, 25, ..., 65 \tag{2-12} $$

donde $\widetilde{AS}_x$ es el valor suavizado a la edad x, de los valores AS_{x_k}, cuando $k = 1, 2, ..., N_x$ recorre todos los salarios de trabajadores de edad x.

Segundo, la senda salarial se calcula de la forma

$$ \widehat{S}_x = \prod_{a=e}^{x-1} \overline{Y}_a, \qquad x = e, ..., r \tag{2-13} $$

[1] Datos tomados en 2014, https://www.cgm.gov.co

Todo esto permite definir la senda salarial calculada con base en las fórmulas (2-4) y (2-13) como

$$\widetilde{AS}_{x+1} = \frac{\hat{S}_{x+1}}{\hat{S}_x}\widetilde{AS}_x \quad = \quad \overline{Y}_x\widetilde{AS}_x. \tag{2-14}$$

Por lo tanto, tomando como fundamento la teoría actuarial y el modelo de regresión no lineal para la función salarial de Carriere y Shand (1998), este trabajo define la senda salarial como una función que depende de la experiencia y la inflación, pues se supone que los incrementos salariales de los individuos de un periodo a otro son explicados por mejoras en su cualificación profesional por experiencia (educación) o por aumentos de la inflación.

Así mismo, este trabajo toma la aplicación propuesta por Montenegro et al. (2013), en la que mostró que el sistema público de pensiones necesita de los subsidios por parte del Estado, para argumentar un caso opuesto al de los autores mencionados, demostrando que el sistema público funciona sin necesidad de subsidios. Aunque este trabajo recurre a la metodoloía de Montenegro et al. (2013), es necesario aclarar que se tiene dos diferencias fundamentales: primero, Montenegro et al. (2013) hace una distinción de individuos por género, mientras en esta tesis no se hace esta diferencia; y segundo, Montenegro et al. (2013) realiza sendas en intervalos inter-decílicos para cada edad y género, mientras este trabajo realiza variaciones salariales entre tres diferentes periodos de tiempo para el mismo individuo identificado con cédula.

Para el caso aplicado, el salario inicial AS_e se toma arbitrario para lograr valores de acuerdo al nivel salarial actual (2016), iniciando con $AS_e = \$100.000$.

Con base en lo anterior, la propuesta metodológica es usar modelos de regresión no paramétrica, específicamente el Loess y el Nadaraya-Watson, para estimar la senda salarial de un empleado genérico del Municipio de Medellín para el periodo 2008-2011 y 2011-2014 , y estimar si el capital acumulado más los rendimientos es suficiente para adquirir un renta vitalicia o, en caso contrario, otorgarle un subsidio por parte del Estado.

3. Estimación de los Modelos

3.1. Análisis preliminar de los datos de los salarios

Los datos utilizados corresponden a los salarios por edad y cédula de empleados del Municipio de Medellín, durante los años 2008, 2011 y 2014 [1]. La revisión descriptiva de los datos de salario y edad tiene dos componentes: el individual, a través de medidas descriptivas, y el multivariado, por medio de test y gráficos como histogramas, Q-Q plot, Boxplot, violín plot, gráficos de dispersión, estimación de la correlación y finalmente se efectua la prueba de colas pesadas del salario y el estudio de contornos.

La Tabla 3-1 muestra el análisis individual de los datos del salario, de los cuales se resalta: primero, el indicador Skewness (asimetría) muestra valores positivos o valores más separados de su media al lado derecho, lo que supone una cola derecha más larga en la distribución; y segundo, el resultado de la Kurtosis es positivo que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y muestra colas más estrechas hasta los valores extremos de la distribución , en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Las anteriores dos medidas brindan indicios acerca de la no normalidad de los datos. Adicionalmente, medidas como el mínimo y el máximo muestran las grandes diferencias en materia salarial y dan a entender que se está ante una población bastante heterogénea, así mismo, la desviación estándar muestra una gran dispersión entre los datos, confirmando la presencia de grandes diferencias salariales.

Tabla 3-1.: Estadísticas Descriptivas

	Salario		
n	10653	Mean	2120492.804
Std. Dev	1044949	Median	1762764
Min	569992	Max	7396808
25th	1312282	75th	2674580
Skewness	1.122	kurtosis	0.898

Pasando a la variable edad, pese a que la Tabla 3-2 señala una distribución con

[1] Datos tomados de https://www.cgm.gov.co, durante el año 2014

asimetría negativa, lo que hace referencia a que la cola izquierda es la más larga de la distribución, esto da cuenta de individuos con edades más separadas de la media de la población. De igual forma, la kurtosis hace alusión a una distribución leptocúrtica de los datos. Esto no es necesariamente cierto, pues revisando la Figura 3-2 y teniendo en cuenta el tamaño de la población (n=10653), se puede afirmar que la variable edad presenta una distribución normal. Sumado a lo anterior, se tiene una población con valores extremos en materia del rango de edad, pues el valor del máximo supera la edad de jubilación establecida por ley, esto se presenta en algunos casos de trabajadores que no se retiran a tiempo y buscan aumentar sus ingresos, pues tienen la posibilidad de recibir dos, jubilación y salario.

Tabla 3-2.: Estadísticas Descriptivas

	Edad		
n	10653	Mean	46.873
Std. Dev	7.338	Median	47
Min	20	Max	70
25th	43	75th	52
Skewness	-0.489	kurtosis	0.333

El análisis multivariado de los datos comienza con las pruebas de normalidad bivariada por medio del Test de Mardia, que consta de una prueba de hipótesis fundamentada en la revisión de simetría, kurtosis y un análisis gráfico. Los resultados de la Tabla 3-3 reafirman los resultados anteriores, y se desecha la idea de normalidad multivariada de los datos.

Tabla 3-3.: Test de Normalidad Multivariada de Mardia

	Datos
g1p	1.5786
chi.skew	2802.913
p.value.skew	0
g2p	8.8820
z.kurtosis	11.3797
p.value.kurt	0
chi.small.skew	2804.22
p.value.small	0
Result	Data are not multivarite normal

El análisis gráfico del Test del Mardia es consistente con la conclusión anterior y evidencia la presencia de valores extremos en los datos, pues se presenta una serie de puntos que muestran un enorme distanciamiento del comportamiento medio de los

datos. En la revisión de los gráficos Boxplot, dispersión y análisis de contornos se profundizará acerca de la razón de su presencia.

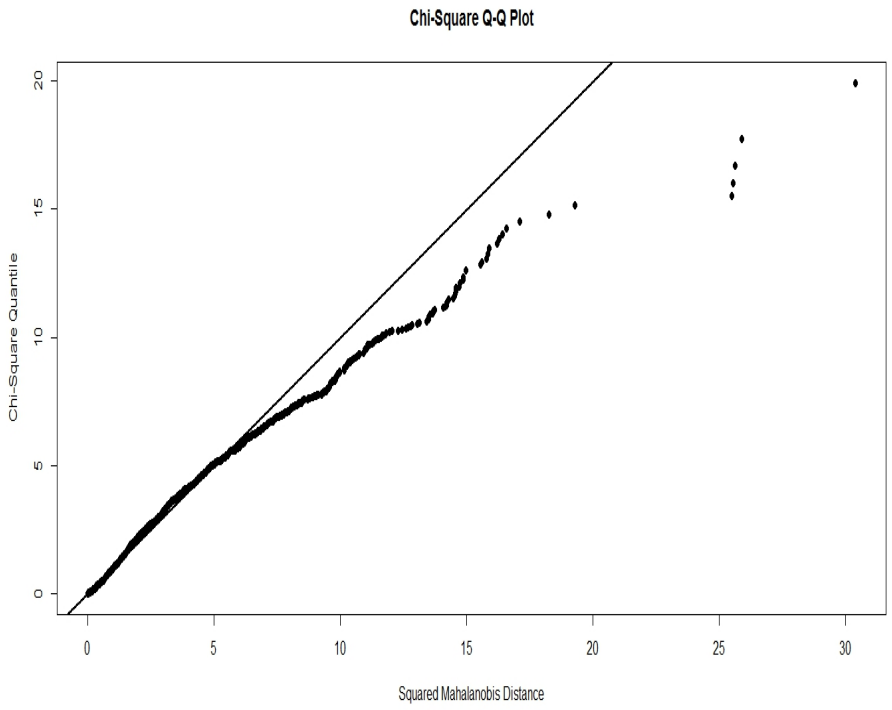

Figura 3-1.: Mardia Test

Prosiguiendo, en los histogramas de la Figura 3-2, se pueden apreciar la distribución del salario y la edad respectivamente. En el primero, se muestra una distribución con una gran acumulación al lado izquierdo, lo que significa que existe una mayor concentración de individuos que reciben un salario entre 1 y 2.5 millones de pesos mensuales; razón por la cual, la distribución del salario no satisface las condiciones de normalidad. En el segundo, se presenta una distribución con una ligera asimetría al lado izquierdo, dando cuenta de una mayor concentración de individuos jóvenes que se ubican por debajo de la media. Finalmente, es interesante anotar que la concentración de salarios en los menores rangos puede estar asociado con el hecho de tener una población joven que está iniciando su vida laboral y sus salarios son relativamente bajos por la falta de experiencia o educación.

En la Figura 3-3, el Q-Q Plot reafirma que los datos del salario y la edad no proceden de una distribución normal, esto se puede apreciar en la falta de un comportamiento lineal cerca de la distribución teórica señalada por la línea oblicua. Al mismo tiempo, se confirma el salario como la variable más alejada de la normalidad.

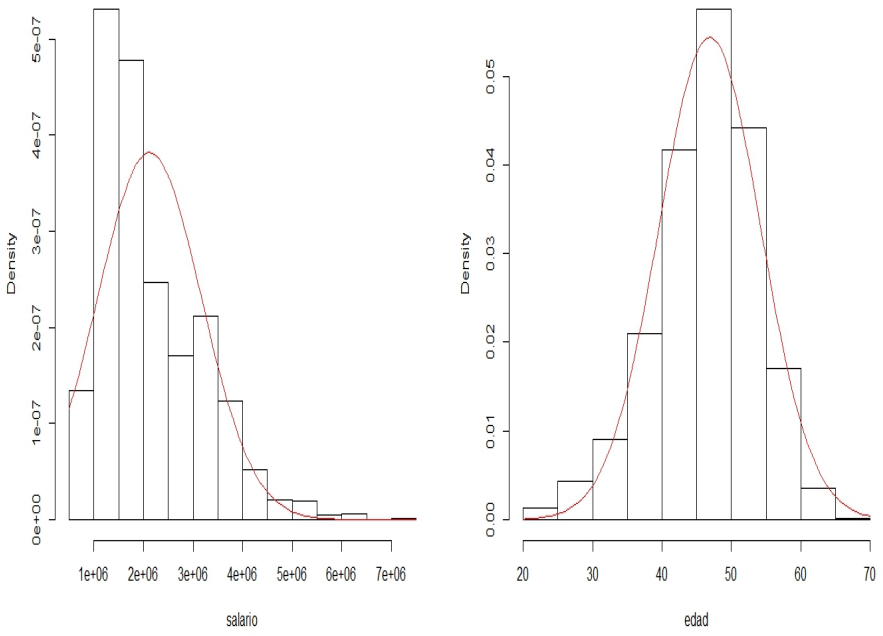

Figura 3-2.: Histogramas Salario y Edad

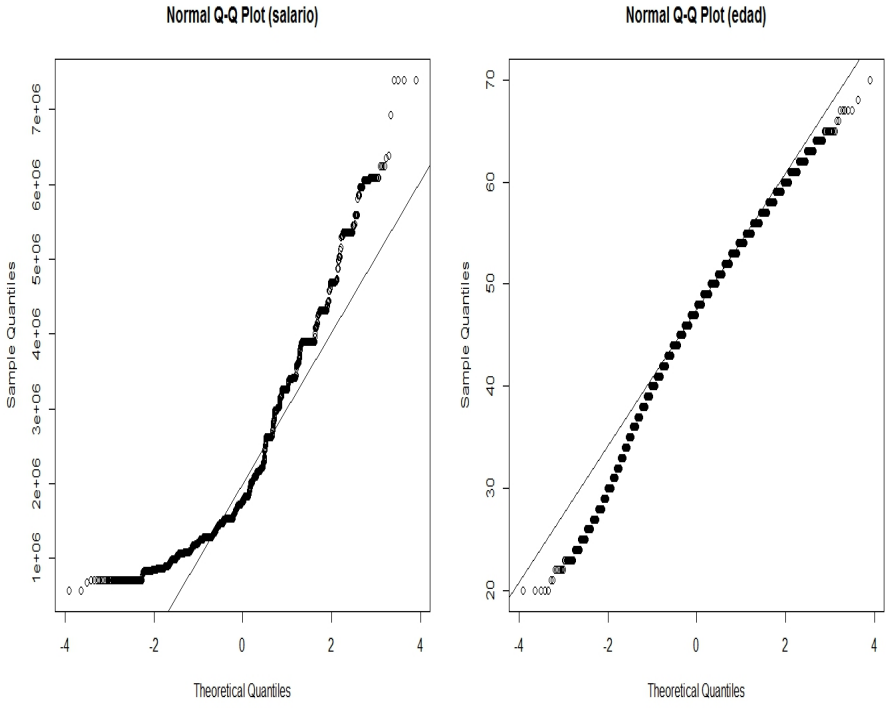

Figura 3-3.: Normal Q-Q Salario-Edad

Para confirmar el supuesto de normalidad de las variables analizadas, se realiza el Test Kolmogorov-Smirnov y, a través de los resultados expresados en la Tabla 3-4, se confirma el supuesto acerca de que no se presenta una distribución normal bivariada.

Tabla 3-4.: Test de Normalidad: Kolmogorov-Smirnov

Variable	Statistic	p-value	Normality
Salario	0.1584	0	NO
Edad	0.0656	0	NO

El Boxplot de la Figura 3-4 permite analizar la distribución de los datos entorno a su mediana, observando la existencia de ciertos valores extremos. En materia del salario, la mediana tiende hacia el límite inferior del rango y hay presencia de datos extremos en el límite superior, lo que indica la presencia de pocos individuos con altos salarios. En cuanto a la edad, la mediana se ubica en el centro del rango, y se observan valores extremos en ambos límites, mayormente en el límite inferior, siendo consecuente con el resultado anterior donde evidencia más concentración de trabajadores jóvenes. En consecuencia, la presencia de estos valores extremos es un factor fundamental en la elección de la metodología para hacer inferencia de los datos.

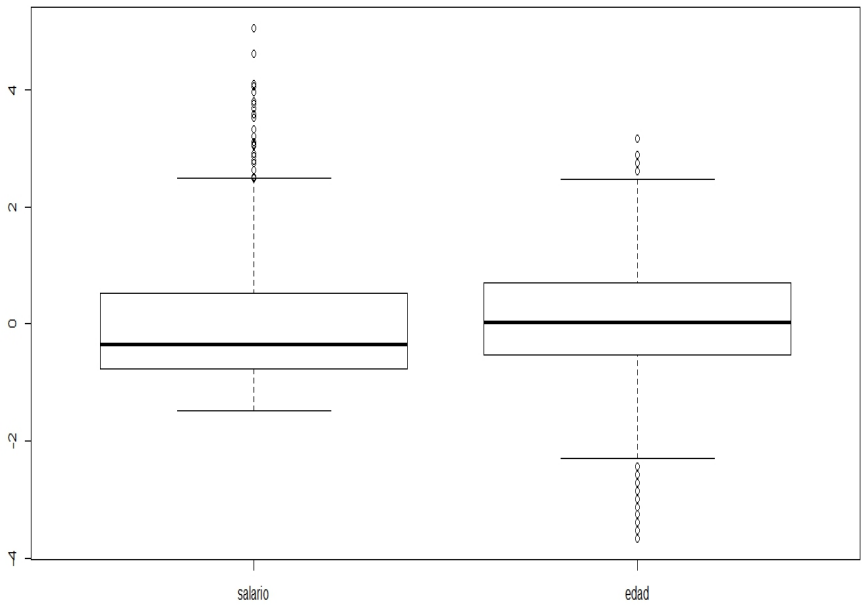

Figura 3-4.: Boxplot

En el Violín Plot de la Figura 3-5 se muestra que tanto el salario como la edad tienen una distribución unimodal; en otras palabras, no hay presencia de otros picos o acumulaciones en la distribución, indicando que las dos variables, de forma individual, son un sólo conjunto de individuos con algunos problemas de valores extremos. En lo que respecta a la forma de su distribución, se percibe que el salario tiene un comportamiento asimétrico con una cola más pronunciada (parte superior); mientras que en la edad es menos evidente la asimetría, observando detalladamente que la parte inferior es ligeramente más extensa.

En la Figura 3-6 se efectua una revisión de la relación de las dos variables por medio

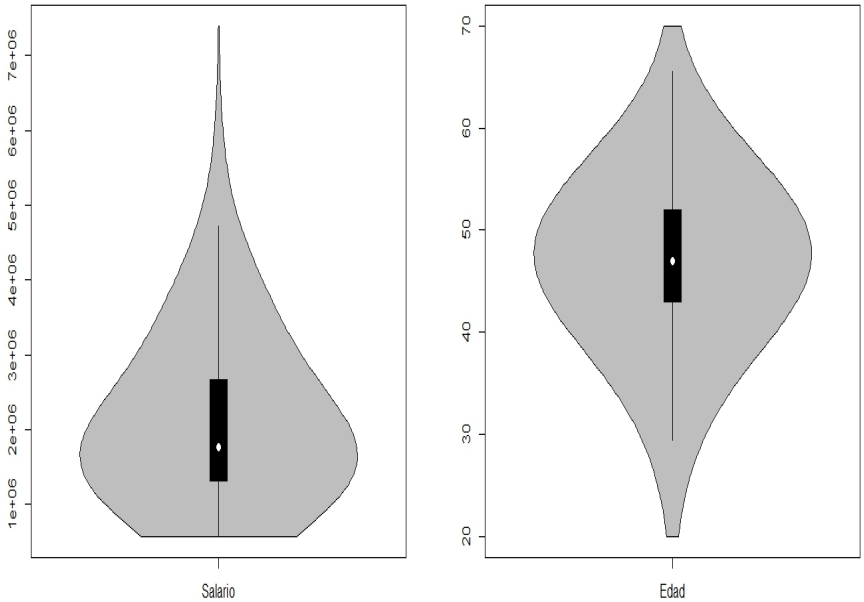

Figura 3-5.: Violín plot

del gráfico de dispersión que brinda información de las condiciones de relación año a año, y de lo cual se perciben varios fenómenos a examinar. Primero, a medida que se pasa de un periodo a otro, se constata la separación entre los individuos de mayor salario y los individuos de bajo salario. Segundo, un hecho común a los tres periodos es la presencia de un gran grupo de individuos entre los 40 y 55 años que corresponden a un nivel salarial entre 1 y 2 millones de pesos. Tercero, en la parte superior de los gráficos se contempla una serie de valores que se pueden considerar como extremos, correspondientes al mayor nivel salarial. Cuarto, se intuye una especie de líneas horizontales entre los puntos, lo que brinda una idea de relativa estabilidad entre el nivel de salario y la edad; es decir, para algunos individuos se mantiene constante el salario

(aumento del salario sólo por inflación) a pesar de que aumente la edad. Y quinto, la relación entre las variables no parece lineal, pero si sugiere una relación cuadrática.

