

3. Metodología

A continuación se presenta la metodología que se seguirá para efectos del trabajo. El objetivo de este estudio es observar varios fenómenos relacionados con la difusión de tecnología; en específico, la difusión de computadoras en el municipio de Puebla. Este trabajo pretende analizar los fenómenos sociales relacionados a las dos principales corrientes de la literatura, los modelos epidemiológicos y de umbral ya que ambos enfatizan ciertos aspectos del mecanismo de difusión que pueden ser observados empíricamente. Como fue mencionado en el capítulo anterior, una de las virtudes de los modelos de umbral es que su especificación econométrica está abierta a la inclusión de variables de otras corrientes. Es por eso que con frecuencia se utilizan características de los modelos epidemiológicos, relevantes a las redes de comunicación y flujo de información dentro de un planteamiento de modelo umbral, que por lo general utiliza los modelos Probit como principal herramienta econométrica. Un modelo Probit es aquel que se utiliza para explicar un hecho cualitativo (no cuantitativo) donde se explica un resultado binario, es decir 1 ó 0. Se trata de una regresión lineal múltiple con una variable dependiente binaria que recibe el nombre de Modelo de Probabilidad Lineal (dado que la probabilidad de respuesta medida por los parámetros β_j es lineal):

$$(7) \quad P(y = 1 | x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

A continuación se presenta el planteamiento teórico de este modelo básico, que en esencia, es un modelo híbrido donde se mezclan los modelos Epidemiológico y de Umbral al incluir factores de flujo de información en un modelo Probit. Más adelante se presentan las regresiones utilizadas y sus explicaciones.

Se comenzará con el planteamiento general del modelo donde primero se observa el esquema de precio de reserva, después se añaden las variables de flujo de información. La especificación empírica del modelo es la siguiente: para el hogar i en el año t existe la decisión Y_{it} donde $Y_{it} \in \{0,1\}$, donde 1 significa que el hogar tiene computadora y 0 significa que no la tiene. Si el precio de reserva de un hogar es p_{it}^* en el año t y p_{it} es el precio de mercado con el que se topa el hogar i en el año t , entonces:

$$(8) \quad \begin{array}{ll} Y_{it} = 1 & \text{si } p_{it}^* \geq p_{it} \\ Y_{it} = 0 & \text{si } p_{it}^* < p_{it} \end{array}$$

Nótese que este planteamiento, hasta ahora, sigue estrictamente al modelo umbral puesto que la decisión de compra se dicta por medio de la relación costo-beneficio entre el precio de la tecnología, y el precio de reserva. Si un hogar compra una computadora en el año t , y ésta es su primera compra, éste es el primer año en el que su precio de reserva está por encima del precio de mercado, entonces se especifica:

$$(9) \quad P(Y_{it} | x) = \beta_0 x_i^0 + x_{it}^U + x_{it}^R + u_{it}$$

Las variables x_i^0 son un vector de características de los hogares; éstas son las socio-económicas y demográficas (en el siguiente apartado se menciona cuáles son esas variables). Las variables x_{it}^U son las variables de sofisticación y actitud en los hogares; éstas sirven para encontrar posibles sesgos provenientes de características comunes entre los hogares que no fueron observadas. El uso de las variables de sofisticación es para verificar la situación del sistema social, sea éste donde los hogares son muy propensos a tener artículos de vanguardia tecnológica o no. Si se tratase de un sistema ‘sofisticado tecnológicamente’, entonces es de esperarse que los hogares sean más propensos a comprar computadoras; por el contrario, si es un sistema ‘no sofisticado tecnológicamente’, entonces los hogares serán menos propensos a comprar una computadora. Estas variables de actitud medirán la postura de los hogares frente a las computadoras; es decir, su percepción en cuanto a estos artículos, si los consideran caros, les gustan, los consideran necesarios, etc. El término x_{it}^R es para las variables que miden los efectos de red. Finalmente, la variable u_{it} será el término de error. En pocas palabras, se tiene que la decisión de compra es estrictamente determinada por consideraciones de umbral, pero la probabilidad de esta compra (es decir, la probabilidad de que el precio de reserva supere al precio de la tecnología) se determina por consideración de ambos modelos como también por factores no-observables.

Más adelante se plantea el análisis por separado de cada regresión, por el momento se mencionará brevemente cómo se alcanzarán los objetivos de este estudio. Para probar el primer objetivo del trabajo se partirá de la información recopilada de:

cuándo compró su primera computadora el individuo, de esta manera tendremos una serie de datos a lo largo del tiempo que permitirán graficar la curva. El segundo objetivo se verificará buscando patrones similares de uso entre individuos a través de características similares socio-económicas, actitudes y preferencias el uso de Internet y vínculos del usuario con otros individuos (familiares y amigos). El tercer objetivo se alcanzará cuando se evalúe la significancia de las variables que involucran a ambos modelos (Epidemiológico y Umbral) para verificar cuales tiene mayor peso.

