

# Latent structures of the care economy in Colombia

An approach with mixture models,  
econometric models and statistical learning

**Autors:**

Jose Miguel León   
Laura Estefany Pardo 

Department of Statistics  
Faculty of Science

National University of Colombia



# Contenido

## 1 Introducción

- Contextualización
- Encuesta ENUT - DANE

## 2 Metodología

- Tratamiento de Datos
- Selección de Variables

## 3 Hallazgos Preliminares

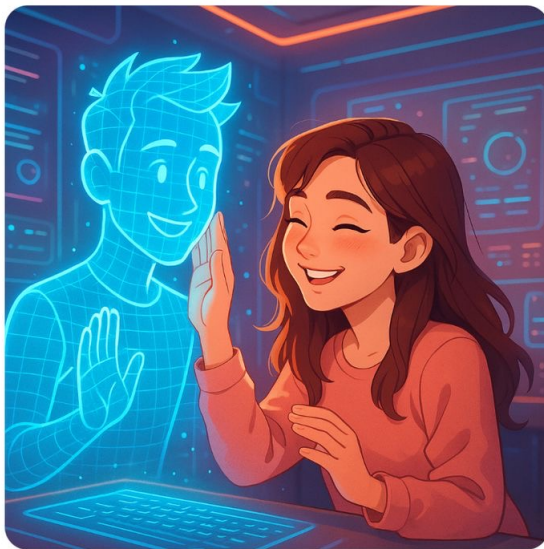
## 4 Modelos Implementados

- Desarrollo
- Comparativa

## 5 Conclusiones

- Recomendaciones

# ¿Existe algo que la IA no haga mejor?



# La Economía del Cuidado y el Trabajo de Cuidado No Remunerado

- **¿Qué es?**

- Actividades no remuneradas en el hogar:
  - Mantenimiento del hogar
  - Cuidado de personas
  - Apoyo a la fuerza laboral remunerada

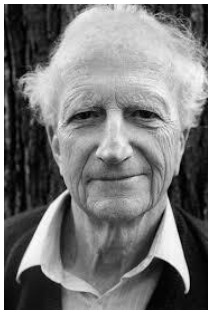
- **Importancia económica:**

- Representa aproximadamente el **20% del PIB**
- Si fuera remunerado, el **TCNR sería el mayor sector económico** del país

- **Relevancia social:** Clave para el bienestar individual y colectivo

- **Motivación del estudio:**

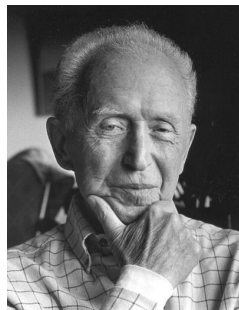
- Comprender la complejidad del TCNR para formular estrategias que:
  - Aumenten la productividad
  - Alivien la carga del cuidado



Gary S. Becker



Margaret G. Reid



Jacob Mincer

## Ley 1413 de 2010

Obliga la inclusión de la **Economía del Cuidado** en el **Sistema de Cuentas Nacionales (SCN)**.

Reconoce la contribución del **trabajo no remunerado** al desarrollo económico y social.

## Artículo 4º

Instrumento principal: **Encuesta Nacional de Uso del Tiempo (ENUT)**.

Diseñada y aplicada por el **DANE** como organismo responsable.

Periodicidad: cada **3 años**.

## Encuesta Nacional del Uso del Tiempo en Colombia



## Tratamiento de Datos

- Se usó el capítulo 8 (Uso del Tiempo) y otros capítulos clave.
- Se filtraron personas mayores de 13 años.
- Se armonizaron datos sobre actividades de cuidado y personales.

## Selección de Variables

- Se inició con 397 variables → reducción progresiva.
- Se eliminaron variables con alto porcentaje de NA.
- Selección final con Boruta, Lasso y Elastic Net.
- Baja colinealidad:  $\rho < 0.5$  y Cramér's V moderado.

## Imputación de Datos

- Variables: educación (14%), salud (6.7%), estrato (0.76%).
- Método: MICE      Modelos: multinomial y logit ordenado.

# Hallazgos Preliminares

- **Cobertura poblacional**  
≈ **90%** de las mujeres mayores de 13 años reportaron realizar trabajo no remunerado en el hogar.
- **Brecha de tiempo por sexo**  
En promedio, las mujeres dedican **4.4 horas/semana**, mientras que los hombres dedican **2 horas/semana**.
- **Relación de la educación**  
A mayor **nivel educativo** se reporta una menor carga de labores de cuidado [3-4 horas/semana]
- **Responsable de las compras habituales**  
Se realiza la identificación a partir de tiempo de viaje al trabajo, tiempo dedicado al cuidado personal, actividades frecuentes, etc.



