

#### *Capítulo IV. Estrategia Econométrica*

La metodología utilizada en esta tesis es la idónea para evaluar la hipótesis planteada desde el inicio, la cual indica que las remesas enviadas por los mexicanos en Estados Unidos están relacionadas de forma negativa con el tiempo de estancia que los migrantes permanecen en dicho país de acogida, es decir, el pasar de ser un migrante temporal a ser uno permanente afecta el envío de dinero a su país de origen, al verse mermados los lazos y las necesidades que tiene éste con sus familiares.

Para evaluar la temporalidad del migrante en Estados Unidos, se usan diferentes variables *proxys*, tales como la experiencia laboral en E.U, tanto en su término lineal como cuadrático, para valorar un posible punto de quiebre. Además se utilizan variables que denotan cierto establecimiento en el país de acogida, como puede ser el hecho de que al menos uno de sus hijos esté estudiando en escuela pública dentro del territorio estadounidense. Otro de los factores que nos indica un arraigamiento del migrante es la obtención de algún beneficio o asistencia de parte del gobierno estadounidense para el migrante que tiene relativamente mayor tiempo viviendo en la unión americana.

El hecho de incorporar la variable que nos muestra si al menos uno de los hijos del jefe de familia está estudiando en escuela pública de Estados Unidos, fue una aportación de Balderas (2003) y a su vez es de suma importancia para demostrar y sustentar la hipótesis de que la temporalidad y el factor tiempo es determinante en el envío de remesas. Tal variable es una Proxy que nos indica si el migrante tomó la decisión de permanecer de forma temporal o permanente en Estados Unidos, ya que el

hecho de tener consigo a por lo menos uno de sus hijos, nos da la idea que los vínculos que éste tiene con su país de origen son mas escasos, aminorando el envío de remesas.

Las variables que muestran si el migrante tiene algún beneficio o asistencia pública por parte del gobierno de Estados Unidos, nos hace deducir que el migrante mexicano tiene un mayor lazo con el país al cual migró, haciéndolo más propenso a estabilizarse de forma permanente. Por tales motivos se espera que las variables que denotan temporalidad arrojen signos negativos.

Aunada a esas variables se encuentra de igual forma la que denota la experiencia laboral del migrante en Estados Unidos. Algunos estudios, como el de Amuedo-Dorante y Pozos (2006b) agregan esta variable y su término cuadrático para rescatan la relación entre remesas y experiencia como un patrón temporal de “U” invertida. En su trabajo afirman que las remesas aumentan al inicio de la estancia del migrante, después de haber cubierto costos de readaptación, para caer paulatinamente a medida que la estancia migratoria rebasa un umbral determinado y los lazos con la comunidad de origen se ven debilitados. En este trabajo se esperan resultados similares en la variable que denota experiencia laboral del migrante dentro del territorio estadounidense, de esa forma se llegaría a importantes conclusiones respecto a la hipótesis base de esta tesis.

Para dar paso al análisis de nuestra hipótesis y con ello a la estrategia econométrica, se estiman dos modelos econométricos, por un lado un modelo *Probit*, tomando como variable dicotómica el hecho de que manden o no remesas; eso nos arroja la probabilidad que existe en el migrante de enviar dinero a su país de origen

relacionándolas con sus características personales, familiares, de temporalidad, localidad de origen y permanencia en el país de alojamiento. Por otro lado, y tomando en cuenta la naturaleza de este estudio, la cual nos lleva inevitablemente a enfrentarnos a un problema de auto selección<sup>1</sup>, se utiliza la técnica propuesta por Heckman (1976) en dos etapas, la cual corrige el sesgo de selección, tomando en la primera etapa la variable dependiente la condición de ser igual a uno que el migrante envíe remesas positivas e igual a cero que no envíe dinero a su país de origen. En la segunda etapa se toma como variable dependiente la cantidad de remesas enviadas en forma de dólares mensualmente por parte del migrante.

Al hacer una estimación por MCO no se estaría teniendo en cuenta el total de la muestra, porque sólo es posible observar las remesas de los individuos que realmente lo hacen; por consiguiente los estimadores estarían sesgados ya que la muestra (individuos que no envían remesas) no es representativa de la población en la que se está interesado (todos los individuos que envían cantidades iguales a cero o positivas).