3.1 Análisis de Variables

El primer grupo de variables se refieren a aspectos demográficos de los individuos. La primera de estas es la zona de la ciudad en la que vive el individuo, esta variable se medirá de 1 a 5 siendo 1 el estrato de menores ingresos y 5 el estrato más rico, así se medirá si el ingreso influye en la decisión de compra de una computadora (Por razones de confidencialidad se decidió utilizar esta variable como *proxy* de ingreso). El sexo del encuestado servirá para buscar diferentes patrones de comportamiento entre hombres y mujeres. Las variables edad (3), número de habitantes en el hogar (4) y años de educación (5) son continuas y se medirán en años; la existencia de personas menores de 18 años se considera porque al tener menores de edad, se intuye que éstos puedan estar asistiendo a la escuela y que la computadora sea necesaria para sus tareas, la pregunta 4 es para obtener la cantidad de computadoras por habitante. Usar computadora en su trabajo o escuela y el llevar a casa trabajo desde éstas, pueden influir en la decisión de tener o no el bien estudiado. Estas preguntas 6, 7 y 8 son atributos del hogar y se miden como 1 ó 0 al ser variables dicotómicas. La pregunta 9, la variable dicotómica sobre si adoptó o no el individuo, es utilizada como la variable dependiente en todas las regresiones.

En cuanto a los hogares que poseen una computadora, se les hacen una serie de preguntas específicas tales como el número de computadoras que tienen (10), los años en que se compraron la primera y última computadora en ese hogar (11 y 12), así mismo, con los datos de la pregunta 11 se podrá construir la curva de difusión pues indican la fecha de adopción inicial de cada hogar. La siguiente es para medir la intensidad de uso (13), medida como un rango. Estas cuatro preguntas son para medir la

afinidad hacia la tecnología en estudio y se miden también de manera continua. Finalmente, la pregunta número 14 pretende analizar las razones principales por las que una persona compró la tecnología.

Para la sección 'Otras', se tienen preguntas que se harán a ambos bloques (los adoptantes y no-adoptantes). La pregunta 15 contiene las variables de actitud de los hogares hacia la tecnología, y como ya se expuso anteriormente, son el primer bloque de variables que buscarán posibles sesgos por características comunes entre los hogares que no fueron observadas. Las preguntas 16 y 17 miden las externalidades de red; es decir, la influencia de amigos, familiares y conocidos que tienen o no tienen computadora, y que se conectan o no se conectan a Internet; éstas se miden por un rango.

Las tres siguientes preguntas (18, 19, 20) son relevantes al Internet y van encaminadas a conocer los posibles efectos de red de esta tecnología. La pregunta 18 ofrece tres opciones según desde dónde se conecte el individuo; cada opción se mide de manera independiente como 1 ó 0; si contestan en las tres opciones un 'no', eso quiere decir que no se conecta a Internet. La pregunta 19 se mide por un rango que va desde 1 hasta 5 dependiendo del tiempo que se navegue en Internet mientras que la número 20 se mide con un rango que va de 0 a 6 según el tiempo que el entrevistado lleve conectándose.

La pregunta 21 mide la posibilidad de comprar una computadora nueva a través de un rango que va de 0 a 2. La pregunta número 22 mide por medio de un rango que va de 1 a 5 la disponibilidad del hogar para pagar por una computadora nueva (precio de reserva); estas dos preguntas se encuentran en el cuestionario para medir los efectos de umbral. Finalmente, la pregunta 23 contiene al segundo bloque de variables que ayudarán a buscar los sesgos por características comunes entre los hogares que no fueron observadas; estas variables son las variables de sofisticación e indicarán el nivel de aceptación que tienen los hogares por accesorios de alta tecnología. En una ciudad tecnológicamente sofisticada donde la mayor parte de la población tiene computadora, es lógico que el resto de la población se vea influenciada por esta sofisticación y tenga una mayor inclinación a comprar una computadora; de igual manera una ciudad con poca sofisticación donde la mayoría de pobladores no tienen computadora, los que no

poseen una no tendrán gran inclinación a comprarla. De ahí la necesidad de usar estas variables *proxy* (Goolsbee y Klenow, 1999).