## HDBSCAN

Grupo	Número de observaciones	Observaciones
5	69,502	Grupo principal
-1	47,928	Ruido o outliers
3	1,432	Subgrupo significativo
1	732	Subgrupo pequeño
2	379	Subgrupo pequeño
0	350	Subgrupo pequeño
4	330	Subgrupo pequeño

Table: Distribución de observaciones por grupo según HDBSCAN

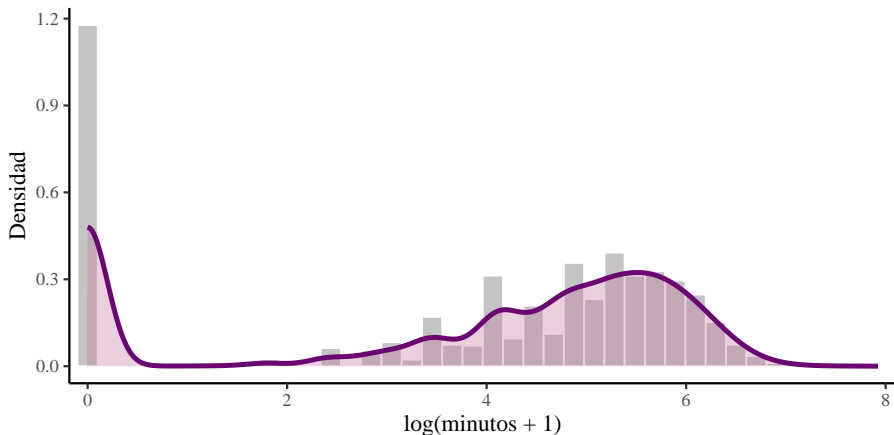
## Kmeans

Clúster	Minutos promedio	Horas promedio
1	341.6	5.69
2	106.5	1.78
3	0.005	0
4	23.6	0.39

Table: Resumen del tiempo en cuidado por clúster

# Distribución del Tiempo de Cuidado No Remunerado (log)

Transformación logarítmica de minutos diarios reportados



# Hurdle Model

**Objetivo:** Modelar una variable  $y$  que toma valores positivos solo para algunos individuos (participantes).

**Supuestos clave:**

- $d = 1$  si el individuo participa ( $y > 0$ ),  $d = 0$  si no ( $y = 0$ ).
- Se estima  $\Pr(d = 1|x)$  con un modelo logit/probit.
- Para participantes,  $f(y|d = 1, x)$  modela los valores positivos.

El *modelo de dos partes* para  $y$  se expresa entonces como

$$f(y|x) = \begin{cases} \Pr[d = 0|x] & \text{si } y = 0, \\ \Pr[d = 1|x]f(y|d = 1, x) & \text{si } y > 0. \end{cases}$$

La estructura general de un modelo de dos partes es:

$$E[\mathbf{Y}|\mathbf{X}] = \Pr(\mathbf{Y} > 0|\mathbf{X}) \times E[\mathbf{Y}|\mathbf{Y} > 0, \mathbf{X}]$$

# Hurdle Model<sup>1</sup>

## 1. Elección del modelo para participación:

- Logit o Probit, con formulación latente:  $d = 1$  si  $x'\beta + \varepsilon > 0$ .

## 2. Distribución de $y$ para los que participan:

- Distribuciones positivas: Log-normal, Gamma, NB truncada, etc.
- Se garantiza  $y > 0$  (normal truncada desde cero).

## 3. Estimación por máxima verosimilitud:

- Separada por partes:
  - Parte 1: Todas las observaciones (modelo binario).
  - Parte 2: Solo observaciones con  $y > 0$ .

---

<sup>1</sup>Cameron, A. C., & Trivedi, P. K. (2005).