Otra alternativa sería la estimación de un modelo *Tobit*, el cual sí toma en cuenta el total de la población, tanto los valores discretos como continuos de la variable dependiente, sin embargo, tiene ciertas desventajas. Una sería que la probabilidad de remitir y el monto finalmente enviado están íntimamente ligados. Una segunda desventaja que posee el modelo *Tobit*, radica en la aseveración de normalidad y heteroscedasticidad en las variables latentes, a pesar de este hecho, como lo afirma Wooldridge (2003), esto no afecta a la inconsistencia de los parámetros, por tal motivo, a manera de comparación, se exhibe en el apéndice de este trabajo un modelo *Tobit* que

---

<sup>1</sup> Los individuos deciden remitir parte de su salario o no, es decir, se autoseleccionan como remitentes o no remitentes de acuerdo a varias características.

corrobore y sustente nuestros resultados realizados con los modelos principales, descritos de forma específica a continuación.

#### 4.1. Modelo Probit

Primero se estima un modelo *Probit*, el cual especifica la probabilidad condicional. Donde  $p$  es la función de distribución acumulada normal estándar, que se expresa como una integral<sup>2</sup>:

$$(4.1) \quad p = \Phi(x' \beta) = \int_{-\infty}^{x' \beta} \Phi(z) dz$$

Donde  $\Phi(\cdot)$  es la función de densidad acumulada (*fda*) de la normal estándar. Los efectos marginales del modelo *Probit* son  $\partial p_i / \partial x_{ij} = \Phi(x_i' \beta) \beta_j = \Phi(\Phi^{-1}(p_i)) \beta_j$ , en el que se especifica los factores que determinan que los migrantes mexicanos remitan o no su ingreso a México. Los modelos *Probit* se emplean en el caso de que la variable dependiente sea binaria, como en el caso de este trabajo donde se quiere evaluar los factores que influyen en la probabilidad de que el migrante remita o no parte de su ingreso. Por tal motivo se emplean los modelos de respuesta binaria.

Denotamos con  $y$  a una variable aleatoria que puede tomar solo dos valores, uno o cero, asociada a la ocurrencia ó probabilidad de un evento (en este caso 1 si remite remesas y 0 lo contrario).

$$y = \begin{cases} 1 & \text{remite remesas } r \\ 0 & \text{no remite remesas } 1-r \end{cases}$$

---

<sup>2</sup> COLIN Cameron A (2005). *Microeconometrics Methods and Applications*, Cambridge. Pp.445-451.

No hay pérdida de generalidad en el establecimiento de los valores de 1 y 0 todo lo que se está estableciendo es la probabilidad de remitir o no. El modelo de elección binaria es un modelo de probabilidad de ocurrencia del evento denotado por  $y_i$ , condicional en el conjunto de información:

$$(4.2) \quad Prob(y_i = 1/X) = Prob(y = 1 / X_1, X_2, \dots, X_n)$$

Donde  $y_i$  es la variable aleatoria dicotómica,  $i$  denota el  $i$ -ésimo individuo,  $X_k$  es el conjunto completo de variables explicativas, vector del cual depende la probabilidad del evento, que en este caso son: el sexo, edad, edad al cuadrado, años de escolaridad, años de escolaridad al cuadrado, estado marital y número de miembros en la familia, son integrados en la mayoría de modelos econométricos que a remesas se refiere (Lucas y Stark, 1985; Taylor, 1987; Funkhouser, 1995; Durand, Kandel, Parrado, Massey, 1996; Hassan, Zeller, Meliczeck, 2001).

Si modificamos (4.2) en la siguiente distribución G del conjunto de variables exógenas  $X_k$ :

$$(4.3) \quad Prob(y_i = 1/X) = G(\beta_0 + \beta_1 X_1 + \dots, \beta_k X_k) = G(\beta_0 + X\beta_k)$$

Donde G es una función que asume valores que se hallan estrictamente entre cero y uno  $0 < G(z) < 1$ , para todos los números reales  $z$ . Estos aseguran que las probabilidades de

respuesta del estimador se hallen estrictamente entre cero y uno. El modelo *Probit*,  $G$ , es la función de distribución acumulada (*FDA*) de la normal estándar.

En la mayor parte de las aplicaciones de los modelos de respuesta binaria, el objetivo principal es explicar los efectos de  $X_j$  en la probabilidad de respuesta de  $Prob(y = 1/X)$ . Para obtener la dirección del efecto de las  $X_j$  en  $y^*$ , se calculan los efectos de los cambios marginales de  $\beta_j$ ; las cuales permiten estimar el efecto de  $X_j$  sobre la probabilidad de éxito de  $Prob(y = 1/X)$ , contemplando la naturaleza no lineal de  $G(\cdot)$ . Por tal motivo se busca obtener el efecto parcial de variables aproximadamente continuas en la probabilidad de respuesta, si  $X_j$  es una variable continua, su efecto parcial sobre  $p(X) = Prob(y = 1/X)$  se obtiene el cambio marginal de:

$$(4.4) \quad \frac{\partial p(x)}{\partial x_j} = g(\beta_0 + X\beta_k) \beta_j, \quad \text{donde } g(z) = \frac{dG}{dz}$$

