



UNIVERSIDAD NACIONAL DE COLOMBIA

**Dinámica del empleo informal en Colombia: Una
aproximación desde cadenas de Markov y funciones de riesgo.
Primer trimestre de 2010.**

Benjamín Eduardo Venegas Venegas

Universidad Nacional de Colombia
Facultad de Ciencias Económicas
Bogotá, Colombia
2011

**Dinámica del empleo informal en Colombia: Una
aproximación desde cadenas de Markov y funciones de riesgo.
Primer trimestre de 2010.**

Benjamín Eduardo Venegas Venegas

Tesis presentada como requisito parcial para optar el título de:
Magister en Ciencias Económicas

Director:
Gustavo Adolfo Junca

Universidad Nacional de Colombia
Facultad de Ciencias Económicas
Bogotá, Colombia
2011

Agradecimientos

Agradezco a todas las personas que de una u otra manera colaboraron en la realización de esta tesis, especialmente agradezco la valiosa colaboración de mis buenos amigos Amanda, Carolina, Nury y César, a mi director Gustavo Junca y a todo el equipo de trabajo de Econometría Consultores S.A.

Resumen

El presente documento presenta un análisis de la duración del empleo informal en Colombia a partir de la Gran Encuesta Integrada de Hogares sustentado en un análisis de funciones de riesgo y cadenas de Markov. Como principal resultado se destaca el hecho de que la informalidad en Colombia ha sido un fenómeno estable a través del tiempo y se espera que se consolide al correr de los años, además que son los más jóvenes y educados los que menos duran en la informalidad y que en términos de condiciones económicas son los empleados informales los que menos ICV presentan comparativamente. Se concluyó que existen evidencia estadística para argumentar que el componente espacial es un factor trascendental a hora de explicar los determinantes de la informalidad.

Palabras Clave: Funciones de Riesgo, Cadenas de Markov, Matrices Ergódicas, Weibull, Modelos paramétricos, econometría espacial, ICV

Abstract

This paper presents an analysis of the duration of informal employment in Colombia from the Integrated Household Survey (Gran Encuesta Integrada de Hogares) underpinned by a risk function and Markov chains analysis. The main result highlights the fact that informality in Colombia has been a stable phenomenon over time and is expected to consolidate over the years. Furthermore the young and the educated last shorter periods in the informal sector. Additionally, the informal sector workers have the lowest quality of life index comparatively. It is concluded that there is statistical evidence to argue that the spatial component is a significant factor in explaining the determinants of informality.

Word Keys: Hazard Functions, Markov Chains, Matrix ergodic, Weibull, parametric models, espacial econometric, ICV

Contenido

INTRODUCCION	9
1. ESTADO DEL ARTE	11
2. MARCO TEÓRICO	12
2.1. Cadenas de Markov	12
2.1.1. Cadenas de Markov Ergódicas	13
2.2. Funciones de riesgo	14
2.3. Distribuciones	16
3. MARCO METODOLÓGICO	18
3.1. Definición empleo informal	18
3.2. Gran Encuesta Integrada de Hogares (GEIH)	20
3.2.1. Características encuesta	21
4. EVOLUCIÓN INFORMALIDAD	21
5. CARACTERIZACIÓN SOCIOECONOMICA	23
6. ESTIMACIÓN	26
6.1. Cadena de Markov	26
6.2. Funciones de Riesgo	27
6.2.1. Estimación no paramétrica	27
6.2.2. Estimación paramétrica	32
6.3. Análisis Espacial	33
A. Tests y regresiones	39
Bibliografía	54

Lista de figuras

1.	Objetivo del uso de Cadenas de Markov	12
2.	Estados de transición	13
3.	Aproximación de la Función de Riesgo	15
4.	Distribución de probabilidad Exponencial	16
5.	Distribución de probabilidad Weibull	17
6.	Distribución de probabilidad Gompertz	18
7.	Distribución de probabilidad Log-Normal	18
8.	Definiciones de Informalidad Fuente: Presentación Banco de la República. Hugo Lopez 2011	19
9.	Evolución informalidad Fuente: DANE, cálculos de la XXI Reunión de presidentes de organizaciones empresariales Iberoamericanas	22
10.	Distribución ocupados informales antes	22
11.	Distribución ocupados informales ahora	22
12.	Distribución poblacional informales	23
13.	ICV: Puntajes mínimos aceptables.	24
14.	ICV empleados informales, formales y desempleados	25
15.	Ciclo laboral	26
16.	Matriz de probabilidades	27
17.	Vector de probabilidades de largo plazo	27
18.	Estimador Kaplan-Meier (función de supervivencia)	28
19.	Estimador Kaplan-Meier (función de supervivencia) por género	29
20.	Estimador Kaplan-Meier (función de supervivencia) por edades	29
21.	Estimador Kaplan-Meier (función de supervivencia) por nivel educativo	30
22.	Estimador Kaplan-Meier (función de supervivencia) por división administrativa municipal	30
23.	Estimador Kaplan-Meier (función de supervivencia) por estado civil	31
24.	Criterios de información ajuste modelo	32
25.	Análisis paramétrico: Distribución Weibull	32
26.	Duración promedio informalidad para algunos municipios de Colombia	34
27.	Regresión espacial	35
28.	ICV: Puntajes componentes	39
29.	ICV: Puntajes componentes	40
30.	ICV: Puntajes componentes	41
31.	Test de igualdad funciones supervivencia por sexo	42
32.	Test de igualdad funciones supervivencia por edad	42
33.	Test de igualdad funciones supervivencia por nivel educativo	43
34.	Test de igualdad funciones supervivencia por estado civil	43
35.	Análisis paramétrico: Distribución exponencial	44
36.	Análisis paramétrico: Distribución Gompertz	45
37.	Análisis paramétrico: Distribución Log-Logistic	46
38.	Análisis paramétrico: Distribución Log-Normal	47
39.	Relaciones espaciales: Matriz W	48
40.	Áreas planificación estratégica Plan Nacional de Desarrollo	49
41.	Tabla duración informalidad	50
42.	Tabla duración informalidad-1	51

43.	Tabla duración informalidad-2	52
44.	Tabla duración informalidad-3	53

INTRODUCCIÓN

Desde la década del noventa en Colombia y en otros países de América Latina se ha desarrollado una serie de reformas legislativas (con diversos resultados) que han ocasionado un conjunto de impactos sociales en términos de las condiciones laborales que debe afrontar la población.

El mayor grado de privatización del sistema de seguridad social más la mayor flexibilidad en términos de la legislación laboral ha generado que en Colombia se desarrolle una dinámica laboral adversa y prolongada en el tiempo, que se caracteriza por generar incentivos para no realizar contrataciones bajo un sistema de seguridad integral ¹ y/o por las relaciones contractuales de corto plazo que para ciertos sectores de la población no resulta tan difícil de alcanzar.

Bajo estas condiciones, las únicas alternativas para aquellos que no pueden acceder a algún tipo de empleo de características ideales son por una lado la de ejercer actividades de baja productividad o por el otro crear empresa de baja capacidad productiva con lo que se pueden generar situaciones en las cuales se cree un círculo vicioso en términos de contrataciones y/o de obligaciones contractuales. Es así que, el trabajo informal (independiente de la definición utilizada) se convierte tanto en escape como en exclusión (Banco Mundial 2007). Escape debido tanto a las características propias de los individuos como a las condiciones económicas y exclusión gracias al conjunto de instituciones que regulan las condiciones laborales.

Es por eso que surgen las siguientes preguntas: ¿Cuál es la duración del empleo informal en Colombia? y más aún ¿Cuáles son los factores que hacen que un individuo dure más que otro en el sector informal? Dada las características actuales de la economía Colombiana, ¿Cuál será la dinámica de largo plazo de este sector?

Para tal fin, y teniendo en cuenta que en Colombia la mayoría de trabajos de duración se han centrado en el desempleo, el presente trabajo busca entender la dinámica de la duración del empleo informal en Colombia a partir de información de la Gran Encuesta Integrada de Hogares (GEIH) tomando como punto de referencia la definición planteada por el Departamento Nacional de Estadística (DANE) y mediante la utilización de herramientas tanto econométricas (análisis de funciones de riesgo (*Hazard Functions*) y econometría espacial) como estadísticas (cadenas de Markov).

Dentro de los principales resultados se encuentran: La dinámica de la informalidad es un fenómeno que se ha mantenido y se espera se mantenga en el tiempo; se ha desarrollado una recomposición de la estructura de la informalidad en Colombia que ha tendido a desarrollar actividades cuenta propia como alternativa de salida; son los empleados informales los que menor ICV presentan comparativamente, además existe una relación negativa entre la duración de la informalidad y el desarrollo económico municipal y son las personas más jóvenes y más educadas las que menos duran en la informalidad.

El documento se encuentra estructurado en 8 partes donde la introducción es la primera. En la segunda se aborda el estado del arte en torno a los estudios de duración. En la tercera, se desarrolla toda la estructura teórica utilizada para las estimaciones tanto del análisis de duración como de las cadenas de Markov. En la cuarta, se detallan los aspectos relacionados al marco metodológico: fuentes de información y aspectos relacionados a la definición de informalidad utilizada. En la quinta parte, se desarrolla una breve presentación de la dinámica de la informalidad durante los últimos

¹Se entiende por sistema de seguridad integral aquel que implica cotización al sistema de seguridad social en términos de salud y pensión

años. En la sexta, se estructura la caracterización socio-económica de los estados empleado informal, empleados formal y desempleados para finalizar con los principales resultados de la estimación en donde se agrega un numeral adicional para llevar a cabo un breve análisis de econometría espacial y con las conclusiones.

1. ESTADO DEL ARTE

Aunque durante los últimos años la producción investigativa sobre el mercado laboral ha abordado muchos frentes de análisis, este se ocupará específicamente de aquellos trabajos que hayan centrado su atención sobre los determinantes de la duración del empleo y específicamente del empleo informal. De todas maneras y pese al objetivo planteado, la mayoría de los trabajos realizados hasta el momento se han ocupado en especificar los determinantes, duración, estimación y análisis del desempleo.

En el año de 1986 y gracias a la misión Chenery ² se llevó a cabo en Colombia la primera estimación de la duración del desempleo. Posteriormente surgen una serie de trabajos que hacen una crítica a dicha estimación.

Lopez (1994) realiza el cálculo de la probabilidad mensual de salir del desempleo a través de la estimación de un modelo paramétrico con una especificación de distribución Weibull; Nuñez y Bernal (1997) llevan a cabo, entre otras cosas, una aproximación a partir de Funciones de Riesgo de los determinantes de la duración del desempleo (cálculo del estimador Kaplan-Meier, modelo paramétrico con distribución Weibull y modelo semiparamétrico). Dentro de las variables utilizadas en dicho estudio se encuentran: edad, rama de actividad económica, localización geográfica (se utilizan los principales cascos urbanos), niveles de educación, entre otras.

Tenjo y Ribero (1998) realizan una contrastación empírica a través de la especificación de un modelo Probit y de un modelo Weibull para hallar en su orden la probabilidad de estar desempleado y la duración del desempleo. Florez (2002), Uribe, Ortiz y Correa (2004) u Uribe, Castro y Ortiz (2004) realizan estudios de segmentación del mercado laboral (empleo formal e informal) a partir de modelos probit, multinomial y de ecuaciones mincerianas respectivamente. Castellar y Uribe (2003) también llevan a cabo un modelo econométrico para hallar los determinantes de la duración del desempleo pero para el área metropolitana de Cali; en ese mismo año Martinez realiza una estimación, con las mismas variables utilizadas por Nuñez y Bernal, no solo de la duración del desempleo sino del empleo a través de modelos de especificación paramétrica (Exponencial, Weibull, etc) como no paramétrica. Mayorga (2005) da cuenta de los determinantes y dinámica del mercado laboral por rama de actividad económica. Para ello y siguiendo a Tenjo y Ribero (1998) calculó un modelo probit para estimar la probabilidad por sector de estar empleado, adicionalmente realizó el cálculo de matrices de transición tanto en el corto plazo como en el largo plazo. Para terminar, Viáfara y Uribe (2006) realizan estimaciones de la duración del desempleo y el efecto sobre éste de los canales de búsqueda utilizados, Sánchez, Duque y Ruiz (2009) analizan la duración del desempleo a la luz del aumento de los costos laborales y Sánchez (2011) realiza un análisis de la duración del desempleo en Bucaramanga 2006.

Como podrá observarse, la mayoría de trabajos presentados realizan los análisis teniendo como base estructural el desempleo y en algunos de ellos el empleo informal, pero ninguno de estos últimos se centran a fondo en cuestiones relacionadas con las *dinámicas y duración del empleo informal*.

Es por esto que para lograr una comprensión adecuada de esta problemática, se recurre a herramientas que permitan responder a las preguntas que motivan la presente investigación y son a saber: ¿Cuál es la duración y determinantes del empleo informal en Colombia? y dadas las características actuales ¿Cuál es su dinámica de largo plazo?

²Ver Ocampo y Ramírez (1986)

Entender la dinámica de largo plazo implica un conocimiento detallado de la distribución de la población en los distintos segmentos laborales, esto es dimensionar el tamaño del sector informal en un punto en el cual dicha situación se estabilice en el tiempo. Para tal fin, se recurre a una herramienta estadística-matemática denominada Cadenas de Markov.

Por su parte, los análisis de duración se llevan a cabo mediante el uso de una herramienta denominada Funciones de Riesgo que además permiten, en su versión paramétrica y bajo un determinado tipo de distribución de probabilidad, la identificación de las variables que inciden en la probabilidad de salida de un determinado estado ³. Este tipo de herramientas son de uso estándar en los modelos de duración del desempleo.