# Negative Binomial Zero Inflated Model

Sea  $Y$  una variable aleatoria discreta (cantidad de minutos semales realizando trabajo de cuidado) y un proceso estructural (ser cuidador), entonces el modelo NBZI se especifica como

$$\begin{cases} Y_i^{\text{ind}} \sim \text{ZINB}(\mu_i, \phi, \tau, \pi_i), \\ g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \\ h(\pi_i) = \gamma_0 + \gamma_1 z_{i1} + \cdots + \gamma_q z_{iq} \end{cases} \quad \text{Nota: Adecuado en presencia de sobredispersión } (\approx 1.4)$$

Donde

$$Y \mid \nu = \begin{cases} 0 & \text{if } \nu = 1 \\ \text{Neg. Binomial}(\mu, \phi, \tau) & \text{if } \nu = 0 \end{cases} \quad \text{y } \nu \sim \text{Bernoulli}(\pi)$$

# Negative Binomial Zero Inflated Model<sup>2</sup>

La función de masa de probabilidad de  $Y$  es

$$f_Y(y; \mu, \phi, \pi) = \begin{cases} \pi + (1 - \pi) \left( \frac{\frac{\mu}{\phi}}{\mu + \frac{\mu}{\phi}} \right)^{\frac{\mu}{\phi}} & \text{if } y = 0 \\ (1 - \pi) \frac{\Gamma(y + \frac{\mu}{\phi})}{\Gamma(\frac{\mu}{\phi}) \Gamma(y + 1)} \left( \frac{\mu}{\mu + \frac{\mu}{\phi}} \right)^y \left( \frac{\frac{\mu}{\phi}}{\mu + \frac{\mu}{\phi}} \right)^{\frac{\mu}{\phi}} & \text{if } y = 1, 2, \dots \end{cases}$$

- $\mu$ : Número esperado del conteo dado que no es 0.
- $\phi$ : Controla la sobredispersión del conteo.
- Funciones de enlace:  $g(\mu_i) = \log(\mu_i)$  y  $h(\pi_i) = \text{logit}(\pi_i)$

---

<sup>2</sup>Garay et al., 2011

# GAM Model with Splines

El GAM es una extensión flexible del modelo lineal que permite capturar relaciones no lineales entre las variables independientes y la variable dependiente. En lugar de usar coeficientes lineales simples, emplea funciones suaves estimadas no paramétricamente (por ejemplo, splines) para cada predictor. Además, utiliza una función de enlace  $g(\cdot)$  que conecta la media esperada de la respuesta con la suma de estas funciones suaves:

$$g(\mu) = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p).$$

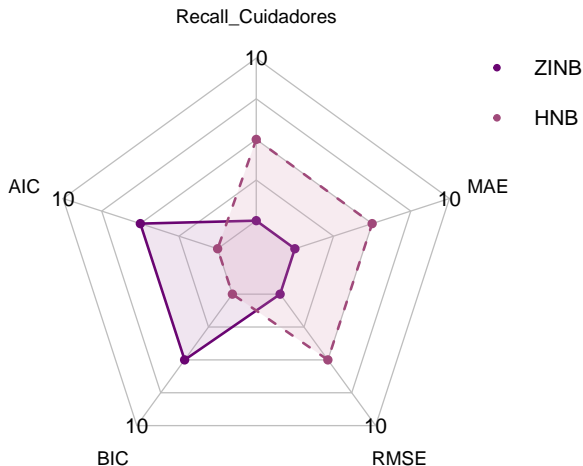
# Modelo completo

Edad	Sexo	Parentesco	Etnia
Régimen de Salud	Actividad Semana Pasada	Percepción del Cuidado	Percepción del Tiempo
Clase	Región	Estrato	Vivienda
Servicio Doméstico	Total del Hogar	Subsidio	Tiempo de Ayuda Recibida
Tiempo Trabajado	Tiempo de Viaje al Trabajo	Nivel Educativo	Tiempo Personal

Table: Variables incluidas en el modelo completo



## Evaluación ZINB vs HNB



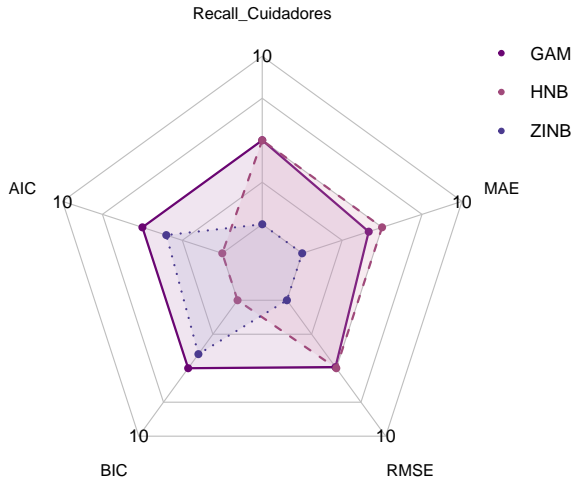
# Modelo reducido con interacción

Edad	Sexo $\times$ Parentesco	Sexo $\times$ Percepción del Cuidado
Percepción del Tiempo	Clase $\times$ Sexo	Estrato
Total del Hogar	Tiempo de Ayuda Recibida	Nivel Educativo
Sexo $\times$ Tiempo Personal		