Por tal motivo, si la variable objeto de estudio es una medición continua que se distribuye según una ley normal, en la que existen uno o varios puntos de truncamiento y/o censura, no es posible utilizar los habituales modelos de regresión lineal estimados por mínimos cuadrados ordinarios (MCO), porque proporcionan estimaciones incorrectas del efecto y de su variabilidad.<sup>3</sup>

#### **4.2. Sesgo de selección**

El sesgo de selección surge de estimar modelos estructurales con respuestas potenciales observadas de forma parcial. Sin embargo, el problema del sesgo de selección puede

---

<sup>3</sup> Véase Wooldridge (2001).

surgir cuando una regla que no sea el muestreo aleatorio simple se emplea para obtener la población subyacente que es objeto de estudio. La representación distorsionada de la población real en una muestra constituye la esencia del problema de selección. Tal problema de identificación se encuentra en la recuperación, dentro de la misma muestra, de las características de una población hipotética, la que puede referirse a las respuestas potenciales de cualquier problema de elección para el cual sólo se observan las elecciones que son afirmativas o poseen cierta cualidad. El sesgo de selección puede surgir de las decisiones que se toman al hacer las encuestas de muestreo o de las decisiones económicas de auto-selección, donde como consecuencia, se observa únicamente subconjuntos de una población de respuestas potenciales (Heckman, 1976). La población hipotética en el caso de esta tesis se refiere tanto a las remesas que los migrantes mexicanos en E.U. envían a su país de origen (que son las observadas) como a las remesas que no se enviaron, es decir, el problema de selección recae en evaluar tanto a los migrantes que tomaron la decisión de enviar remesas como a los que se omitieron de hacerlo.

Hay dos puntos que caracterizan el problema de selección los cuales son de gran importancia hondar. El primero, que tiene su origen en la estadística, conlleva a caracterizar la regla de muestreo, como la aplicación de un sistema ponderado a distribuciones hipotéticas de la población para producir distribuciones observadas. El segundo, de origen econométrico, trata explícitamente el problema de selección como un problema de ausencia de datos, y en su esencia emplea variables observables para imputar valores a las no observables (Heckman, 1976).

#### ***4.3. Corrección de Heckman bietápica***

La estrategia econométrica se basa en la metodología propuesta por Heckman (1976), que considera, en su estudio para medir la oferta laboral de las mujeres, como endógena la decisión del individuo de trabajar o no, a fin de corregir el posible sesgo de selección. Sin embargo el primer autor que toma en cuenta la metodología propuesta por Heckman y la añade a estudios de remesas en la india es Banerjee (1983); el cual examina la importancia de los determinantes y la decisión de remitir, además del monto remitido por los migrantes en Nueva Delhi a sus regiones de origen.

Para la corrección del sesgo de selección, inherente a la muestra utilizada en esta tesis, se utiliza la técnica de Heckman en dos etapas, que consiste primeramente en estimar una función Probit de la totalidad de la muestra, tanto de los migrantes que envían como de los que omiten hacerlo. Una vez estimada la ecuación y a partir de los residuos de la misma se calcula el inverso de la razón de Mills ( $\lambda$ )<sup>4</sup>. En la segunda etapa de Heckman, se calcula un modelo de Mínimos Cuadrados Ordinarios, introduciéndose como un regresor adicional el término de corrección del sesgo de selección (la inversa del ratio de Mills). Las ecuaciones que se calculan de forma simultánea se expresan a continuación<sup>5</sup>:

$$(4.5) \quad \text{Ecuación de regresión en M.C.O: } y_i = X_i \beta + u_{1i}$$

$$(4.6) \quad \text{Ecuación de Selección: } s_i = 1[Z_i \gamma + u_2 \geq 0]$$

Donde la ecuación de selección contiene como variable dependiente el hecho de que  $s = 1$  si observamos  $y$ , y cero de no ser así. Suponemos que los elementos de  $X$  y  $Z$ , se

---

<sup>4</sup> La inversa de la Razon de Mills ( $\lambda$ ) es una función inversa monótonica de la probabilidad de enviar remesas y se expresa como el cociente entre la función de densidad de dicha probabilidad sobre su función acumulada (Green, 2000).