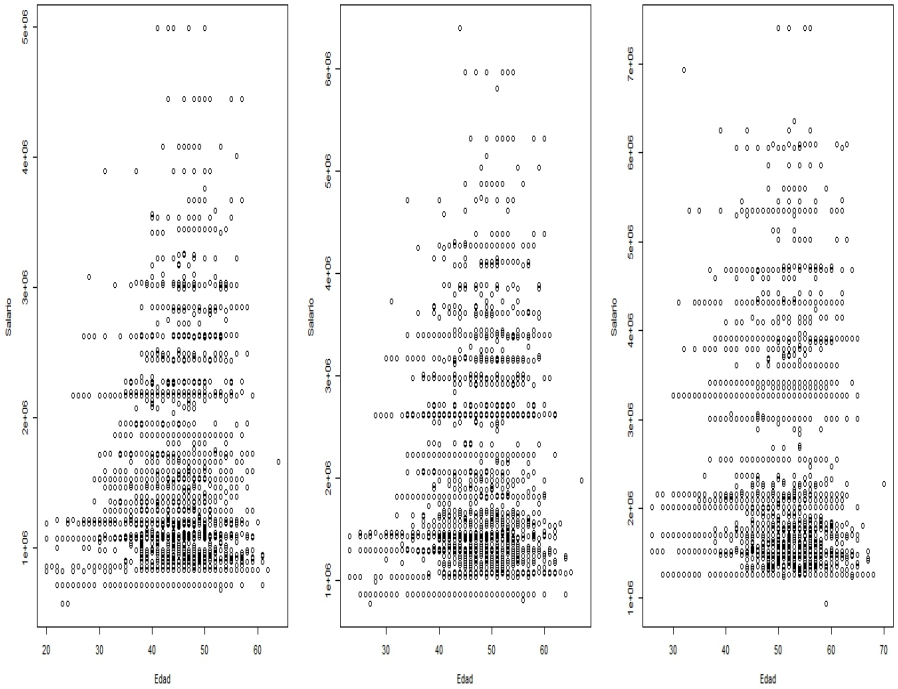

Figura 3-6.: Gráfico de dispersión: Salario vs Edad

Prosiguiendo, en la Tabla 3-5 se expone el resultado del valor de la correlación entre el salario y la edad, cuyo porcentaje afirma que sólo en el 20 % de las veces las dos variables se encuentran en la misma fase (crecimiento o decrecimiento), demostrando una relación lineal relativamente baja entre los datos. En consecuencia, la baja asociación brinda información acerca de la relación de dependencia de las variables y se puede afirma que no se presenta una dependencia fuerte.

Tabla 3-5.: Estimación de la Correlación Salario vs Edad

	salario	edad
salario	1.00	0.20
edad	0.20	1.00

Por otra parte, para comprobar la hipótesis de colas pesadas se utiliza un estadístico no paramétrico que permite examinar la vida media residual. Este estadístico es un estimador de la función $e(t) = \mathbf{E}(X - t|X > t)$, $t > 0$, para una variable X positiva. El criterio para decidir si una muestra $X_j,\ j = 1, ..., n$ proviene de una variable X

con distribución de cola pesada, es que $e(t)$ sea una función creciente, y en tal caso, al mantener una línea constante (horizontal) ó decreciente, se clasifica de *cola liviana*. En la Figura 3-7 se graficaron las vidas medias residuales de los incrementos anuales para los periodos 2008, 2011 y 2014 respectivamente. En la librería de R, la función *ismev* cuenta con el estimador *mrl.plot*().

```
require(ismev)
par(mfrow=c(1,3))
mrl.plot(S08)
mrl.plot(S11)
mrl.plot(S14)
```

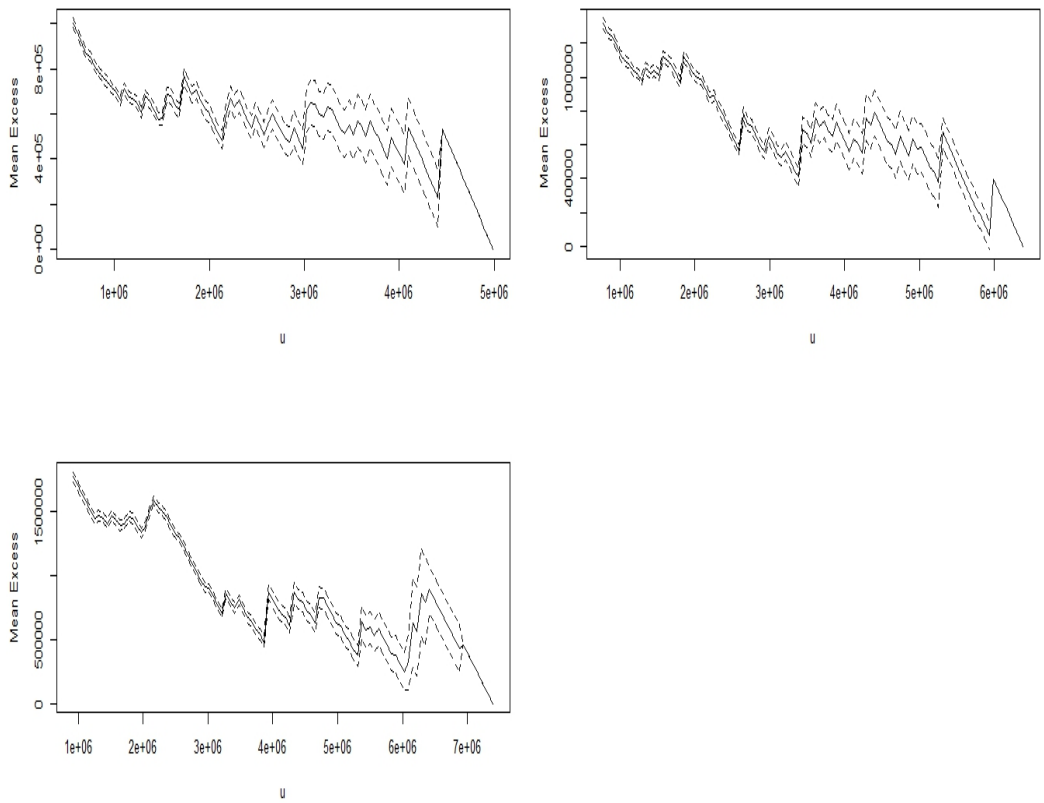

Figura 3-7.: Vida media residual: (arriba)2008-2011 y (abajo) 2014

Con lo anterior, se percibe que la vida media residual tiene leves variaciones en la sección central de los datos, además, los extremos derechos no determinan el diagnóstico ya que, al ser pocos, generan mayor variabilidad, por tal razón, los incrementos no tiene una distribución de cola pesada. Por tanto, se puede concluir que no son necesarios

métodos generalizados para analizar la relación $Y_{x_k,t}$ versus x_k.

Ahora, se realiza un estudio de contornos para la distribución bivariada de las variables (salario, edad) para los tres periodos 2008, 2011 y 2014, ilustrado en la Figura 3-8, realizado a través de la librería 'MASS' de R-Project. En ellos se evidencia la presencia de dos cluster con condiciones salariales diferentes, hecho que se puede explicar por varias circunstancias. Primero, por la rigidez de la carrera administrativa que genera gran dificultad para ascender o cambiar de puesto debido a que esto únicamente se puede hacer por concurso público o por factor político. Y segundo, por la existencia de un sistema piramidal de rangos profesionales mutuamente excluyentes, pues en un equipo de trabajo conformado por líder de programa, líder de proyecto, profesional especializado, profesional universitario, tecnólogos y técnicos sólo se permite la existencia de varios profesionales en las tres categorías de menor rango y para ascender se tiene que generar la vacante de un rango superior.

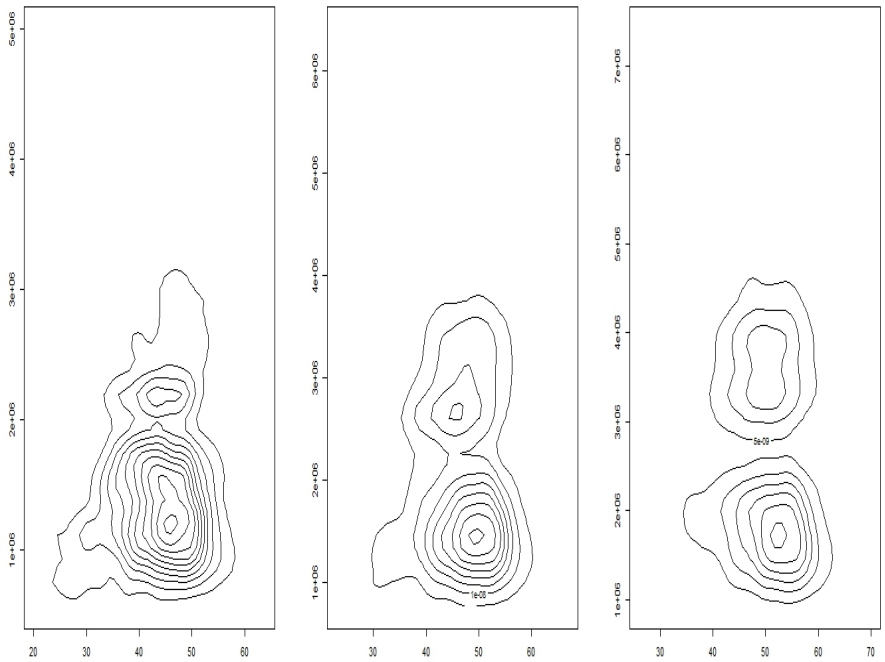

Figura 3-8.: Contornos

De la Figura 3-8 se concluye que la distribución conjunta es una mezcla de dos distribuciones bivariadas que han evolucionado y marcado en el tiempo, y aunque anteriormente se dieron algunos supuestos, se desconoce la razón exacta del por qué el cluster superior ha aumentado tanto en tan corto tiempo. Sin embargo, se puede resaltar dos aspectos, primero se evidencia una posible polarización del ingreso sin transición

entre los grupos. Y segundo, revisando la forma de los contornos (circunferencias) se puede afirmar que no existen una relación lineal entre las variables, esto significa que la edad explica muy poco la variabilidad del salario.

En resumen, es importante resaltar que suponer que la variable salario depende sólo de la edad es muy restrictivo, pues este supuesto es el causante de muchos de los problemas observados en las pruebas realizadas en los datos y desencadena impresición en la modelación. En otras palabras, el salario depende de un sin número de variables como: educación, experiencia, capacidad intelectual, factores políticos, relaciones sociales, la universidad que otorga el título, hasta suerte en algunos casos. Al intentar explicar el salario por una sola variable se dejan de lado muchos factores importantes que causan efectos como la baja relación lineal, cluster y existencia de datos extremos. De hecho, en la modelación se pueden presentar dos situaciones problematicas: primero, el error puede asimilar todas las variables faltantes, lo que ocasionaría volatilidad y, en algunos casos, heteroscedasticidad; y segundo, la presencia de cluster puede generar cambios estructurales entre las relaciones de las variables.

Pese a todo, en esta coyuntura, surgen dos posibilidades, la primera utilizar tranformaciones de Box-Cox (1964) y logarítmica, ver Apéndice A, y la segunda usar métodos robustos como la Regresión Loess y el estimador de Nadaraya-Watson, puesto que son estimadores no paramétricos con base en regresiones lineales locales y con capacidad de ajustar diversas relaciones no lineales.

3.2. Observaciones en la implementación de métodos paramétricos

Se optó por eliminar los análisis con este modelo debido a que la distribución conjunta empírica de las variables $(X, Y) = (edad, salario)$, obtenida a partir del examen de los contornos de nivel, durante los períodos 2008, 2011 y 2014, muestra dos cluster tales que uno se aleja progresivamente del otro y genera un efecto de contaminación en la distribución de la variable dependiente, el cual afecta los estadísticos t y F, a causa de su falta de robustez.

Una regresión lineal ó incluso no lineal, no tendría la flexibilidad suficiente para detectar el movimiento progresivo del cluster de salarios observado en la Figura 3-8. Una alternativa podría ser una regresión polinómica dado que permite estimar cambios de concavidad, pero debido a los problemas de multicolinealidad que este tipo de regresión introduce, y a que, evidentemente los datos distorsionaron los estadísticos t, y F,

tampoco se consideró.

Además se intento utilizar la opción de transformación de las variables por medio de Box-Cox (1964) y transformación logarítmica, para mayores detalles ver Apéndice A, sin embargo, los resultados no fueron satisfactorios pues persiste la distorsión de los estadístio t, F, y baja signficancia del modelo general (R-cuadrado), por lo cual, los métodos paramétricos no se consideran una opción para hacer las estimaciones.

3.3. Estimación mediante regresión no paramétrica

Una posible transformación de la ecuación propuesta por Carriere y Shand (1998, ecuación (3.4), pág. 9), consiste en un procedimiento matemático que permite modificarla de forma que se pueda expresar como una ecuación no paramétrica, obteniendo como resultado

$$Y_{x_k} = exp(m(x_k) + \epsilon_x^{(k)}). \tag{3-1}$$

Esta ecuación será la base de las regresiones No-paramétricas Loess y Nadaraya-Watson (Basu y Ullah, 1992, pág. 154), modelos propuestos como la mejor opción dada su flexibilidad para la estimación de las relaciones de los datos y la robustez frente a valores extremos.

3.3.1. Supuestos de la regresión no paramétrica

Los supuestos que se tendrán en cuenta para este modelos son: una muestra de tamaño n de una variable aleatoria bidimensional (X_i, Y_i), $i = 1, 2, ..., n$. Por el momento se asumirán continuas, pero más adelante se tratará el caso de X discreta y Y continua. Se asumen que estos datos son generados por el modelo

$$Y_i = m(X_i) + \epsilon_i. \tag{3-2}$$

Donde los residuos (ϵ_i) son una sucesión de variables aleatorias (independiente e idénticamente distribuidas), tales que

$$E(\epsilon_i|X_i) = 0, \tag{3-3}$$

$$Var(\epsilon_i|X_i) = \delta^2(X_i). \tag{3-4}$$

La función $m(.)$ se denomina media condicional y se asume que posee al menos $max\{p, r\}$ derivadas continuas en una vecindad del punto de interés, donde $r \geq 2$ es el

orden del núcleo, y $p \geq 0$ es el grado del polinomio local. Se asume que la densidad de X tiene al menos r derivadas continuas en la misma vencidad (el caso X discreta no requiere este supuesto). Adicionalmente, se supone una situación general de heterocedasticidad, en otras palabras, las varianzas de los errores se suponen distintas.

En la regresión no paramétrica, la única restricción que se le impone a la función m es que sea suave, y se entiende la suavidad en términos de derivabilidad. Es por este aspecto que a estos métodos se les denomina métodos de suavizamiento o suavizado.

3.3.2. Estimación del modelo de regresión no paramétrica

Los estimadores iniciales de regresión no paramétrica introducidos por Nadaraya (1964) y Watson (1964) fueron los estimadores de tipo núcleo. En términos del análisis la atención se centra en la elección apropiada de los parámetros decisivos en el comportamiento de los estimadores de regresión resultantes, como son: el ancho de banda ó parámetro de suavizado y el grado de 'bondad de ajuste' del polinomio. En la práctica, se determinan unas bandas centradas en cada punto y de un ancho h, y se calculan las estimaciones en el punto utilizando tan sólo las observaciones que caen dentro. Adicionalmente, para conseguir la curva de regresión estimada, la banda determinada por el parámetro h recorrerá todo el diagrama de dispensión de izquierda a derecha. Con base en esta intención se han creado diversas técnicas, de las cuales se destacarán algunas de las más importantes:

Estimadores tipo núcleo: el estimador localmente constante ó estimador de Nadaraya-Watson, $m(x)$, realiza un promedio de las observaciones que caen en cada banda de la siguiente manera

$$\widehat{m}_(x) = \sum_{i=1}^{n} w_{i,n} y_i, \tag{3-5}$$

donde,

$$w_{i,n} = \frac{K_h(x_i - x)}{\sum_{i=1}^{n} K_h(x_i - x)} \tag{3-6}$$

para cierto ancho de banda $h > 0$. En esta ecuación, la función núcleo, denominado K, determina la forma de los pesos asociados a cada observación dentro de la banda definida, por lo cual, juega un papel importante al momento de realizar la estimación.

Los tipos de núcleos más utilizados son:

$$K(\mu) = \begin{cases} (3/4)(1-\mu^2)1\{|\mu| \leq 1\}, & Epanechnikov \\ (1-|\mu|)1\{|\mu| \leq 1\}, & Triangular \\ (2\pi)^{-1/2}exp(-\mu^2/2), & Gaussiano \\ (1-\mu^3)^3 1\{|\mu| \leq 1\}, & Tri-cubo \end{cases}$$

Adicionalmente, una función núcleo $K : \Re \to (0, \infty)$ satisface tres condiciones

$$\int_{\Re} K(\mu)du = 1, \tag{3-8}$$

$$\int_{\Re} \mu K(\mu)du = 0, \tag{3-9}$$

$$\int_{\Re} \mu^2 K(\mu)du > 0, \tag{3-10}$$

Definición 3.1 Un núcleo K se dice de orden $k = 2, 3, ...,$ si satisface

$$\int_{\Re} x^{\nu} K(x)dx = 0, \quad \nu = 2, 3, ..., k-1 \tag{3-11}$$

$$\int_{\Re} x^k K(x)dx = \beta_k \neq 0. \tag{3-12}$$

El estimador N-W (3-5) se puede definir de manera equivalente en términos de una minimización como

$$\hat{m}(x) = \arg_{\beta} min \sum_{i=1}^{n} K_h(x_i - x)(y_i - \beta)^2 \tag{3-13}$$

que es un estimador de Mínimos Cuadrados Ponderados, con pesos $w_i = K_h(x_i - x)$. Este estimador de $m(x)$ es una constante β. De ahí el nombre de "localmente constante".

Regresión polinomial local: efectua un ajuste polinomial con las observaciones que caen en la banda. En Cleveland (1979) y se puede definir como:

$$\beta(x) = min \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)\left[y_i - \sum_{j=0}^{p}(\beta_j/j)\left[\frac{(x_i - x)}{h}\right]^j\right] \tag{3-14}$$

donde, $\beta = (\beta_0, \beta_1, ..., \beta_P)'$. El estimador de $m(x)$ es $\hat{m}(x) = \hat{\beta}_0(x)$, donde $\beta(x)$ es un vector que depende de x. El estimador de la j-ésima derivada es $\hat{m}^j(x) = \hat{\beta}_j(x)h^{-j}$.