Table: Variables e interacciones incluidas en el modelo reducido

# Modelo reducido con interacción

## Evaluación GAM HNB ZINB



# Resultados

## Proceso Estructural: No realizar cuidado

- Las mujeres tienen 96% más chances de realizar tareas de cuidado que los hombres.
- Por cada año adicional en la edad, el chance de no realizar cuidado aumenta en un 8.9%.
- A mayor número de personas en el hogar, el chance de no realizar cuidado aumenta en un 10%.
- El chance de realizar cuidado siendo mujer y pareja del jefe de hogar aumenta en 58%.
- Ser hija/hijo o nieta/nieto del jefe del hogar aumenta el chance de no realizar cuidado en 44% y 66% respectivamente.

# Resultados

## Intensidad de tiempo realizando cuidado

- El tiempo esperado dedicado a labores de cuidado por las mujeres es aproximadamente 2.22 veces mayor que los hombres.
- Por cada año adicional de edad, el tiempo esperado realizando cuidado disminuye en un factor de aproximadamente 0.13% por año.
- A mayor estrato el tiempo esperado realizando actividades de cuidado disminuye en un factor de 3.9%.
- El alcanzar un mayor nivel educativo reduce la media esperada de tiempo de cuidado en un factor de 29.8%

# Conclusiones

- La feminización persistente revela cómo el cuidado se naturaliza como extensión del rol femenino.
- Variables demográficas como la edad y la composición del hogar (número de integrantes) aumentan las chances de no realizar cuidado, lo que sugiere que la participación en estas tareas puede desplazarse hacia otros miembros en hogares más grandes y a medida que las personas envejecen.
- Una vez inmersas en tareas de cuidado, las mujeres dedican más del doble de tiempo que los hombres, reflejando no solo un mayor involucramiento, sino también una mayor carga de trabajo.

# Conclusiones

- El envejecimiento tiene un efecto de reducción en la intensidad del cuidado, aunque marginal (0.13% menos por cada año), indicando que las responsabilidades tienden a disminuir con la edad.
- La condición socioeconómica (estrato) actúa como un factor protector: personas de estratos más altos dedican menos tiempo al cuidado, lo que puede estar vinculado a la posibilidad de tercerizar estos servicios.
- Un mayor nivel educativo está fuertemente asociado con una disminución en la intensidad del tiempo de cuidado, lo que sugiere que las trayectorias educativas podrían reducir las barreras de género y redistribuir las cargas de cuidado.

Aún hay mucho camino por recorrer, algunas líneas de investigación adicionales para aquellos interesados en involucrarse a trabajar con estos datos son:

- Análisis de Clases Latentes
- Análisis Conjunto de las Ediciones de la Enut
  - Datos Panel
- Modelos Mixtos
  - Efectos Aleatorios
- Aplicaciones de Muestreo
  - Estimadores, CV, IC
  - Calibración
  - Uso de variables auxiliares
- Análisis con la desagregación de las actividades de cuidado



# Referencias I



Congreso de la República de Colombia. (2010). *Ley 1413 de 2010: Por la cual se regula la inclusión de la economía del cuidado en el sistema de cuentas nacionales*. Diario Oficial No. 47.890.



Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., & Lachos, V. H. (2011). *On estimation and influence diagnostics for zero-inflated negative binomial regression models*. Computational Statistics & Data Analysis, 55(3), 1304–1318.  
<https://doi.org/10.1016/j.csda.2010.09.019>



Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.



Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.



Kursa, M. B., & Rudnicki, W. R. (2010). *Feature Selection with the Boruta Package*. Journal of Statistical Software, 36(11), 1–13.  
<https://doi.org/10.18637/jss.v036.i11>



van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *mice: Multivariate Imputation by Chained Equations in R*. Journal of Statistical Software, 45(3), 1–67.  
<https://doi.org/10.18637/jss.v045.i03>



Zeileis, A., Kleiber, C., & Jackman, S. (2008). *Regression Models for Count Data in R*. Journal of Statistical Software, 27(8), 1–25.  
<https://doi.org/10.18637/jss.v027.i08>