

Factores que condicionan la Probabilidad de Cotizar a las pensiones de Retiro en México

Marco Hernando Albarrán Núñez¹

Resumen

El presente trabajo intenta mostrar evidencia empírica de los determinantes del ahorro en el sistema de pensiones en México. Para esto, se realiza un ejercicio de regresión de un modelo no lineal, Probit. Dicho modelo cuantifica la probabilidad de cotizar para una pensión de retiro (variable dependiente), es decir, los factores que condicionan la mayor o menor probabilidad de que los individuos ahorren en el sistema de pensiones, sin importar si es público o privado. Los resultados de la estimación del modelo Probit, sugieren que el bajo nivel de aportaciones al sistema de pensiones está relacionado con una serie de características socio-económicas, en particular con la educación, el género y el tipo de trabajo que se realiza. Además, el trabajo contiene una serie de pruebas que se realizan al modelo Probit que se propone, con el fin de comprobar que es un modelo adecuado, capaz de cuantificar de forma correcta la probabilidad de que los individuos decidan cotizar al sistema de pensiones.

¹ Facultad de Ciencias, Universidad Nacional Autónoma de México, marcoalbarran@ciencias.unam.mx

I. Introducción

En el caso de los niveles de cotización al sistema de pensiones, existen estudios (Hernández, 2001; Murillo & Venegas, 2011; Tuesta, 2014; OCDE, 2014; Carpeta & Peña, 2015) que sugieren que factores como la educación, la edad, el número de integrantes de las familias y ciertas características económicas, sociales y demográficas, así como factores laborales, son relevantes para explicar el nivel de ahorro que realiza un individuo para su pensión de retiro.

El objetivo del presente trabajo es identificar los factores que afectan a la probabilidad de ahorrar en un sistema de pensiones formal, ya sea público o privado.

Las variables del modelo propuesto se obtienen de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)² 2014, que lleva a cabo el INEGI. La ENIGH 2014 proporciona información sobre la distribución, monto y estructura del ingreso y gasto de los hogares; adicionalmente, ofrece información sobre las características sociodemográficas y ocupacionales de los integrantes del hogar. El conjunto de variables exógenas que se utilizan en el modelo Probit que se propone en este trabajo son variables de tipo personal, demográfico, familiar y laboral. Con dichas variables se pretende encontrar la probabilidad de que los individuos ahorren para una pensión de retiro.

² La ENIGH es una encuesta de corte transversal, con representatividad nacional y para el conjunto de las áreas urbanas y rurales del país.

II. Datos

A continuación se describen las variables del modelo que se propone.

Cuadro 1. Descripción de las variables seleccionadas

VARIABLE	DESCRIPCIÓN
Ahorro_AFORE	Variable dependiente del modelo. Es una variable binaria, que toma los valores 0 y 1. Donde Ahorro_AFORE = 1 indica que el individuo ahorra en una AFORE y/o fondo de pensiones y Ahorro_AFORE = 0 si no ahorra.
Mujer	Variable binaria, donde Mujer=1 indica que es mujer y Mujer=0 que es hombre.
Edad	Variable de tipo numérica que indica el número de años de las personas.
NE_2	Variable binaria, que toma el valor de 1 si el individuo cuenta con educación básica completa y 0 en otro caso.
NE_4	Variable binaria, que toma el valor de 1 si el individuo cuenta con educación superior completa y 0 en otro caso.
Etnia	Variable binaria, donde Etnia =1 indica que la persona pertenece a algún grupo étnico y Etnia = 0 en caso contrario.
Casado	Variable binaria, donde Casado =1 significa que la persona vive Casado/a o en pareja y Casado =0 en caso contrario.
tot_integ	Variable de tipo numérica que indica el número total de integrantes en el hogar.
FamMen	Variable binaria, que toma el valor de 1, sí la familia tiene menores de edad o el valor de 0 en caso de que no.
Fam65mas	Variable binaria, que toma el valor de 1, sí la familia tiene integrantes mayores de 65 años; o el valor de 0 cuando no los tenga.
FamTransfer	Variable binaria, que indica si la familia recibe transferencias de cualquier tipo (becas, donativos, remesas y beneficios del gobierno).
FamconJubil	Variable binaria, que toma el valor de 1, sí la familia tiene jubilados o el valor de 0 en caso de que no.
Informal_trab	Variable binaria, donde Informal_trab = 1 indica que el trabajador no tiene un contrato de trabajo formal; Informal_trab = 0 en otro caso.
TamEmp_1	Variable binaria, que toma el valor de 1 si el trabajador labora en una empresa pequeña y 0 en otro caso.
TamEmp_3	Variable binaria, que toma el valor de 1 si el trabajador labora en una empresa grande y 0 en otro caso.
TipoEmpresa_2	Variable binaria, que toma el valor de 1 si el trabajador labora en una empresa privada no familiar y 0 en otro caso.

III. Metodología. Modelo Probit

En un modelo Probit binomial la variable dependiente puede tomar sólo dos valores: 0 y 1. Éstos pueden representar el éxito o el fracaso de un evento, por ejemplo: ahorrar para una pensión de retiro, puede significar la ocurrencia del evento; dado un conjunto de variables explicativas.

De acuerdo con Cameron y Trivedi (2005), la probabilidad de que ocurra el evento en cuestión, está determinada por la siguiente expresión

$$P_i = \Pr(\mathbf{y}_i = 1|\mathbf{x}') = F(\mathbf{x}'\beta) \quad (1)$$

La cuál contiene una matriz \mathbf{x} de variables explicativas del modelo, β el conjunto de coeficientes. Además, F es una función paramétrica de $\mathbf{x}'\beta$, generalmente es una función de distribución acumulada que asegura el cumplimiento de $0 \leq P \leq 1$. Es decir, que la probabilidad esté dentro del intervalo cerrado 0,1.

Al utilizar un modelo Probit, de la expresión (1), la función de distribución acumulada $F(\cdot)$ se convierte en la FDA de la normal estándar la cual se suele denotar como: Φ , de esta forma la expresión (1) se transforma en

$$P_i = \Pr(\mathbf{y}_i = 1|\mathbf{x}') = \Phi(\mathbf{x}'\beta) \quad (2)$$

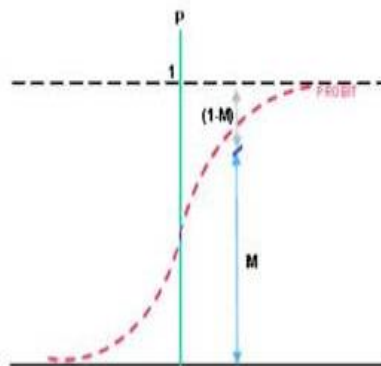
Los modelos Probit, parten de la función de probabilidad condicional:

$$P_i = \Pr(\mathbf{y}_i = 1|\mathbf{x}') = \Phi(\mathbf{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z) dz \quad (3)$$

donde $\Phi(\cdot)$ es la función de distribución de una variable aleatoria normal estándar, con derivada $\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$, la cuál es la función de densidad normal estándar.