2. MARCO TEÓRICO

2.1. Cadenas de Markov

Como se mencionó brevemente, el objetivo del uso de las Cadenas de Markov como herramienta de análisis se sustenta en la necesidad de entender cual será el comportamiento esperado-probabilidad estable o de largo plazo- (a partir de las condiciones actuales) de los estados laborales que se definen en la presente investigación: empleado formal, empleado informal y desempleado.

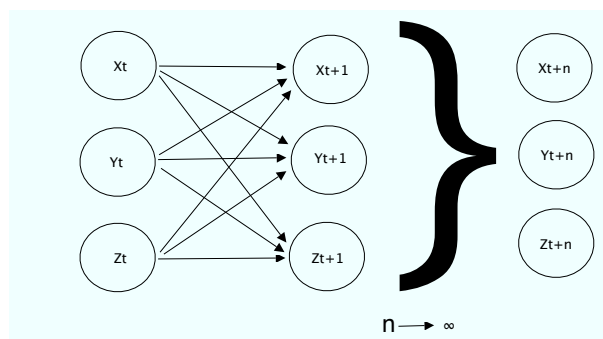


Figura 1: Objetivo del uso de Cadenas de Markov

La gráfica 1 resume de manera adecuada los resultados del uso del procedimiento estadístico: todos los estados se relacionan entre sí de manera que la probabilidad de movilizarse de un estado a otro es independiente del pasado, la probabilidad de pertenecer a alguno de los estados en mayor que cero y existe un vector de largo plazo a partir del cual las probabilidades de tránsito entre estados se estabiliza en el tiempo. Formalmente:

Una secuencia de variables aleatorias (X_0, X_1, \dots) es una cadena de Markov con estado espacios Ω y matriz de transición P si para todo $x, y \in \Omega$, todo $t \geq 1$, y todos los eventos $H_{t-1} = \cap_{s=0}^{t-1} \{X_s = x_s\}$ satisfaciendo $P(H_{t-1} \cap \{X_t = x\}) > 0$ se tiene que:

$$P\{X_{t+1} = y | H_{t-1} \cap \{X_t = x\}\} = P\{X_{t+1} = y | X_t = x\} = P(x, y)^4 \quad (1)$$

³Con salida se entiende el tránsito de un estado a cualquier otro definido.

⁴También llamada propiedad de Markov

Es decir, la probabilidad condicional de proceder del estado x al estado y es la misma, sin importar cual es la secuencia de estados precedentes x_0, x_1, \dots, x_{t-1} al estado actual x . Esto es exactamente porque la matriz de probabilidades $P|\Omega|x|\Omega|$ basta para describir las transiciones. La x -th columna o fila de P es la probabilidad fija de trasladarse al estado siguiente $P(x, \cdot)$. Así, P es estocástica y todas las entradas son no negativas y

$$\sum_{y \in \Omega} P(x, y) = 1 \quad (2)$$

para todo x que pertenezca a Ω .

Ahora bien, considerese el caso representado en la figura 1 en donde existen tres estados de transición

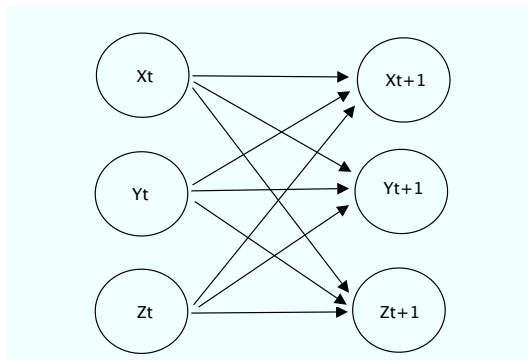


Figura 2: Estados de transición

y en donde la matriz de probabilidades de transición viene dada por:

$$P = \begin{Bmatrix} P(X_t, X_{t+1}) & P(X_t, Y_{t+1}) & P(X_t, Z_{t+1}) \\ P(Y_t, X_{t+1}) & P(Y_t, Y_{t+1}) & P(Y_t, Z_{t+1}) \\ P(Z_t, X_{t+1}) & P(Z_t, Y_{t+1}) & P(Z_t, Z_{t+1}) \end{Bmatrix} = \begin{Bmatrix} p & q & r \\ t & u & v \\ 1-p-t & 1-q-u & 1-r-v \end{Bmatrix}$$

Es precisamente a partir de esta representación que se puede obtener el estado a partir del cual las probabilidades ya no dependen del tiempo, es decir existe algún valor de convergencia P_{ij}^n a partir del cual hay independencia del estado inicial.

Sin embargo, el alcance del análisis del presente se va a restringir a aquellas cadenas de Markov que se caracterizan por tener una recurrencia positiva y aperiódica.

2.1.1. Cadenas de Markov Ergódicas

Una Cadena de Markov se dirá ergódica si existe un t tal que para todo $x, y \in \Omega$, $P^t(x, y) > 0$. Para cadenas de Markov finitas las siguientes ⁵ par de condiciones son equivalentes a ergodicidad:

- Irreducible: Para todo $x, y \in \Omega$, existe $t = t(x, y)$ tal que $P^t(x, y) > 0$.
- Aperiódica: Para todo $x \in \Omega$, el divisor común más grande $\{t : P(x, y) > 0\} = 1$.

⁵La finitud responde a la condición de no nulidad.

En otras palabras, una cadena de Markov es ergódica si en algún momento del tiempo no periodico pero recurrente toma el valor de alguno de sus Ω estado espacios y si todos ⁶ ⁷ sus estados se comunican (cada estado puede ser alcanzado por todos los demás).

La principal propiedad de las cadenas de Markov Ergódicas viene dada por:

Para una cadena de Markov Ergódica existe una única distribución π tal que

$$\text{para todo } x, y \in \Omega \quad \lim_{x \rightarrow \infty} P^t(x, y) = \pi(y) \quad (3)$$

Las probabilidades límites π se denominan probabilidad de estado estable. Adicionalmente, las probabilidades de estado estable $\pi(y)$ satisfacen las ecuaciones de estado estable

$$\pi(y) = \sum_{x=0}^M \pi(x)P(x, y) \quad (4)$$

donde

$$\sum_{x=0}^M \pi(x) = 1 \quad (5)$$

Son en total $M + 1$ ecuaciones con M incógnitas. Una de las primeras $M + 1$ ecuaciones depende de las otras.

2.2. Funciones de riesgo

Las función de riesgo o *Hazard Functions* son un método estadístico que se encuentran dentro de los análisis de supervivencia que buscan explicar la distribución del tiempo hasta la ocurrencia de algún evento específico o *riesgo*. En otros términos, la duración del evento no es el objetivo último de este tipo de análisis sino la probabilidad de salida de éste sujeta a la duración acumulada del mismo. Para el caso particular, las funciones de riesgo van a permitir explicar las probabilidades de salida, tiempo de duración, del empleado informal.

Sea $F(t)$ la función de distribución que especifica la probabilidad de distribución de la duración,

$$F(t) = P(T < t) \quad (6)$$

la cual especifica la probabilidad de que la variable aleatoria T se realice con una duración menor a t . La correspondiente función de densidad es

$$f(t) = \frac{dF(t)}{dt} \quad (7)$$

que es la probabilidad de realización en cada momento del tiempo. Ahora bien, a partir de éstas definiciones se puede obtener la de la función de supervivencia como aquella que caracteriza la probabilidad de realización del evento aleatorio con una duración mayor que t :

$$S(t) = 1 - F(t) \quad (8)$$

⁶Un estado j es periodico con periodo t si es posible un retorno sólo en los pasos $t, 2t, 3t, \dots$

⁷Un estado j es recurrente si la probabilidad de que alguna vez el sistema retorne al estado j es igual a 1.

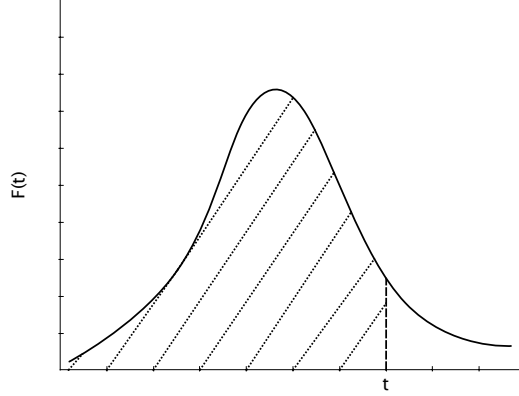


Figura 3: Aproximación de la Función de Riesgo

$$S(t) = P(T \geq t) \quad (9)$$

Es así que se puede definir la función de riesgo como:

$$\lambda(t) = \frac{f(t)}{S(t)} \quad (10)$$

esto es la velocidad en el que el evento puede ser completado con duración de t dado que el periodo de duración se ha extendido t periodos, o lo que es lo mismo

$$\lambda(t) = \lim_{h \rightarrow 0} P(t \leq T < t + h | T \geq t) / h \quad (11)$$

Un ejemplo claro de las definiciones presentadas anteriormente se presenta a continuación:

- Asuma que un individuo ha durado en el empleo informal durante 3 periodos, así la probabilidad de estar empleado en el sector informal con una duración inferior a 3 periodos es F (función de distribución); equivalentemente la probabilidad de estar empleado en dicho sector durante 3 periodos o más es S (función de supervivencia). La probabilidad de que el periodo termine, salir del empleo informal, entre 3 y $3 + h$ condicionado a haber durado 3 periodos es $\lambda * h$ (función de riesgo).

A partir de la definición de la función de riesgo se obtiene:

$$\lambda(t) = -\frac{dS/dt}{S} \quad (12)$$

de donde

$$\lambda(t) = -d \ln S(t) / dt. \quad (13)$$

La Función de Riesgo provee una conveniente definición de dependencia de duración. *Dependencia positiva en la duración* existe en el punto t^* si $d\lambda(t)/dt > 0$ en $t = t^*$. Significa que la probabilidad de que el periodo termine rápidamente incrementa a medida de que la duración del periodo aumenta. Por el contrario, *dependencia negativa en la duración* existe en el punto t^* si $d\lambda(t)/dt < 0$. A cada momento la probabilidad de que el periodo termine disminuye.

2.3. Distribuciones

A continuación se hace una breve descripción de los tipos de distribuciones utilizadas en el análisis de duración para su posterior elección sobre la base del mejor ajuste a los datos.

- a. Distribución exponencial. Para la distribución exponencial con parámetro $\gamma > 0$,

$$F(t) = 1 - \exp(-\gamma t) \quad (14)$$

$$S(t) = \exp(-\gamma t) \quad (15)$$

$$f(t) = \gamma \exp(-\gamma t) \quad (16)$$

$$\lambda(t) = \gamma \quad (17)$$

Este tipo de distribuciones se caracterizan porque la función de riesgo es constante y no refleja la dependencia en la duración.

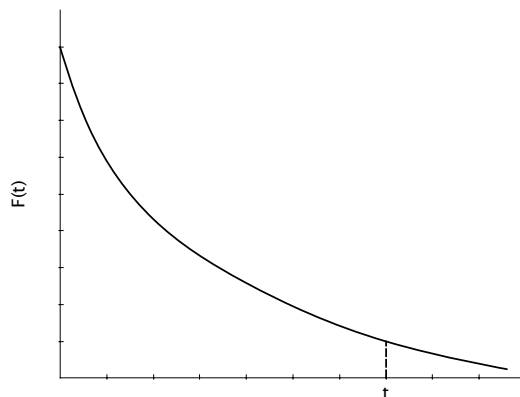


Figura 4: Distribución de probabilidad Exponencial

- b. Distribución Weibull. La distribución Weibull maneja dos parámetros ($\gamma > 0$ y $\alpha > 0$) con una función de riesgo de la familia

$$\lambda(t) = \gamma \alpha t^{\alpha-1} \quad (18)$$

si se observa, ésta es una generalización de la distribución exponencial cuando $\alpha = 1$. Es así que:

$$F(t) = 1 - \exp(-\gamma t^\alpha) \quad (19)$$

$$S(t) = \exp(-\gamma t^\alpha) \quad (20)$$

$$f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^\alpha) \quad (21)$$

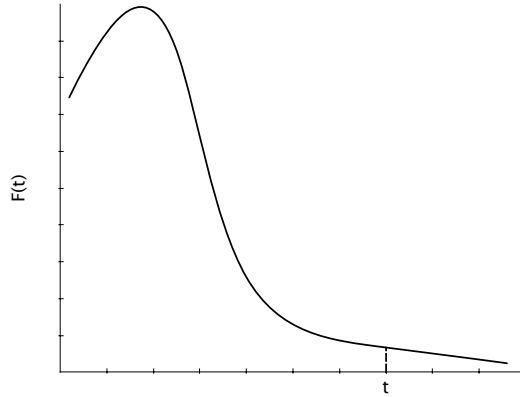


Figura 5: Distribución de probabilidad Weibull

- c. Distribución log-logística. La distribución log-logística maneja dos parámetros ($\gamma > 0$ y $\alpha > 0$) y función de riesgo

$$\lambda(t) = \gamma\alpha t^{\alpha-1}/(1 + t^\alpha\gamma) \quad (22)$$

y aplicando las fórmulas se tiene que:

$$F(t) = 1 - [1/(1 + t^\alpha\gamma)] \quad (23)$$

$$S(t) = 1/(1 + t^\alpha\gamma) \quad (24)$$

$$f(t) = \gamma\alpha t^{\alpha-1}/(1 + t^\alpha\gamma)^2 \quad (25)$$

- d. Distribución Gompertz. En este tipo de modelos la función de riesgo viene caracterizada por:

$$\lambda(t) = \alpha e^{\gamma t} \quad (26)$$

donde $\gamma = e^{X\beta}$ y γ es el parámetro de la pendiente. La función de supervivencia viene dada por:

$$S(t) = e^{-\alpha\gamma^{-1}(e^{\gamma t}-1)} \quad (27)$$

- e. Distribución Log-Normal. En el modelo Log-Normal la función de supervivencia es:

$$S(t) = 1 - \Phi\left[\frac{\ln(t) - \mu}{\sigma}\right] \quad (28)$$

donde Φ es la función de distribución estandar $\mu = X\beta$. La función de riesgo es:

$$\lambda(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}}e\left[\frac{-1}{2\sigma^2}(\ln(t) - \mu)^2\right]}{1 - \Phi\left[\frac{\ln(t) - \mu}{\sigma}\right]} \quad (29)$$

Desarrollado ya, y de manera breve, el instrumental teórico que será utilizado se procede a identificar los aspectos metodológicos necesarios para llevar a cabo las estimaciones.