La regresión polinomial local inicia con el supuesto de la función de regresión en el que m tiene p derivadas en un punto x_0, momento en el que, por medio del teorema de Taylor, se tiene una aproximación de este tipo para los valores en un entorno de x_0.

$$m(x) \approx m(x_0) + m^{'}(x_0)(x - x_0) + \frac{m^{''}(x_0)}{2!}(x - x_0)^2 + ... + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p \quad (3\text{-}15)$$

Posteriormente, esto fundamenta que se puede aproximar localmente m por funciones polinómicas de grado p

$$P_p(x) = \sum_{j=0}^{p} \beta_j (x - x_0)^j \quad (3\text{-}16)$$

De este modo, se llega a las estimaciones de los coeficientes $\hat{\beta}_j$ con $j = 0, ..., p$ y entonces, revisando la expresión, se observa que la estimación del término independiente $\hat{\beta}_0$ será un estimador de m en x_0 y el restante de los coeficientes de $\hat{\beta}_j$ producirán estimaciones de sus derivadas.

Por lo tanto, con el propósito de estimar m localmente mediante polinomios de grado p se examina un problema de mínimos cuadrados ponderados

$$\beta(x) = min \sum_{i=1}^{n} k_h \left(\frac{X_i - x_0}{h} \right) \left\{ Y_i - \sum_{j=0}^{p} (\beta_j)(X_i - x_0)^j \right\}^2 \quad (3\text{-}17)$$

donde:

1. h es un parámetro denominado ancho de banda o parámetro de suavizado que controla las observaciones que caen en cada entorno.
2. $K_h(u) = h^{-1}K(\frac{u}{h})$, donde la función $k(.)$ se denomina función núcleo. Esta función define las ponderaciones que se asignan a cada observación en el entorno local considerado. Por lo general, se supone una densidad simétrica.
3. p es el grado de ajuste polinomial local.

3.3.3. Selección del núcleo, K

Según Marron and Nolan (1988), citado en el manual de software Xplore, ver Härdle, Klinde and Müller (2012, sec. 6.1.3, pág. 71):

The difference between two kernel functions K is almost negligible when the bandwidths are appropriately rescaled. Kernel density estimation requires two parameters: the kernel function K and the bandwidth parameter h. In practice, the choice of the

kernel is not nearly as important as the choice of the bandwidth. The theoretical background of this observation is that kernel density estimates using two different kernels is almost negligible.

Sin embargo, según Gasser and Müller (1979, pág. 33), la afirmación anterior es válida solamente si el núcleo es de orden k. De hecho, afirman que se debe utilizar núcleos de orden $k = 2$.

En cualquier caso, la elección del núcleo no afecta el buen comportamiento de las estimaciones resultantes, por lo tanto, utilizar una función u otra obedece fundamentalmente a razones netamente técnicas.

3.3.4. Selección del polinomio, p

En la elección del grado de los ajustes polinomiales p se tiene una discusión por la existencia de un trade-off que busca una equivalencia óptima entre sesgo y varianza. Para la elección de ajustes de grado cero (Nadaraya-Watson) ó uno (que alcanza un estimador lineal local, ver Cleveland, 1979), se producen estimaciones con poca variabilidad (muy suaves), pero con sesgos muy elevados. El caso contrario ocurre al utilizar ajustes con grados mayores como dos ó tres, donde se generan menores sesgos (mayor adaptabilidad), pero con mayor varianza. En esta tesis, la propuesta se enfocará en un ajuste de grado cero o Nadaraya-Watson (N-W).

3.3.5. Métodos de selección de banda, h

El ancho de banda se determina a través de un parámetro positivo cuyo rango inicial es la amplitud del intervalo de estimación. La elección del parámetro, aunque fundamental para el procedimiento de estimación, presenta un trade-off entre flexibilidad y complejidad, esto significa dos posibles escenarios. Si h es pequeño indica mayor flexibilidad, al generar una curva con mayor variación; y menor complejidad, al examinar detalladamente comportamientos locales con pocos datos. Si h es grande señala menor flexibilidad en las estimaciones, causado por observaciones distantes que generan una curva con menor variación; y mayor complejidad debido a la dificultad de aglutinar el comportamiento local. Por lo tanto, es preciso buscar un balance entre flexibilidad y complejidad, ó en otras palabras, entre sesgo y varianza, con el propósito de obtener la mejor estimación posible. Para esto se presentan una serie de críterios que determinan la bondad de las estimaciones, como

-El Error Cuadrático Medio Condicional (MSE):

$$\widehat{MSE}(x) = \mathbf{E}[(\hat{m}(x) - m(x))^2] \tag{3-18}$$

donde la función $m(x)$ se asume con ciertas condiciones de suavidad y MSE((x)) $\rightarrow$ 0 cuando n $\rightarrow$ 0, ver Gasser and Müller (1979, pág 33, Theorem 2).

- El Error Cuadrático Medio (MISE):

$$MISE(\hat{(}M)) = \int_{\Re} \mathbf{E}[(\hat{m}(x) - m(x))^2]dx \tag{3-19}$$

Descripciones de estos procedimientos se pueden consultar en Härdle, Klinke and Müller (2012, pág. 178).

La elección entre un criterio u otro se debe al objetivo que se pretende: MSE cuando se busca optimalidad en cada punto, o MISE cuando se busca optimalidad conjunta en todo el intervalo de estimación.

Existe una serie de métodos para selección del ancho de ventana ó h, entre los cuales se destaca:

Métodos de selección basados en los datos Se fundamentan en la minimización de criterios de error como MSE y MISE, pero desconocidos, entre los que se encuentra la suma residual de cuadrados con el fin de lograr la elección del parámetro. Entre estos se destacan:

1. La validación cruzada
2. La validación cruzada generalizada
3. El criterio de información de Akaike
4. El error de predicción finito
5. La T de Rice.

De los anteriores, se utilizará la validación cruzada porque es un método utilizado para la selección de parámetros que determinan el equilibrio entre sesgo y varianza. La técnica radica en tres pasos: primero, de forma consecutiva, elegir de la muestra cada una de las observaciones x_i; segundo, estimar el modelo con los datos restantes y predecir el dato ausente con ese estimador; y, tercero, comparar esa predicción con el dato real. El procedimiento se realiza con cada posible valor de h, lo que posibilita construir la función

$ECMP_{(CV)}(h) = \frac{1}{n}\sum_{i=}^{n}(y_i - \widehat{m}_i(x_i))^2,$

que mide el error de predicción del estimador fuera de la muestra para cada h. El valor que minimice esa función, h_{CV}, será el valor del parámetro de suavizado elegido.

Adicionalmente, se pueden consultar otra serie de métodos para estimar h como: métodos plug-in y métodos basados en Bootstrap, para mayores detalle ver Parra (2011, pág. 25-30).

3.3.6. Efecto de borde en N-W

El efecto de borde se refiere a un problema de sesgo en el estimador $\overline{m}(x)$ cuando X es una variable aleatoria de soporte $[a, b]$ y x está cercano a los límites a,b.

En Gasser and Müller (1979, pág. 63), se definen núcleos especiales (polinomios) para tratar el efecto de borde, denominados 'boundary kernels'. Los núcleos se asumen de soporte compacto, para garantizar resultados asintóticos, y con segunda derivada continua. Pero los resultados son muy técnicos y no es fácil implementarlos numéricamente, por lo que se optó por permitir el sesgo del efecto de borde, aunque, la naturaleza del problema de la tesis, que consiste en estimar una diferencia promedio entre dos regresiones no paramétricas, afectadas posiblemente por el mismo tipo de sesgo en los borde, puede que no represente un sesgo en la diferencia misma. Es decir, puede que el sesgo se cancele (aproximadamente) al realizar la diferencia entre las dos curvas.

3.3.7. Relación Loess y regresión lineal local Nadaraya-Watson

El concepto de regresión lineal móvil involucra una ventana móvil de cada punto X_j indicada por $B_N(X_j)$. Esta ventana móvil se define como una vecindad de X_j que contiene $2N + 1$ puntos cercanos a X_j, distribuidos simétricamente a ambos lados del punto. Si se recorre la regresión con todos los $2N+1$ de $Y_i = \alpha+\beta X_i, \quad X_i \in B_N(X_j)$, se obtiene un estimador $\hat{Y}_j = \hat{\alpha}_j + \hat{\beta}_j X_j$.

Cuando se utilizan Mínimos Cuadrados Ponderados en una regresión lineal móvil, y los pesos son

$$w_{i,j} = \frac{C_j}{\lambda} K\left(\frac{X_j - X_i}{\lambda}\right) \tag{3-20}$$

entonces se trata de una minización similar a (3-13), de la siguiente forma

$$\left(\hat{\alpha}_j, \hat{\beta}_j\right) = arg_{\alpha,\beta} min \sum_{X_i \in B_N(X_j)} w_{i,j}\left(Y_i - (\alpha + \beta X_i)\right)^2 \tag{3-21}$$

Al aplicar en Loess la familia = "simétrica "es posible realizar una estimación robusta calculada por medio del algoritmo de los mínimos cuadrados reponderados, utilizando iterativamente la función bi-peso de Tukey para la reponderación. En cuanto a $k(.)$, núcleo completamente soportado (en pesos locales), funciona dando pesos bajos a los

x_i que están lejos de x, usando el núcleo "tri-cubo ", para mayores detalles ver Lee y Cox (2008, pág. 5).

3.3.8. Estimaciones

El proceso de estimación de las escalas salariales y las sendas salariales comprende una serie de pasos que permiten encontrar las sendas óptimas. La estimación no paramétrica se basa en la elección de los parámetros adecuados para hacer inferencia sobre los datos.

1. Estimación de la regresión Nadaraya-Watson

La estimación de la regresión Nadaraya-Watson se realizó con la librería stats y la función ksmooth de R, la cual es una estimación de la regresión con el kernel Nadaraya-Watson.
El proceso inicia definiendo el parámetro de ancho de banda h. En el trabajo se decide utilizar los métodos de selección basados en los datos, con atención especial en la validación cruzada. Entre las librerías de R que utilizan este método para realizar la estimación se tienen: sm y np.

a. La función h.select de la librería sm selecciona un parámetro de suaviazación para la densidad estimada en una ó dos dimensiones, y para la regresión no paramétrica con una ó dos covariables.

```
h3 <- h.select(x08, Yk1, method = "cv")
```

Resultado: valor estimado adecuado de $h = 0.8209049$

b. Las funciones hcv y sm.regression de la Librería sm calculan h óptimo con base en el método CV mediante la función hcv, que se utiliza en el estimador N-W con la función sm.regresion.

```
h.cv1 <- hcv(x08, Yk1, display="lines", ngrid=32)
sm.regression(x08, Yk1, h=hcv(x08, Yk1))
```

Tabla 3-6.: Resultados de la validación cruzada

h	cv
0.6497	5.59e-07
0.9967	1.01e-07
1.529	2.63e-06
2.3459	7.70e-05
3.5990	2.31e-05
5.5213	6.42e-05
8,4704	1.43e-04
12.9948	2.17e-04

Resultado: valor estimado adecuado de $h = 0.64$, de acuerdo con la metodología de CV.

c. La función npreg de la librería np determina el estimador N-W y permite escoger h utilizando el método de validación cruzada (CV). Esta función calcula h óptima en caso de datos mixtos: discretos y continuos, por lo que está recomendada para este estudio. De la ayuda de esta función se lee:

npregbw computes a bandwidth object for a p-variate kernel regression estimator defined over mixed continuous and discrete (unordered, ordered) data using expected kullback-leibler cross-validation, or least-squares cros validation using the method of Racine and Li (2004) and Racine, Li and Zhu (2004).
En consecuencia, este método es el óptimo para realizar las estimaciones de la regresión N-W. Pero las restantes metodologías son tomadas como referentes para hallar un rango de variaciación del ancho de ventana óptimo.

```
model.np08 <- npreg(S08 ~ x,regtype = "ll",bwmethod = "cv.aic",
gradients = TRUE)
# muestra el valor del ancho de banda optimo
summary(model.np08)
Resultado: valor estimado adecuado de $h=0.65554$
```

valor Bandwidth: 0.65554
Regression Data: 3551 training points, in 1 variable(s)
Kernel Regression Estimator: Local-Linear
Bandwidth Type: Fixed
Residual standard error: 643610.9
R-squared: 0.05076852

ontinuous Kernel Type: Second-Order Gaussian
NO. Continuos Explanatory Vars.:1

Luego de realizar la estimación del valor óptimo del parámetro h, se presentan algunos resultados de la estimación N-W.

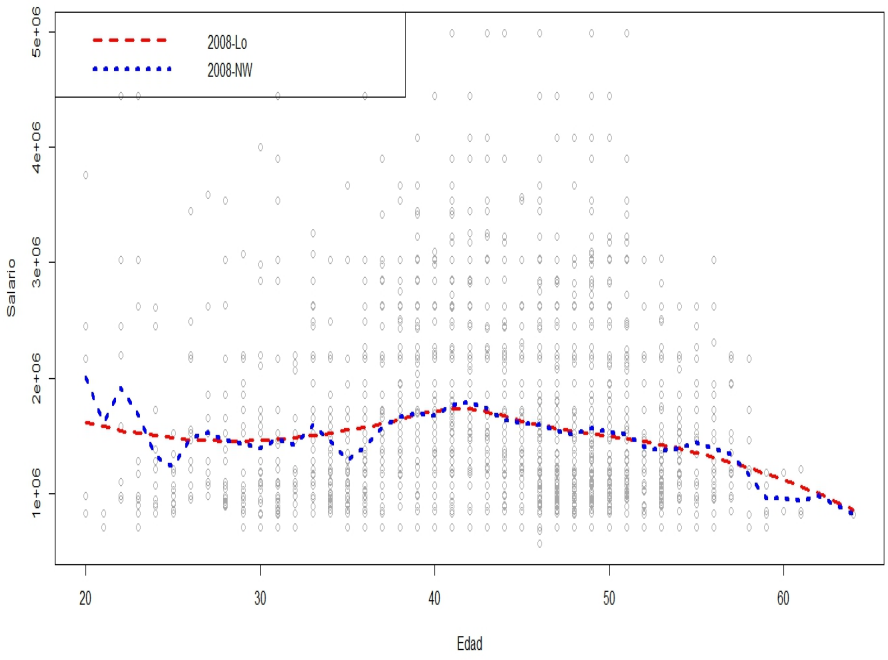

Figura 3-9.: Estimación escalas salariales N-W y Loess, 2008

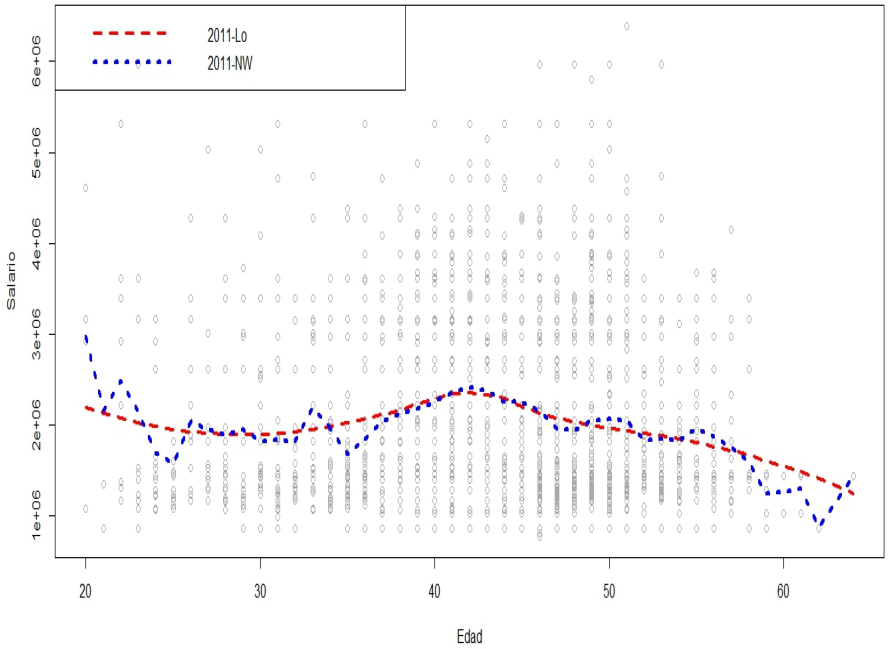

Figura 3-10.: Estimación escalas salariales N-W y Loess, 2011

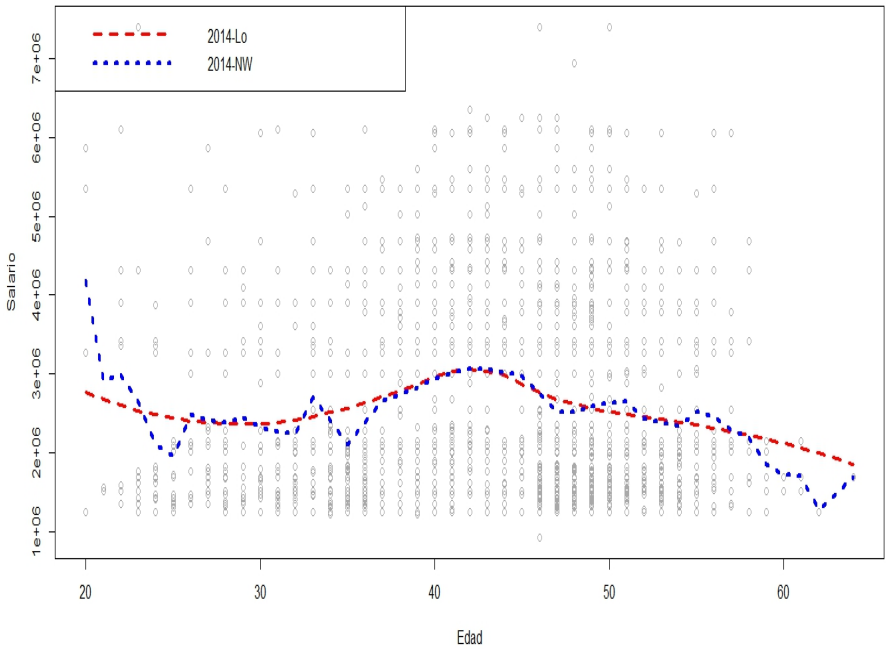

Figura 3-11.: Estimación escalas salariales N-W y Loess, 2014

En la Figura 3-9, 3-10 y 3-11 se presentan los resultados de la escala salarial para la estimación con la regresión Nadaraya-Watson, observando algunas variaciones en toda la dimensión de la gráfica pero manteniendo la consistencia de la curva en torno a la estimación de la escala salarial de la regresión Loess; por lo cual, se aprecian dos estimaciones cercanas en su interpretación general, aunque con diferencias en términos de su variación local, pues el N-W captura gran variación en materia local. Es importante aclarar que en este gráfico se incluye la regresión Loess para hacer un comparativo de la resistencia de las dos regresiones al efecto borde, donde se evidencia que la regresión Loess logra los mejores resultados frente a este problema.