El modelo Probit, produce estimaciones que se encuentran en el intervalo $[0,1]$. Como se puede apreciar en la gráfica 1.

Gráfica 1. Modelo Probit



Fuente: Gujarati, 2003.

Al no ser no lineal en los parámetros, el modelo Probit es estimado por máxima verosimilitud. Dada una muestra (y_i, x_i) , donde $i = 1, \dots, N$ en la cual se asume independencia entre las i observaciones, la función de densidad del i -ésimo individuo de la muestra, se representa por

$$f(y_i | \mathbf{x}_i) = p_i^{y_i} (1 - p_i)^{1 - y_i}, \quad y_i = 0, 1 \quad (4)$$

Dado que $p_i = F(\mathbf{x}_i' \beta)$ y que las N observaciones de la muestra son independientes, la función del logaritmo de verosimilitud es

$$L_N(\beta) = \sum_{i=1}^N [y_i \ln F(\mathbf{x}_i' \beta) + (1 - y_i) \ln \{1 - F(\mathbf{x}_i' \beta)\}] \quad (5)$$

Al diferenciar la expresión (5) con respecto a β , se obtienen las condiciones de primer orden, que permiten estimar los parámetros del vector $\hat{\beta}$. Sin embargo, al no existir una solución explícita para $\hat{\beta}$, es necesario emplear métodos iterativos como el Newton-Raphson que converge rápidamente para el modelo Probit.

III.1. Estimación del modelo

El fenómeno de estudio de este trabajo, puede ser modelado con un modelo Probit, ya que la variable explicada es cualitativa y binaria, de esta forma se puede encontrar la probabilidad de que un individuo ahorre para su pensión de retiro.

La variable dependiente del modelo propuesto es el ahorro que realizan los individuos para su pensión de retiro, es decir, si una persona ahorra en una AFORE y/o fondo de pensión. Esta variable es discreta y toma el valor de 0 cuando la persona no ahorra para su pensión de retiro y el valor de 1, sí la persona ahorra para su pensión de retiro.

El modelo propuesto indica que la decisión de ahorrar para pensiones depende de un conjunto de variables exógenas, incluidas en el vector \mathbf{x}' , de modo que

$$P_i = \Pr(\mathbf{y}_i = 1 | \mathbf{x}') = \text{"Probabilidad de ahorrar en el sistema de pensiones"}$$

Con el objetivo de identificar las variables explicativas más relevantes para entender la decisión de un individuo de cotizar al sistema de pensiones³, se han incluido dos grupos de variables explicativas: en primer lugar, características sociodemográficas de la población como género, edad, estado civil, estudios, tipo de familia y en segundo lugar, variables del mercado de trabajo (sector, tipo de compañía, relación laboral, tipo de contrato, empleo formal).

Entonces, el conjunto de todas las variables explicativas, viene dado por el siguiente vector

$$\mathbf{x} = [\text{Mujer}_i, \text{Edad}_i, \text{NE}_2_i, \text{NE}_4_i, \text{Etnia}_i, \text{Casado}_i, \text{tot_integ}_i, \text{FamMen}_i, \text{Fam65mas}_i, \text{FamTransfer}_i, \text{FamconJubil}_i, \text{Informal_trab}_i, \text{TamEmp}_1_i, \text{TamEmp}_3_i, \text{TipoEmpresa}_2_i]$$

Donde el subíndice i denota la i -ésima observación para cada una de las variables independientes.

³ Sin importar si el ahorro es a un sistema voluntario u obligatorio.

Se estiman dos modelos Probit, en ambos modelos la variable dependiente es la misma: el ahorro que realizan los individuos al sistema de pensiones, la cual es una variable binaria que toma los valores 0 y 1. En el primer modelo, denominado “Modelo_1” se consideran solo variables sociodemográficas como explicativas, el número de observaciones de la muestra con el que se efectúa la estimación de este modelo es de 58,902 individuos; si también se incluyen variables del mercado de trabajo como se hace en el segundo modelo llamado “Modelo_2”, la muestra se reduce a 21,087 observaciones, este último modelo es más completo y produce mejores estimaciones para conocer la probabilidad de cotizar al sistema de pensiones.⁴ Las estimaciones en STATA arrojan los siguientes resultados:

Cuadro 2. Estimación de parámetros del Modelo_1

```

Iteration 0: log likelihood = -39111.884
Iteration 1: log likelihood = -32839.753
Iteration 2: log likelihood = -32815.44
Iteration 3: log likelihood = -32815.431
Iteration 4: log likelihood = -32815.431

Probit regression                               Number of obs   =       58902
                                                LR chi2(11)     =       12592.91
                                                Prob > chi2     =         0.0000
Log likelihood = -32815.431                    Pseudo R2      =         0.1610

```

Ahorro_AFORE	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Mujer	-.4815093	.0114103	-42.20	0.000	-.503873 - .4591456
Edad	.0131116	.0004173	31.42	0.000	.0122938 .0139295
NE_2	-.3586964	.0129008	-27.80	0.000	-.3839814 -.3334113
NE_4	.6001477	.0187894	31.94	0.000	.5633211 .6369742
Etnia	-.2931164	.0127634	-22.97	0.000	-.3181322 -.2681005
Casado	.425323	.0130728	32.54	0.000	.3997009 .4509451
tot_integ	-.0617182	.0034134	-18.08	0.000	-.0684083 -.0550281
FamMen	.2190085	.014048	15.59	0.000	.191475 .2465419
Fam65mas	-.3849298	.0173211	-22.22	0.000	-.4188786 -.3509811
FamTransfer	-.3270451	.0133625	-24.47	0.000	-.353235 -.3008552
FamconJubil	.6711335	.0183412	36.59	0.000	.6351854 .7070817
_cons	-.2197752	.0250475	-8.77	0.000	-.2688673 -.1706831

Cuadro 3. Estimación de parámetros del Modelo_2

```

Iteration 0: log likelihood = -13731.751
Iteration 1: log likelihood = -9070.1727
Iteration 2: log likelihood = -8968.5717
Iteration 3: log likelihood = -8968.4619
Iteration 4: log likelihood = -8968.4619

Probit regression                               Number of obs   =       21087
                                                LR chi2(15)     =       9526.58
                                                Prob > chi2     =         0.0000
Log likelihood = -8968.4619                    Pseudo R2      =         0.3469