3.2 Muestreo

Para obtener la información, se aplicará una encuesta en el municipio de Puebla. La población de este municipio según el Censo General de Población Vivienda de INEGI del años de 2000 es de 1,346,916 habitantes. También se obtuvo de ese mismo Censo la población por colonias y el número de hogares con ingresos mayores a 5 salarios mínimos; se obtuvieron los porcentajes y con base en éstos se ordenaron las colonias y se dividieron en 5 zonas, del más pobre al más rico. El número de colonias en el Municipio de Puebla es de 532 y eso hace que en cada bloque haya 108 colonias en promedio; es decir están organizadas en quintiles, del más pobre al más rico. De esta manera hay una desviación estándar de 2,978 y una media de 2,527. Por cuestiones de disponibilidad de tiempo y recursos, fue necesario manejar un nivel de confianza de 90% y un nivel de error de 10%; así, el tamaño óptimo de la muestra es de 356 individuos los cuales se dividieron entre las 5 zonas de colonias. Debido a que solo interesa entender el nivel de adopción por medio de cada zona de ingreso, el numero de hogares en la muestra está repartido equitativamente entre las cinco zonas, con cada zona con 71 hogares encuestados (una de ellas con 72).

El cuestionario está basado en la encuesta del trabajo de Goolsbee y Klenow (1999), la cual fue encargada a la empresa Technographics. La encuesta original es bastante amplia e involucra algunas preguntas que no fueron utilizadas; los autores simplemente utilizaron las preguntas que consideraron necesarias. De la misma manera, en este trabajo se utilizaron las variables que fueron necesarias; resulta obvio que habrá algunas preguntas que deberán modificar su formato debido a factores culturales y sociales. La experiencia obtenida en las realización de encuestas piloto determinó que el punto focal en el formato del cuestionario era hacerlo entendible a todos los estratos de la población aún cuando esto pudo implicar una ligera pérdida de profundidad en el trabajo. Es decir, no se pudo obtener tanta información como se quisiera, pero eso fue preferible a que los entrevistados respondieran mal o no respondiera porque eso hubiera implicado perder observaciones e incurrir en posibles sesgos.

3.3 Análisis de Regresiones

Se plantean las siguientes regresiones:

3.1 Demografía

$$COMP = \beta_1 + \delta_1 SEXO + \beta_2 EDAD + \beta_3 HAB + \beta_4 EDUCAJ + \delta_5 MENOR + \delta_6 CASA + \delta_7 TRAB + \sum_{i=1}^n \delta_8 ZONA_i + \mu$$

Este es un modelo Probit debido a que se utiliza una variable dependiente dicótoma como lo es *COMP*, se correrá la regresión contra las variables demográficas (variables de interés) verificando la hipótesis de que este tipo de variables influyen en la decisión de adopción de los individuos de la misma manera que lo hicieron Goolsbee y Klenow. En este caso se usa la variable *ZONA* dividida en sus 5 categorías previamente explicadas. Es importante precisar que para efectos econométricos, no se utilizó la variable de educación sino una variable de educación ajustada (*EDUCAJ*) para corregir para aquellas personas que debido a su menor edad no alcanzaban cierto número de años de escolaridad. La ecuación para obtener el ajuste es la siguiente:

$$(10) \quad \begin{aligned} EDUCAJ &= EDUC_i - 9 && \text{si } EDAD_i \geq 15 \\ EDUCAJ &= (15 - EDAD_i) + (EDUC_i - 9) && \text{si } EDAD_i < 15 \end{aligned}$$

Para el ajuste, se toma 9 años de escolaridad como los mínimos obligatorios por lo que cada año adicional de escolaridad queda como saldo positivo, mientras que un número menor a 9 años se registra como negativo. Sin embargo, para aquellos menores de 15 años solamente se toma como negativo si es que está rezagado con respecto a su edad (por ejemplo un encuestado de 14 años con 8 años de escolaridad tendría un valor de 0 por no tener rezago, sin embargo, ese mismo encuestado con 7 años de escolaridad tendría un valor de -1).

3.2 Actitud y Sofisticación

$$COMP = \beta_1 + \sum_{i=1}^n \delta_1 ACTIT_i + \sum_{i=1}^n \delta_2 SOFIST_i + \mu$$

Debido a que se esperan sesgos en esta ecuación como resultado de características comunes entre los hogares no observadas, se correrá una nueva regresión Probit con las variables de dos preguntas, la 15 y la 23, para medir a través de estas *proxy* la sofisticación y la actitud del individuo hacia la tecnología. Para este rubro se corre una regresión adicional tomando la sofisticación como un rango, es decir, como la suma de todos los artículos de tecnología siendo un índice del nivel de sofisticación.