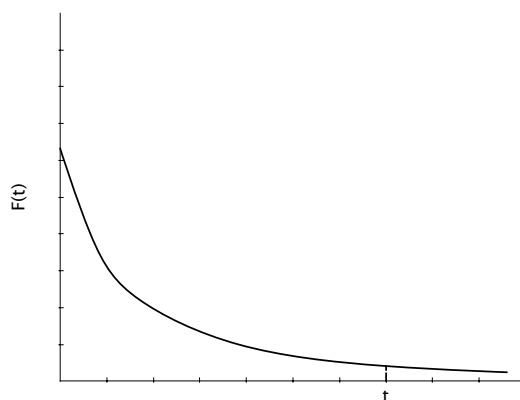


Figura 6: Distribución de probabilidad Gompertz

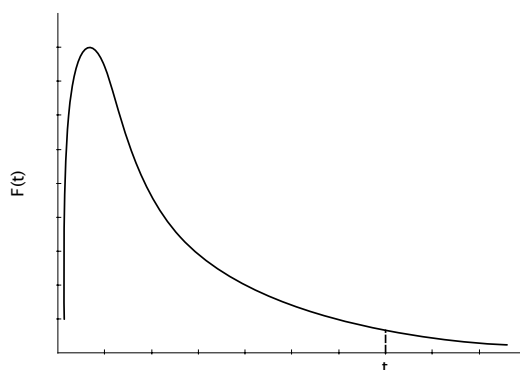


Figura 7: Distribución de probabilidad Log-Normal

3. MARCO METODOLÓGICO

3.1. Definición empleo informal

A partir del criterio del proyecto PREALC (Programa Regional de Empleo para América Latina y El Caribe 1978) de la OIT y de ciertas recomendaciones del grupo de DELHI ⁸ es que el DANE adopta una definición de empleo informal que se remite en últimas a la resolución de la OIT de 1993.

Se define empleado informal a los asalariados o patronos vinculados a empresas de hasta 5 trabajadores, a quienes trabajan como empleados familiares sin remuneración, a los trabajadores del servicio doméstico y a los trabajadores por cuenta propia excepto los independientes profesionales. Obsérvese como para el propósito del presente trabajo se asume como definición de empleado informal la T-5 siendo diferente de la definición T-10 por el número de trabajadores máximo por

⁸Grupo creado en 1997 en el marco de los grupos de las ciudades de la Comisión de Estadísticas con el fin de abordar diversos asuntos metodológicos relativos al tratamiento del sector no estructurado.

empresa. Como bien lo plantea Lopez (2011):

	OIT-10	OIT-5
Trabajadores cuenta propia diferentes de profesionales independientes	Todos	Todos
Patron de empresa	Hasta 10 trabajadores	Hasta 5 trabajadores
Obreros o empleados de empresas privadas	Hasta 10 trabajadores	Hasta 5 trabajadores
Empleados de servicios doméstico	Todos	Hasta 5 trabajadores
Jornal y otros trabajadores	Todos	Hasta 5 trabajadores

Figura 8: Definiciones de Informalidad Fuente: Presentación Banco de la República. Hugo Lopez 2011

Operativamente, la definición de informalidad se resumen en la siguientes categorías:

- I. Los empleados particulares y los obreros que laboran en establecimientos, negocios o empresas que ocupen hasta cinco personas en todas sus agencias y sucursales, incluyendo al patrono y/o socio;
- II. los trabajadores familiares sin remuneración;
- III. Los trabajadores familiares sin remuneración en negocios o empresas de otros hogares;
- IV. Los empleados domésticos;
- V. Los jornaleros o peones;
- VI. Los trabajadores por cuenta propia que laboran en establecimientos de hasta cinco personas, excepto los independientes profesionales ;
- VII. Los patrones o empleadores en empresas de cinco trabajadores o menos;
- VIII. Se excluye los obreros o empleados del gobierno.

La definición de la informalidad está explicada a partir de las unidades productivas y no sobre las características de los ocupados.

En contraposición a ésta definición se encuentran aquellas que están basadas en la afiliación al sistema de seguridad social, ya sea en términos de pensión o de salud, como proxy del grado de desarrollo de la calidad del empleo. Guataquí (2011) lleva a cabo un completo estado del arte en cuanto a la evolución de las definiciones utilizadas en los trabajos académicos que intentan caracterizar el empleo informal en Colombia. Adicionalmente, para él la continua referencia a la definición utilizada por el DANE se debe a aspectos puramente históricos y no a aspectos estructurales que respondan a las cambiantes dinámicas de la economía Colombiana.

Sin embargo, para el propósito del presente trabajo no se tienen en cuenta aquellas definiciones que están basadas en la exclusión del sistema de seguridad social porque, y al contrario de lo que

argumenta Guataquí, se cree que el grado de desarrollo institucional se debe ver reflejado en su capacidad de revertir aquellas condiciones estructurales que permiten el surgimiento de actividades que por sí mismas garanticen el cumplimiento de las condiciones legalmente aceptables para el total de la población y no en su capacidad de lograr asegurar un mínimo de condiciones a unos pocos trabajadores. Como se verá, este argumento está sustentado en la comprensión real de la dinámica, en el tiempo constante, de la informalidad en Colombia.

3.2. Gran Encuesta Integrada de Hogares (GEIH)

Es a partir de la recomendación de la OIT de 1982 que varios países desarrollaron una metodología y subsecuente recolección de datos del sector informal cuyo origen conceptual fue la carencia de un sistema adecuado que “captara adecuadamente el comportamiento de una porción considerable del mercado de trabajo, en gran parte asociada con las relaciones laborales pre-modernas que persistían en los mercados de trabajo” (Guataquí, García y Rodríguez 2011).

En Colombia, sólo hasta el año de 1986 se empieza a desarrollar el módulo de informalidad en la Encuesta Nacional de Hogares (ENH) basado en la información de las 7 principales ciudades, incluidas sus áreas metropolitanas (Bogotá, Cali, Medellín, Barranquilla, Manizales, Pasto y Bucaramanga) con una periodicidad bianual hasta el año 2000 ⁹. En el año 2001 entra en vigencia la Encuesta Continua de Hogares (ECH) con lo que la aplicación del módulo de informalidad se llevaba a cabo los segundos trimestres de cada año para las 13 principales ciudades, incluidas sus áreas metropolitanas (Bogotá, Medellín-Valle de Aburrá, Cali-Yumbo, Barranquilla-Soledad, Bucaramanga-Girón-Piedecuestras-Floridablanca, Manizales-Villamaría, Pasto, Pereira-Dos quebradas-La Virginia, Cúcuta-Villa del Rosario- Los Patios-El Zulia, Ibagué, Montería, Cartagena y Villavicencio). Durante los años 2001 y 2002 el DANE realizó la medición del sector informal mediante la denominada *Encuesta 1 2 3*. Esta encuesta fue realizada en 2 fases:

- Fase 1: Se aplicó una encuesta para identificar los miembros del hogar, patronos o trabajadores con cuenta propia, de una unidad económica informal visible, no visible o móvil. A partir de esta información se construyó un marco muestral de unidades económicas de los hogares.
- Fase 2: Se llevó a cabo una encuesta a las unidades económicas informales, para identificar las características generales de la unidad económica tales como actividad económica, edad, formas de financiamiento, número y características del personal ocupado, etc.
- Fase 3: A partir de una encuesta de ingresos y gastos, se indagó sobre la proporción de gastos en establecimientos informales y las razones para ello.

Durante la vigencia de la Encuesta Continua de Hogares (2001-2006) el módulo de informalidad se siguió aplicando los segundos trimestres de cada año hasta cuando se dio paso a la Gran Encuesta Integrada de Hogares (GEIH).

La Gran Encuesta Integrada de Hogares amplía la muestra de hogares y su aplicación aumenta la cobertura para 24 ciudades a nivel nacional. En la aplicación de dicha encuesta se decide mantener la batería de preguntas del módulo de informalidad pero con una aplicación continua para el total nacional, que incluye además de las áreas metropolitanas, cabeceras y centros poblados.

⁹Específicamente en el mes de Junio

3.2.1. Características encuesta

La Gran Encuesta Integrada de Hogares (GEIH) es una ¹⁰ encuesta por muestreo probabilístico, multietápico, estratificado, de conglomerados desiguales, ordenada en módulos o registros (8 en total) y auto-ponderado (para las 24 ciudades capitales con sus áreas metropolitanas) en donde el universo objetivo es la población civil no gubernamental cuyos resultados permiten obtener cobertura nacional con resultados a nivel urbano y rural, cinco grandes regiones, total departamental y 13 grandes ciudades ¹¹ con sus áreas metropolitanas y 11 ciudades intermedias ¹². Para todos los casos, la unidad básica de observación es el hogar.

Para el análisis llevado a cabo se utilizaron 3 de los registros (Ocupados, Desocupados y Características generales) con los cuales se obtuvo una muestra de 203.624 registros de personas repartidos entre urbano (181.596) y rural (22.028). El análisis nacional es posible por la utilización de los factores de expansión ¹³.

En esa medida, se propone realizar las siguientes actividades:

- I. A partir de la gran encuesta de hogares y de la definición de empleo informal se construye la matriz de probabilidades que relaciona el tránsito entre ser empleado formal, empleado informal y desempleado.
- II. Con base en los resultados del numeral anterior, llevar a cabo un análisis de los determinantes de la duración del empleo informal. Específicamente se recurre a las estimaciones tanto paramétricas como semiparamétricas, y
- III. Llevar a cabo el mismo análisis del numeral II pero introduciendo el componente espacial. Así se tendrá un marco conceptual más amplio que sirva como base explicativa de la dinámica del empleo informal.

4. EVOLUCIÓN INFORMALIDAD

Entender el comportamiento de la evolución de las tasas de informalidad a partir de la definición adoptada permite comprender las características de la informalidad en Colombia. En esa medida, a continuación se presenta un cuadro que resume la proporción de empleados informales en las 13 principales áreas metropolitanas (por tamaño de empresa) para los segundos trimestres de cada año, desde el 2001 al 2010.

Es claro que, independientemente del tipo de definición utilizada, la tasa de informalidad en Colombia siempre presenta niveles por encima del 50% con lo que se puede argumentar que durante el periodo de análisis no se han llevado a cabo medidas que contrarresten este comportamiento.

Es de resaltar que todas las estimaciones de la tasa de informalidad que realiza el DANE son de carácter urbano y para el alcance del presente trabajo se lleva a cabo una estimación de la tasa

¹⁰Encuesta primer trimestre de 2010: 154 municipios

¹¹Bogotá, Medellín-Valle de Aburrá, Cali-Yumbo, Barranquilla-Soledad, Bucaramanga-Floridablanca-Girón-Piedecuestras, Manizales-Villamaría, Pasto, Pereira-Dosquebradas-La virginia, Ibagué, Cúcuta-Villa del Rosario-Los Patios-El Zulia, Villavicencio, Montería y Cartagena

¹²Tunja, Florencia, Popayán, Valledupar, Quibdó, Neiva, Riohacha, Santa Marta, Armenia, Sincelejo y San Andrés

¹³Peso o representación que le corresponde, a los datos muestrales, en el universo investigado

Concepto	Abr-Jun									
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total 13 áreas										
Ocupados informal T5	54,5	54,7	54,2	52,3	52,3	51,4	50,1	49,6	52,6	51,6
Ocupados informal T10	60,7	61,4	60,7	58,7	58,9	58,5	56,5	55,6	58,3	57,6

Figura 9: Evolución informalidad Fuente: DANE, cálculos de la XXI Reunión de presidentes de organizaciones empresariales Iberoamericanas

de *informalidad nacional* que para el primer trimestre de 2010 asciende al 61.37%.

Es así que a partir de las preguntas retrospectivas del módulo de informalidad (que indaga entre otras cosas por su situación laboral inmediatamente anterior a la actual) se analiza en detalle la distribución de la proporción de los informales en las categorías definidas:

OCUPADOS INFORMALES ANTES	Porcentaje
Empleados particulares	43.14
Empleado doméstico	6.84
Trabajador cuenta propia	33.61
Patrón	1.45
Trabajador familiar sin remuneración	2.93
Trabajador sin remuneración en empresas o negocios de otros hogares	0.22
Jornalero o peón	11.79

Figura 10: Distribución ocupados informales antes

OCUPADOS INFORMALES AHORA	Porcentaje
Empleados particulares	15.06
Empleado doméstico	5.43
Trabajador cuenta propia	58.84
Patrón	6.6
Trabajador familiar sin remuneración	7.41
Trabajador sin remuneración en empresas o negocios de otros hogares	0.79
Jornalero o peón	6.07

Figura 11: Distribución ocupados informales ahora

Se ha generado una recomposición a través del tiempo que ha tendido a favorecer el desarrollo generalizado de aquellas actividades que poseen un carácter más individual. Es decir, ante la imposibilidad de encontrar empleo en actividades que implican una relación laboral subordinada se ha recurrido al desarrollo de aquellas en las cuales el propio individuo se convierte en el coordinador de las actividades que implica la creación de un negocio propio. Esto es tanto actividades de alta y baja productividad.