```

Ahorro_AFORE	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Mujer	-.1308873	.0237919	-5.50	0.000	-.1775185 -.0842561
Edad	.0205707	.0009458	21.75	0.000	.018717 .0224245
NE_2	-.1918119	.0253278	-7.57	0.000	-.2414535 -.1421703
NE_4	.1547051	.0341963	4.52	0.000	.0876816 .2217286
Etnia	-.2382944	.0235655	-10.11	0.000	-.2844819 -.1921069
Casado	.2362565	.0255345	9.25	0.000	.1862097 .2863032
tot_integ	-.0569719	.0061266	-9.30	0.000	-.0689797 -.044964
FamMen	.2041087	.0263682	7.74	0.000	.1524279 .2557895
Fam65mas	-.1411458	.0341197	-4.14	0.000	-.2080192 -.0742724
FamTransfer	-.1599127	.0244501	-6.54	0.000	-.207834 -.1119914
FamconJubil	.4162997	.0414044	10.05	0.000	.3351486 .4974508
Informal_t~b	-1.235255	.0269561	-45.82	0.000	-1.288088 -1.182422
TamEmp_1	-.2827004	.0247934	-11.40	0.000	-.3312945 -.2341063
TamEmp_3	.3116405	.0342064	9.11	0.000	.2445972 .3786839
TipoEmpres~2	.4953935	.0242013	20.47	0.000	.4479598 .5428272
_cons	.5727589	.0550714	10.40	0.000	.464821 .6806968

⁴ Para mayor detalle, ver el apartado sobre la Bondad de Ajuste del Modelo.

Cuadro 4. Estimación de parámetros del Modelo_1 y Modelo_2

Variable	Modelo_1	Modelo_2
Mujer	-.4815093***	-.13088727***
Edad	.01311164***	.02057072***
NE_2	-.35869636***	-.19181189***
NE_4	.6001477***	.15470512***
Etnia	-.29311638***	-.23829442***
Casado	.42532299***	.23625646***
tot_integ	-.06171822***	-.05697189***
FamMen	.21900847***	.20410869***
Fam65mas	-.38492983***	-.14114581***
FamTransfer	-.32704511***	-.15991272***
FamconJubil	.67113355***	.41629968***
Informal_t~b		-1.2352549***
TamEmp_1		-.28270039***
TamEmp_3		.31164054***
TipoEmpres~2		.4953935***
_cons	-.21977517***	.57275889***
N	58902	21087
chi2	12592.907	9526.5787
r2_p	.16098568	.34688142

Legend: * p<0.05; ** p<0.01; *** p<0.001

Como se puede apreciar en la Cuadro 4, todas las variables del modelo son estadísticamente significativas al 99%. Además, al incluir las variables del mercado de trabajo (Modelo_2), el número de observaciones de la muestra se reduce y aumenta el Pseudo R² (pasa de 16% a 35%), es decir, el modelo es mejor, ya que aumenta su capacidad explicativa.

Como señalan Cameron y Trivedi (2005), la interpretación de los coeficientes en el modelo Probit no se puede hacer directamente, ya que este modelo no es lineal.⁵ Sin embargo, se puede interpretar el sentido de los signos, para mostrar si existe relación directa o inversa entre la variable explicativa asociada al parámetro y la probabilidad.

De acuerdo con la estimación final del modelo Probit propuesto (Cuadro 4), las variables explicativas sociodemográficas que tienen un efecto negativo sobre la variable dependiente (el ahorro personal en el sistema de pensiones) son, el género, el bajo nivel educativo, la pertenencia a algún grupo étnico, el tamaño de la familia, si la familia recibe transferencias de cualquier tipo, si la familia tiene integrantes mayores de 65 años. Por el lado del mercado de trabajo, las variables que más afectan la cotización al sistema de pensiones en México son, si el tipo de trabajo que se realiza es informal y si la empresa donde se labora es pequeña.

Así mismo, de la estimación del modelo (Cuadro 4), se puede apreciar que las variables independientes que inciden de forma positiva sobre la probabilidad de ahorrar en el sistema de pensiones son: tener un nivel educativo alto, estar casado/a o en pareja, vivir en familias con menores y familias donde algún integrante sea jubilado. Si el trabajador labora en una empresa grande y/o una empresa que sea privada no familiar, la probabilidad de que ahorre en el sistema de pensiones es mayor.

De los parámetros del modelo estimado, podemos decir que las cuestiones de género inciden sobre la probabilidad de ahorrar para una pensión de retiro. Según los resultados del modelo, las mujeres tienen menos probabilidad de cotizar al sistema de pensiones que los hombres. Esto puede ser explicado por la situación de desventaja en la que se encuentran las mujeres al momento de participar en el mercado laboral, donde existen problemas de desigualdad.

⁵ Ver el apartado de Efectos Marginales para la interpretación de los coeficientes del modelo propuesto.

De acuerdo con los resultados del modelo, la variable Edad tiene un impacto positivo en la probabilidad de ahorrar en el sistema de pensiones. Tomando como referencia el Modelo del Ciclo de Vida (MCV), a medida que pasa el tiempo, el ingreso y el ahorro del individuo promedio, van creciendo hasta que llegan a un nivel máximo, a partir del cual el ahorro decrece porque se empieza a gastar para hacerle frente a la vejez y/o a la imposibilidad de permanecer en el mercado laboral.

La probabilidad de ahorrar en el sistema de pensiones, es más alta para individuos que tienen un nivel de educación mayor. Este resultado está ligado con el signo positivo de la variable del modelo NE_4, que indica si un individuo posee educación superior completa; en contraste las personas que sólo tienen educación básica completa (variable NE_2), tienen una probabilidad menor de cotizar al sistema de pensiones. Según Hernández (2001), en México un nivel educativo mayor incentiva a participar en el sistema de pensiones. La mayor escolaridad, está ligada con la percepción de un mayor ingreso, lo que se traduce en un mayor ahorro para la jubilación.

El signo negativo en el parámetro Etnia, significa que pertenecer a un grupo étnico reduce la probabilidad de cotizar a un sistema de pensiones. Este hecho puede ser explicado por la dificultad que tienen las personas que pertenecen a etnias de acceder al mercado laboral y obtener un empleo formal que permita cotizar a un fondo de pensiones. Además, está el factor de la desinformación de los beneficios del ahorro para la vejez.