2. Estimación con la regresión Loess

Al momento de realizar este proceso se tendrán en cuenta los siguientes factores para elegir el ancho de ventana óptimo para las estimaciones: primero, realizar ejercicios con diversos anchos de ventana que permitan definir cuál brinda los mejores resultados; segundo, considerar las metodologías de estimación del ancho de ventana; y tercero, examinar los valores usados en funciones Loess de otros trabajos que pretendían obtener una estimación robusta. A continuación, se empieza con la revisión de las estimaciones:

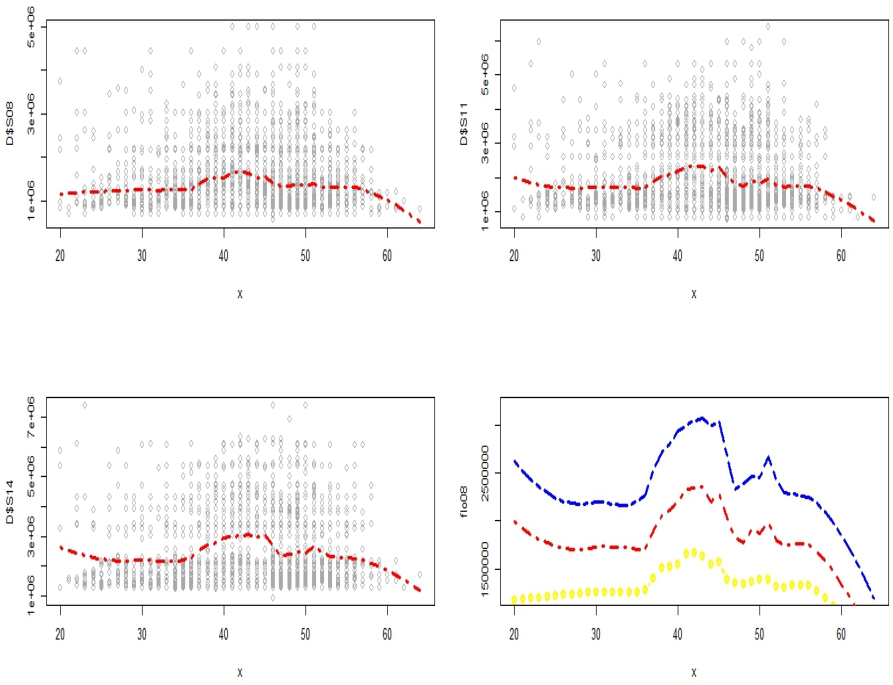

Figura 3-12.: Estimación de la escala salarial Loess, h=0.20

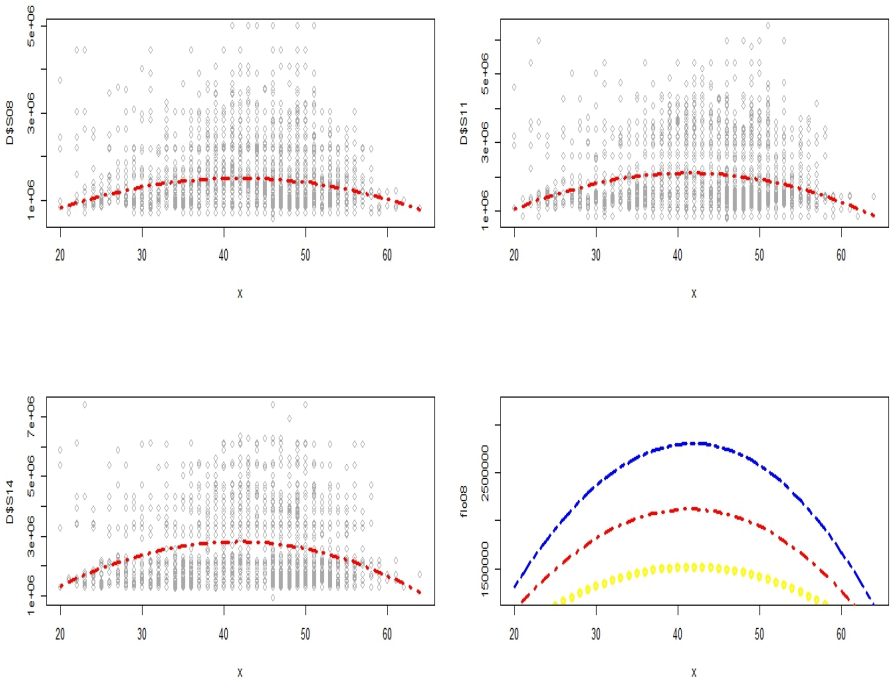

Figura 3-13.: Estimación de la escala salarial Loess, h=2

En los gráficos 3-12, 3-13 y 3-14, se evidencian tres resultados diferentes: en el primero se utiliza un parámetro demasiado pequeño que estima muy bien en términos locales, pero que en general tiene demasiada variación. En el segundo se encuentra el efecto contrario, pues los resultados locales se ven afectados por valores muy lejanos, lo que hace difícil resumir el comportamiento local, ocasionando grandes sesgos que desencadenan poca variabilidad en términos generales. En el tercero se halla un valor intermedio que alcanza a capturar algo de variabilidad local sin perder generalidad.

Ahora, examinando los valores del ancho de ventana sugeridos en la sección anterior, se puede definir un intervalo de valores entre 0.64 y 0.82, rango congruente que alberga el valor elegido para la estimación de Loess de 0.75.

La regresión Loess es comunmente usada con el valor de 0.75 en trabajos que prentenden una estimación con características robustas. Coherente con Lee y Cox (2008, pág. 7) "el método robusto LOESS por defecto para suavizar la selección de parámetros es el valor fijo, $\lambda = 3/4$. ".

En síntesis, la elección del parámetro ($h = 0.75$) es respaldado por la pruebas realizadas, por el rango de variación óptimo hallado con los métodos usados (sm, np) y por la recomedación práctica de otros trabajos que pretendían una estimación robusta. De

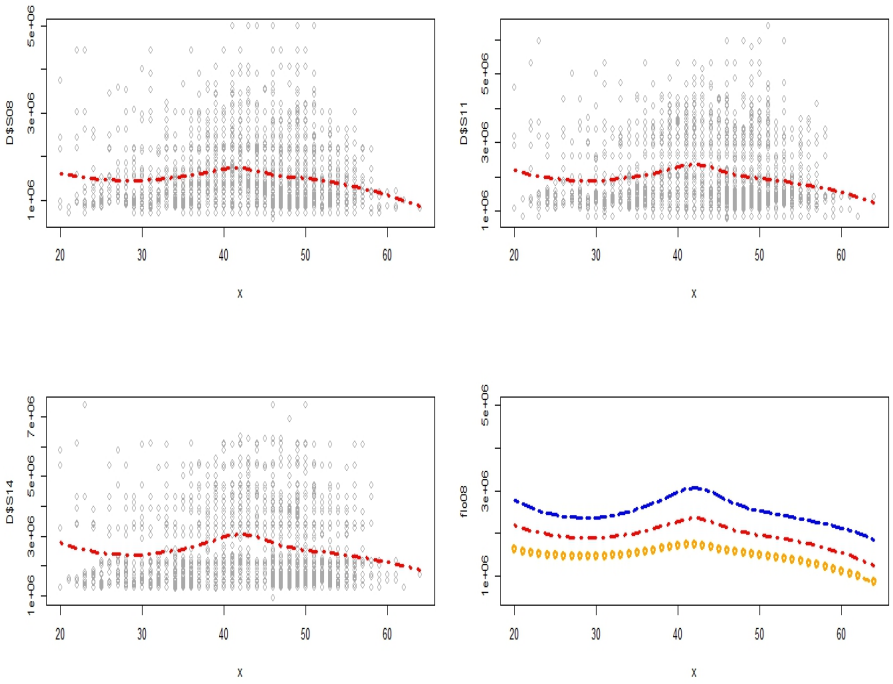

Figura 3-14.: Estimación de la escala salarial Loess, h=0.75

hecho, lo más importante en términos prácticos, es que la estimación es consistente con la teoría económica y actuarial en la construcción de la senda salarial. A continuación se exponen los resultados de la estimación en la senda salarial, para mayores detalles de la estimación ver Apéndice B.

En la Figura 3-15 se evidencian los resultados de la estimación de los incrementos con la regresión Loess; se presentan los incrementos salariales por edad, que tienen un origen diferente debido a que los incrementos del periodo 2008-2011 son mayores a los del periodo 2011-2014, y se destaca el aumento notable de las dos curvas en la mitad del periodo, cercano a los 40 años. También se observan diferencias en las variaciones de las dos curvas, hasta el punto de que la curva del periodo 2008-2011 muestra fuertes cambios en el transcurso del tiempo con una tendencia decreciente, entre tanto, la curva para los años 2011-2014, evidencia un crecimiento leve y una variación muy pequeña en el periodo de tiempo revisado.

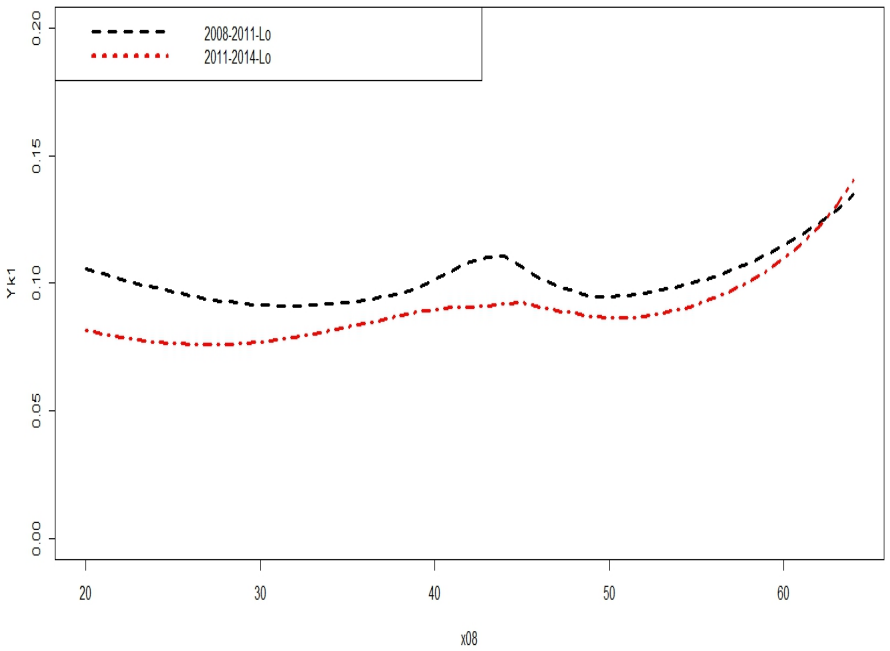

Figura 3-15.: Estimación de los incrementos salariales

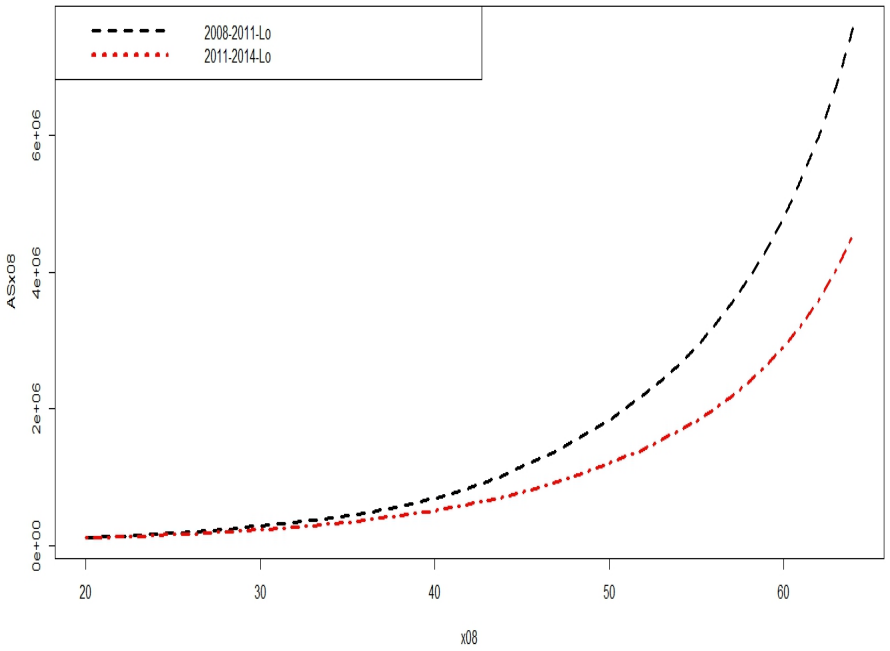

Figura 3-16.: Estimación de la senda salarial Loess

En la Figura 3-16 se observan las sendas salariales de los dos periodos, en la que destaca que ambas sendas muestra un comportamiento creciente, aunque es mucho mayor la tasa de crecimiento de la senda del periodo 2008-2011.

En la Figura 3-17 se señalan los incrementos salariales, cuyas variaciones se encuentran entre un 6 % y 10 % acordes con los aumentos por ley. Ahora, se corrobora que los incrementos son mayores para los individuos de los 40 años en adelante, lo que puede estar relacionado a un aumento por experiencia o por educación. También, se ratifica de forma notable que los aumentos son mayores para el periodo 2008-2011, este efecto esta relacionado con el ciclo político, pues fue una administración de mayores aumentos salariales.

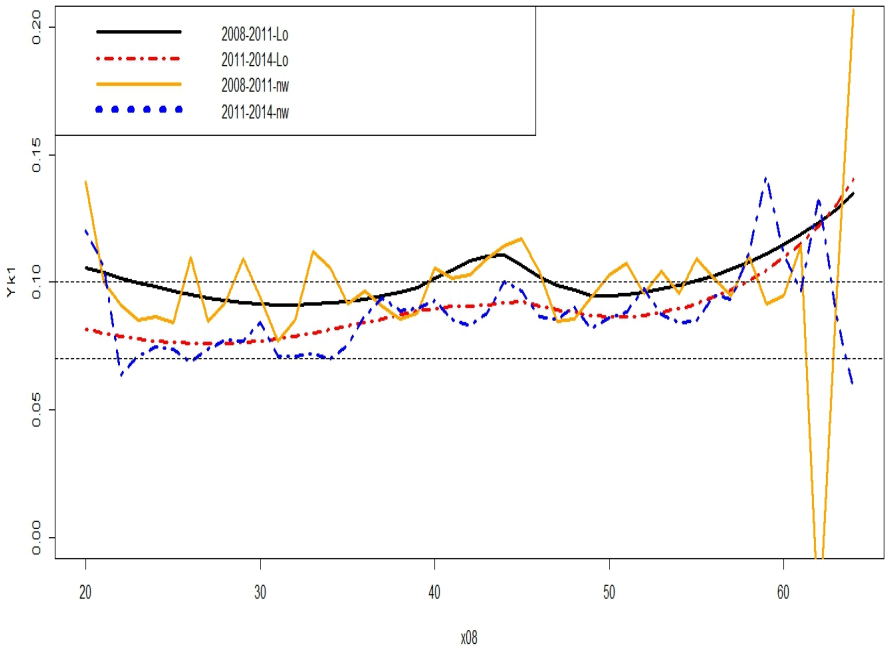

Figura 3-17.: Incrementos salariales Loess y N-W

A continuación, en la Figura 3-18 se presentan las sendas salariales con ambos métodos de estimación. Se observan que ambas regresiones, para los mismos periodos de tiempo, presentan resultados muy similares, y aunque existen diferencias después de los 60 años, estas no son significativas.

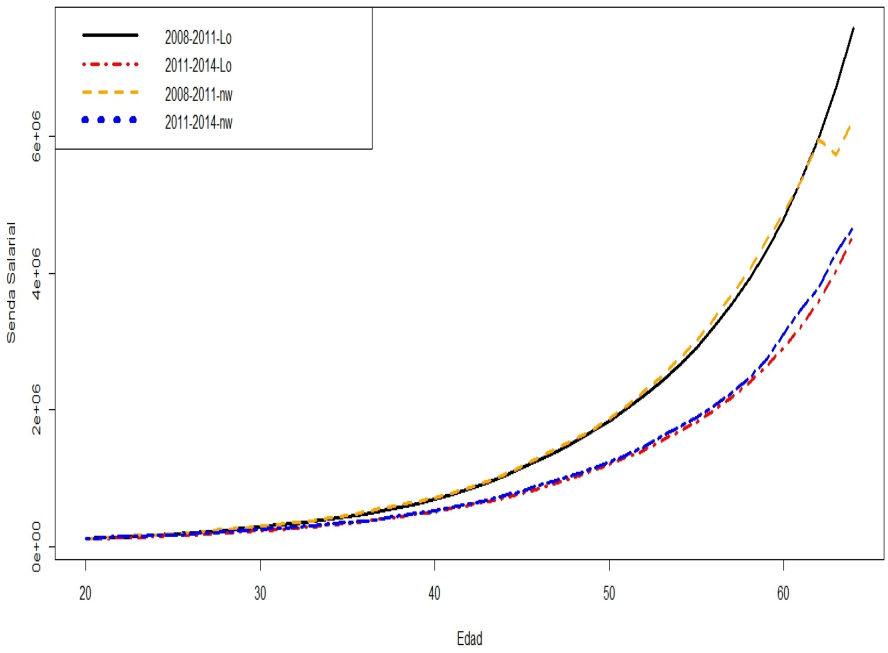

Figura 3-18.: Estimación de la senda salarial

3.4. Conclusiones acerca de las estimaciones

1. Realizando una comparación de la metodologia de cálculo de la senda salarial propuesta por Montenegro et al (2013), ecuación (2-11) y la propuesta en el presente trabajo en la ecuación (2-14), esta última se estima mediante Y_x, que se calcula a través de métodos no paramétricos en las ecuaciones (3-2), (3-5) y (3-6). Las estimaciones para cada uno de estos dos métodos son diferentes: en el primero se genera un promedio simple del intervalo inter-decílico de la forma $\frac{1}{{}_xk_g}$. El segundo se efectua mediante $m(X_i)$ (ecuación (3-5)), que realiza un promedio de las observaciones que caen en cada banda, pero este promedio se ve influenciado por $w_{i,n}$ (ecuación (3-6)), que determina el ancho de la banda y la forma de asociación de cada observación dentro de la banda.

   En el primer método, la otra parte de la ecuación (2-11), $\sum_{j\in {}_xI_{g,i}} AS^{(g)}_{x,j}$, es una sumatoria de los valores que cumplen los patrones de edad y género. Mientras que en el segundo método, la otra parte de la ecuación (3-2) es ϵ_i es un componente de error.

En resumen, se tienen tres diferencias básicas: primero, aunque las dos metodologías parten de realizar un promedio, uno es más refinado que otro desde el punto de vista estadístico. Segundo, el método utilizado por Montenegro et al (2013) se hace posterior a la agrupación de género y edad de cada decil, mientras que la metodología propuesta en este trabajo no requiere ninguna separación previa de los datos, pues los individuos son identificados con cédula. Tercero, la propuesta de Montenegro et al (2013) genera 10 sendas, una por cada decil, que son resultado de separar la población por género y edad; es decir, se calcularon como sendas individuales y luego se presentan como un conjunto, mientras, que la propuesta de este trabajo estima una sola senda para toda la población.