3.3 Comunicaciones e Internet

$$COMP = \beta_1 + \sum_{i=1}^n \delta_1 LUGCON_i + \sum_{i=1}^n \delta_2 HORAS_i + \sum_{i=1}^n \delta_3 TIEMPO_i + \sum_{i=1}^n \delta_4 FACOMP_i + \sum_{i=1}^n \delta_5 FAINT_i + \mu$$

Esta regresión se incluye para verificar si las redes y las externalidades influyen en la situación de tener o no una computadora por parte del individuo. El primer grupo de variables, LUGCON, HORAS y TIEMPO son referentes al Internet mientras que el segundo grupo, FACOMP y FAINT buscan efectos de red entre la comunidad inmediata del encuestado. Con estas variables se probará la hipótesis de que las redes personales y las externalidades de red, vinculadas principalmente al Internet, tienen injerencia en la decisión de compra del individuo.

3.4 Precio de Reserva

$$COMP = \beta_1 + \sum_{i=1}^n \delta_1 RES_i + \sum_{i=1}^n \delta_2 ZONA_i + \mu$$

Con esta regresión se plantea analizar los efectos de umbral, la variable *RES* da información sobre la capacidad de pago del individuo. Como se menciona anteriormente, el individuo adoptará una si recibe los suficientes beneficios por adoptarla comparada con el costo de adquisición. Se incluye la variable *ZONA* debido a que el precio que una persona está dispuesta a pagar está ligado a su ingreso, de esta manera se controla para el ingreso con esta variable *proxy*.

Tabla 3.1 Lista de Variables

#	NOMENC.	VARIABLE	DESCRIPCIÓN
1	ZONA	Zona de la ciudad en la que vive	= I muy pobre = II pobre = III media = IV media alta = V alta
2	SEXO	Sexo del individuo	= 1 Hombre = 0 Mujer
3	EDAD	Edad del individuo	Continua en años
4	HAB	Número de habitantes en el hogar	Continua en Años
5	EDUC	Años de educación del individuo	Continua en años
6	MENOR	Menores de 18 años	= 1 Si = 0 No
7	CASA	Trabajo de oficina a casa	= 1 Si = 0 No
8	TRAB	Computadora en trabajo	= 1 Si = 0 No
9	COMP	Computadora en casa	= 1 Si = 0 No
10	NUM	Computadoras que tiene	Continua
11	PRIM	¿Cuándo compró su primera Computadora?	Continua
12	ULT	¿Cuándo compró su última computadora?	Continua
13	INTENS	¿Qué tan seguido usa su computadora?	Continua de 0 a 6 según escala
14	PQSI	¿Por qué compró o no compró una computadora?	a) = 1 Hubo recursos para compra inmediata b) = 1 Hubo opciones de crédito para compra a plazos c) = 1 Es necesaria para la escuela de algún familiar d) = 1 Es necesaria para el trabajo de algún familiar e) = 1 Aligera mi carga de trabajo f) = 1 Porque mucha gente ya tiene una g) = 1 Por entretenimiento (juegos) h) = 1 Para estar conectado por Internet i) = 1 Para comunicarme con otras personas
15	ACTIT	Decisión de Compra o No Compra	a) = 1 Las Computadoras son caras b) = 1 Es importante tener una Comp. c) = 1 Me intimidan las Computadoras d) = 1 Me gustan las Computadoras e) = 1 Se usar las computadoras f) = 1 Arraigo de la Computadora g) = 1 Aligera carga de Trabajo h) = 1 Influye publicidad i) = 1 Internet = importante TV y Radio

16	FACOMP	Familiares y Amigos con Computadora	Continuo 1 a 3 (1 menos de la mitad, 2 la mitad, 3 más de la mitad)
17	FAINT	Familiares y Amigos con Internet	Continuo 1 a 3 (1 menos de la mitad, 2 la mitad, 3 más de la mitad)
18	LUGCON	Desde donde se conecta	=1 Casa =1 Oficina =1 Café Internet
19	HORAS	Horas a la semana que navega	Continua de 0 a 5 según escala
20	TPO	Tiempo que lleva conectado	Continua de 0 a 6 según escala
21	POSIB	Posibilidad de comprar una Computadora	Continua de 0 a 2
22	RES	Precio de reserva	=1 \$0 a \$5 (miles) =2 \$5 a \$10 (miles) =3 \$10 a \$15 (miles) =4 \$15 a \$20 (miles) =5 +\$20 (miles)
23	SOFIST	Accesorios de vanguardia tecnológica	a) =1 Teléfono Celular b) =1 Palm c) =1 DVD d) =1 Videocámara e) =1 Cámara Digital f) =1 TV Satelital o Cable g) =1 Videojuegos