Los resultados de la estimación, indican que las personas que están casadas o viven en pareja, tienen una probabilidad mayor de ahorrar para una pensión de vejez que aquellas que están solteras. Asimismo, si el individuo pertenece a una familia donde vivan menores o una familia que ya incluya una persona jubilada, tienen una mayor probabilidad de ahorrar en el sistema de pensiones. Este último factor, vivir en familias con jubilados, indica que existen fuertes incentivos para ahorrar para la vejez cuando un miembro de la familia está disfrutando de una pensión de retiro. En contraste, la probabilidad de cotizar al sistema de pensiones se reduce cuando el individuo pertenece a una familia que recibe transferencias o a una familia que incluye a mayores de 65 años.

Por el lado de las variables del mercado laboral, el modelo estimado sugiere que si el trabajador labora en una empresa de tamaño pequeña (TamEmp_1), tiene una probabilidad de ahorrar en el sistema de pensiones menor que si trabaja en una empresa grande (TamEmp_3). Asimismo, la probabilidad de cotizar en pensiones es mayor cuando se trabaja en una empresa de tipo privada no familiar (TipoEmpresa_2). Un factor importante que inhibe el ahorro en el sistema de pensiones es el hecho de ser un trabajador informal (Informal_trab), ya que el signo negativo de la estimación del modelo, nos dice que los individuos que no tienen un empleo formal tienen una probabilidad menor de ahorrar para pensiones que aquellos trabajadores que tienen un contrato laboral formal.

Al utilizar un modelo Probit, la interpretación de los coeficientes, no se puede realizar directamente, ya que el modelo estimado no es lineal. Sin embargo, en el apartado de efectos marginales se explicará detalladamente cómo se usan los parámetros en la interpretación del modelo.

En el siguiente apartado, se realizarán diversas pruebas al modelo Probit que se propone, con el fin de comprobar que es un modelo adecuado, capaz de cuantificar de forma correcta la probabilidad de que los individuos decidan cotizar al sistema de pensiones.

III.2. Pruebas de hipótesis y correcta especificación

En este apartado se consideran diversas pruebas sobre la correcta especificación del modelo Probit. Para probar restricciones de exclusión, se realizarán las principales pruebas de significancia, el test de Wald y la prueba de razón de verosimilitud (*LR* por sus siglas en inglés *likelihood ratio*). Además, se demostrará que los errores del modelo propuesto son homocedásticos.

III.2.1. Prueba de Wald

La prueba de Wald sirve para probar la significancia global del modelo. El Test de Wald, permite probar restricciones múltiples sobre el modelo.

Consideremos la siguiente ecuación,

$$y = X\beta + u$$

Sea $R\beta = r$ el conjunto de las q restricciones lineales a examinar conjuntamente, por medio de la siguiente prueba de hipótesis:

$$H_0: R\beta = r$$

La hipótesis nula de la prueba es que los parámetros de las variables omitidas son cero.

Para un conjunto de restricciones de la forma $R\beta = r$, se construye el estadístico de Wald:

$$W = \frac{1}{s^2}(R\beta - r)'[R(X'X)^{-1}R']^{-1}(R\beta - r)$$

$$W \sim \chi^2(q)$$

Los resultados del modelo completo en STATA son los siguientes,

Cuadro 5. Prueba de Wald del Modelo_2

```
( 1) [Ahorro_AFORE]Mujer = 0
( 2) [Ahorro_AFORE]Edad = 0
( 3) [Ahorro_AFORE]NE_2 = 0
( 4) [Ahorro_AFORE]NE_4 = 0
( 5) [Ahorro_AFORE]Etnia = 0
( 6) [Ahorro_AFORE]Casado = 0
( 7) [Ahorro_AFORE]tot_integ = 0
( 8) [Ahorro_AFORE]FamMen = 0
( 9) [Ahorro_AFORE]Fam65mas = 0
(10) [Ahorro_AFORE]FamTransfer = 0
(11) [Ahorro_AFORE]FamconJubil = 0
(12) [Ahorro_AFORE]Informal_trab = 0
(13) [Ahorro_AFORE]TamEmp_1 = 0
(14) [Ahorro_AFORE]TamEmp_3 = 0
(15) [Ahorro_AFORE]TipoEmpresa_2 = 0

      chi2( 15) = 6713.42
      Prob > chi2 = 0.0000
```

El p -value es 0.00, entonces se rechaza la hipótesis nula y se demuestra que el modelo es significativo de manera conjunta.

III.2.2. Prueba de Razón de Verosimilitud

Un método alternativo para probar la significancia global del modelo, es la prueba de razón de verosimilitud (LR por sus siglas en inglés *likelihood ratio*), la cual es asintóticamente equivalente al test de Wald si el modelo está correctamente especificado.

La prueba de razón de verosimilitud (LR) se utiliza para comprobar la significatividad conjunta del vector de coeficientes $(\beta_1, \beta_2, \dots, \beta_k)$. Es decir, las variables dependientes del modelo probit.

La prueba LR , básicamente consiste en comparar dos modelos, uno restringido y uno no restringido o inicial. Se tienen las siguientes hipótesis:

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$, Los coeficientes no son significativos conjuntamente.

$H_1: \beta_j \neq 0, j = 1, 2, \dots, k$, Los coeficientes son significativos conjuntamente (al menos uno es distinto de cero).

Para llevar a cabo esta prueba, se estima (por máxima verosimilitud) el modelo con todos los condicionantes X 's y el modelo restringido con una constante solamente como variable explicativa, se puede construir un test de razón de verosimilitud comparando el valor de la función de verosimilitud en el máximo para ambos modelos, valor que se distribuye como una χ^2 con $k - 1$ grados de libertad (siendo k el número total de variables independientes, incluida la constante). Por tanto,

$$LR = -2(\ln L_{NR} - \ln L_R) \sim \chi^2_{k-1}$$

Donde $\ln L_{NR}$ es el valor del logaritmo de la función de verosimilitud en el máximo para el modelo general, y $\ln L_R$ es el valor del logaritmo de la función de verosimilitud en el máximo para el modelo restringido.

La decisión de este contraste es: elegido un nivel de significancia α , comparamos el valor muestral de la χ^2 con el valor de tablas con $k - 1$ grados de libertad. Si el valor muestral es mayor que el de tablas, es decir, $LR > \chi^2_{k-1}$ rechazamos la hipótesis nula y por tanto, concluimos que las variables independientes son conjuntamente explicativas.

Para realizar la prueba de Razón de Verosimilitud, es necesario estimar dos modelos: uno restringido y uno no restringido. Al estimar ambos modelos en STATA y aplicar la prueba LR , se obtuvieron los siguientes resultados,

Cuadro 6. Prueba de Razón de Verosimilitud

Likelihood-ratio test (Assumption: <u>restringido</u> nested in <u>completo</u>)	LR chi2(11) = 1510.43 Prob > chi2 = 0.0000
--	---

Como $LR > \chi^2_{k-1}$ rechazamos la hipótesis nula y se demuestra que el modelo es significativo globalmente hablando.