2. Los métodos de regresión no paramétricos utilizados para realizar la estimación de las sendas salariales producen resultados prácticamente idénticos para el mismo periodo de tiempo, sin embargo, se observan algunas diferencias en las escalas salariales, pues la regresión Nadaraya-Watson muestra una mayor variación en términos locales.

3. Las sendas salariales que comprenden el periodo 2008-2011 son mucho mayores a las del periodo 2011-2014, este hecho puede estar vinculado a mayores aumentos en los factores salariales en ese periodo de la administración pública, pues el gobierno nacional fija un mínimo de incremento salarial y los entes territoriales pueden aumentar varios puntos por encima de ese límite inferior basados en su disponibilidad de recursos económicos.

4. Aplicación: la cuestión de los subsidios en el RPM-PD

4.1. Cálculo de la mesada pensional

Los elementos para el cálculo de la mesada pensional en el RPM-PD se definen a continuación: Se asume un afiliado genérico al RPM-PD que ingresó al sistema a la edad e y se jubila a la edad r. La variable x indica su edad, tal que, durante su vida activa $e \leq x \leq r$. Su salario a la edad x es AS_x, dado por la senda salarial, donde la edad máxima para un ser humano, legalmente admitida se indica por ω, luego $x \leq \omega$, para el caso, colombiano es de $\omega = 110$.

Se asume una tasa media i_δ de aumento del salario mínimo anual y además, se supone una tasa efectiva anual de capitalización de aportes denotada por i. Se asume que se cumple $i > i_\delta$. Adicionalmente, se define la tasa técnica anual i_r tal que

$$1 + i_r = \frac{1+i}{1+i_\delta}. \tag{4-1}$$

Las variables de ingeniería financiera básicas en las pensiones del RPM-PD se definen a continuación.

El Ingreso Base de Liquidación a la edad x, $IBL(x)$, se define como el promedio de los últimos n salarios anuales actualizados por inflación

$$IBL(x) = \frac{1}{n}\sum_{j=1}^{n}(1+i_\delta)^{j-1}AS_{x-j}, \tag{4-2}$$

donde $n \geq 1$ indica número de salarios anteriores para calcular el promedio ($n = 10$ es el caso de la Ley 100).

El valor capitalizado de los aportes mensuales está dado por

$$C = 12(0.115)\sum_{j=e}^{r} AS_j(1+i)^{r-j} \tag{4-3}$$

En la expresión que utilizan Montenegro et al. (2013) se utiliza un porcentaje de 13 %, pero en ambos regímenes sólo el 11.5 % va a capitalización.

La tasa de reemplazo según el art. 34 de la Ley 100 de 1993, modificada por el art. 10 de la Ley 797 de 2003, está dada por

$$b(r-e) = 0.80 - \left(0.145 + 0.005\,IBL - 0.015\left\lfloor \frac{52(r-e)-1300}{50} \right\rfloor\right)_+, \tag{4-4}$$

donde $\lfloor x \rfloor$ es la función parte entera (floor), y $x_+ = I(x > 0)$ es la función parte positiva. La expresión para la función $b(x-e)$ según el art. 34 de la Ley 100 de 1993 es

$$b(r-e) = \begin{cases} 0.25 + 0.0004(52)(r-e) & \text{si} \quad 0 < 52(r-e) \leq 1200 \\ 0.01 + 0.0006(52)(r-e) & \text{si} \quad 1200 < 52(r-e) \leq 1400 \\ 0.85 & \text{si} \quad 1400 < 52(r-e) \end{cases} \tag{4-5}$$

La mesada pensional a la edad de retiro r es B_r definida por:

$$B_r = b(r-e)IBL(r), \tag{4-6}$$

donde $b(r-e)$ es la tasa de reemplazo a la edad r en función del número de años laborados $r-e$.

4.2. Cálculo de las rentas vitalicias

Para el cálculo del valor de una renta vitalicia ó anualidad de vida se requieren las expresiones actuariales correspondientes. El objetivo de esta sección es proporcionar tales expresiones y definiciones adicionales necesarias.

La probabilidad de supervivencia de una persona que está viva a la edad x, durante $k > 0$ años más, se denota por ${}_kp_x$, hay diferentes modelos para calcular ${}_kp_x$. El primero es un modelo no paramétrico consiste en las Tablas de Vida, que es el método legalmente aceptado en el país que consiste en las Tablas de Vida, usadas por la SuperFinanciera Resolución 1555 de 2010. En este caso ${}_kp_x = l_{x+k}/l_x$, donde l_x es el total de supervivientes de edad x, de un total inicial l_0, tomado por ejemplo $l_0 = 1000000$. El segundo método es utilizar una distribución paramétrica como la siguiente:

La distribución Gompertz-Makeham para la variable aleatoria $T(x) \in [0, \omega - x]$, que representa el número de años y fracción que alcanza a sobrevivir una persona viva de edad x, se define a partir de la fuerza de mortalidad

$$\mu_{x+t} = -\frac{d \ln \P(T(x) > t)}{dt},$$

colocando

$$\mu_{x+t} = a + bc^{x+t}, \tag{4-7}$$

con $a > 0$, $a + b > 0$ y $c > 1$.

El parámetro a mide la fuerza constante de mortalidad por riesgo de accidente y el término bc^{x+t} es la fuerza de mortalidad por envejecimiento. La probabilidad de (x) sobrevivir hasta la edad $x + t$ es:

$$ {}_tp_x = e^{-\int_0^t \mu_{x+s} ds} = s^x g^{c^x(c^t-1)} \tag{4-8} $$

donde $s = e^{-a}$, $g = \exp(-b/\ln(c))$. En la Tabla **4-1** siguiente se muestran valores estimados de los parámetros (s, g, c) con base en las dos últimas Tablas de Vida vigentes en el país.

	s	g	c
H(80-89)	0.9984022	0.9998469	1.1428834
M(80-89)	0.9995144	1.1071748	0.9929047
H(00-08)	1.0968159	0.9955661	0.9999789
M(00-08)	0.9950944	0.9999962	1.1257614

Tabla 4-1.: Parámetros Gompertz-Makeham por género según Tablas de Vida para los periodos 1980-1989 y 2000-2008. Cálculos propios.

El valor actuarial de una renta vitalicia para un jubilado (no se diferencia el género aquí) de edad x, con pagos anuales anticipados, que se incrementan anualmente a partir del segundo año en la tasa i_δ, está dada por $\ddot{a}_{x|i_r}$, donde $v_r = 1/(1 + i_r)$

$$ \ddot{a}_{x|i_r} = \sum_{k=0}^{\omega-x-1} v_r^k \, {}_kp_x, \tag{4-9} $$

```
#-------programa para valor
#-------de una renta vitalicia
#-------con base en Gompertz-Makeham
aax = function(s,g,C,x,i){
v = 1/(1+i)
k = seq(0,110-x-1,1)
kpx = tpx(s,g,C,x,t)
vk = (1+i)^(-t)
p = sum(vk*kpx)
return(p)
}
```

En la legislación pensional anterior a 2015, el valor de la tasa técnica i_r, ver (4-1), por Ley era 0.04. A partir de la Resolución 3099 de 2015 este valor se modificará

semestralmente por el Ministerio de Hacienda y Crédito Público. Ver la función `aax.u` para el cálculo de $\ddot{a}_{x|i_r}$, en el Apéndice D. Nótese que, equivalentemente

$$\ddot{a}_{x|i_r} = \sum_{k=0}^{\omega-x-1} \left(\frac{1+i_\delta}{1+i}\right)^k {}_kp_x, \tag{4-10}$$

La anualidad-salario anticipada, para una persona de edad x, temporal a n periodos, se define como el capital necesario para garantizar a esta persona una serie de pagos anticipados, empezando con el valor 1, con incrementos anuales iguales a los que se asumen para los salarios y por mejoramiento profesional. La expresión de este capital es

$${}^s\ddot{a}_{x:\overline{n}|} = \sum_{j=0}^{n-1} v_r^j \left(\frac{IS_{x+j}}{IS_x}\right) {}_jp_x,\ x+n \leq 72, \tag{4-11}$$

donde $v_r = (1+i_\delta)/(1+i)$, con i_δ como la tasa efectiva anual de reajuste del salario mínimo legal vigente, i como la tasa efectiva anual de cálculo de la anualidad y IS_x como el salario medio nacional, establecido en el decreto 2779 de 1994, Ministerio de Hacienda y Crédito Público, pág.1-2.

Se hace el siguiente supuesto, necesario para poder utilizar las estimaciones de las sendas salariales en el Capítulo 3, y no restringirse a la senda salarial con base en los salarios medios nacionales IS_x y la tasa i_δ, como en el Decreto 2779 de 1994. Dada una senda salarial específica S_x, la anualidad salario (4-11) se define como

$${}^s\ddot{a}_{x:\overline{n}|} = \sum_{j=0}^{n-1} v^j \frac{S_{x+j}}{S_x} {}_jp_x. \tag{4-12}$$

Nótese que en (4-12) aparece el factor de descuento $v = (1+i)^{-1}$ en lugar del factor v_r en (4-11). Esto es debido a que en la senda salarial S_x están incluídos los dos tipos de incrementos: por reajuste anual de salarios y por mejoramiento profesional. En cambio, en (4-11) los salarios medios nacionales IS_x solamente se consideran los incrementos por mejora profesional.

Para poder incluír las mesada adicionales de diciembre y mediados de año es necesario definir rentas vitalicias con pagos periodo vencido. Una renta vitalicia que provee m pagos vencidos al año, con $m = 12, 13, 14$, con incrementos anuales a la tasa i_δ, tiene valor presente

$$\nabla a_{x|i_\delta}^{(m)} = \sum_{j=1}^{m(\omega-x)} (1+i)^{-\frac{j}{m}} (1+i_\delta)^{\lfloor \frac{j}{m} \rfloor} {}_{\frac{j}{m}}p_x. \tag{4-13}$$

El símbolo para esta renta vitalicia no es estándar en la literatura. Se usa el símbolo "gradiente": $\nabla a_{x|i_\delta}^{(m)}$ para indicar el crecimiento de pagos a una tasa fija i_δ, ya que en ingeniería financiera se denomina un sistema tal de pagos como "gradiente geométrico". La fórmula (4-13) se puede modificar para llevarla a una forma conveniente

para los cálculos.

$$\begin{aligned}
\nabla a^{(m)}_{x|i_\delta} &= \sum_{j=1}^{m(\omega-x)} (1+i)^{-\frac{j}{m}}(1+i_\delta)^{\lfloor \frac{j}{m} \rfloor} \, {}_{\frac{j}{m}}p_x \\
&= \sum_{k=1}^{\omega-x} \left(\frac{1+i_\delta}{1+i}\right)^{k-1} \sum_{t=1}^{m} \frac{1}{(1+i)^{\frac{t}{m}}} \, {}_{k-1+\frac{t}{m}}p_x \\
&= \sum_{k=1}^{\omega-x} \left(\frac{1}{1+i_r}\right)^{k-1} \sum_{t=1}^{m} v^{\frac{t}{m}} \, {}_{k-1+\frac{t}{m}}p_x \\
&= \sum_{k=1}^{\omega-x} v_r^{k-1} \sum_{t=1}^{m} v^{\frac{t}{m}} \, {}_{k-1+\frac{t}{m}}p_x \qquad (4\text{-}14)
\end{aligned}$$

Una fórmula de aproximación para $\nabla a^{(m)}_{x|i_\delta}$, en (4-15), con base en (4-14), se obtiene definiendo las cantidades siguientes.

$$\begin{aligned}
v_\delta &= 1/(1+i_\delta) \\
d_\delta &= 1 - v_\delta \\
d_m &= 1 - v^{1/m} \\
i_m &= v^{-1/m} - 1 \\
\alpha(m) &= \frac{id}{m\, i_m d_m} \\
\beta(m) &= \frac{1}{d_m}\left(\frac{i}{m\, i_m} - 1\right)
\end{aligned}$$

$$\nabla a^{(m)}_{x|i_\delta} = [\alpha(m) - d_\delta(1+\beta(m))]\ddot{a}_{x|i_s} - (1+\beta(m))v_\delta. \qquad (4\text{-}15)$$

El código siguiente en R permite calcular (4-15). Con base en este código se puede calcular el capital necesario para financiar una renta vitalicia, con pagos mes vencido, más dos mesadas adicionales, una en junio y otra en diciembre, por valor de B_r el primer año, como

$$C = B_r(\nabla a^{(12)}_{x|i_\delta} + \nabla a^{(2)}_{x|i_\delta}) \qquad (4\text{-}16)$$

El capital C es el costo de una renta vitalicia, sin recargos. Esta cantidad depende de los factores actuariales como la probabilidad de supervivencia ${}_kp_x$, las tasas de incremento y capitalización i_δ, i, pero también del Ingreso Base de Liquidación (IBL). Para calcularlo de manera genérica se requiere una senda salarial representativa de la evolución de los cotizantes.

```
geo.avx = function(s,g,C,x,m,i,id){
if( 1+id >= 1+i ) stop("tasa inflacion mayor que tasa capitalizacion")
v = 1/(1+i)
d = 1 - v
ir = (1+i)/(1+id)-1
dr = 1-1/(1+ir)
aa.ir = aax(s,g,C,x,ir)

dm = 1-v^(1/m)
im = v^(-1/m)-1

f1 = d/dm - d + ds*(1/dm - 1)*(m*d/im-v)*(1+ir)
f2 = (1+ir)*(1/dm-1)*(m*d/im-v)

aax = f1*aa.ir-f2
return(aax)
}
```

4.3. Cálculo de los subsidios

En esta sección se definen varios conceptos, como la tasa de cotización y la condición de equilibrio entre aportes y costo de la anualidad, la expresión para el capital acumulado por aportes y una fórmula para el cálculo de los subsidios.

Tasa de cotización. Se define $\alpha \in (0,1)$ como un porcentaje sobre el salario anual que define el monto de la cotización anual de un trabajador de edad x. Esta cotización se indica por CE_x, y como siempre en este estudio se identifica cada periodo anual con una edad. Entonces

$$\begin{aligned} CE_x &= \alpha AS_x \\ &= \alpha AS_e(S_x/S_e). \end{aligned} \tag{4-17}$$

donde S_x es la escala salarial. Utilizando la expresión con base en los salarios medios nacionales Decreto en 2779 de 1994, se obtiene de manera similar

$$CE_x = \alpha AS_e(1+i_\delta)^{x-e} IS_x/IS_e.$$

En la teoría de los sistemas pensionales de prestación definida, con financiamiento, la tasa α se define a partir de una relación de equilibrio, como sigue. El caso de los sistemas públicos es similar, pero no se trata aquí. Se denota B_r la mesada mensual pensional dada en (4-6), ${}_kp_x$ una distribución de supervivencia determinada, $i \in (0,1)$ la tasa efectiva anual de interés de las anualidades, $i_\delta \in (0,1)$ la tasa promedio de

incremento anual de salarios, $i_r = (1+i)/(1+i_\delta) - 1$, la tasa técnica, con $i > i_\delta$. Dada una senda salarial $(AS_x, x = e, \ldots, r)$, una ley de mortalidad ${}_kp_x$, la tasa α en (4-18), para un afiliado genérico, se define mediante una ecuación de equilibrio entre el capital acumulado y el precio de la renta vitalicia con mesada anual anticipada B_r según (4-6), dada por

$$\alpha AS_e {}^s\ddot{a}_{e:\overline{r-e}|} = v^{r-e}{}_{r-e}p_e[B_r\ddot{a}_{r|i_r}]. \tag{4-18}$$

En (4-18) el valor presente de los aportes del empleado está en $AS_e {}^s\ddot{a}_{e:\overline{r-e}|}$ y el pasivo actuarial está en el miembro derecho, representado por el costo de la renta vitalicia $B_r\ddot{a}_{r|i_r}$, calculada a la edad de jubilación r, pero llevada a valor presente con el factor de descuento actuarial $v^{r-e}{}_{r-e}p_e$: el valor presente financiero v^{r-e}, condicionado a la supervivencia desde la edad inicial e hasta la edad de jubilación r, ${}_{r-e}p_e$. La única posibilidad de déficit sería un valor $\alpha > 1$.

En un sistema de prestación definida financiado, con cuentas individuales, un supuesto fundamental es que la senda salarial represente la evolución de los salarios de manera aproximada para todos los afiliados. En el Capítulo 3 se analizó este supuesto mostrando las dificultades que puede presentar la estimación de la senda salarial. El valor de α en (4-18) depende de la senda salarial a través de B_r y de la anualidad salario ${}^s\ddot{a}_{e:\overline{r-e}|}$.

4.4. Ecuación de capitalización de aportes

La capitalización de los aportes se hace a una tasa efectiva anual i_a, que depende del manejo financiero de la Entidad responsable, por ejemplo, una Adminstradora de Fondos de Pensiones (AFP). La relación entre i_a e i no se discutirá aquí. Pero i_a que debe ser superior a la tasa asumida para valorar la anualidad, i, $i_a > i$, para garantizar que la ecuación de equlibrio (4.18). Se asume que i es una tasa compuesta $i = (1+i_\delta)(1+i_r) - 1$, pero este supuesto ha sido muy criticado por no ser realista, para mayores detalles ver Caro(2006). Dadas i_r por parte del Ministerio de Hacienda y Crédito Público, e i_δ por parte del DANE, como el IPC, las fórmulas de rentas vitalicias no utilizan i_δ para ajuste anual de las mesadas sino el decreto para ajuste del salario mínimo, por lo que las reservas pueden ser insuficientes. Además, componer estas tasas no siempre resulta en una tasa que se pueda garantizar en el mercado financiero. Estos problemas no puede resolverse aquí y se realizarán los cálculos asumiendo unas tasas i_t, i_δ lo más realistas posibles. En cambio, i_a depende ampliamente de cada tipo de portafolio utilizado por cada AFP y puede tener valores mayores que i.

Si V_x denota el valor del saldo de la cuenta de ahorros pensional, en un régimen de prima media financiado, ó en el Régimen de Ahorro Individual con Solidaridad (RAIS), se tienen al menos dos maneras de plantear el flujo de caja que se genera mediante la capitalización de las cotizaciones CE_x a la tasa i_a efectiva anual, tal que $v_a = (1+i_a)^{-1}$.