<sup>5</sup> Véase Wooldridge (2000).

observan siempre y los parámetros estimados serían  $X\beta = \beta_0 + \beta_1X_1 + \dots + \beta_kX_k$ , así como  $Z\gamma = \gamma_0 + \gamma_1Z_1 + \dots + \gamma_mZ_m$ .

La ecuación de interés fundamental es la (4.5) y estimamos la  $\beta$  por MCO dada una muestra aleatoria. La ecuación de selección, (4.6), depende de las variables observadas,  $Z_h$ , y de un error inobservable,  $u_2$ . La suposición que rige el modelo será que  $Z$  es exógena en (4.6):

$$(4.7) E(u_1|X,Z) = 0$$

De hecho, para que funcionen correctamente los métodos anteriormente propuestos necesitamos que  $X$  sea un subconjunto escrito de  $Z$ : cualquier  $X_j$  es también un elemento de  $Z$ , y se cuenta con algunos elementos de  $Z$  que no están en  $X$ . Si bien, en contados casos tiene sentido excluir elementos de la ecuación de la selección, incluir todos los elementos de  $X$  en  $Z$  no es muy costoso, excluirlos lleva a inconsistencias si se hace de manera incorrecta.

Lo anterior nos lleva a las reglas de identificación para un modelo con ecuaciones simultáneas, según Gujarati (2001). Antes debemos tener en cuenta que el problema de identificación pretende establecer si las estimaciones numéricas de los parámetros de una ecuación estructural pueden ser obtenidas de los coeficientes estimados de la forma reducida, si realmente se puede lograr este cometido, se dice que la ecuación particular está identificada, de lo contrario se cae en problemas de identificación.

Una vez que se da por entendido de que se trata el problema de identificación en un modelo con ecuaciones simultaneas, se da paso a citar las dos condiciones que nos permiten resolver dicho problema. Por un lado está la condición de orden de identificación y por otro la de rango de identificación.

La primera tiene que ver con el hecho de excluir una o más variables en la ecuación de MCO, que estén presentes en la ecuación de selección, aunque es una condición necesaria mas no suficiente, es crucial para no caer en problemas de identificabilidad. Tener un elemento de  $Z$  que no está en  $X$ , significa que se necesita una variable que influya en la selección pero que no ejerza un efecto parcial en  $y$ . No es absolutamente necesario aplicar el procedimiento, pero los resultados son comúnmente poco convincentes a menos que contemos con una restricción de exclusión en (4.5). El motivo es que si bien la razón inversa de Mills es una función de  $Z$ , a menudo se aproxima mediante una función lineal. Si  $Z=X$ ,  $\lambda_i$  estimado, puede estar altamente correlacionado con los elementos de  $X_i$ . Tal multicolinealidad lleva a errores estándares muy elevados para la  $\beta_j$  estimada.

Este criterio es el más utilizado para asegurar o determinar la identificación de una ecuación. De igual forma, hay que tomar en cuenta que el criterio de exclusión de una o varias variables está basado en expectativas *a priori* o teóricas acerca de la ausencia de ciertas variables en una ecuación determinada.

Para este trabajo las dos variables de selección que no se encuentran en el modelo de MCO y que tienen presencia en el modelo *Probit*, son por un lado, la que nos indica el tamaño de la familia del migrante y por otro la variable dicotómica que nos

dice si el migrante paga o no impuestos federales. Estas dos variables son de crucial importancia ya que el número de miembros del hogar nos proporciona cierto grado de dependencia con el país de origen, además del compromiso y los lazos que mantiene el migrante con su familia. La variable dicotómica que denota el pago de impuestos federales nos da la suposición de que el migrante está en el extranjero de manera formal, hecho relevante para proporcionarnos probabilidades de envío de remesas al país de origen (Durand, *et. al.*, 1996).

La segunda condición es la de rango, la cual es necesaria y suficiente para comprobar la identificabilidad del modelo. Se logra solamente cuando el coeficiente de la variable excluida en la función de  $Z$  no es cero, es decir, si la variable excluida efectivamente hace parte de la función  $Z$ , no solo en forma probable sino en forma real. Aunque, como afirma Harvey en Gujarati (2001), la condición de orden por lo general es suficiente para asegurar la identificabilidad y aunque es importante tener presente la condición de rango, la no verificación de su cumplimiento raramente generará problemas en el modelo.