III.2.3. Prueba de heteroscedasticidad en la regresión probit

El modelo Probit asume homoscedasticidad en los errores.⁶ Para demostrarlo, se construye una prueba de hipótesis, asumiendo el siguiente modelo para la hipótesis nula:

$$H_0: \Pr(y_i = 1|\mathbf{x}) = \Phi(\mathbf{x}'_i\beta/\sigma),$$

Con una varianza $\sigma^2 = 1$ y, como la hipótesis alternativa

$$H_1: \Pr(y_i = 1|\mathbf{x}) = \Phi(\mathbf{x}'_i\beta/\sigma_i)$$

donde los errores son heteroscedasticos con varianza $\sigma_i^2 = \exp(\mathbf{z}'_i\delta)$.

Las variables exógenas ($\mathbf{z}_1, \dots, \mathbf{z}_m$) no contienen una constante, porque la restricción $\delta = 0$, produce $\sigma_i^2 = 1$ como en el modelo de la hipótesis nula H_0 . Incluyendo una constante en \mathbf{z} haría el modelo no identificado.

Al realizar la prueba de heteroscedasticidad del *modelo completo*, se realizó la estimación correspondiente con el test de razón de verosimilitud sobre la variable *Edad*, ya que las demás son variables binarias; se obtienen los siguientes resultados:

Cuadro 7. Prueba de heteroscedasticidad

Heteroskedastic probit model		Number of obs = 21087				
		Zero outcomes = 7511				
		Nonzero outcomes = 13576				
Log likelihood = -8920.643		wald chi2(15) = 870.31				
		Prob > chi2 = 0.0000				
Ahorro_AFORE	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Ahorro_AFORE						
Mujer	-.2098759	.0363298	-5.78	0.000	-.2810811	-.1386708
Edad	.0382534	.0028655	13.35	0.000	.0326371	.0438697
NE_2	-.2885107	.0389954	-7.40	0.000	-.3649403	-.2120811
NE_4	.194098	.049504	3.92	0.000	.0970719	.2911241
Etnia	-.3491064	.0373766	-9.34	0.000	-.4223632	-.2758496
Casado	.32701	.039051	8.37	0.000	.2504714	.4035486
tot_integ	-.0771528	.0092532	-8.34	0.000	-.0952888	-.0590168
FamMen	.2756656	.0397234	6.94	0.000	.1978091	.3535222
Fam65mas	-.1599723	.0518579	-3.08	0.002	-.2616119	-.0583326
FamTransfer	-.232116	.036977	-6.28	0.000	-.3045897	-.1596423
FamconJubil	.5428733	.0624395	8.69	0.000	.4204942	.6652524
Informal_t~b	-1.784702	.0763758	-23.37	0.000	-1.934396	-1.635008
TamEmp_1	-.4239357	.0401794	-10.55	0.000	-.5026859	-.3451855
TamEmp_3	.4692162	.0534864	8.77	0.000	.3643848	.5740475
TipoEmpres~2	.6998535	.0427936	16.35	0.000	.6159796	.7837274
_cons	.6228162	.0807446	7.71	0.000	.4645598	.7810727
lnsigma2						
Edad	.0112632	.0011814	9.53	0.000	.0089477	.0135786
Likelihood-ratio test of lnsigma2=0: chi2(1) = 95.64				Prob > chi2 = 0.0000		

La prueba LR indica que a un nivel de 5% de significancia, no existe una mejora estadísticamente significativa del modelo resultado de la generalización al modelo homocedastico debido a que $p = 0.0000$, es decir, no hubo evidencia estadística suficiente para rechazar la hipótesis nula. Por lo tanto, el modelo es homocedástico.

⁶ Esta es una condición que debe de cumplir el modelo Probit que se propone. La homocedasticidad en los errores, significa que la varianza de cada término de error u_i , es un número constante igual a σ^2 .

III.3. Bondad de ajuste del modelo

Al tratarse de un modelo no lineal, la bondad de ajuste del modelo Probit estimado, puede ser medida por el Pseudo R^2 y el porcentaje de predicciones correctas.⁷ Estos indicadores permiten conocer que tan bueno es el modelo que se propone para explicar el fenómeno de estudio.

III.3.1. Pseudo R^2 o R^2 de MacFadden

El Pseudo R^2 , indica la bondad de ajuste del modelo a los datos. Es una aproximación basada en una comparación de la verosimilitud del modelo sólo con la constante (modelo restringido), con la verosimilitud del modelo con todos los parámetros estimados (modelo no restringido).

El Pseudo R^2 está definido como:

$$Pseudo R^2 = R_{MCF}^2 = 1 - \frac{\ln L_{NR}}{\ln L_R}, \quad 0 \leq R_{MCF}^2 \leq 1$$

Donde, $\ln L_{NR}$ es igual al máximo del logaritmo natural de la función de verosimilitud (FV) del modelo no restringido y $\ln L_R$ al logaritmo natural de la FV del modelo restringido.

El Pseudo R^2 , está acotado en el intervalo 0,1. Un Pseudo R^2 creciente puede indicar un mejor ajuste del modelo. De hecho esta razón de verosimilitudes indica el nivel de mejora que tiene el modelo no restringido (completo) respecto al modelo restringido (con solo el intercepto).

Además, el Pseudo R^2 puede servir para comparar la capacidad explicativa de modelos distintos. Así, notamos que de los modelos estimados, el Modelo_2 tiene un Pseudo R^2 más alto que el Modelo_1 (0.3468 vs 0.1609). Esto indica que al incluir las variables del mercado de trabajo, tenemos un mejor modelo, que explica el ahorro en el sistema de pensiones (Ver Cuadro 10).

Cuadro 8. Ajuste del modelo a los datos

Measures of Fit for probit of Ahorro_AFORE			
Log-Lik Intercept Only:	-13731.751	Log-Lik Full Model:	-8968.462
d(21071):	17936.924	LR(15):	9526.579
		Prob > LR:	0.000
McFadden's R2:	0.347	McFadden's Adj R2:	0.346
ML (Cox-Snell) R2:	0.364	Cragg-Uhler(Nagelkerke) R2:	0.499
Mckelvey & Zavoina's R2:	0.536	Efron's R2:	0.401
Variance of y*:	2.154	Variance of error:	1.000
Count R2:	0.796	Adj Count R2:	0.428
AIC:	0.852	AIC*n:	17968.924
BIC:	-191854.634	BIC':	-9377.232
BIC used by Stata:	18096.226	AIC used by Stata:	17968.924

El cuadro 8, muestra diferentes estadísticos de diagnóstico de la regresión Probit, de los cuales la mayoría se han comentado anteriormente. La fila que merece mención es donde aparece la medida de ajuste más importante del modelo Probit: Pseudo R^2 o R^2 de MacFadden. Como ya se mencionó este estadístico nos permite conocer la bondad de ajuste del modelo a los datos. Además, es preciso comentar que junto con el Pseudo R^2 estándar aparece el R^2 ajustado, que simplemente corrige el hecho de que al añadir nuevas variables el R^2 aumenta artificialmente, restando al numerador del R^2 el número de parámetros del modelo. Por esta razón, el ajustado es preferible al R^2 estándar. Los resultados del modelo arrojan una R^2 de MacFadden de 0.347 y una ajustada de 0.346.