Asumiendo que las cotizaciones son anticipadas, se tiene una primera ecuación, (4-19), que se denominará ecuación para la reserva actuarial. Con $v = (1+i)^{-1}$, e i la tasa para cálculo de anualidades, se forma la ecuación para el flujo de caja

$$V_x = \frac{V_{x-1} + CE_{x-1}}{vp_{x-1}},\ x = e, e+1, \ldots, r. \tag{4-19}$$

La solución de la ecuación (4-19), con $V_{e-1} = 0$ está dada por

$$V_x = \frac{CE_e\ {}^s\ddot{a}_{e:\overline{x-e}|}}{v^{x-e}\ {}_{x-e}p_e}. \tag{4-20}$$

Con la ayuda del software MAPLE v.2016.1, se programan la instrucciones

$$\{f(n) = \frac{f(n-1) + C(n-1)}{vp(n-1)}\};$$
$$rsolve(\%, f(k));$$

y se obtiene la expresión

$$\begin{aligned}&\left(\frac{1}{v}\right)^k \prod_{n=0}^{k-1} \frac{1}{p(n)} \left[\sum_{m=0}^{k-1} \frac{C(m)\left(\frac{1}{v}\right)^{-m}}{p(m)\prod_{n=0}^{m}\frac{1}{p(n)}}\right]\\ &= \frac{1}{v^k \prod_{n=0}^{k-1} p(n)} \sum_{m=0}^{k-1} \frac{C(m)v^m}{p(m)\prod_{n=0}^{m}\frac{1}{p(n)}}\end{aligned} \tag{4-21}$$

Reemplazando en los sumandos y productorias en (4-21), n por $n+e$, m por $m+e$, k por $x-e$

$$\frac{1}{v^{r-e}\prod_{n=0}^{x-e-1} p_{n+e}} \sum_{n=0}^{x-e-1} \frac{CE_{n+e}v^{n+e}}{p_{n+e}\prod_{j=0}^{n}\frac{1}{p_{j+e}}} \tag{4-22}$$

Esta expresión (4-22) se simplifica utilizando las identidades

$$\prod_{n=0}^{x-e-1} p_{n+e} = p_e p_{e+1} \cdots p_{x-1} = {}_{x-e}p_e$$
$$\frac{p_{n+e}}{\prod_{j=0}^{n} p_{j+e}} = \frac{p_{n+e}}{p_e p_{e+1} \cdots p_{n+e}} = \frac{p_{n+e}}{{}_{n+1}p_e} = \frac{1}{{}_np_e}.$$

para obtener

$$\frac{1}{v^{x-e}\ {}_{x-e}p_e} \sum_{n=0}^{x-e-1} CE_{n+e}v^{n+e}\ {}_np_e \tag{4-23}$$

pero, de la definición de cotización de ejercicio CE_x, (4-17) y de anualidad salario ${}^s\ddot{a}_{x:\overline{n}|}$, (4-11), se tiene

$$\begin{aligned}\sum_{j=0}^{x-e-1} CE_{e+j} v^j {}_jp_e &= \sum_{j=0}^{x-e-1} CE_{e+j} v^j {}_jp_e \\ &= \alpha AS_e \sum_{j=0}^{x-e-1} (S_{e+j}/S_e) v^j {}_jp_e \\ &= CE_e \, {}^s\ddot{a}_{e:\overline{x-e}|}.\end{aligned}$$

por tanto, la solución se puede escribir como

$$V_x = \frac{CE_e \, {}^s\ddot{a}_{e:\overline{x-e}|}}{v^{x-e} \, {}_{x-e}p_e}.$$

con lo cual se completa la demostración. La consecuencia importante es que la solución (4-20) cumple, reeemplazando $x = r$, aplicando la ecuación (4-18), que

$$V_r = B_r \ddot{a}_{r|i_r}, \tag{4-24}$$

es decir, el capital ahorrado es el necesario para financiar una renta vitalicia que paga inicialmente la mesada B_r, con incrementos anuales a una tasa media i_δ. Y capitalizando a la tasa i, no a la tasa i_a.

Una observación importante es que (4-20) se puede re-escribir como

$$V_r = \sum_{j=0}^{r-e-1} \left(\frac{\alpha \, {}_jp_e}{{}_{x-e}p_e} \right) AS_{j+e}(1+i)^{r-e-j} = \sum_{j=0}^{r-e-1} \alpha_j AS_{j+e}(1+i)^{r-e-j},$$

donde $\alpha_j = \left(\frac{\alpha \, {}_jp_e}{{}_{x-e}p_e} \right)$ es un porcentaje de cotización que va cambiando con los años.

Una segunda expresión para el flujo de caja consiste en capitalizar con la tasa i_a, que es una tasa de mercado, utilizada en los portafolios de la AFP. Utilizando la misma notación que en (4-19) (sin generar confusión ya que el caso que se aplicará es el que sigue),

$$V_x = \frac{V_{x-1} + CE_{x-1}}{v_a}, \; x = e, e+1, \ldots, r. \tag{4-25}$$

La solución es

$$V_r = \alpha \sum_{j=0}^{r-e-1} AS_{j+e}(1+i_a)^{r-e-j} \tag{4-26}$$

Aplicar un porcentaje fijo α puede producir un déficit, ya que la ecuación (4-24) no se cumple. Pero si la tasa i_a es suficientemente alta, se tendrá la condición (4-24). De manera que la existencia de los subsidios dependerá de qué tan alta sea la tasa de

capitalización durante la fase de acumulación ó de ahorro. Una tasa óptima será el mayor valor i_a tal que

$$V_r(i_a) \leq B_r \ddot{a}_{r|i_r}. \tag{4-27}$$

De otra forma, se presentarán subsidios en el caso $V_r(i_a) < B_r \ddot{a}_{r|i_r}$. Por lo que en este estudio interesará encontrar los mayores valores i_a para los cuales se cumpla (4-27), dada la senda salarial AS_x, la ley de mortalidad ${}_tp_x$, y la tasa i_δ.

En el caso $V_r(i_a) < B_r \ddot{a}_{r|i_r}$ el valor del subsidio se define como el porcentaje

$$SBD = \frac{B_r \ddot{a}_{r|i_s}}{V_r(i_a)} - 1.0. \tag{4-28}$$

4.5. Ejemplos numéricos

En esta sección se va a ilustrar numéricamente el cálculo de las mesadas pensionales con la fórmula del régimen RPM-PD, el costo de una renta vitalicia y el capital acumulado de las cotizaciones, con base en las sendas salariales estimadas por los métodos siguientes

1. Mediante las fórmulas de salarios medios nacionales en el Decreto en 2779 de 1994. (Es solamente un referente, no se calculan)

2. Mediante las estimaciones con regresiones no parámetricas

Con base en el costo de la pensión se encontrarán valores de las tasas i_a a las cuales se capitalizan los aportes para conseguir un menor valor de posibles subsidios.

Todos los resultados siguientes asumen un afiliado genérico. En un primer caso se analizará el caso de género masculino de 62 años, que empezó a cotizar al sistema de RPM-PD, a la edad de 20 años, con un salario inicial de $100.000. Los aportes serán el 11.5 % de los salarios que crecerán de acuerdo a una senda salarial.

La renta vitalicia se calcula como en el sistema de RPM-PD, con 14 mesadas al año. Se utilizarán las tasas $i_r = 0.04$ y una tasa de crecimiento de la mesada de $i_\delta = 0.04$. La ley de mortalidad será una distribución Gompertz-Makeham correspondiente a los casos hombre-mujer, ajustada para el caso de la Tabla de Vida, ver SuperFinanciera Resolución 1555 de 2010.

Los cálculos determinarán una tasa i_a mediante la cual el capital acumulado sea menor pero aproximado al valor del costo de la renta vitalicia.

Una pregunta que interesa responder hace referencia a si ¿el costo de la renta vitalicia y el monto del capital acumulado dependen del método utilizado para calcular la senda salarial?. En la Tabla **4-3** se pueden ver los resultados para el caso de Tabla de Vida Hombres 2000-2008.

Género	s	g	c	Edad	Tasa i	Tasa i_δ
H	0.9929043	0.9999789	1.1260000	62	0.0816	0.04
M	0.9955659	0.9999962	1.1430000	57	0.0816	0.04

Tabla 4-2.: Parámetros de la Renta Vitalicia

Senda	Periodo	Tasa i_a	Pensión B_r	Costo $B_r\ddot{a}_{r\|i_r}$
Reg. N-W	2008-2011	0.1651	\$4'194.504	\$776'823.488
Loess	2008-2011	0.1651	\$4'115.847	\$762'256.174
Reg. N-W	2011-2014	0.1651	\$2'637.924	\$488'544.456
Loess	2011-2014	0.1651	\$2'513.067	\$465'420.783

Tabla 4-3.: Estimaciones para afiliado hombre, edad = 62

En la Tabla **4-4** se pueden ver los resultados para el caso de Tabla de Vida Mujeres 2000-2008.

Senda	Periodo	Tasa i_a	Pensión B_r	Costo $B_r\ddot{a}_{r\|i_r}$
Reg. N-W	2008-2011	0.1648	\$2'595.693	\$529'669.153
Loess	2008-2011	0.1648	\$2'517.405	\$513'693.405
Reg. N-W	2011-2014	0.1648	\$1'650.098	\$336'713.859
Loess	2011-2014	0.1648	\$1'593.766	\$325'219.064

Tabla 4-4.: Estimaciones para afiliado mujer, edad = 57

5. Análisis de la metodología en Montenegro, Jiménez y Hurtado (2013)

5.1. Análisis de la metodología para sendas salariales

El punto de partida es la metodología de Montenegro et al. (2013) para analizar los datos del año 2014. La Figura 5-1 muestra los resultados de los datos estudiados en

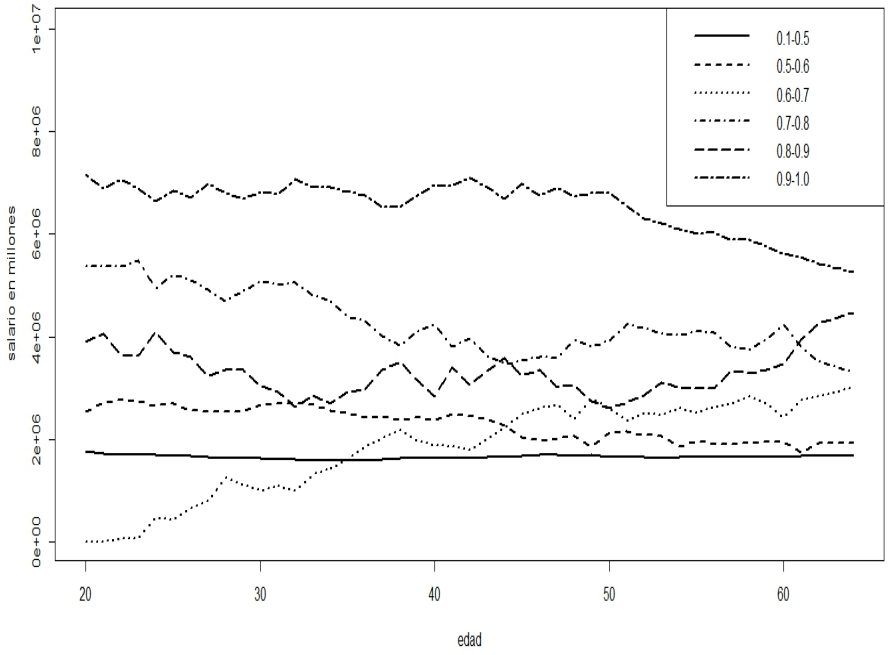

Figura 5-1.: Senda Salariales 2014. Metodología de Montenegro et al (2013)

esta tesis bajo esta metodología, cuya generación de sendas salariales no parece ser coherente, ya que se observan sendas decrecientes, lo cual contradice el hecho de que las sendas contienen un factor que depende del porcentaje de ajuste salarial anual. Además, presentan intersecciones entre ellas. (Ver el programa en R en el Apéndice C).

5.1.1. Cálculo del capital acumulado

Con los elementos del capítulo anterior se procederá a revisar las fórmulas en el estudio de Montenegro et al. (2013) para el cálculo de posibles subsidios. La expresión para el cálculo del capital acumulado con base en la capitalización de los aportes como un porcentaje de las sendas salariales se define como

$$VC_g^i = 12aS_{min}\, d_g \sum_{e=20}^{\bar{e}_{min}^g} {}_e s_g^i (1+tc)^{\bar{e}_{min}^g - e + 1} \tag{5-1}$$

Los indicadores $g \in \{h, m\}$, e $i = 1, 2, \ldots, 10$, definen el género y el decil respectivamente. El porcentaje $a = 0.135$ es la cotización del salario mínimo mensual ${}_e s_g^i S_{min}$, calculado con la senda salarial. La tasa de capitalización efectiva anual es $tc = 0.04$. El indicador d_g de fidelidad se define como un porcentaje relacionado con el número de semanas cotizadas; para 1300 semanas es $d_m = 0.67$ y $d_h = 0.59$, que corresponde a una edad de jubilación de 57 años para mujeres y 62 para hombres.

Para una revisión de la fórmula (5-1) se procede a re-escribirla con la notación del presente trabajo. Ignorando la diferencia de género y el decil, se denota el capital VC_g^i por V_r. Reemplazando en (5-1) el contador de la sumatoria e por j, e iniciando desde la edad $e = 20$ hasta la edad $\bar{e}_{min}^g = r$, con $tc = i$, $a = \alpha$, $S_{min} = AS_e$, ${}_e s_g^i = \frac{S_x}{S_e}$, y eliminando el factor 12, que es sólo una escala, se obtiene la siguiente ecuación

$$V_r = \alpha AS_e\, d_g \sum_{j=e}^{r} \frac{S_x}{S_e}(1+i)^{r-j+1}$$

En estas expresiones, multiplicar el capital por una constante a modo de penalización no es justificable, por tanto, la constante d_g será reemplazada por $d_g = 1$, cuyo papel representa una reducción del capital acumulado proporcional al número de semanas cotizadas.

Recordando que $CE_x = \alpha \frac{S_x}{S_e} AS_e = \alpha AS_x$ es la cotización mensual durante el año en el cual el cotizante tiene edad x, se obtiene

$$V_r = \sum_{j=e}^{r} CE_j (1+i)^{r-j+1} = \sum_{j=1}^{r-e+1} CE_{r-j+1}(1+i)^j = \sum_{j=0}^{r-e-1} CE_{j+e}(1+i)^{r-e-j} \tag{5-2}$$

Si se observa con detalle, en la primera suma existe un error dado que a la edad r se capitaliza la cotización por un año más, lo cual no es cierto, por tal razón es necesario cambiar $\sum_{j=e}^{r} CE_j(1+i)^{r-j+1}$ por $\sum_{j=e}^{r-1} CE_j(1+i)^{r-j}$.

La expresión (5-2) es el valor futuro de las cotizaciones, capitalizadas año anticipado, a la tasa i. El flujo de caja V_x correspondiente para el capital a la edad x, es

$$V_x = \frac{V_{x-1} + CE_x}{v},\ x = e, e+1, \ldots, r. \tag{5-3}$$

donde $v = 1/(1+i)$.

La expresión tiene solución, con $V_{e-1} = 0$, convirtiéndose en

$$V_x = \sum_{j=e}^{x} CE_j (1+i)^{x-j+1}. \tag{5-4}$$

Crítica. La metodología en Montenegro et al. (2013) utiliza una tasa real i en la ecuación de flujo en (5-4) que es la misma tasa técnica i_r del cálculo de las anualidades. Ver Montenegro et al. (2013, pág. 9):

> "En ese sentido se supone que los aportes realizados por el afiliado tienen un rendimiento real igual a la tasa de interés técnico. [...] Para capitalizar los aportes de un afiliado se utiliza una tasa real de largo plazo $tc = 4\,\%$ siendo ésta similar a la tasa de negociación de los TES UVR emitidos a tasa fija. Se quiere resaltar que, pese a que los niveles promedio de estas tasas reales han estado en un 5.8 %, se considera más relevante ajustar el ejercicio a la situación actual de promedio en un 4 %, dadas las perspectivas macroeconómicas del país."

La interpretación de este supuesto de utilizar tasas de rendimientos reales es que pueden presentarse problemas teóricos y prácticos al momento de implementarse con los datos reales. En primer lugar, la descomposición de la tasa de rendimiento = tasa real + tasa de inflación, puede no darse en rendimientos generados por portafolios de inversión, en los cuales es posible tener tasas de rendimientos muy bajas ó inclusive negativas. En segundo lugar, usar tasas reales para capitalizar valores generados a partir de sendas salariales no parece coherente toda vez que las sendas, por su definición, deben incluír un efecto por inflación. En consecuencia, para utilizarlas deberían deflactarlas con respecto a la inflación ó a la tasa de ajuste del salario mínimo. En todo caso, la capitalización de aportes se realiza en portafolios de inversión que, en el RAIS, han demostrado altos rendimientos. Por tanto, la posibilidad de que se generen transferencias por insuficiencia en los capitales ahorrados depende críticamente del nivel medio de los rendimientos de los portafolios.

Por la razón anterior, este estudio considera que la hipótesis de existencia de subsidios en el RPM-PD se prueba en Montenegro et al. (2013) de forma poco satisfactoria. Como se señaló, no usaron tasas realistas, de valor medio alto, en los portafolios de inversión, sino tasas reales. Además, la metodología utilizada para generar sendas salariales parece que, al aplicarla a datos reales, produce resultados no coherentes: las sendas no aumentan con la edad sino que son incluso decrecientes ó se intersectan entre ellas.

5.1.2. Fórmulas de renta vitalicia

Montenegro et al. (2013, ecuación 3, pág. 11) proporciona la fórmula utilizada en ese estudio para calcular el valor de la renta vitalicia de la siguiente forma

$$VPN_{g,i} = \sum_{j=\bar{e}_g+1}^{110} 12.\overline{max}\{1, IBL_{g,1}.TR_{g,i}\}.\, Smin._{{}^g_e P_{\bar{e}_g+1}}.\,(1+tit)^{\bar{e}_g-e+1}. \quad (5\text{-}5)$$

para $i = 1, 2, ..., 10, \quad g = h, m$

La ecuación (5-5) permite conocer el valor descontado de los pagos futuros de pensión de un individuo promedio diferenciando por el género y el decil del ingreso. A continuación se explican los componentes de la ecuación. $\overline{max}\{\}$ indica la presencia de la cota superior, que según la constitución son 25 SMMLV. A su vez, se tomaron en cuenta parámetros como el IBL, la tasa de reemplazo TR y la senda salarial $Smin$. Los parámetros de ponderación de los flujos de pensión denotan la probabilidad de estar vivo en los siguientes años, distinguiendo por el género y la edad del afiliado, entre estos se incluye e para una edad fija; k para una edad mayor o igual que e; y p_j^g para la probabilidad de estar muerto a la edad j en el género g. En este caso, ${}^g_e P_{\bar{e}_g}$ hace referencia a la probabilidad condicional de un afiliado de un género específico g que esté vivo a la edad k dado que tiene la edad e. También, se incorpora la tasa de interés de capitalización de los aportes $(1+tit)^{\bar{e}_g-e+1}$.

5.1.3. Fórmulas del Ingreso Base de Liquidación (IBL)

La ecuación IBL en Montenegro et al. (2013, pág. 10) muestra el cálculo según el género y el decil de ingreso, de la forma

$$IBL_{g,i}(t) = \frac{1}{[t/d_g]} \sum_{k=1}^{[t/d_g]} s_{g,\bar{e}_g+1-k,i} \quad i = 1, ..., 10; \quad g = h, m. \quad (5\text{-}6)$$

La revisión de la ecuación (5-6) permite definir algunos elementos como: el parámetro t hace referencia al número de años laborales; d_g es la edad diferenciada por género; y $s_{g,\bar{e}_g+1-k,i}$ es la senda salarial de género g, edad e que pertenece al decil de ingreso i.