5. CARACTERIZACIÓN SOCIOECONOMICA

Con el fin de tener una idea preliminar de cual puede ser la situación socioeconómica de los empleados informales en Colombia se procede a hacer una breve descripción de la distribución poblacional y del desarrollo económico, este último medido a través del cálculo del Índice de Calidad de Vida (ICV).

Como puede observarse, la distribución poblacional de los empleados informales es muy simétrica (normal) entre géneros (hombres y mujeres) con ligeras diferencias en la magnitud entre rangos de edades. Los hombres presentan mayor distribución de población en el rango de entre los 45 a los 49 años seguido del rango de entre 25 a los 29 años mientras que las mujeres concentran mayor población en los rangos de entre 40 a los 44 y entre los 45 a los 49 años.

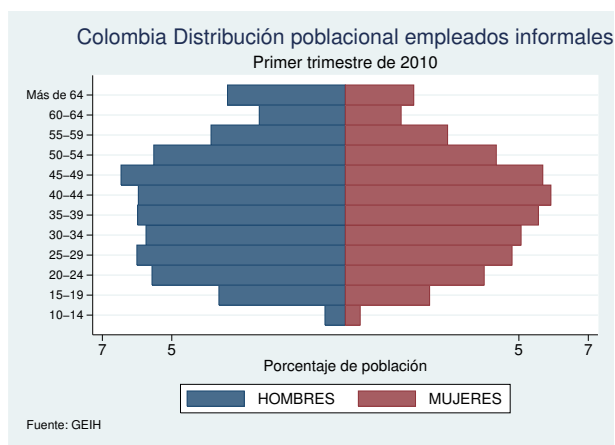


Figura 12: Distribución poblacional informales

De todas maneras se puede inferir que las mujeres presentan una mayor concentración de la informalidad en el rango de entre los 40 a los 49 años mientras que para los hombres el rango de entre los 25 a los 29 años adquiere gran relevancia.

En términos de las características económicas se recurrió al ICV. Éste mide el grado de desarrollo económico a través de la evaluación de una serie de factores como: acceso a servicios, calidad de la vivienda, educación y capital humano y tamaño y composición del hogar en donde constitucionalmente se definen puntajes mínimos aceptables (para mayor detalle ver anexo)¹⁴. Para tener una idea clara de cual es la situación real de los empleados informales, se procede a calcular el índice tanto para los empleados formales como para los desempleados.

¹⁴Fuente: Encuesta de Calidad de Vida 2007 para Bogotá ECVB 2007.

Variable de referencia	Puntaje para un mínimo estándar de vida	Puntaje máximo de cada variable
Acumulación colectiva de bienes materiales		
Abastecimiento de agua	4	6.99
Eliminación de excretas	3	7.14
Recolección de basuras	6	6.62
Material con que cocinan	5	6.67
Acumulación individual de bienes materiales		
Material de los pisos	3.2	6.79
Material de las paredes	2.3	6.11
Capital social básico		
No hacinamiento	7.8	12.8
Proporción de niños (niñas) menores de 6 años en el hogar	---	7.45
Acumulación de capital humano		
Educación del jefe o jefa del hogar	9.4	11.52
Educación mayores de 12 años	9.66	12.31
Asistencia de jóvenes a la escuela	5.66	5.66
Asistencia de niños (niñas) entre cinco y once años	5.69	9.95
Total	67	100

Figura 13: ICV: Puntajes mínimos aceptables.

ICV	FACTOR 1. ACCESO A LA CALIDAD DE LOS SERVICIOS				FACTOR 2. EDUCACION Y CAPITAL HUMANO				FACTOR 3. TAMAÑO Y COMPOSICIÓN DEL HOGAR		FACTOR 4. CALIDAD DE LA VIVIENDA				
	Total	Eliminación de excretas	Abastecimiento de agua	Combustible para cocinar	Recolección de basuras	Total	Escolaridad máxima del jefe del hogar	Escolaridad promedio de personas de 12 años y más	Proporción de jóvenes de 12 a 18 años que asisten a secundaria, técnica o universitaria	Proporción de niños de 5 a 11 años que asisten a un establecimiento educativo	Total	Hacinamiento	Total	Material paredes	Material pisos
Ocupados formales	81.86	27.17	6.99	6.91	6.64	6.59	6.54	6.55	5.08	8.59	18.01	12.8	12.21	5.99	6.21
Desempleados	79.71	26.62	6.74	6.75	6.48	6.5	5.08	5.73	5.05	8.43	18.07	12.8	11.39	5.72	5.66
Ocupados informales	78.49	26.54	6.62	6.82	6.41	6.51	4	5.11	4.9	8.28	17.91	12.8	11.24	5.69	5.54

Figura 14: ICV empleados informales, formales y desempleados

Es de resaltar que los empleados informales son los que menor ICV presentan, aunque no sea significativamente diferente de los demás, por lo que son los que menor desarrollo económico presentan y no sólo en el indicador global sino en todos y cada uno de los componentes.

Florez (2004) desarrolla un argumento que se ajusta acertadamente a este hecho. Cuando los jóvenes (menores de 30 años) ingresan al mundo laboral, lo hacen directamente a un mercado de baja productividad (informal), pero luego de adquirir cierto grado de experiencia y de destrezas ingresan al mercado laboral formal. Después de haber adquirido más experiencia y en ciertos casos de acumular capital se retorna al mercado laboral informal.

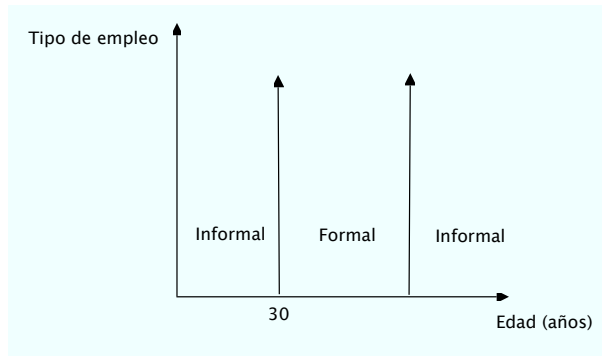


Figura 15: Ciclo laboral

Es así que después de haber abandonado el mercado laboral formal es difícil que se logre un nuevo enganche en actividades laborales de este tipo. Por el contrario, se observará que es en el empleo informal en donde se presenta una menor tasa de rotación inter-temporal.

6. ESTIMACIÓN

Como se mencionó inicialmente, las estimaciones del presente documento se llevan a cabo a partir de cifras nacionales (urbano-rural).

6.1. Cadena de Markov

Teniendo en cuenta el marco metodológico ya desarrollado y especificando como estado espacios Ω el ser empleado informal, empleado formal y desempleado ¹⁵, tanto para el periodo en cuestión como para el anterior ¹⁶ y a partir del total nacional (Cabecera y Resto), se llevo a cabo la estimación del vector de probabilidades de largo plazo para cada uno de estos estados.

En esa medida los resultados obtenidos son los siguientes ¹⁷:

¹⁵La discriminación de la población se hizo a partir de las definiciones del DANE.

¹⁶Como la GEIH no es un panel, los estados en el periodo anterior se reconstruyeron a partir de preguntas retrospectivas.

¹⁷cifras en porcentaje

P	Empleo informal	Empleo formal	Desempleo
Empleo informal	97.12	16.06	33.39
Empleo formal	2.04	82.21	59.05
Desempleo	0.835	1.73	7.55

Figura 16: Matriz de probabilidades

En su orden, los valores de la diagonal representan las probabilidades de permanecer empleado en el sector informal, en el sector formal y de permanecer desempleado. A partir de esta matriz de probabilidades y siguiendo el procedimiento, ya mencionado en el marco teórico, se hallan las probabilidades de estado estable:

Empleo informal	Empleo formal	Desempleo
85.72	13.24	1.02

Figura 17: Vector de probabilidades de largo plazo

Es interesante ver como los resultados muestran una tendencia de largo plazo hacia la ocupación informal. Además, se puede argumentar que la tasa de rotación de la población ocupada es relativamente baja (2,88 % y 17,79 %) mientras que la tasa de rotación de los desocupados es del 92,45 %, siendo la ocupación formal el porcentaje de recepción más alta. A partir de éstos resultados se continuará con una descripción más detallada de las características del empleo informal sustentados en el análisis de funciones de riesgo.

6.2. Funciones de Riesgo

6.2.1. Estimación no paramétrica

Este tipo de estimación esta basada en un análisis gráfico que permite entender la dinámica de la variable a analizar. Concretamente, el estimador Kaplan-Meier ¹⁸ permite estimar la función de supervivencia (probabilidad de duración en el evento mayor o igual a t periodos) a partir de un ordenamiento ascendente del tiempo de salida (con j tiempos de permanencia distintos) del estado en el que se encuentre. De esta forma, el estimador empírico tendría la forma:

$$\widehat{S}(t_j) = \prod_{i=1}^j (n_i - h_i)/n_i \quad (30)$$

donde n_i es el número de individuos que permanece en el estado hasta después de j (conjunto de riesgo) y h_i los individuos que no permanecen después de j . Adicionalmente, y a partir de lo anterior,

¹⁸Este análisis se realiza a partir de las observaciones no censuradas, es decir a partir de aquellas que tuvieron cotas de duración.

se puede realizar una estimación de la tasa hazard:

$$\hat{\lambda}(t_j) = h_j/n_j \quad (31)$$

que es la probabilidad de completar el periodo con una duración de t_j condicionado a que el alcance de la duración de los periodos sea t_j .

Para el caso específico se realizó la estimación de la función de supervivencia, complemento de la probabilidad de *salir del estado*¹⁹ empleado informal, a partir de este estimador obteniendo los siguientes resultados: a medida que avanza el tiempo disminuye la probabilidad de supervivencia en el estado de empleado informal, obteniéndose un valor de 50 % en el periodo 15 hasta estabilizarse en aproximadamente 26 % en el periodo 20.

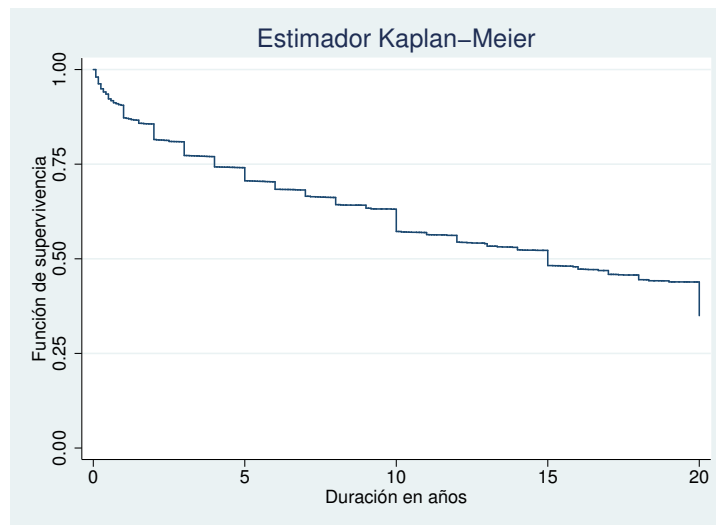


Figura 18: Estimador Kaplan-Meier (función de supervivencia)

Ahora bien, a partir de esta primera aproximación, se procede a caracterizar las funciones de supervivencia por género, edad, nivel educativo, división administrativa municipal (cabecera y resto) y estado civil.

¹⁹Se define la salida de un estado como el tránsito a otro estado.