⁷ El Pseudo R^2 es un estadístico análogo al R^2 del modelo de regresión lineal, que indica la bondad de ajuste del modelo a los datos.

Otra medida sobre la bondad de ajuste del modelo, es la del Count R^2 . Esta medida está basada en la comparación de los valores observados en la muestra y los estimados por el modelo. El modelo Probit predice la probabilidad de ocurrencia de un suceso, en consonancia con ello, en todos aquellos casos en los que el modelo prediga más de 0.5 de probabilidad de ocurrencia, la predicción será que ocurra; y en todos los casos en los que el modelo dé una probabilidad inferior a 0.5, se pronosticará que no sucederá. El Count R^2 es simplemente la proporción de predicciones correctas según este criterio, es decir, en qué porcentaje de casos la predicción derivada del modelo de regresión Probit acierta. Para el modelo Probit que se propone, el modelo acierta en el 79.6% de los casos, lo que parece un porcentaje bastante alto de acierto.

III.3.2. Porcentaje de predicciones correctas

El porcentaje de predicciones de las observaciones clasificadas correctamente respecto a los datos observados, es otra medida sobre la bondad de ajuste del modelo. El procedimiento consiste en comparar las probabilidades estimadas con los valores observados, por medio de una función simétrica, es decir, si $F(\mathbf{x}'\beta) \geq 0.5$ y coincide con su valor observado $\mathbf{y}_i = 1$, entonces las predicciones son correctas, de igual forma cuando $F(\mathbf{x}'\beta) < 0.5$ y sus observaciones son $\mathbf{y}_i = 0$. En el caso de que, el valor estimado no coincida con el valor observado, es decir, $F(\mathbf{x}'\beta) \geq 0.5$ cuando $\mathbf{y}_i = 0$ o $F(\mathbf{x}'\beta) < 0.5$ y $\mathbf{y}_i = 1$, la probabilidad estimada es incorrecta.

Esta medida sobre la bondad de ajuste, basada en la correcta clasificación de los datos, se puede realizar en STATA y se obtienen los siguientes resultados,

Cuadro 9. Valores ajustados y reales

Probit model for Ahorro_AFORE			
Classified	True		Total
	D	~D	
+	11237	1959	13196
-	2339	5552	7891
Total	13576	7511	21087
Classified + if predicted Pr(D) >= .5 True D defined as Ahorro_AFORE != 0			
Sensitivity	Pr(+ D)		82.77%
Specificity	Pr(- ~D)		73.92%
Positive predictive value	Pr(D +)		85.15%
Negative predictive value	Pr(~D -)		70.36%
False + rate for true ~D	Pr(+ ~D)		26.08%
False - rate for true D	Pr(- D)		17.23%
False + rate for classified +	Pr(~D +)		14.85%
False - rate for classified -	Pr(D -)		29.64%
Correctly classified			79.62%

Del cuadro anterior, se puede mencionar que el porcentaje de los valores correctamente especificados fue de 79.62%. De las 21,087 observaciones del modelo, 1,959 están incorrectamente clasificadas con 1 cuando su correcta clasificación es 0, y 2,339 valores están mal clasificados como 0 cuando su valor correcto debe ser 1. Los restantes valores están correctamente clasificados.

La razón de 11,237/13,576, *la medida de sensibilidad*, nos dice la porción de observaciones correctamente especificadas como $\mathbf{y}_i = 1$. La razón de 5,552/7,511, *medida de especificidad* nos dice la porción de observaciones correctamente especificadas como $\mathbf{y}_i = 0$. Las razones 1,959/7,511 y 2,339/13,576 son conocidas como *las tasas de clasificación incorrecta de falsos positivos y falsos negativos*.

III.4. Efectos Marginales

El impacto de cada variable explicativa sobre la probabilidad de ahorrar para una pensión de retiro, no se puede interpretar directamente de la estimación del modelo Probit propuesto. En un modelo no-lineal, los efectos marginales son más ilustrativos que los propios coeficientes.

Para obtener los efectos marginales de un modelo Probit, se parte de la distribución acumulada de probabilidad de la normal estándar,

$$P_i = \Pr(\mathbf{y}_i = 1|\mathbf{x}') = \Phi(\mathbf{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz \quad (3)$$

Considerando la ecuación (3), el efecto marginal de un cambio de una unidad en la j-ésima variable exógena sobre la probabilidad de ocurrencia del evento, se puede establecer como la derivada parcial de la probabilidad P_i con respecto a la variable independiente j-ésima X_{ij} evaluada en su valor promedio, esto es,

$$\frac{\partial P_i}{\partial X_{ij}} = \frac{\partial \Pr(\mathbf{y}_i = 1|\mathbf{x}')}{\partial X_{ij}} = \frac{\partial \Phi(\mathbf{x}'\beta)}{\partial X_{ij}} = \phi(\mathbf{x}'\beta)\beta_{ij} \quad (6)$$

La ecuación (6) muestra el efecto marginal que experimenta la variable dependiente ante cambios en alguna de las variables explicativas, manteniendo constantes las demás variables independientes. Al obtener los efectos marginales, se pueden realizar interpretaciones directamente de los coeficientes estimados sobre la variable dependiente del modelo. En este caso, la interpretación de los coeficientes del modelo estimado representa el cambio en la probabilidad de ahorrar para una pensión de retiro ante cambios en alguna variable X_{ij} que pertenece al vector de variables independientes \mathbf{x}' , manteniendo los otros factores fijos.

Cameron y Trivedi (2005), señalan que existen tres variantes de efectos marginales, los cuales son el efecto marginal promedio (AME), efectos marginales a un nivel representativo (MER) y efectos marginales en la media (MEM).

A continuación se presenta la estimación realizada en STATA de los efectos marginales promedio.