Por lo que el modelo por corrección de Heckman queda de la siguiente manera:

$$(4.8) \quad \text{Remit} = \beta_0 + \beta_1 (\text{salario}) + \beta_1 (\text{sex}) + \beta_2 (\text{edad}) + \beta_3 (\text{edad}^2) + \beta_4 (\text{soltero}) + \beta_5 (\text{unionlibre}) + \beta_6 (\text{separado}) + \beta_7 (\text{viudo}) + \beta_8 (\text{edu}) + \beta_9 (\text{edu}^2) + \beta_{10} (\text{usexp}) + \beta_{11} (\text{usexp}^2) + \beta_{12} (\text{hijosest}) + \beta_{13} (\text{desemp}) + \beta_{14} (\text{welfare}) + \beta_{15} (\text{setentas}) + \beta_{16} (\text{ochentas}) + \beta_{17} (\text{noventas}) + \beta_{18} (\text{bajo}) + \beta_{19} (\text{medio_bajo}) + \beta_{20} (\text{medio}) + \beta_{21} (\text{medio_alto}) + u$$

*Ecuación de selección:*

$$(4.9) \quad \text{Remit\_biv} = \beta_0 + \beta_1(\text{sex}) + \beta_2(\text{edu}) + \beta_3 + \beta_4(\text{miembros}) + \beta_5(\text{hijosest}) + \beta_6(\text{desemp}) + \beta_7(\text{fedtx}) + u$$

<b>Variable</b>	<b>Definición</b>
Variables dependientes:	
Remit	Remesas enviadas mensualmente en forma de dólares.
Remit_biv	Variable dicotómica que representa 1= si envía remesas, 0= no envía ninguna cantidad de remesas.
Variables independientes:	
Salario*	Salario mensual en dólares en el último viaje, tomando como base el 2000.
Sex*	Variable dicotómica, que define 1= masculino, 0= femenino.
Edad	Edad del migrante en el momento de la última migración.
Edad2	Variable Edad al cuadrado.
Soltero	Variable discreta igual a 1 si el migrante es soltero, 0 en otro caso.
Unionlibre	Variable discreta igual a 1 si el migrante esta en unión libre, 0 en otro caso.
Separado	Variable discreta igual a 1 si el migrante está separado, 0 en otro caso.
Viudo	Variable discreta igual a 1 si el migrante es viudo, 0 en otro caso.
Edu*	Número de años de educación formal por parte del migrante.
Edu2	Variable años de educación al cuadrado.
Miembros*	Número de miembros del hogar en la familia del migrante.
Usexp*	Años de experiencia laboral del migrante en la Unión Americana.
Usexp2	Variable años de experiencia laboral al cuadrado.
Fedex*	Variable dicotómica, definida como 1 si el migrante paga Impuestos Federales e igual a 0 en otro caso.
Hijosest*	Variable dicotómica, igual a 1 si el migrante tiene por lo menos a uno de sus hijos consigo estudiando en escuela pública dentro de E.U, igual a 0 en otro caso.
Desemp*	Variable dicotómica, igual a 1 si el migrante ha recibido algún bono de desempleo por parte del gobierno de E.U, e igual a 0 en otro caso.
Welfare*	Variable discreta, definida con 1 si el encuestado ha recibido asistencia

pública por parte del gobierno estadounidense, igual a 0 en otro caso.

---

VARIABLES DE TIEMPO:

Setentas	Variable dicotómica, definida con 1 si la fecha en que el migrante viajó por última vez a los E.U. cae en la década de los setentas, e igual a 0 en otro caso.
Ochentas	Variable dicotómica, definida con 1 si la fecha en que el migrante viajó por última vez a los E.U. cae en la década de los ochentas, e igual a 0 en otro caso.
Noventas	Variable dicotómica, definida con 1 si la fecha en que el migrante viajó por última vez a los E.U. cae en la década de los noventas, e igual a 0 en otro caso.

---

VARIABLES DE LUGAR O ESTADO:

Bajo	Variable dicotómica, definida con 1 si el migrante proviene de los estados de Oaxaca, Michoacán, Veracruz y Guerrero e igual a 0 en caso contrario.
Medio_bajo	Variable dicotómica, definida con 1 si el migrante proviene de los estados de Hidalgo, Zacatecas, Puebla y Tlaxcala e igual a 0 en caso contrario.
Medio	Variable dicotómica, definida con 1 si el migrante proviene de los estados de Nayarit, Guanajuato, Sn. Luis Potosí, Mexico e igual a 0 en caso contrario.
Medio_alto	Variable dicotómica, definida con 1 si el migrante proviene de los estados de Sinaloa, Durango, Colima y Jalisco e igual a 0 en caso contrario.
Alto	Variable dicotómica, definida con 1 si el migrante proviene de los estados Chihuahua, Baja California Norte, Aguascalientes y Nuevo León e igual a 0 en caso contrario.

---

\*Variables utilizadas en la ecuación de selección.