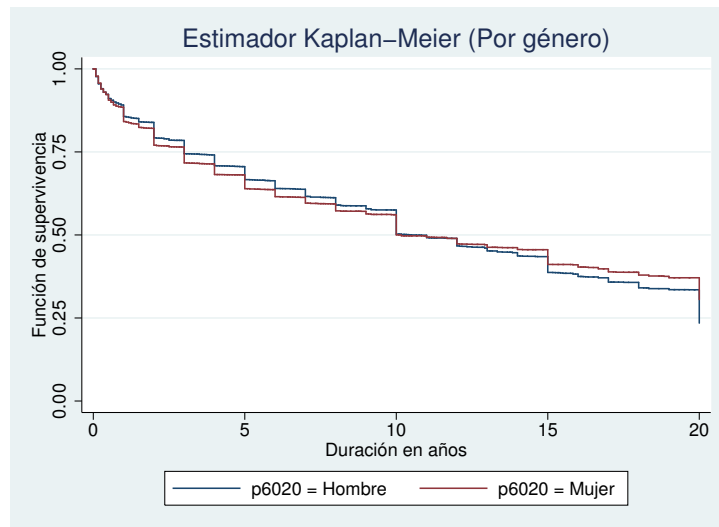


Figura 19: Estimador Kaplan-Meier (función de supervivencia) por género

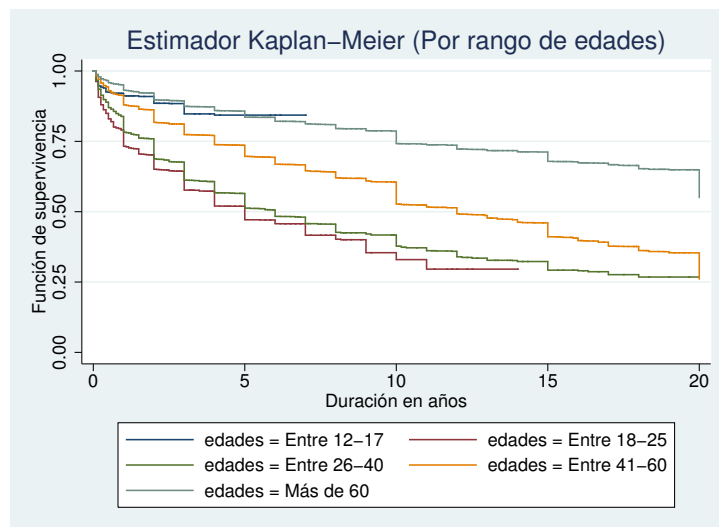


Figura 20: Estimador Kaplan-Meier (función de supervivencia) por edades

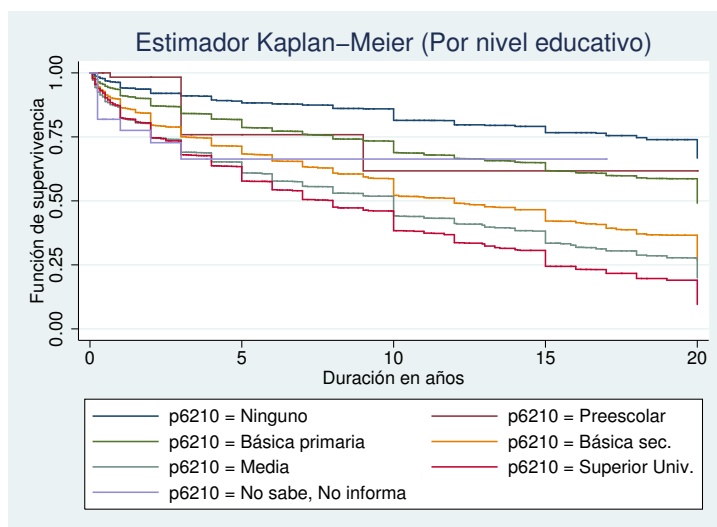


Figura 21: Estimador Kaplan-Meier (función de supervivencia) por nivel educativo

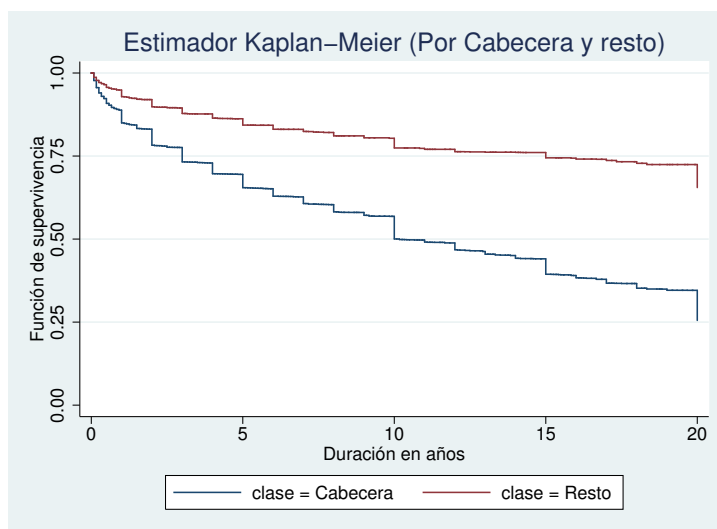


Figura 22: Estimador Kaplan-Meier (función de supervivencia) por división administrativa municipal

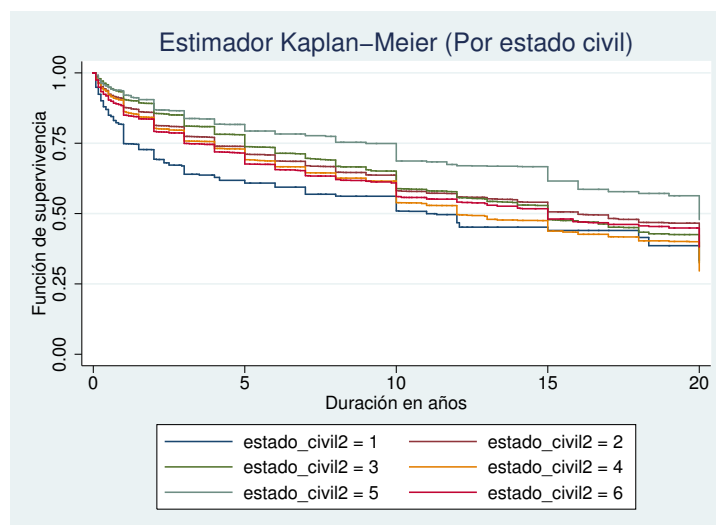


Figura 23: Estimador Kaplan-Meier (función de supervivencia) por estado civil

Teniendo en cuenta que existe evidencia estadística significativa que sustenten diferencias dentro de las categorías de variables utilizadas ²⁰, se puede inferir que:

- I. Las mujeres tienen una mayor probabilidad, aunque no significativa, de salida de la informalidad,
- II. Son los individuos más jóvenes los que tienen una mayor probabilidad de supervivencia pero con menor tiempo de duración,
- III. Los individuos con mayores niveles educativos son los que tienen una mayor probabilidad de salida,
- IV. La población residente en zonas rurales tienen mayor probabilidad de permanecer en el sector informal y
- V. Por estado civil ²¹ los individuos viudos tienen una probabilidad mayor de salir de la informalidad pero no significativa.

Es de resaltar que después de 15 años de permanecer en el empleo informal, la probabilidad de salida es de apenas 50% con lo que se puede argumentar que las características que definen la duración son primordialmente de carácter estructural, las que durante dicho lapso de tiempo no se han podido resolver.

²⁰Ver anexo 2

²¹1" No está casado(a) y vive en pareja <2 años" 2" No está casado(a) y vive en pareja ≥2 años" 3" Esta casado(a)" 4" Esta separado(a) o divorciado(a)" 5" Esta viudo(a)" 6" Esta soltero(a)"

6.2.2. Estimación paramétrica

Este tipo de análisis permiten caracterizar de forma más detallada los determinantes de la duración de la informalidad en Colombia.

Para llevar a cabo este análisis se recurrió a todas las especificaciones de distribuciones definidas en el marco teórico y utilizando criterios de información (tanto AIC como BIC) se obtuvo que el modelo que mejor se ajusta a los datos es el Weibull ²²:

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
exponential	48689	.	-1.02e+07	16	2.04e+07	2.04e+07
weibull	48689	.	-9898365	17	1.98e+07	1.98e+07
gomartiz	48689	.	-1.21e+07	17	2.42e+07	2.42e+07
loglogistic	48689	.	-1.19e+07	17	2.37e+07	2.37e+07
lognormal	48689	.	-1.19e+07	17	2.38e+07	2.38e+07

Note: N=Obs used in calculating BIC; see [\[R\] BIC note](#)

Figura 24: Criterios de información ajuste modelo

Concretamente:

```

failure _d: perdidos == 1
analysis time _t: duracion_emple_infor/12
weight: [pweight=fex_c_def]
Fitting full model:

Iteration 0: log pseudolikelihood = -10410872
Iteration 1: log pseudolikelihood = -9961091.2
Iteration 2: log pseudolikelihood = -9898438.9
Iteration 3: log pseudolikelihood = -9898364.9
Iteration 4: log pseudolikelihood = -9898364.9

Weibull regression -- log relative-hazard form

No. of subjects   = 10829558.97      Number of obs   = 48689
No. of failures  = 3270908.54
Time at risk     = 53761405.19

Log pseudolikelihood = -9898364.9      Wald chi2(16)  = 12496.15
                                          Prob > chi2    = 0.0000

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
_Iclase_2	-1.133782	.0532577	-21.29	0.000	-1.238165 -1.029399
_Irange_ed~2	.1668567	.0581234	2.87	0.004	.052937 .2807764
_Irange_ed~3	.0589061	.0467958	1.26	0.208	-.0328119 .1506241
_Irange_ed~4	-.2843593	.0393305	-7.23	0.000	-.3614458 -.2072729
_Irange_ed~6	-.6830254	.1630219	-4.19	0.000	-1.002543 -.3635082
_Ip6210_2	-.7174223	.7284352	-0.98	0.325	-2.145129 .7102845
_Ip6210_3	-.7123627	.0496561	-14.35	0.000	-.8096869 -.6150384
_Ip6210_4	-.3501114	.0547429	-6.40	0.000	-.4574054 -.2428174
_Ip6210_5	-.2096175	.0520112	-4.03	0.000	-.3115576 -.1076773
_Ip6210_6	-.1363179	.0577878	-2.36	0.018	-.2495799 -.023056
_Ip6210_9	-2.57059	1.161048	-2.21	0.027	-4.846202 -.2949777
_Ip6070_2	-1.502719	.0526007	-28.57	0.000	-1.605815 -1.399624
_Ip6070_3	-1.607344	.0535219	-30.03	0.000	-1.712245 -1.502443
_Ip6070_4	-1.483263	.0576302	-25.74	0.000	-1.596216 -1.370311
_Ip6070_5	-1.993723	.0889116	-22.42	0.000	-2.167987 -1.819459
_Ip6070_6	-1.461983	.0540105	-27.07	0.000	-1.567842 -1.356125
/ln_p	-.3058971	.0100242	-30.52	0.000	-.3255441 -.2862501
p	.7364624	.0073824			.7221343 .7510748
1/p	1.357843	.0136113			1.331425 1.384784

Figura 25: Análisis paramétrico: Distribución Weibull

²²Para mayor detalle ver Anexo 2

Como podrá observarse, se excluyen ciertas categorías de las variables que se utilizaron para el análisis, no solo porque el procedimiento práctico lo requiere sino porque no son estadísticamente significativas para el modelo:

- I. De la variable clase se excluye la categoría Cabecera (1) (*_Iclase_2*). Además del modelo se puede extrapolar que no solo el ser parte de regiones consideradas como rurales es estadísticamente significativo sino que también se reduce el tiempo de permanencia en el sector informal.
- II. Las personas con más de 60 años fueron excluidas del análisis (*_I rango_ed*). Del resto de categorías únicamente el grupo de personas entre los 26 y 40 años no es estadísticamente significativo. Adicionalmente, pertenecer a los grupos de personas entre los 12 y 17 años y entre los 41 y los 60 años reduce el tiempo de permanencia en la informalidad.
- III. De la variable que mide el nivel educativo máximo alcanzado (*_Ip6210_*) se excluyen los niveles "Ninguno" y "Preescolar", este último incluido en el modelo. De resto todos los niveles educativos reducen el tiempo de permanencia, con mayor relevancia (excluyendo las personas que no informan) aquellas con educación completa en básica primaria y básica secundaria.
- IV. Todas las categorías definidas en el estado civil (*_Ip6070_*) son significativas y reducen la duración en el desempleo (se excluye la categoría 1).
- V. Las categorías de la variable género no son significativas. Este hecho se explica principalmente por la distribución simétrica de la población.

6.3. Análisis Espacial

Para lograr una comprensión más detallada de los determinantes de la duración de la informalidad en Colombia se llevó a cabo un ejercicio que implica la utilización del análisis de econometría espacial bajo el marco de la estructura teórica básica utilizada para el análisis de duración.

Siguiendo el trabajo de Hays y Kachi se estimó (a partir de un estimador de máxima verosimilitud y con la información por municipios disponible) un modelo de rezagos espaciales con el que, a partir de la evidencia que sustenta la correlación espacial, se intenta explicar las interdependencias en términos de duración con las mismas variables explicativas utilizadas hasta ahora.

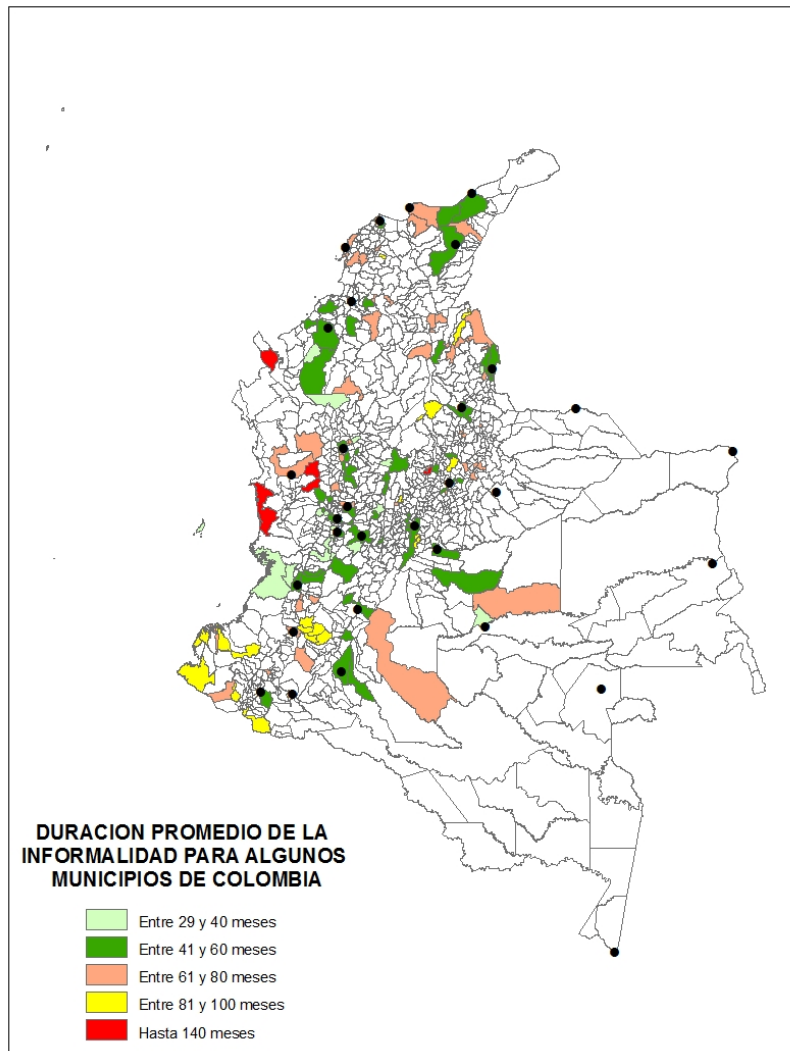


Figura 26: Duración promedio informalidad para algunos municipios de Colombia

La figura 26 presenta el mapa de la duración promedio de la informalidad censurada (no más de 20 años) de los municipios incluidos en la muestra de la encuesta. Los colores más claros hacen referencia a aquellos municipios en los cuales la informalidad presenta duraciones relativamente bajas (entre 29 y 80 meses) mientras que el resto presentan promedios de duración no mayores a los 140 meses. Estos últimos se encuentran distribuidos en: 1 municipio de Antioquia, Huila, Magdalena y Norte de Santander, 3 de Cauca y Santander, 4 de Cundinamarca y Chocó y 5 de Nariño.

En términos generales, este modelo se caracteriza por una especificación en donde la variable de duración (a nivel municipal) es explicada por ella misma rezagada y ponderada por una matriz que

recoge las dependencias espaciales a través de la matriz W (la cual captura los pesos espaciales en un contexto puramente espacial ²³) y mediante la cual se estima el coeficiente de interdependencia espacial (rho) así como por el conjunto de variables utilizadas en la estimación paramétrica y en donde el coeficiente asociado a la distribución Weibull tiene la misma interpretación dada hasta ahora (lambda).

$$y = Ay + X\beta + \epsilon \quad (32)$$

Como podrá observarse, la estructura del modelo tiene las mismas características de un modelo de rezago espacial convencional. De esta manera se obtuvieron los siguientes resultados:

```

initial:      log likelihood = -498301.39
rescale:     log likelihood = -522.68569
rescale eq:  log likelihood = -428.8483
Iteration 0: log likelihood = -428.8483
Iteration 1: log likelihood = -386.62779
Iteration 2: log likelihood = -380.03071
Iteration 3: log likelihood = -379.67388
Iteration 4: log likelihood = -379.67203
Iteration 5: log likelihood = -379.67203
  
```

Number of obs = 154
Wald chi2(19) = 94.42
Prob > chi2 = 0.0000

Log likelihood = -379.67203

	_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
mu	Hombre	.000146	.0000649	2.25	0.024	.0000188 .0002731
	Entre_1825	-.0000688	.0002222	-0.31	0.757	-.0005043 .0003666
	Entre_2640	.0000774	.0001363	0.57	0.570	-.0001898 .0003447
	Entre_4160	.0001058	.0001581	0.67	0.503	-.000204 .0004157
	Más_de_60	.0001757	.0001592	1.10	0.270	-.0001364 .0004877
	Ninguno	-.0048337	.0028229	-1.71	0.087	-.0103665 .0006992
	Preescolar	-.0044712	.0028776	-1.55	0.120	-.0101112 .0011689
	Basica_pri-a	-.0048965	.0028181	-1.74	0.082	-.0104199 .0006269
	Basica_sec-a	-.004946	.0028237	-1.75	0.080	-.0104803 .0005883
	Media	-.0048929	.0028168	-1.74	0.082	-.0104136 .0006279
	Superior_U-a	-.0050565	.0028157	-1.80	0.073	-.0105752 .0004622
	poblacion	6.80e-06	3.76e-06	1.81	0.071	-5.78e-07 .0000142
	Cabecera	.0047631	.002823	1.69	0.092	-.0007698 .010296
	Resto	.0046804	.0027954	1.67	0.094	-.0007985 .0101592
	civil1	-.0002088	.0001477	-1.41	0.157	-.0004982 .0000806
	civil2	-.00002	.0000625	-0.32	0.749	-.0001424 .0001024
civil3	.0000198	.0000703	0.28	0.778	-.000118 .0001577	
civil4	-.0001739	.0000967	-1.80	0.072	-.0003633 .0000156	
civil5	.0002564	.0002098	1.22	0.222	-.0001547 .0006676	
rho	_cons	.0008077	.0000807	10.01	0.000	.0006496 .0009659
	lambda					
	_cons	.3580739	.0186527	19.20	0.000	.3215153 .3946324

Figura 27: Regresión espacial

- I. El coeficiente espacial es estadísticamente significativo y positivo con lo que se puede argumentar que existe correlación espacial en términos de la duración del empleo informal, o dicho de otra forma que el componente espacial es una variable que explica el comportamiento de la informalidad en Colombia,

²³La especificación en términos espaciales se encuentra en el anexo

- II. La variable género adquiere relevancia en este análisis y el ser hombre aumenta la duración en el empleo informal,
- III. La edad no se convierte en una variable relevante en este análisis,
- IV. Todas las variables que caracterizan el grado de educación de los individuos son significativas y reducen la permanencia en la informalidad. Si bien es cierto que no expresa los resultados esperados, también es cierto que sí recogen la esencia de este tipo de variables y es que a mayor educación menor duración en la informalidad.
- V. A mayor población mayor duración en la informalidad sin diferencias significativas entre cabecera y resto.

Conclusiones y recomendaciones

Conclusiones

Después de realizados los análisis necesarios se puede concluir que:

- La tendencia de la informalidad de los últimos años y el análisis de su estado de largo plazo, permiten argumentar que el trabajo informal se ha convertido en una cuestión estructural que requiere de una base económica e institucional mucho más dinámica y robusta que permitan que la población logre un nivel de desarrollo que les garantice las condiciones necesarias no solo para engancharse en el mercado laboral formal sino para hacer más competitivo su trabajo en términos internacionales,
- Las personas con mayor grado de formación académica tienen mayor probabilidad de salir del estado *empleado informal*,
- En términos de desarrollo económico, los ocupados informales presentan un Índice de Calidad de Vida (ICV) menor que el de los ocupados formales y más aún que el de los desempleados. Esto es debido a que este sector presenta una menor tasa de rotación inter-temporal que se puede observar en la dinámica de las probabilidades entre estados lo cual ocasiona menores oportunidades en términos de ingreso familiar para mejorar las condiciones de vida,
- Existe evidencia estadística significativa para argumentar que el componente espacial es una variable que explica la duración de la informalidad en Colombia. En términos generales, se presenta un efecto de desbordamiento, no necesariamente simétrico, de las condiciones socioeconómicas que generan situaciones en las cuales el ser vecino de uno u otro municipio con altas o bajas tasas de informalidad afecta el mercado laboral en cuanto correlaciona su comportamiento con respecto a este.

Recomendaciones de Política

- Empleo de calidad o calidad del empleo? (Bul09, ul09) debe ser la cuestión central a la hora de diseñar la política laboral colombiana. Lineamientos que incentiven el desarrollo de actividades económicas con un contenido alto de trabajo calificado ²⁴ sobre la base de una política educativa agresiva e inclusiva debe convertirse en la “locomotora” que impulse el crecimiento Colombiano de mediano plazo.

Es importante entender que en el corto plazo es necesario aprovechar mas no perpetuar las ventajas relativas del país para poder lograr las bases económicas, sociales e institucionales que faciliten una estructura de desarrollo sostenible sobre las bases de las dinámicas económicas mundiales.

- Teniendo en cuenta que se ha generado una recomposición de la estructura de la distribución de los empleado informales, se deben generar todas las condiciones que garanticen la disminución de la carga fiscal que implica la creación de la pequeña y micro empresa,
- Si bien es cierto que el plan de desarrollo 2010-2014 intenta ser la guía del desarrollo de mediano plazo a partir de entre otras las mejoras en la competitividad y crecimientos de la

²⁴Hace referencia al trabajo que requiera destrezas o capacidades específicas

productividad a través de desarrollos en términos de las capacidades laborales y empresariales sobre la base de una interacción dinámica entre el sector privado y público, también es cierto que no queda del todo claro cuales pueden ser los resultados de dicha inversión en términos de los alcances de la investigación y del acceso a la educación.

- En la Figura 26 se podrá observar que existen áreas bien definidas de duración promedio de la informalidad. En esa medida, se puede concluir que se hace necesaria una política de desarrollo regional focalizada que tenga en cuenta las características de cada zona para hacer más eficiente los esfuerzos para lograr niveles altos de empleo de calidad. Inclusive, se puede lograr una caracterización total de la situación del país si se asume que existe un continuo adyacente para aquellos municipios para los cuales no se cuenta con información. Para tal fin se puede recurrir a las áreas de planificación estratégica regional del Plan Nacional de Desarrollo 2010-2014 (Figura 33 Anexo). Las cuales, sin ser rigurosos, se ajustan a la geo-referenciación de la duración promedio de la informalidad.

Recomendaciones Metodológicas

- Relacionado con el punto anterior, se concluye que se hace necesario que la metodología de GEIH amplíe la muestra de municipios a encuestar para que no se presenten baches de información entre regiones,
- Además, con el propósito de realizar un análisis de resultados de las políticas se recomienda que la GEIH realice un seguimiento de la población encuestada durante por lo menos tres meses para así tener una idea más clara de los impactos sobre lo que se cree deben ser una serie de políticas focalizadas,

A. Tests y regresiones

FACTOR 1. ACCESO Y CALIDAD DE LOS SERVICIOS		
Factores / variables / categorías		Puntaje
Variable	Eliminación de excretas	
Categorías	No tiene sanitario	0
	Inodoro a pozo, inodoro sin conexión, letrina	2,78
	Bajamar	2,97
	Inodoro con conexión a alcantarilla	7,14
Variable	Abastecimiento de agua	
Categorías	De río, quebrada, manantial, nacimiento, agua embotellada	0
	De pozo, lluvia, aljibe	0,78
	De pila pública, carrotanque, aguatero	4,01
	Acueducto por tubería, otra fuente por tubería	6,99
Variable	Combustible para cocinar	
Categorías	Leña, carbón, materiales de desecho, no cocinan	0
	Petróleo, gasolina	4,83
	Gas, electricidad	6,67
Variable	Recolección de basuras	
Categorías	La tiran a un patio	0
	La queman o entierran	1,59
	La tiran al río	2,59
	Recolección pública	6,62

Figura 28: ICV: Puntajes componentes

FACTOR 2. EDUCACIÓN Y CAPITAL HUMANO		
Variable	Escolaridad máxima del jefe del hogar	
Categorías	Sin educación	0
	Primaria incompleta	3,46
	Primaria completa	7,37
	Secundaria incompleta	9,41
	Secundaria completa	10,53
	Superior incompleta	11,42
	Superior completa	11,52
	Posgrados y doctorados	11,52
Variable	Escolaridad promedio personas de 12 y más años	
Categorías	0= escolaridad	0
	0<escolaridad<=4 años	2,39
	4<escolaridad<=5 años	6,54
	5<escolaridad<=10 años	9,66
	10<escolaridad<=11 años	11,54
	11<escolaridad<=15 años	12,11
	escolaridad <=16	12,31
Variable	Proporción de jóvenes de 12-16 años que asisten a secundaria / univ.	
Categorías	0= proporción asistencia	0
	0<proporción asistencia<1	4,37
	proporción asistencia=1	5,66
	Hogares sin jóvenes de 12-18 años	5,66
Variable	Proporción de jóvenes entre 5-11 años que asisten a establecimiento educativo	
Categorías	0= proporción asistencia	0
	1<proporción asistencia<1	0
	proporción asistencia=1	5,69
	Hogares sin jóvenes de 5-11 años	9,95

Figura 29: ICV: Puntajes componentes

Factor 3	FACTOR 3. TAMAÑO Y COMPOSICIÓN DEL HOGAR	
Variable	Proporción de niños y niñas menores de 5 años en el hogar	
Categorías	0,65<proporción de niños y niñas	0
	0,00<proporción de niños y niñas<0,65	0,72
	Proporción de niños y niñas = 0	7,45
Variable	Hacinamiento en el hogar (# personas por cuarto)	
Categorías	7<=hacinamiento	0
	6<=hacinamiento<7	2,47
	5<=hacinamiento<6	3,73
	4<=hacinamiento<5	5,02
	3<=hacinamiento<4	5,84
	2<=hacinamiento<3	7,87
	0<=hacinamiento<2	12,8
	FACTOR 4. CALIDAD DE VIVIENDA	
Variable	Material predominante de los pisos de la vivienda	
Categorías	Tierra y arena	0
	Madera burda, tabla	3,18
	Cemento	4,33
	Baldosín, ladrillo / madera pulida, mármol/alfombra, tapete	6,79
Variable	Material predominante de las paredes de la vivienda	
Categorías	Sin paredes	0
	Guadua, caña, esterilla	0
	Madera burda	0,59
	Bahareque	0,71
	Zinc, tela, cartón	1,64
	Adobe o tapia pisada	2,29
	Bloque, ladrillo	6,11

Figura 30: ICV: Puntajes componentes

```
. sts test p6020,
      failure _d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
```

Cox regression-based test for equality of survival curves

p6020	Events observed	Events expected	Relative hazard
Hombre	1996399.05	2074204.49	0.9634
Mujer	1275354.01	1197548.59	1.0667
Total	3271753.06	3271753.06	1.0000

wald chi2(1) = **13.87**
 Pr>chi2 = **0.0002**

Figura 31: Test de igualdad funciones supervivencia por sexo

```
. sts test rango_edades,
      failure _d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
```

Cox regression-based test for equality of survival curves

rango_edades	Events observed	Events expected	Relative hazard
Entre 12-17	25998.13	36060.05	0.8614
Entre 18-25	329677.51	153001.71	2.5490
Entre 26-40	666199.51	370244.95	2.0719
Entre 41-60	1514699.02	1348438.46	1.2457
Más de 60	734334.37	1363163.38	0.5968
Total	3270908.54	3270908.54	1.0000

wald chi2(4) = **1062.39**
 Pr>chi2 = **0.0000**

Figura 32: Test de igualdad funciones supervivencia por edad