Cuadro 10. Promedio de los Efectos Marginales

```

Marginal effects after probit
y = Pr(Ahorro_AFORE) (predict)
= .7197898

```

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	x
Mujer*	-.0446105	.00819	-5.44	0.000	-.060669	-.028552	.332148	
Edad	.0069271	.00032	21.65	0.000	.0063	.007554	35.7849	
NE_2*	-.0645164	.0085	-7.59	0.000	-.081185	-.047848	.501399	
NE_4*	.0506711	.01087	4.66	0.000	.029369	.071973	.20994	
Etnia*	-.0826098	.00838	-9.86	0.000	-.099025	-.066194	.267226	
Casado*	.080599	.00881	9.15	0.000	.063335	.097863	.608527	
tot_in~g	-.019185	.00206	-9.29	0.000	-.023232	-.015138	4.46009	
FamMen*	.0691054	.00896	7.71	0.000	.05154	.086671	.555888	
Fam65mas*	-.0488839	.01212	-4.03	0.000	-.072648	-.02512	.13487	
FamTra~r*	-.0529855	.00796	-6.66	0.000	-.068584	-.037387	.662968	
Famcon~l*	.1248054	.01079	11.56	0.000	.103648	.145963	.092759	
InForm~b*	-.3907821	.00745	-52.46	0.000	-.405381	-.376183	.531892	
TamEmp_1*	-.0971305	.0087	-11.17	0.000	-.114175	-.080086	.358799	
TamEmp_3*	.0994106	.01018	9.76	0.000	.079455	.119366	.232987	
TipoEm~2*	.1623899	.00767	21.16	0.000	.147349	.177431	.430644	

(*) dy/dx is for discrete change of dummy variable from 0 to 1

De la estimación anterior de los efectos marginales podemos notar que todas las variables resultan ser estadísticamente significativas. El género influye en la probabilidad de cotizar para una pensión de retiro, ya que según los resultados de la estimación anterior, el ser mujer disminuye la probabilidad de ahorrar para una pensión de retiro en 4.5 puntos porcentuales. Otros factores que inciden de forma negativa sobre la probabilidad de cotizar son tener sólo la educación básica (disminuye la probabilidad en 6.5%) y el hecho de pertenecer a algún grupo étnico, reduce la probabilidad de ahorrar en el sistema de pensiones en más de 8 puntos porcentuales.

En contraste, tener educación superior completa y estar casado o en pareja son factores que influyen de forma positiva en la probabilidad de ahorrar en el sistema de pensiones. En el caso de tener educación superior completa, la probabilidad aumenta en 5% y el estar casado o en pareja incrementa la probabilidad de cotizar a las pensiones en 8%.

De acuerdo con los resultados anteriores, a medida que las familias son más grandes, es decir, el número total de integrantes en la familia es mayor, la probabilidad de ahorrar para una pensión de retiro se reduce en 2%. Además, familias que reciben transferencias de cualquier tipo y en donde algún integrante de la misma sea mayor de 65 años, tienen una probabilidad menor de cotizar al sistema de pensiones, la probabilidad se reduce en alrededor del 5% en ambos casos. Del otro lado se tiene que si en la familia existen menores de edad, la probabilidad de cotizar para una pensión de retiro se eleva en 7 puntos porcentuales. Un hecho que resulta interesante, es que aquellas familias donde algún integrante sea jubilado, tienen una probabilidad más alta de ahorrar para una pensión de retiro, es decir, si la familia tiene a algún jubilado la probabilidad se incrementa en 12%, esto podría significar que existen incentivos para ahorrar para la vejez cuando algún miembro de la familia disfruta de los beneficios de una pensión de retiro.

Del cuadro 10 también se puede decir, que el hecho de estar en la informalidad reduce notoriamente la probabilidad de ahorrar en el sistema de pensiones, en un 40%. A pesar de que un trabajador informal puede cotizar al sistema de pensiones, los resultados sugieren que no cotizan. Este constituye un serio problema, si tenemos en cuenta que mucha gente se emplea en el sector informal de la economía y no están ahorrando para la vejez. Otro factor que reduce la probabilidad de cotizar en el sistema de pensiones es trabajar en una empresa pequeña (-9%); en contraste, si el trabajador labora en una empresa grande o una empresa privada, la probabilidad de ahorrar para una pensión de retiro se incrementa en 10 y 16 puntos porcentuales, respectivamente.

Ahora bien, si consideramos a un individuo representativo de la muestra, podemos encontrar la probabilidad que tiene dicho individuo de cotizar al sistema de pensiones. Por ejemplo, si tomamos un individuo hombre de 40 años, casado, con educación superior completa, y que labore en una empresa grande y privada, su probabilidad de ahorrar para una pensión de retiro es mayor que la de una mujer con la misma edad, también casada o en pareja, que sólo tenga estudios básicos, que pertenezca a algún grupo étnico y que trabaje en el sector informal.

Según las estimaciones realizadas en STATA, para el primer caso, se tiene un 97% de probabilidad de ahorrar en el sistema de pensiones con las características mencionadas (Edad=40 Mujer=0 NE_4=1 Etnia=0 Casado=1 Informal_trab=0 TipoEmpresa_2=1). Para el segundo caso, la probabilidad de cotizar para una pensión de retiro es de 43% para el individuo representativo que posee las siguientes características: Edad=40 Mujer=1 NE_2=1 Etnia=1 Casado=1 Informal_trab=1. De esta forma, notamos diferencias muy marcadas en cuanto a los factores que inciden para el ahorro en el sistema de pensiones.

Los resultados de estas estimaciones particulares se muestran a continuación,

Cuadro 11. Efectos Marginales a un nivel representativo (MER).

(Edad=40 Mujer=0 NE_4=1 Etnia=0 Casado=1 Informal_trab=0 TipoEmpresa_2=1)

Marginal effects after probit
y = Pr(Ahorro_AFORE) (predict)
= .97318839

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
Mujer*	-.0091993	.00182	-5.05	0.000	-.012768 -.005631	0
Edad	.0012747	.00011	11.64	0.000	.00106 .001489	40
NE_2*	-.0119297	.00205	-5.83	0.000	-.015938 -.007921	.501399
NE_4*	.0111239	.00235	4.73	0.000	.006514 .015733	1
Etnia*	-.0185521	.00244	-7.62	0.000	-.023327 -.013777	0
Casado*	.018358	.00252	7.29	0.000	.013421 .023294	1
tot_in~g	-.0035304	.00047	-7.52	0.000	-.004451 -.00261	4.46009
FamMen*	.012989	.00196	6.62	0.000	.009141 .016838	.555888
Fam65mas*	-.0096686	.00268	-3.61	0.000	-.014924 -.004413	.13487
FamTra~r*	-.0094486	.0016	-5.91	0.000	-.01258 -.006317	.662968
Famcon~l*	.0187833	.00201	9.35	0.000	.014848 .022719	.092759
Inform~b*	-.2168374	.01051	-20.62	0.000	-.237444 -.196231	0
TamEmp_1*	-.0190678	.00254	-7.52	0.000	-.024038 -.014097	.358799
TamEmp_3*	.016592	.00209	7.96	0.000	.012504 .02068	.232987
TipoEm~2*	.0489069	.00369	13.24	0.000	.041669 .056145	1

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Cuadro 12. Efectos Marginales a un nivel representativo (MER).