El punto de partida para la comparación, entre la ecuación expuesta por Montenegro et al. (2013) y la propuesta en este trabajo, es dejar de lado las diferencias de género. Luego, se debe iniciar por la distinción de las características particulares de cada método, por ejemplo, es mucho más sencillo utilizar una fracción $1/n$ para contar el número de salarios anteriores para calcular el promedio (n=10 según la Ley 100), que la fracción $\frac{1}{[t/d_g]}$. En cuanto a la sumatoria, ambas dan cuenta de la acumulación de salarios durante un periodo de tiempo (en Colombia son los 10 últimos años de

vida laboral). Sin embargo, la ecuación en Montenegro et al. (2013) sólo suma la senda salarial y no considera la tasa de capitalización (por lo menos no de forma implícita), mientras que en este trabajo se considera la senda salarial múltiplicada por una tasa de capitalización de $(1+i_\delta)^{j-1}$, generando la ecuación planteada en (4-2)

$$IBL(x) = \frac{1}{n}\sum_{j=1}^{n}(1+i_\delta)^{j-1}AS_{x-j}$$

Crítica Respecto a este cálculo, existe divergencia en dos puntos. Primero, la presentación del IBL resultante del trabajo de Montenegro et al. (2013) no muestra el valor de capitalización, ó por lo menos no de forma implícita. Siendo así, omitir esta tasa de capitalización es suponer que no es esencial para el cálculo del IBL, posición que este trabajo no considera adecuada, debido a que no es correcto ignorar los incrementos del ajuste anual del salario mínimo. Segundo, es muy compleja la fracción que utilizan para expresar el promedio del número de salarios anteriores, por lo cual este trabajo propone $\frac{1}{n}$.

En cuanto al cálculo del IBL, la mayor diferencia entre los autores mencionados y la propuesta de este trabajo radica en la metodología para cuantificar la senda salarial. En el trabajo de Montenegro et al. (2013) la senda salarial $s_{g,\bar{e}_g+1-k,i}$ se construye como un promedio del salario de muchos individuos con una misma edad y un mismo decil de ingreso, generando un efecto directo en el cálculo del IBL, ya sea por sobreestimación ó subestimación del salario real que se gana un individuo. Para evitar estos efectos de dispersión en el cálculo del IBL, este trabajo propone una metodología en la elaboración de una senda salarial AS_x para un individuo genérico, identificado con cédula.

5.1.4. Estimaciones

En las siguientes tablas se realizan los cálculos utilizando los datos del estudio, pero con la metodología de Montenegro et al (2013), se utilizan únicamente tres deciles para ilustrar los resultados.

Senda	Periodo	Tasa i_a	Pensión B_r	Costo $B_r\ddot{a}_{r\|i_r}$	Déficit
Decil 1-5	2014	0.277	\$1'737.651	\$73'503.356	\$ -48'273.879
Decil 7	2014	0.277	\$3'233.121	\$94'036.143	\$ -47'093.476
Decil 9	2014	0.277	\$4'665.915	\$162'150.869	\$ -94'405.035

Tabla 5-1.: Estimaciones para afiliado hombre, edad = 62

En la siguiente Tabla se pueden ver los resultados para el caso de Tabla de Vida Mujeres 2000-2008.

Senda	Periodo	Tasa i_a	Pensión B_r	Costo $B_r\ddot{a}_{r\|i_r}$	Déficit
Decil 1-5	2014	0.396	\$1'537.600	\$61'807.320	\$ -35'273.382
Decil 7	2014	0.396	\$3'033.121	\$83'102.859	\$ -33'733.073
Decil 9	2014	0.419	\$4'464.900	\$144'294.166	\$ -73'045.609

Tabla 5-2.: Estimaciones para afiliado mujer, edad = 57

En síntesis, los resultados de las estimaciones con la metodología de Montenegro et al (2013) reflejan una baja acumulación de capital producto de sendas decrecientes y por ende se generan una déficit, que se convierte en subsidios por parte del estado.

5.2. Conclusiones

1. Con relación a la metodología para el cálculo de sendas salariales en Montenegro et al. (2013), se encontró que en la aplicación realizada para el caso analizado en este trabajo, no es coherente con la definición de senda salarial, por lo que no es recomendable.

2. La hipótesis planteada en Montenegro et al. (2013) en lo referente a que los subsidios representan un faltante con respecto a un costo de una renta vitalicia con relación al capital generado por los aportes y sus rendimientos, resulta cuestionable. Lo que esta tesis encontró es que estos subsidios pueden generarse si los aportes se capitalizan con tasas de rendimientos muy bajas. En todo caso, los resultados encontrados muestran que tasas en promedio de 16 % anual son suficientes para eliminar la posibilidad de un descalce entre el capital acumulado y el valor de la renta vitalicia, calculada con los parámetros del RPM-PD.

3. Las estimaciones realizadas permiten observar una serie de valores negativos o déficit en todos los deciles utilizados y en los dos géneros, por lo cual, es necesario que aparezca el Estado como un ente que respalde el sistema pensional con recursos.

6. Conclusiones

Los trabajos teóricos y empíricos que se han construido en el área de economía laboral para explicar los determinantes del salario enfatizan en una cantidad de variables como posibles causantes de las diferencias salariales, entre los cuales se tiene: el mercado laboral (oferta y demanda de empleos), el costo de vida, la productividad, el nivel de educación (incluyendo en algunas ocasiones el status de la universidad que emite el título), la experiencia profesional, la edad y las habilidades cognitivas del individuo, entre otros. Siendo así, determinar el salario implica definir cuáles variables juegan un papel clave en el mercado laboral y en la economía de un territorio particular. En este trabajo se hacen una serie de complejos y rigurosos supuestos desde el punto de vista económico intentando explicar el salario en términos de la edad, suponiendo que a mayor edad se tiene mayor experiencia o mayor nivel educativo. Por lo tanto, los métodos estadísticos son sometidos a una gran exigencia para precisar las relaciones entre las variables, y es claro que no se alcanza a capturar todos los efectos en una sola ecuación.

Un resultado del estudio radica en descartar los métodos paramétricos para la estimación de las sendas salariales, basados en la premisa de que no son los más adecuados ante la presencia de valores extremos y su estimación genera grandes distorsiones. Ante este hecho, para la población estudiada, resulta conveniente utilizar métodos no paramétricos para calcular la senda salarial, como la regresión Loess y la regresión Nadaraya-Watson.

El punto de partida para la construcción de la senda salarial son las definiciones teóricas dadas por Carriere y Shand (1998). Esta iniciativa garantizaría una metodología estadística y actuarial rigurosa que podría mejorar las estimaciones del recorrido salarial de los individuos analizados y así tener fundamentos claros para determinar si realmente son necesarios los subsidios a los afiliados del RPM. La claridad teórica de la senda salarial es fundamental al momento de evaluar la eficiencia del sistema pensional de carácter público.

Teniendo en cuenta lo anterior y con relación a la metodología para el cálculo de sendas salariales dadas por Montenegro, Jiménez y Hurtado (2013), se encontró que esta metodología no es coherente con la definición teórica de senda salarial por varias razones, primero, produce sendas decrecientes, que no cumplen ni siquiera con el

supuesto de crecimiento por inflación. Segundo, se interceptan entre ellas, lo cual es una propiedad indeseable, pues estaría afirmando que los individuos pueden cambiar de sendas. Por consiguiente no es metodología recomendable para este caso particular. A su vez, resulta cuestionable la hipótesis planteada en Montenegro et al. (2013) sobre la inherente necesidad de los subsidios para cubrir el faltante respecto al costo de una renta vitalicia debido a que el capital generado entre los aportes y los rendimientos no son suficientes para adquirirla. No obstante, este trabajo encontró que esta hipótesis es relativa, pues depende de las tasas de capitalización de los aportes, es decir, si los aportes se capitalizan con tasas de rendimiento bajas son necesarios los subsidios. Este trabajo concluye que tasas en promedio del 16 % anual son suficientes para eliminar la posibilidad de un desbalance entre el capital acumulado y el valor de la renta vitalicia, calculada con los parámetros del RPM-PD, en otras palabras, bajo estas circunstancias no serían necesarios los subsidios para el RPM.

Las investigaciones empíricas están enfatizando en la necesidad de adoptar un nuevo sistema pensional que garantice la viabilidad financiera por encima de todo. En particular, el marco estadístico, actuarial y económico no puede ingnorarse a la hora de evaluar la eficiencia del sistema, como tampoco los criterios utilizados por los diversos investigadores para construir el elemento fundamental para el análisis de la senda salarial. Quizás esta visión de la construcción de la senda salarial arroje resultados más sólidos y recomendaciones de política confiables. Sumado a esto, la metodología para realizar el cálculo de transferencias se presenta con base en procesos de capitalización individual, por lo cual resultaría inapropiado aplicarla al caso de RPM-PD, ya que este por definición, es un proceso de reparto colectivo; la capitalización no está relacionada con cuentas individuales.

Por último, este trabajo de tesis pretende ser un aporte a la aplicación de la teoría actuarial y financiera, en la construcción de sendas salariales, así mismo, procura ser un punto de partida para posteriores trabajos de investigación que realicen análisis sobre otro tipo de poblaciones y contribuir a la discusión del tema pensional en Colombia.

A. Aplicaciones con transformaciones Box-Cox y logarítmica

La aplicación de la tranformación de Box-Cox (1964) tradicional no es la más indicada para el caso de poblaciones con presencia de datos extremos, por lo cual, es necesario usar una transformación de Box-Cox Robusta, ver Castaño (2011), que permita afrontar el problema evidenciado anteriormente en las pruebas realizadas a los datos. En este caso, se presentan los siguientes resultados:

Tabla A-1.: Transformación Box-Cox Robusta

	lambda Robusto	lambda BC	ndatos
Salario	-1.65	-0.275	10653

En síntesis la transformación no produce el efecto esperado de normalización de las variables, y pese a que se presenta una leve mejora en el comportamiento de los datos, estos no se acercan a la normalidad, excluyendo la opción de usar los modelos paramétricos. Una posible explicación está relacionada con la presencia de cluster en el salario, debido a que es uno de los posibles causantes de los valores extremos. Por lo tanto, otra opción para el análisis son las regresiones no paramétricas, específicamente la regresión Loess y el estimador de Nadaraya-Watson, ya que son estimadores con base en regresiones lineales locales y con capacidad de ajustar diversas relaciones no lineales. Sin embargo, se realizó el ejercicio de modelación con los datos transformados para demostrar que no se tienen grandes cambios frente a los modelos sin transformar. A continuación se presentan los resultados de este ejercicio.

A.1. Resultados regresión no lineal con la tranformación B-C

En la sección se presentan los resultados de los modelos con la transformación de Box-Cox Robusta sobre la variable salario. En este caso, se utiliza el argumento 'robusto' para indicar que la variable salario tiene un exponente de -1.65. A continuación se

mostrarán los resultados de los modelos no lineales.

Regresión no lineal, modelo 1

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.2, b = 0.19, g = -0.07$). Luego de la estimación del modelo se obtienen los parámetros ($a = -0.14521, b = -0.05165, g = -0.06447$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 19.27 y Error Residual Estándar o RSE = 0.07369, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-2.: Modelo de Regresión No-lineal, modelo 1= $Y^{-1.65}$

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	-0.145207	0.001443	-100.661	2e-16
β	-0.051654	0.047926	-1.078	0.281
λ	-0.064467	0.080775	-0.798	0.425

Regresion no lineal, modelo 2

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.8, b = 0.60, g = -0.10$). Luego de la estimación del modelo se obtienen los parámetros ($a = -0.14521, b = -0.05165, g = -0.06446$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 19.27 y Error Residual Estándar o RSE = 0.07369, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-3.: Modelo de Regresión No-lineal, modelo 2 = $Y^{-1.65}$

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	-0.145207	0.001443	-100.661	2e-16
β	-0.051650	0.047925	-1.078	0.281
λ	-0.064465	0.080776	-0.788	0.431

Regresion no lineal, modelo 3

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.1, b = 0.08, g = -0.01$). Luego de la estimación del modelo se obtienen los parámetros ($a = -0.14521, b = -0.05165, g = -0.06447$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 19.26 y Error Residual Estándar o RSE = 0.07369, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-4.: Modelo de Regresión No-lineal, modelo 3 = $Y^{-1.65}$

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	-0.1452	0.001443	-100.015	2e-16
β	0.051651	0.047927	-1.078	0.281
λ	-0.68536	0.080781	-0.798	0.425

A.2. Resultados regresión no lineal con la tranformación logarítmica

En la sección se presentan los resultados de los modelos con la transformación logarítmica sobre la variable salario.

Regresión no lineal, modelo 4

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.2, b = 0.19, g = -0.07$). Luego de la estimación del modelo se obtienen los parámetros ($a = 0.186903, b = 0.06731962, g = -0.06996$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 13.72 y Error Residual Estándar o RSE = 0.1439, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-5.: Modelo de Regresión No-lineal, modelo 4= Y^2

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	0.186904	c	0.186904	2e-16
β	0.067320	0.064673	1.041	0.298
λ	-0.069963	0.094307	-0.742	0.458

Regresion no lineal, modelo 5

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.9, b = 0.80, g = -0.25$). Luego de la estimación del modelo se obtienen los parámetros ($a = 0.1869, b = 0.06732, g = -0.06999$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 14.22 y Error Residual Estándar o RSE = 0.1439, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-6.: Modelo de Regresión No-lineal, modelo 5 = Y^2

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	0.186904	0.002019	92.579	2e-16
β	0.067324	0.064674	1.041	0.298
λ	-0.069966	0.094308	-0.742	0.458

Regresion no lineal, modelo 6

Los parámetros iniciales utilizados para el modelo no lineal son valores estándar empleados en el trabajo ($a = 0.09, b = 0.65, g = -0.09$). Luego de la estimación del modelo se obtienen los parámetros ($a = 0.1868, b = 0.06730, g = -0.06992$) y otros resultados del modelo son: la suma de cruadrados residuales RSS = 12.26 y Error Residual Estándar o RSE = 0.1439, con 3548 grados de libertad. En la siguiente tabla se pueden observar los resultados del modelo.

Tabla A-7.: Modelo de Regresión No-lineal, modelo 6 = Y^2

	Estimate	Std. Error	t value	Pr(>|t|)
ξ	0.186904	0.002019	92.579	2e-16
β	0.067306	0.064656	1.041	0.298
λ	-0.06993	0.094259	-0.742	0.458

En resumen, los resultados de los modelos no muestran un cambio notable pues, pese a la transformación de las variables, siguen presentando resultados de baja significancia en los parámetros. Esto da a entender que únicamente la constante (inflación) ejerce un efecto en el aumento salarial de los individuos. Por tanto, se considera que la regresión no lineal con las transformaciones de Box-Cox-robusta y la logarítmica, no son un método viable para realizar los cálculos de la senda salarial con los datos de estudio de esta tesis.

A.3. Resultados regresión lineal múltiple con la tranformación B-C

En la sección se presentan los resultados de los ejercicios con los las transformaciones Box-Cox robusta y la logarítmica. El modelo utilizado es un modelo cuadrático, pues los datos sugieren una relación de este estilo. A continuación se mostrarán los resultados de los modelos de regresión lineal múltiple.

Regresión lineal múltiple, robusta 2008-2011

En la siguiente Tabla A-8 se pueden observar los resultados del modelo 2008-2011.

Tabla A-8.: Modelo de Regresión lineal múltiple, modelo 1= $Y^{-1.65}$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8972	0.0313	28.64	0.0000
x	-0.0015	0.0015	-0.98	0.3263
x2	0.0000	0.0000	0.92	0.3600

Residual standard error: 0.0737 on 3548 degrees of freedom
Multiple R-squared: 0.0003419, Adjusted R-squared: -0.0002216
F-statistic: 0.6068 on 2 and 3548 DF, p-value: 0.5452

Regresión lineal múltiple, robusto 2011-2014
En la siguiente Tabla A-9 se pueden observar los resultados del modelo.

Tabla A-9.: Modelo de Regresión lineal múltiple, modelo 2 = $Y^{-1.65}$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.9596	0.0320	29.94	0.0000
x	-0.0035	0.0015	-2.29	0.0222
x2	0.0000	0.0000	1.92	0.0549

Residual standard error: 0.0754 on 3548 degrees of freedom
Multiple R-squared: 0.003908, Adjusted R-squared: 0.003346
F-statistic: 6.959 on 2 and 3548 DF, p-value: 0.0009629

A.4. Resultados regresión lineal múltiple con la tranformación logarítmica

Regresión lineal múltiple, logarítmico 2008-2011
En la siguiente Tabla A-10 se pueden observar los resultados del modelo.

Tabla A-10.: Modelo de Regresión lineal múltiple, modelo 3 = Y^{2}

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.1387	0.0612	18.62	0.0000
x	0.0030	0.0029	1.01	0.3119
x2	-0.0000	0.0000	-0.92	0.3582

Residual standard error: 0.1439 on 3548 degrees of freedom
Multiple R-squared: 0.0004327, Adjusted R-squared: -0.0001307

F-statistic: 0.7679 on 2 and 3548 DF, p-value: 0.464

Regresión lineal múltiple, logarítmico 2011-2014
En la siguiente Tabla A-11 se pueden observar los resultados del modelo.