```
. sts test p6210,
      failure _d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
```

Cox regression-based test for equality of survival curves

p6210	Events observed	Events expected	Relative hazard
Ninguno	95915.51	254175.34	0.4083
Preescolar	521.28	1154.80	0.4858
Básica primaria	931651.52	1377039.22	0.7347
Básica secundaria	708736.98	616962.63	1.2590
Media	1056849.99	729784.86	1.5920
Superior Universitaria	477546.06	291910.46	1.8015
No sabe, No informa	531.73	725.78	0.7876
Total	3271753.06	3271753.06	1.0000

Wald chi2(6) = 821.49
Pr>chi2 = 0.0000

Figura 33: Test de igualdad funciones supervivencia por nivel educativo

```
. sts test p6070,
      failure _d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
```

Cox regression-based test for equality of survival curves

p6070	Events observed	Events expected	Relative hazard
1	111861.99	75540.26	1.4933
2	1024744.70	1064690.27	0.9685
3	849591.49	912367.94	0.9365
4	456456.12	415654.66	1.1053
5	76510.77	113262.57	0.6794
6	752587.98	690237.37	1.1000
Total	3271753.06	3271753.06	1.0000

Wald chi2(5) = 67.71
Pr>chi2 = 0.0000

Figura 34: Test de igualdad funciones supervivencia por estado civil

```

. xi:streg i.clase i.rango_edades i.p6210 i.p6070,dist(exp) noconstant nohr
i.clase      _Iclase_1-2      (naturally coded; _Iclase_1 omitted)
i.rango_edades  _Irange_eda_1-6  (naturally coded; _Irange_eda_1 omitted)
i.p6210      _Ip6210_1-9      (naturally coded; _Ip6210_1 omitted)
i.p6070      _Ip6070_1-6      (naturally coded; _Ip6070_1 omitted)

      failure _d:  perdidos == 1
analysis time _t:  duracion_emple_infor/12
weight:  [pweight=fex_c_def]

Iteration 0:  log pseudolikelihood = -12212870
Iteration 1:  log pseudolikelihood = -10887976
Iteration 2:  log pseudolikelihood = -10209567
Iteration 3:  log pseudolikelihood = -10189690
Iteration 4:  log pseudolikelihood = -10189658
Iteration 5:  log pseudolikelihood = -10189658

Exponential regression -- log relative-hazard form

No. of subjects      = 10829558.97      Number of obs      = 48689
No. of failures      = 3270908.54
Time at risk         = 53761405.19

Log pseudolikelihood = -10189658      Wald chi2(16)     = 32928.70
                                          Prob > chi2       = 0.0000

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
_Iclase_2	-1.281574	.0596386	-21.49	0.000	-1.398464	-1.164685
_Irange_ed~2	.2907749	.0683156	4.26	0.000	.1568788	.4246709
_Irange_ed~3	.0550069	.0546119	1.01	0.314	-.0520304	.1620442
_Irange_ed~4	-.409485	.0475443	-8.61	0.000	-.5026701	-.3162999
_Irange_ed~6	-.4192004	.1691844	-2.48	0.013	-.7507956	-.0876051
_Ip6210_2	-.8894717	.830622	-1.07	0.284	-2.517461	.7385175
_Ip6210_3	-.9372293	.0578219	-16.21	0.000	-1.050558	-.8239005
_Ip6210_4	-.5455736	.0646203	-8.44	0.000	-.6722272	-.41892
_Ip6210_5	-.3844218	.0615195	-6.25	0.000	-.5049977	-.2638459
_Ip6210_6	-.3016626	.0690322	-4.37	0.000	-.4369632	-.166362
_Ip6210_9	-3.293565	1.179427	-2.79	0.005	-5.6052	-.9819303
_Ip6070_2	-1.795438	.0579474	-30.98	0.000	-1.909013	-1.681864
_Ip6070_3	-1.921475	.0598313	-32.11	0.000	-2.038742	-1.804207
_Ip6070_4	-1.78048	.0636853	-27.96	0.000	-1.905301	-1.655659
_Ip6070_5	-2.390162	.0924209	-25.86	0.000	-2.571303	-2.20902
_Ip6070_6	-1.710189	.0599441	-28.53	0.000	-1.827677	-1.5927

Figura 35: Análisis paramétrico: Distribución exponencial

```

. xi:streg i.clase i.rango_edades i.p6210 i.p6070,dist(gomp) noconstant nohr
i.clase      _Iclase_1-2      (naturally coded; _Iclase_1 omitted)
i.rango_edades  _Irange_eda_1-6  (naturally coded; _Irange_eda_1 omitted)
i.p6210      _Ip6210_1-9      (naturally coded; _Ip6210_1 omitted)
i.p6070      _Ip6070_1-6      (naturally coded; _Ip6070_1 omitted)

      failure_d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
Fitting full model:

Iteration 0: log pseudolikelihood = -12785955
Iteration 1: log pseudolikelihood = -12167591
Iteration 2: log pseudolikelihood = -12109381
Iteration 3: log pseudolikelihood = -12109221
Iteration 4: log pseudolikelihood = -12109221

Gompertz regression -- log relative-hazard form

No. of subjects      = 10829558.97      Number of obs      = 48689
No. of failures     = 3270908.54
Time at risk        = 53761405.19

Wald chi2(16)      = 11156.04
Prob > chi2       = 0.0000

Log pseudolikelihood = -12109221

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
_Iclase_2	-1.198509	.0554026	-21.63	0.000	-1.307096	-1.089922
_Irange_eda_2	.2041474	.0631999	3.23	0.001	.0802779	.3280168
_Irange_eda_3	.029304	.0496644	0.59	0.555	-.0680364	-.1266444
_Irange_eda_4	-.3390457	.0414198	-8.19	0.000	-.4202271	-.2578644
_Irange_eda_6	-.5714599	.1675538	-3.41	0.001	-.8998593	-.2430604
_Ip6210_2	-.7158293	.7591245	-0.94	0.346	-2.203686	.7720274
_Ip6210_3	-.8100917	.0519974	-15.58	0.000	-.9120046	-.7081787
_Ip6210_4	-.4414128	.057451	-7.68	0.000	-.5540148	-.3288108
_Ip6210_5	-.2963333	.0547513	-5.41	0.000	-.4036438	-.1890227
_Ip6210_6	-.2223093	.0616343	-3.61	0.000	-.3431103	-.1015083
_Ip6210_9	-2.822848	1.170537	-2.41	0.016	-5.117058	-.5286382
_Ip6070_2	-1.637	.0554206	-29.54	0.000	-1.745622	-1.528377
_Ip6070_3	-1.749252	.0563932	-31.02	0.000	-1.859781	-1.638723
_Ip6070_4	-1.617375	.0605956	-26.69	0.000	-1.73614	-1.49861
_Ip6070_5	-2.172385	.0908821	-23.90	0.000	-2.35051	-1.994259
_Ip6070_6	-1.576373	.0570034	-27.65	0.000	-1.688098	-1.464649
/gamma	-.0583443	.0039576	-14.74	0.000	-.066101	-.0505876

Figura 36: Análisis paramétrico: Distribución Gompertz

```

. xi:streg i.clase i.rango_edades i.p6210 i.p6070,dist(loglogistic) noconstant
i.clase      _Iclase_1-2      (_Iclase_1 for clase==1 omitted)
i.rango_edades  _Irange_eda_1-6      (naturally coded; _Irange_eda_1 omitted)
i.p6210        _Ip6210_1-9      (naturally coded; _Ip6210_1 omitted)
i.p6070        _Ip6070_1-6      (naturally coded; _Ip6070_1 omitted)

      failure_d: perdidos == 1
      analysis time _t: duracion_emple_infor/12
      weight: [pweight=fex_c_def]
Fitting full model:

Iteration 0: log pseudolikelihood = -12856029 (not concave)
Iteration 1: log pseudolikelihood = -12059866
Iteration 2: log pseudolikelihood = -11959936
Iteration 3: log pseudolikelihood = -11863543
Iteration 4: log pseudolikelihood = -11862379
Iteration 5: log pseudolikelihood = -11862378

Loglogistic regression -- accelerated failure-time form

No. of subjects      = 10829558.97      Number of obs      = 48689
No. of failures      = 3270908.54
Time at risk         = 53761405.19

Log pseudolikelihood = -11862378      Wald chi2(16)      = 14929.86
                                          Prob > chi2         = 0.0000

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
_Iclase_2	1.151504	.0774585	14.87	0.000	.9996878	1.303319
_Irange_ed~2	-.5829534	.0871989	-6.69	0.000	-.7538602	-.4120467
_Irange_ed~3	-.4199095	.071443	-5.88	0.000	-.5599352	-.2798837
_Irange_ed~4	.1357452	.0590858	2.30	0.022	.0199392	.2515513
_Irange_ed~6	.5876807	.2148981	2.73	0.006	.1664881	1.008873
_Ip6210_2	.4154373	.9589163	0.43	0.665	-1.464004	2.294879
_Ip6210_3	.5340363	.0831571	6.42	0.000	.3710514	.6970211
_Ip6210_4	.0491417	.0892966	0.55	0.582	-.1258765	.2241599
_Ip6210_5	-.1275738	.0870979	-1.46	0.143	-.2982825	.043135
_Ip6210_6	-.160959	.0925	-1.74	0.082	-.3422556	.0203377
_Ip6210_9	2.60769	3.259596	0.80	0.424	-3.781001	8.996381
_Ip6070_2	2.198626	.0859104	25.59	0.000	2.030245	2.367007
_Ip6070_3	2.36718	.0856847	27.63	0.000	2.199241	2.535119
_Ip6070_4	2.124033	.0920096	23.08	0.000	1.943698	2.304369
_Ip6070_5	2.664386	.1286514	20.71	0.000	2.412233	2.916538
_Ip6070_6	2.13896	.0884903	24.17	0.000	1.965522	2.312398
/ln_gam	.118957	.0102469	11.61	0.000	.0988734	.1390406
gamma	1.126321	.0115413			1.103927	1.149171

Figura 37: Análisis paramétrico: Distribución Log-Logistic

```

. xi:streg i.clase i.rango_edades i.p6210 i.p6070,dist(lognormal) noconstant
i.clase      _Iclase_1-2      (_Iclase_1 for clase==1 omitted)
i.rango_edades  _Irange_eda_1-6      (naturally coded; _Irange_eda_1 omitted)
i.p6210        _Ip6210_1-9      (naturally coded; _Ip6210_1 omitted)
i.p6070        _Ip6070_1-6      (naturally coded; _Ip6070_1 omitted)

      failure_d:  perdidos == 1
      analysis time _t:  duracion_emple_infor/12
      weight:  [pweight=fex_c_def]
Fitting full model:

Iteration 0:  log pseudolikelihood = -12755174 (not concave)
Iteration 1:  log pseudolikelihood = -12038847
Iteration 2:  log pseudolikelihood = -11947135
Iteration 3:  log pseudolikelihood = -11889378
Iteration 4:  log pseudolikelihood = -11889008
Iteration 5:  log pseudolikelihood = -11889008

Lognormal regression -- accelerated failure-time form

No. of subjects      = 10829558.97      Number of obs      = 48689
No. of failures      = 3270908.54
Time at risk         = 53761405.19

Log pseudolikelihood = -11889008      Wald chi2(16)      = 12195.73
                                          Prob > chi2         = 0.0000

```

_t	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
_Iclase_2	1.126374	.0789884	14.26	0.000	.9715598	1.281189
_Irange_ed~2	-.5130834	.0901786	-5.69	0.000	-.6898303	-.3363365
_Irange_ed~3	-.3700685	.0757065	-4.89	0.000	-.5184506	-.2216865
_Irange_ed~4	.1580901	.0641831	2.46	0.014	.0322935	.2838867
_Irange_ed~6	.5188026	.2131962	2.43	0.015	.1009456	.9366595
_Ip6210_2	.5417808	.7714673	0.70	0.483	-.9702673	2.053829
_Ip6210_3	.5206136	.0864659	6.02	0.000	.3511434	.6900837
_Ip6210_4	.0239964	.0935061	0.26	0.797	-.1592721	.207265
_Ip6210_5	-.1627759	.0917966	-1.77	0.076	-.3426939	.017142
_Ip6210_6	-.1677093	.0977641	-1.72	0.086	-.3593234	.0239047
_Ip6210_9	1.718299	2.2026	0.78	0.435	-2.598719	6.035317
_Ip6070_2	2.232165	.0881928	25.31	0.000	2.05931	2.40502
_Ip6070_3	2.450084	.08898	27.54	0.000	2.275687	2.624482
_Ip6070_4	2.165884	.0952757	22.73	0.000	1.979147	2.352621
_Ip6070_5	2.711954	.1362523	19.90	0.000	2.444904	2.979003
_Ip6070_6	2.183075	.0912806	23.92	0.000	2.004168	2.361981
/ln_sig	.7294137	.0097105	75.12	0.000	.7103815	.7484459
sigma	2.073864	.0201382			2.034767	2.113712