(Edad=40 Mujer=1 NE_2=1 Etnia=1 Casado=1 Informal_trab=1)

Marginal effects after probit
y = Pr(Ahorro_AFORE) (predict)
= .43074107

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
Mujer*	-.0518704	.00937	-5.53	0.000	-.07024 -.0335	1
Edad	.0080825	.00037	21.71	0.000	.007353 .008812	40
NE_2*	-.0761699	.01004	-7.59	0.000	-.095844 -.056496	1
NE_4*	.0611422	.01363	4.49	0.000	.034435 .08785	.20994
Etnia*	-.0946968	.0093	-10.19	0.000	-.112918 -.076476	1
Casado*	.090111	.00968	9.31	0.000	.071137 .109085	1
tot_in~g	-.0223851	.00241	-9.29	0.000	-.027108 -.017662	4.46009
FamMen*	.0798988	.01022	7.81	0.000	.059859 .099938	.555888
Fam65mas*	-.0548457	.01307	-4.20	0.000	-.080464 -.029227	.13487
FamTra~r*	-.0630314	.00968	-6.51	0.000	-.08201 -.044053	.662968
Famcon~l*	.164886	.01618	10.19	0.000	.133175 .196597	.092759
Inform~b*	-.4248611	.00935	-45.46	0.000	-.443178 -.406544	1
TamEmp_1*	-.1098693	.0096	-11.45	0.000	-.128683 -.091055	.358799
TamEmp_3*	.1233147	.01358	9.08	0.000	.0967 .149929	.232987
TipoEm~2*	.1937582	.00943	20.55	0.000	.175275 .212241	.430644

(*) dy/dx is for discrete change of dummy variable from 0 to 1

IV. Conclusiones

En este trabajo se mostró evidencia empírica sobre los determinantes, que inciden en la decisión de los individuos de ahorrar en el sistema de pensiones en México. Se propuso un modelo de regresión no lineal Probit, dicho modelo resulta adecuado para cuantificar la probabilidad que tiene un individuo con determinadas características, de ahorrar para una pensión de retiro.

La base de datos utilizada, fue la ENIGH 2014. En dicha encuesta, se identifican preguntas ligadas al tema de la seguridad social, las prestaciones que posee una persona en su trabajo, el tipo de trabajo que realizan los individuos. Además, se distinguen aquellos individuos que ahorran en una AFORE y/o fondo de pensión y a los que no lo hacen. Estos factores son la base para cuantificar la probabilidad de que los trabajadores coticen a un sistema de pensiones ya sea público o privado.

Las características sociodemográficas y laborales de la población explican en buena medida a la variable dependiente (el ahorro en el sistema de pensiones). Es por eso que se utilizó un modelo microeconómico, que cuantifica la probabilidad de que el individuo cotice en pensiones. Según las estimaciones del modelo Probit, el bajo nivel de aportaciones al sistema de pensiones está relacionado con una serie de características socio-económicas, en particular con la educación, el género y el tipo de trabajo que se realiza.

Los resultados del modelo estimado, permiten corroborar algunos puntos que intuitivamente tienen sentido, también permiten ratificar algunas hipótesis de las variables utilizadas. Por ejemplo, en el tema de la educación el modelo predijo que a mayor educación es más probable que el individuo representativo ahorre en el sistema de pensiones. También el género condiciona la probabilidad de ahorrar para una pensión de retiro, la desigualdad laboral evidencia que las mujeres tienen menos probabilidad de ahorrar que los hombres para la jubilación. Así mismo, la actividad laboral que se realiza es un factor que explica si el individuo es más propenso a ahorrar en el sistema de pensiones; es decir, si el individuo trabaja en una empresa privada no familiar o en una empresa grande, tiene mayor probabilidad de cotizar para una pensión de retiro.

Resulta claro que existe una relación muy estrecha entre las pensiones y el mercado de trabajo. El modelo predice que las personas que trabajan en la informalidad, tienen una probabilidad mucho menor de cotizar al sistema de pensiones que aquellos individuos que cuentan con un empleo formal. En la actualidad, una proporción alta de la Población Económicamente Activa (PEA) no tiene un contrato laboral formal, por lo que es menos probable que ahorren para una pensión de retiro. Esto constituye un serio problema a mediano y largo plazo, si consideramos que la población envejece y no está preparándose para enfrentar el riesgo de la vejez.

Bibliografía

Cameron, A., & Trivedi, P. (2005). Logit and Probit Models & Latent Variable Models. En *Microeconometrics: Methods and Applications* (págs. 465-478). Cambridge University Press, NY.

Carpeta, W., & Peña, J. (Septiembre de 2015). *Cotizaciones en el sistema general de pensiones colombiano: Aproximación con un modelo de respuesta binaria*. Econografos, No. 82. FCE, Universidad Nacional de Colombia.

Gujarati, D. (2003). *Econometría*, 4ª ed., McGraw-Hill, México.

Hernández, G. (2001). *Políticas para promover una ampliación de la cobertura de los sistemas de pensiones: el caso de México*. Serie Financiamiento del Desarrollo, núm. 107, CEPAL, Santiago, Chile.

Murillo, S., & Venegas, F. (2011). *Cobertura de los sistemas de pensiones y factores asociados al acceso a una pensión de jubilación en México*. Papeles de Población, vol. 17, núm. 67, UAEM, pp. 209-250.

OCDE, Banco Mundial, BID (2014). *Panorama de las Pensiones: América Latina y el Caribe*.

Rosales, R., Perdomo, J., Morales, C. y Urrego, J. A. (2013). Modelos de Probabilidad Lineal, Probit y Logit. En *Fundamentos de Econometría Intermedia. Teoría y Aplicaciones* (págs. 107-158). Facultad de Economía, Universidad de los Andes, Colombia.

Solís, F., & Villagómez, A. (1999). Las Pensiones, en: Solís, F., & Villagómez, A. (Eds.), *La Seguridad Social en México* (págs. 103-159). Fondo de Cultura Económica, México.

Tuesta, D. (Agosto de 2014). *La economía informal y las restricciones que impone sobre las cotizaciones al régimen de pensiones en América Latina*. Documento de Trabajo, N° 14/20, BBVA Research.