Tabla A-11.: Modelo de Regresión lineal múltiple, modelo 4 = Y^2

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.1387	0.0612	18.62	0.0000
x	0.0030	0.0029	1.01	0.3119
x2	-0.0000	0.0000	-0.92	0.3582

Residual standard error: 0.1439 on 3548 degrees of freedom
Multiple R-squared: 0.0004327, Adjusted R-squared: -0.0001307
F-statistic: 0.7679 on 2 and 3548 DF, p-value: 0.464

En síntesis, los modelos evidencian resultados similares en varios parámetros como: primero, los valores de significancia general son muy bajos, lo que sugiere que los modelos no alcanzan a capturar la relación entre las variables. Segundo, se encuentran parámetros significativos como el intercepto y en algunos casos una de las variables, lo que indica que los modelos capturan el aumento por inflación. Por consiguiente, los resultados del modelo lineal múltiple con las transformaciones no logran satisfacer los críterios estadísticos básicos para realizar la estimación de las sendas salariales.

```
# Analisis de modelos con transformaciones

D = read.table(file.choose(), header = T)
D = as.data.frame(D)

S08 =subset(D, D$periodo == 2008,
select = c(cedula,salario,edad))


S11 =subset(D, D$periodo == 2011,
select = c(cedula,salario,edad))


S14 =subset(D, D$periodo == 2014,
select = c(cedula,salario,edad))
```

```
#-------------------------------------
index08 <- with(S08, order(cedula, salario,edad))
S08 = S08[index08, ]

index11 <- with(S11, order(cedula, salario,edad))
S11 = S11[index11, ]

index14 <- with(S14, order(cedula, salario,edad))
S14 = S14[index14, ]

#-------------------------------------

x08 = order(S08$edad,S08$salario)

x = S08$edad[x08]

Yk1 = (S11$salario/S08$salario)^(1/3)
Yk2 = (S14$salario/S11$salario)^(1/3)

################################
####Parametros de transformacion

#####Transformacion Box-Cox-Robusta
yr1 = Yk1^-1.65
yr2 = Yk2^-1.65

####Transformacion logaritmica
yl1 = Yk1^2
yl2 = Yk2^2



#---regresion nolineal------------------------------

#--- regresion nolineal, robusta


b0 = list(a=0.2, b=0.19, g=-0.07)
```

```
modelor = nls(yr1 ~ exp(a + b*exp(g*x08)),
data=data.frame(y=yr1,x=x08),
algorithm = "port",
start = b0,
control=list(maxiter=5550,
tol=0.01, warnOnly=TRUE))

summary(modelor)
print(xtable(print(modelor)))

print((pars = modelor$m$getPars()))


Yk1.est = as.numeric(modelor$m$fitted())

#------------------------ regresion nolineal, logaritmica
b1 = list(a=0.2, b=0.19, g=-0.07)

modelol = nls(yl1 ~ exp(a + b*exp(g*x08)),
data=data.frame(y=ynormal,x=x08),
algorithm = "port",
start = b1,
control=list(maxiter=5550,
tol=0.01, warnOnly=TRUE))

summary(modelol)

(pars = modelol$m$getPars())


#----- regresion lineal t=0 (Ejercicios de la tesis reg lineal, robusta)
x2 = x*x

m.reg08r = lm( yr1 ~ x + x2)
summary(m.reg08r)
freg08r = exp(m.reg08r$fitted)

print(xtable(m.reg08r))

r08 = residuals(m.reg08r)
```

```
t = seq(1,length(r08))

par(mfrow=c(2,2))
plot(t,r08,type='o',ylab='residuo',col='gray')
abline(h=0,lty=2)
plot(density(r08),xlab='x',main= '')
qqnorm(r08)
qqline(r08,col=2)

g.reg08r=glm(formula = yr1 ~  x + x2, family = Gamma)
summary(g.reg08r)

greg08r = exp(g.reg08$fitted)

plot(x,yr1,type='p')
lines(x,freg08r,col='red',lwd=2)
lines(x,g.reg08r,col='blue',lwd=2)

#----- regresion lineal t=0 (Ejercicios de la tesis reg lineal, logaritmica)

m.reg08n = lm( yl1 ~ x + x2)
summary(m.reg08n)
freg08n = exp(m.reg08n$fitted)

print(xtable(m.reg08n))

r08 = residuals(m.reg08n)
t = seq(1,length(r08))

par(mfrow=c(2,2))
plot(t,r08,type='o',ylab='residuo',col='gray')
abline(h=0,lty=2)
plot(density(r08),xlab='x',main= '')
qqnorm(r08)
qqline(r08,col=2)

g.reg08n=glm(formula = yl1 ~  x + x2, family = Gamma)
summary(g.reg08n)

greg08n = exp(g.reg08$fitted)
```

```
plot(x,yl1,type='p')
lines(x,freg08n,col='red',lwd=2)

lines(x,greg08n,col='blue',lwd=2)
```

B. Estimación de modelos no paramétricos

```
## Estimacion Regresion Loess #########



#--------------regresion loess salarios vs edad

x08 = order(edad08,D$S08)
x = edad08[x08]
x08 = seq(20,64)
#--------------------suavizados por edad


m.lo08=loess(D$S08 ~ x,span=0.75, family="symmetric",
             control = loess.control(surface = "direct"))

flo08 = m.lo08$fitted

m.lo11=loess(D$S11 ~ x,span=0.75, family="symmetric",
             control = loess.control(surface = "direct"))

flo11 = m.lo11$fitted

m.lo14=loess(D$S14 ~ x,span=0.75, family="symmetric",
             control = loess.control(surface = "direct"))

flo14 = m.lo14$fitted

#--------------------suavizados por cada edad diferente

s11 = predict(m.lo11,data.frame(x=x08))
s08 = predict(m.lo08,data.frame(x=x08))
```

```
s14 = predict(m.lo14,data.frame(x=x08))

#-------------graficas salario suavizado versus datos

par(mfrow=c(2,2))

plot(x,D$S08,type='p',col='darkgray')
lines(x,flo08,lwd=3,lty=4,col='red')

plot(x,D$S11,type='p',col='darkgray')
lines(x,flo11,lwd=3,lty=4,col='red')

plot(x,D$S14,type='p',col='darkgray')
lines(x,flo14,lwd=3,lty=4,col='red')

plot(x,flo08,type='b',lwd=3,col='yellow', ylim=c(1200000,3200000))
lines(x,flo11,lwd=3,lty=4,col='red')
lines(x,flo14,lwd=3,lty=6,col='blue')

#legend("topright",  c("2008","2011","2014"),
#col=c(7,2,4), lty = c(3,4,6),lwd=c(3,3,3), inset = .01)



#-------------estimar los incrementos por edad

Yk1 = (s11/s08)^(1/3)-1

Yk2 = (s14/s11)^(1/3)-1

par(mfrow=c(1,2))


plot(x08,Yk1,type='l',lty=2,lwd=3,ylim=c(0.0,0.2))
lines(x08,Yk2,lwd=3,lty=4,col='red')

#--------------sendas salariales
n = length(x08)
ASx08 = double(n)
ASx08[1] = (1+Yk1[1])*1.0e05
```

```
for(j in 2:n){
        ASx08[j] = (1+Yk1[j-1])*ASx08[j-1]}

ASx11 = double(n)
ASx11[1] = (1+Yk2[1])*1.0e05
for(j in 2:n){
        ASx11[j] = (1+Yk2[j-1])*ASx11[j-1]}

#--------------graficas sendas salariales 2008-2011 y 2011-2014
plot(x08,ASx08,type='l',lty=2,lwd=3)
lines(x08,ASx11,lwd=3,lty=4,col='red')

legend("topleft",  c("2008-2011","2011-2014"),
       col=c(1,2), lty = c(2,4),lwd=c(3,3))


#################################################################


###### Estimacion del parametro h  ################################


## Primer metodo#################################################

h3 <- h.select(x08, Yk2, method = "cv")
h3
fit3 <- locpoly(x08,Yk2,bandwidth = h3, degree=0)
lines(fit3,col="blue",lty=4,lwd=2)

h4<-h.select(x08, Yk1,method = "cv")
h4
fit4 <- locpoly(x08,Yk1,bandwidth = h4, degree=0)
lines(fit4,col="blue",lty=4,lwd=2)


## Segundo metodo################################################

h.cv <- hcv(x, display="lines", ngrid=32)
sm.density(x, h=hcv(x))
```

```
h.cv <- hcv(x, Yk1, display="lines", ngrid=32)
sm.regression(x, Yk1, h=hcv(x, Yk1))


## Tercer metodo#####################################################


bw <- npregbw(Yk1~x08, regtype="ll", bwmethod="cv.aic")
bw
npplot(bws=bw, plot.errors.method="asymptotic")
points(x08, Yk1, cex=.2, col="red")


#################################################################
#################################################################

## Estimacion Nadaraya-Watson ####################################

Yx08.nw = ksmooth(x08, Yk1,
                  kernel="normal",
                  bandwidth=0.89)$y

Yx11.nw = ksmooth(x08, Yk2,
                  kernel="normal",
                  bandwidth=0.89)$y


#------------------------ graficas

par(mfrow=c(1,2))
plot(x08,Yk1,type='p',col='darkgray',
     ylab = "incremento anual 2008-2011")
lines(x,Yx08.nw,lty=3,lwd=3)

plot(x08,Yk2,type='p',col='darkgray',
     ylab = "incremento anual 2011-2014")
lines(x,Yx11.nw,lty=3,lwd=3)

#----------------------senda salarial
```

```
ASx08.nw = double(n)
ASx08.nw[1] = 1.0e05
for(j in 2:n){
        ASx08.nw[j] = Yx08.nw[j-1]*ASx08.nw[j-1]}


ASx11.nw = double(n)
ASx11.nw[1] = 1.0e05
for(j in 2:n){
        ASx11.nw[j] = Yx11.nw[j-1]*ASx11.nw[j-1]}


plot(x08,ASx08.nw,type='l',lwd=2,lty=1,
     ylab="Senda Salarial",xlab="Edad",ylim=c(1.0e05,7.0e06))
lines(x08,ASx11,lwd=2,lty=3)
lines(x08,ASx08,lwd=2,lty=4)
lines(x08,ASx08.nw,lwd=2,lty=6)
lines(x08,ASx11.nw,lwd=5,lty=8)

legend("topleft",  c("2011-2014-Lo","2008-2011-Lo",
                      "2008-2011-nw","2011-2014-nw"),
       lty = c(3,4,6,8),lwd=c(2,2,2,5))
```

C. Estimación de las sendas salariales bajo la metodología de Montenegro et al. (2013)

```
#cargue de los datos y modificaciones
S08 =subset(D, D$periodo == 2008,
select = c(cedula,salario,edad))

S11 =subset(D, D$periodo == 2011,
select = c(cedula,salario,edad))

S14 =subset(D, D$periodo == 2014,
select = c(cedula,salario,edad))

#-------------------------------------

index08 <- with(S08, order(cedula, salario,edad))
S08 = S08[index08, ]

index11 <- with(S11, order(cedula, salario,edad))
S11 = S11[index11, ]

index14 <- with(S14, order(cedula, salario,edad))
S14 = S14[index14, ]
#-------------------------------------
x08 = order(S0$edad,S08salario)
# analisis del modelo en
ASx11 = S11salario
ASx08 = S08salario
ASx14 = S14salario
x = S08edad[x08]
edad = as.numeric(levels(as.factor(x)))
```

```
Sx = mat.or.vec(length(edad),6)
for(k in 1:nrow(Sx)){
xTg = ASx14[edad==edad[k]]
xIg = quantile (xTg,
c(0,.5, .6, .7, .8, .9,1.0),
include.lowest = TRUE)
Sx[k,1] = (1/floor(0.1*length(xTg))/5)*
sum(xTg[which(xTg >= xIg[1] & xTg < xIg[2])])
for(j in 2:6){
Sx[k,j] = (1/floor(0.1*length(xTg)))*
sum(xTg[which(xTg >= xIg[j] & xTg < xIg[j+1])])
}}
#las sendas salariariales se suavizan con el metodo ewma
require(qcc)
par(mfrow=c(1,1))
plot(ewmaSmooth(edad,Sx[,1],lambda=0.1),type="l"
,lty=1,lwd=2,ylim=c(5.0e+04,10.0e+06),
ylab="salario en millones",
xlab="edad")
for(j in 2:6){
lines(edad,Sx[,j],lwd="2",col="j")}
lines(ewmaSmooth(edad,Sx[,j],lambda=0.1),lty=j,lwd=2)
}
legend("topright", c("0.1-0.5",
"0.5-0.6","0.6-0.7","0.7-0.8","0.8-0.9",
"0.9-1.0"),lty = seq(1,6),lwd=rep(2,6))
```

D. Cálculos de la renta vitalicia

```
# supervivencia GM

tpx = function(s,g,c,x,t){
tpx = s^t*g^(c^x*(c^t-1))
return(tpx)
}

# funcion para la anualidad salario */

saxn.u = function(e,r,i,id){
if( id > i ) stop("i debe ser mayor que id")
t = seq(0,r-e)
vs = ((1+id)/(1+i))^(t)
b = ls(e+t)/ls(e);
kpx = (110-e-t)/(110-e)
a = sum(vs*b*kpx)
return(a)}

aan = function(n,i){
v = 1/(1+i)
p = (1-v^n)/(1-v)
return(p)
}

#--Programa para valor de una renta vitalicia
# con base en Gompertz-Makeham
aax = function(s,g,C,x,i){
        v = 1/(1+i)
        k = seq(0,110-x-1,1)
        kpx = tpx(s,g,C,x,k)
        vk = (1+i)^(-k)
        p = sum(vk*kpx)
```

```
        return(p)
}
geo.avx = function(s,g,C,x,m,i,id){
 if (i + id >= 1+i) stop("tasa inflacion mayor que tasa capitalizacion")
v = 1/(1+i)
d = 1-v
ir = (1+i)/(1+id)-1
dr = 1-1/(1+ir)
aa.ir = aax(s,g,C,x,ir)

dm = 1-v^(1/m)
im = v^(-1/m)-1

f1 = d/dm - d + dr*(1/dm - 1)*(d/(m*im)-v)*(1+ir)
f2 = (1+ir)*(1/dm-1)*(d/(m*im)-v)

aax = f1*aa.ir-f2
return(aax)
}

biss = function(k){
# k en periodos anuales produce tasa reemplazo
n=52*k
b=(0.25+0.0004*n)*ifelse(n <= 1200,1,0)+
(0.01+0.0006*n)*ifelse(n > 1200,1,0)*
ifelse(n <= 1400,1,0)+
0.85*ifelse(n > 1400,1,0)
return(b)
}

ibl = function(AS,e,r,id,n){
# requiere salarios AS(x-10),...,AS(x-1)
edad = seq(e,r)
er = which(edad==r)
t = seq(0,n-1)
s = sum((1+id)^t*AS[(er-(n-1)):er])/n
return(s)
}

Br = function(e,r,AS,id,n){
```

```
br = biss(r-e)*ibl(AS,e,r,id,n)
return(br)
}

# ejemplos calculos de subsidios
# utilizando sendas salariales

source("funciones.mesada.r")

e=20
r=62
id=0.04
ir=0.04
i=(1+id)*(1+ir)-1
i12 = (1+i)^(1/12)-1
v=1/(1+i)
se=1e+05
alfa = 0.115

#----------generar la senda salarial
# ASx08, ASx11, ASx08.lo, ASx11.lo

AS = ASx08.nw


plot(seq(e,r),AS,type='b')

#----------generar IBL, Br y costo renta vitalicia

ibl20 = ibl(AS,e,r,id,10)
(Br20 = Br(e,r,AS,id,10))


C1=1.150402; s1=0.9790224; g1=0.9999980;
C2=1.136262; s2=0.9803407; g2=0.9999916;
C3=1.182592; s3=0.9784421; g3=0.9999999;
C4=1.197560; s4=0.9841095; g4=1.0000000;
# mujeres-80-89:
a = 4.450e-03
b = 2.814e-05
```

```
C = 1.101e+00
# hombres-80-89:
a = 4.918e-03
b = 1.559e-05
C = 1.107e+00
# mujeres-00-08:
a = 4.444e-03
b = 5.017e-07
C = 1.143e+00

#-----------------------------
# hombres-00-08:
a = 7.121e-03
b = 2.504e-06
C = 1.126e+00

s = exp(-a)
g = exp(-b/log(C))

(A = geo.avx(s,g,C,x=64,m=12,i,id))
(B = geo.avx(s,g,C,x=64,m=2,i,id))
(costo20 = Br20*(A+B))/1.0e06


#-----------------------------

# mujeres-00-08:
a = 4.444e-03
b = 5.017e-07
C = 1.143e+00


s = exp(-a)
g = exp(-b/log(C))

(A = geo.avx(s,g,C,x=64,m=12,i,id))
(B = geo.avx(s,g,C,x=64,m=2,i,id))
(costo20 = Br20*(A+B))/1.0e06
```

Bibliografía

Anderson, A. W. (1992). Pension Mathematics for Actuaries. ACTEX Publications, Winsted, Connecticut.

Basu, R.;Ullah, A. (1992) Chinese earnings-age profile: a nonparametric analysis. *The Journal of the Royal Statistical Society, Series B*, Vol. 26, pp. 201-252.

Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations. *Journal of International Trade and Economic Development*,1, 151-165.

Carriere, J.; Shand, J. (1998). New salary functions for pension valuations. *North-American Actuarial Journal*, 2(3), 18-27.

Caro, L. G. (2006). Payout Phase for Fully-Funded Pension Schemes: An Alternative Approach to Risk Distribution.
http://www.actuaries.org/EVENTS/Congresses/Paris/Papers/3067.pdf.

Castaño, E. (2011). Una estimación no paramétrica y robusta de la transformación Box-Cox para el modelo de regresión.*Lecturas de Economía*, No. 75 (julio-diciembre 2011), pp. 89-106.

Cleveland, W.S.(1979). Robust locally-weighted regression and smoothing scatterplots.*Journal of the American Statistical Association*, Vol. 74, 829-836.

Gasser, T. and Müller, H. G. (1979). Kernel estimation of regressions functions. *In Smoothing techniques for curve estimation (pp. 23-68*,Springer Berlin Heidelberg.

Härdle, W., Klinde, S. and Müller, M. (2012). XploRe-Learning Guide: Learning Guide,Springer Science & Business Media

Lee S. J., Cox, D. D. (2008). Robust Smoothing: Smoothing Parameter Selection and Applications to Fluorescence Spectroscopy. *Carnegie Mellon University, Department of Statistics*. Follow this and additional works at: http://repository.cmu.edu/statistics

Marples, W.F. (1962). Salary Scales.*Trans. Soc. Actuaries*, vol XIV, Part I, 2-30.

Marron, J.S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statistics & Probability Letters*, 7(3):195-199.

Maplesoft (2003) Maple v.2016.1. Waterloo Maple Inc. Waterloo On, Canada.

Mincer, J. (1974). Schooling, Experience and Earnings, National Bureau of Economic Research, New York.

Ministerio de Hacienda y Crédito Público (2015) Resolución 3099 de 19 de Agosto de 2015 y Anexo. Por la cual, se determina las fórmulas para establecer el saldo de una pensión.... Min Hacienda y Crédito Público. República de Colombia.

Ministerio de Hacienda y Crédito Público (1994) Decreto 2779 de 1994. Por el cual se adopta la tabla de salarios medios nacionales. Ministerio de Hacienda y Crédito Público. República de Colombia.

Montenegro, S., Jiménez, L. and Hurtado, C. (2013). Los subsidios pensionales en el régimen de reparto colombiano: reformas paramétricas para focalizar correctamente el gasto social del estado *Documentos CEDE N. 30*, Universidad de los Andes, Departamento de Economía.

Nadaraya, E. A., (1964). Some new estimates for distributions functions. Theory Probab. Appl. 15, 497-500.

Parra, J. M., (2011). Estimación de un modelo aditivo no paramétrico. Universidad de Granada, Facultad de Ciencias, Departamento de estadística e investigación operativa. http://masteres.ugr.es/moea/pages/tfm1011/estimacionnoparametricadeunmodeloaditivo/.

Racine JS, Li Q., (2004). Nonparamettric Estimation of Regresion Functions with Both Categorical and Continuous Data. *Journal of Econometrics*, 119(1), 99-130.

Racine JS, Li Q, Zhu X (2004). Kernel Estimation of Multivariate Conditional Distributions. *Annals of Economics and Finance*, 5(2), 211-235.

Superintendencia Financiera de Colombia (2010). Por lo cual se actualizan las Tablas de Mortalidad de Rentistas Hombres y Mujeres. RESOLUCIÓN NÚMERO 1555 DE 2010 (Julio 30).

Watson, G.S., (1964). Smooth regression analysis. Sankhya Ser. A 26, 359-372.