Figura 38: Análisis paramétrico: Distribución Log-Normal



Figura 39: Relaciones espaciales: Matriz W

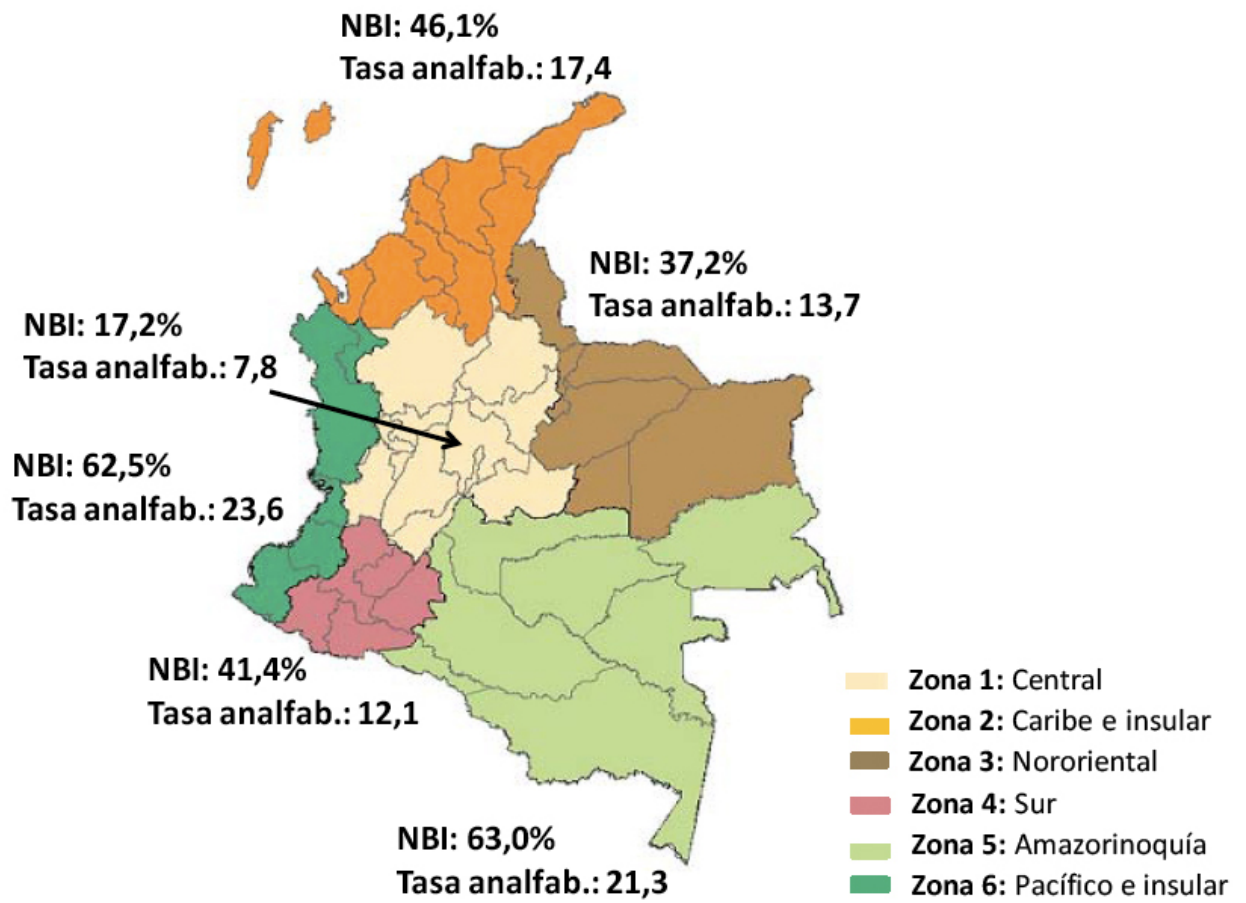


Figura 40: Áreas planificación estratégica Plan Nacional de Desarrollo

RANGO	DURACION	DEPARTAMENTO	MUNICIPIO
Entre 29 y 40 meses	29.25	Antioquia	Barbosa
	33.00704	Valle del Cauca	Buenaventura
	33.04819	Valle del Cauca	El Águila
	33.06452	Risaralda	Balboa
	33.11966	Valle del Cauca	Tuluá
	34.03448	Antioquia	Puerto Triunfo
	34.14118	Valle del Cauca	Dagua
	34.21053	Tolima	Armero
	35.02174	Norte de Santander	Santiago
	35.74074	Boyacá	Chivatá
	36.60656	Quindío	La Tebaida
	37.69792	Quindío	Quimbaya
	38.11579	Meta	Puerto Concordia
	38.29473	Córdoba	Valencia
	38.52	Antioquia	Ituango
39.23595	Tolima	Rovira	
39.65079	Valle del Cauca	Sevilla	

Figura 41: Tabla duración informalidad

Entre 41 y 60 meses	40.19418	Cundinamarca	Madrid
	40.49398	Valle del Cauca	Yumbo
	40.53165	Córdoba	Los Córdoba
	41.39535	Norte de Santander	El Zulia
	41.64179	Tolima	Lérida
	42.40196	Risaralda	La Celia
	43.13861	Risaralda	Belén de Umbria
	43.15022	Cesar	Aguachica
	43.35833	Tolima	Chaparral
	43.4067	Quindío	Calarca
	43.76923	Boyacá	Toca
	43.87479	Valle del Cauca	Calli
	43.91045	Boyacá	Puerto Boyacá
	44.35135	Norte de Santander	San Cayetano
	44.44828	Valle del Cauca	Palmira
	44.45454	Cundinamarca	Girardot
	44.98969	Córdoba	Tierralta
	45.13091	Santander	Girón
	45.18518	Cundinamarca	Soacha
	45.33663	Cundinamarca	Mosquera
	45.90244	Santander	Suaita
	45.98404	Antioquia	Bello
	47.83231	Risaralda	Dosquebradas
	47.88158	Caldas	Villamaría
	48.66766	Santander	Bucaramanga
	49.13445	Caldas	Aguadas
	49.32867	Córdoba	Lorica
	49.33333	Boyacá	Oicatá
	49.58468	Santander	Piedecuesta
	50.54482	Tolima	Ibagué
	50.63326	Risaralda	Perelra
	51.06911	Santander	Floridablanca
	51.71154	Boyacá	Ventaquemada
	51.74648	Santander	Puente Nacional
	52.38281	Antioquia	Abejorral
	52.56667	Antioquia	Marinilla
	52.58647	Meta	San Martín
	52.60255	Norte de Santander	Cúcuta
	52.7234	Antioquia	Itagüí
	53.05781	Meta	Villavicencio
	53.11864	Valle del Cauca	El Cerrito
	53.13044	Antioquia	Sabaneta
	53.70887	Caquetá	Florencia
	53.73643	Caquetá	La Montañita
	54.12861	La Guajira	Riohacha
	54.21905	Norte de Santander	Villa del Rosario
	54.25032	Antioquia	Medellín
	54.39706	Norte de Santander	Los Patios
	54.53409	Risaralda	La Virginia
	54.56667	Cesar	Valledupar
	54.67188	Antioquia	Retiro
	55.17961	Bogotá, D.C.	Bogotá, D.C.
	55.17998	Nariño	Pasto
	55.8705	Atlántico	Malambo
	56	Cundinamarca	El Rosal
	57.10084	Córdoba	Sahagún
	57.47802	Sucre	San Luis de Sincé
	57.74792	Córdoba	Montería
	57.88679	Antioquia	Santa Bárbara
	58.056	Caldas	La Dorada
58.20716	Quindío	Armenia	
58.44407	Huila	Neiva	
58.99206	La Guajira	Dibulla	
59.00496	Atlántico	Soledad	
59.2875	Huila	Gigante	
59.74854	Risaralda	Pueblo Rico	

Figura 42: Tabla duración informalidad-1

Entre 61 y 80 meses	60.46817	Magdalena	Santa Marta
	60.65517	Antioquia	Tarazá
	60.76864	Cauca	Popayán
	60.84211	Caldas	Riosucio
	61.16667	Norte de Santander	San Calixto
	61.26534	Caldas	Manizales
	61.40458	La Guajira	San Juan del Cesar
	61.70968	Meta	Mapiripán
	62.26923	Cauca	Corinto
	62.84127	Antioquia	Urrao
	64.0297	Cauca	Santander de Quilichao
	64.11864	Norte de Santander	Durania
	64.14815	Norte de Santander	Ocaña
	64.36995	Chocó	Quibdó
	64.73077	Antioquia	Caldas
	64.98125	Cesar	Pailitas
	65.67423	Sucre	Sincelejo
	65.77413	Boyacá	Tunja
	65.82758	Santander	Jesús María
	65.84252	Cesar	Tamalameque
	66.00476	Caquetá	San Vicente del Caguán
	66.1203	Cauca	Villa Rica
	66.14103	Bolívar	Morales
	66.20834	Antioquia	Copacabana
	66.63953	Bolívar	Talagua Nuevo
	67.78049	Nariño	Ricaurte
	68.44949	Magdalena	Ciénaga
	68.76471	Antioquia	Girardota
	68.94024	Bolívar	Cartagena
	69.21085	Sucre	San Benito Abad
	69.27178	Atlántico	Barranquilla
	69.29214	Norte de Santander	Tibú
	69.40426	Boyacá	Nuevo Colón
	69.59504	Boyacá	Sogamoso
	70.05479	Antioquia	Envigado
	71.66666	Bolívar	Mahates
	71.86111	Boyacá	Betétiva
	72.32	Cundinamarca	Quebradanegra
	72.45544	Santander	San Gil
	74.0625	Sucre	Buenavista
75.15929	Boyacá	Tasco	
75.33929	Nariño	La Unión	
75.41666	Santander	Málaga	
75.56989	Bolívar	Arjona	
76.35802	Atlántico	Campo de La Cruz	
77.96	Boyacá	Duitama	
79.40741	Nariño	La Tola	
79.90455	Huila	San Agustín	

Figura 43: Tabla duración informalidad-2

Entre 81 y 100 meses	81.97087	Nariño	San Andres de Tumaco
	83.04546	Antioquia	La Estrella
	83.11475	Cauca	Inzá
	83.20513	Santander	Barrancabermeja
	83.22156	Huila	La Plata
	83.65854	Norte de Santander	Teorama
	84.25	Cundinamarca	Une
	85.5	Santander	Gambita
	85.69841	Nariño	El Charco
	88.3	Nariño	Ipiales
	88.47619	Cauca	Totoró
	89.38182	Cundinamarca	La Peña
	92.38298	Cauca	Silvia
	94.12122	Nariño	Mallama
	94.875	Cundinamarca	Nimalma
95.10363	Magdalena	Concordia	
99.01176	Cundinamarca	Chipaque	
99.66154	Nariño	Mosquera	
Hasta 140 meses	100.4744	Chocó	El Carmen de Atrato
	102.8468	Chocó	Unguía
	117.6336	Chocó	Bajo Baudó
	124.2459	Chocó	Bagadó
	138.25	Santander	Albania

Figura 44: Tabla duración informalidad-3

Bibliografía

- Jorge Bula. *Empleo de calidad o calidad del empleo en tiempo de crisis*. Number 328 in Contraloría General de la República. Economía Colombiana, 2009.
- C Castellar and J Uribe. Determinantes de la duración del desempleo en el área metropolitana de cali, 1988-2000. *Archivos de Economía*, (218), 2003.
- DANE. Metodología informalidad gran encuesta integrada de hogares. Technical report, 2009.
- Juan Carlos Guataquí, Andrés García, and Mauricio Rodríguez. El perfil de la informalidad laboral en colombia. *SERIE DE DOCUMENTOS DE TRABAJO*, (95):31, 2011.
- Jude C. Hays and Aya Kachi. Interdependent duration models in political science. Preparado para la reunión anual de la Asociación Americana de Ciencias Políticas, 2009.
- IOE. La economía informal en america latina. respuestas al cuestionario. XXI REUNION DE PRESIDENTES DE ORGANIZACIONES EMPRESARIALES IBEROAMERICANAS, 2010.
- Tony Lancaster. Econometric methods for the duration of unemployment. *Econometrica*, 47(4), 1979.
- Hugo Lopez. Mercado laboral urbano y desempleo friccional y estructural en colombia: el papel del sena. *Planeación y desarrollo*, pages 257–290, 1994.
- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.
- Hermes Fernando Martínez. Cuánto duran los colombianos en el desempleo y el empleo?: Un análisis de supervivencia. *Documentos CEDE*, (35):42, 2003.
- Wilson Mayorga. Determinantes sectoriales del desempleo. Technical Report 293, Departamento Nacional de Planeación, 2005.
- Jairo Nuñez and Raquel Bernal. El desempleo en colombia: tasa natural, desempleo cíclico y estructural y la duración del desempleo (1976-1998). *Ensayos sobre Política Económica*, (32), 1997.
- J Ocampo and M Ramírez. Principales conclusiones y recomendaciones de la misión de empleo en el problema laboral colombiano: diagnóstico, perspectivas y políticas. informe final de la misión de empleo. *Economía Colombiana*, (10), 1986.
- Plan nacional de desarrollo 2010-2014. Technical report, 2011.
- Guillermo Perry, William Maloney, Omar Arias, Pablo Fajnzylber, Andrew Mason, and Jaime Saavedra. *INFORMALIDAD: ESCAPE Y EXCLUSIÓN*. BANCO MUNDIAL, 2008.
- Paola Roldan and Carlos Hernandez. Quienes terminan en la informalidad? impacto de las características y el tiempo de búsqueda. *Revista de Economía del Caribe*, (4):149–180, 2009.
- Fabio Sánchez and Oriana Álvarez. Un diagnostico de la informalidad laboral y los costos laborales en colombia y una propuesta de politica para incentivar la formacion laboral. 2011.
- Cristina Sánchez. Duracion del desempleo en bucaramanga: Un analisis de supervivencia 2006. Master's thesis, UNIVERSIDAD NACIONAL DE COLOMBIA, 2010.
- Fabio Sánchez, Valentina Duque, and Mauricio Ruíz. Costos laborales y no laborales y su impacto sobre el desempleo, la duración del desempleo y la informalidad en colombia, 1980-2007. *Documentos CEDE*, 2009.
- J Tenjo and R Ribera. Participación, desempleo y mercados laborales en colombia. *Archivos de Macroeconomía*, (81), 1998.

Carlos Augusto Viáfara and José Ignacio Uribe. Duración del desempleo y canales de búsqueda de empleo en Colombia, 2006. *Departamento Nacional de Planeación*, (340), 2008.