

MODELOS MULTIECUACIONALES RECURSIVOS PARA LA EXPLICACION DE LA OFERTA DE MANO DE OBRA

Felix Pimentel

Con el asesoramiento de: Héctor J. Apezchea

Aníbal Alvarez

Centro de Informaciones y Estudios del Uruguay
Montevideo - Uruguay.

Este trabajo sobre modelización que presentamos es la primera aproximación de un estudio más general y exhaustivo sobre oferta de mano de obra, desarrollado actualmente por C.I.E.S.U.

Al diseñar la estrategia de modelización de los dos modelos recurrentes que decidió tratar CIESU, consideramos en principio dos problemas previsibles:

- Un costo de procesamiento relativamente alto.
- La estimación de varios modelos cuyas variables endógenas (\hat{p} - Probabilidad de participar en la fuerza de trabajo, \hat{E} - Probabilidad de estar empleado participando en la fuerza de trabajo), eran variables "Dummy", dicotómicas, con valores 0-1.

La bibliografía existente, es absolutamente coincidente en que estos modelos a los que usualmente se los designa como modelos "PROBIT", no cumplen con una de las condiciones del modelo lineal, la condición de homoscedasticidad, es decir $E(u_1' u_1) = \sigma^2 \mathbf{I}_T$ (la matriz de variancias y covariancias de las 1^{er} perturbaciones debe ser una matriz identidad multiplicada por una covarianza constante).

Con estos modelos PROBIT, claramente heterocedásticos, al aplicar mínimos cuadrados corrientes, se obtienen estimadores insesgados, lineales pero que no son de mínima varianza.

Lo ideal habría sido tratar los modelos "PROBIT" con los métodos aconsejados usualmente (ver bibliografía, 1), 2), 5), 6)). Sin embargo, debido a que sociología ha ignorado con bastante éxito este problema de la heterocedasticidad, y a que en definitiva, como ya dijimos, este trabajo sólo pretende una primera aproximación; decidimos realizar las estimaciones como si se cumplieren la totalidad de las hipótesis del modelo lineal clásico, juzgando a posteriori la calidad de las estimaciones \hat{P} y \hat{E} , mediante experimentos numéricos relativos a fenómenos suficientemente conocidos por la sociología.

El problema de la heterocedasticidad que desarrollaremos en detalle más adelante, nos indujo a tratar con mucho cuidado los modelos en que se estimaban \hat{E} y \hat{P} , por ejemplo excluyendo el salario como variable explicativa de \hat{P} , debido a que $P=0$ implica en nuestra muestra salario propio = 0, y a que introducir un salario estimado complicaría aún más la estimación.

Por las previsibles dificultades que hemos expuesto, se diseñó una estrategia de modelización, modular e incluso indeterminada "a priori".

ESTRUCTURA MODULAR E INDETERMINADA DE LOS MODELOS RECURRENTE

La estructura es indeterminada y modular por lo siguiente:

a) Existen ecuaciones que integran más de un modelo recurrente, por ejemplo la ecuación (1.) (ver Cuadro I).

$$1.) \hat{P} = F_h (X_h)$$

(Probabilidad de participar en la población económicamente activa, caso hombres).

b) Se decidió implicar ecuaciones en el modelo recurrente sólo en el caso de que los tests estadísticos y la teoría sociológica justificaran su inclusión.

Ejemplo:

Un posible modelo aplicable a hijos varones pertenecientes a la población económicamente activa, pudo ser:

$$1) \hat{P}_h = F_h (X_h)$$

2) $\ln_h \hat{w}_1 = g_{2,h} (X_{2,h})$ Estimación del salario del varón, PEA, cabeza de hogar.

3) $\ln_m \hat{w}_2 = g_{2,m} (X_{2,m})$ Estimación del salario de la compañera, PEA.

4) $\ln \hat{w}_h = g_{3,h} (X_{3,h})$ Estimación del salario de los hombres, PEA, en general.

5) $\hat{E}_{3,h} = H_{3,h} (X_{3,h}, \ln \hat{w}_h, \ln \hat{w}_1, \ln_m \hat{w}_2)$ Probabilidad de estar ocupado perteneciendo a la PEA, casos hombres

Pero si la realidad estadística y sociológica nos hubiera dicho que $\ln_h w_1$ y $\ln_m w_2$ eran tan poco significativas que podían llegar a estropear la ecuación 5), es muy posible que hubiésemos optado por el modelo:

- 1) $\hat{P}_h = F_h (X_h)$
- 2) $\ln \hat{w}_h = G_{3,h} (X_{3,h})$
- 3) $\hat{E}_{3,h} = H_{3,h} (X_{3,h}, \ln \hat{w}_h)$

Eliminando dos posibles ecuaciones, y modificando una, que por ser la última y por haberse utilizado el método "paso a paso" del sistema "STEP-WISE" no habría representado mayor problema. (Ver cuadro I: Esquema inicial de la modelización).

CONJUNTOS DE VARIABLES EXOGENAS, SU AMBITO Y SU NOTACION

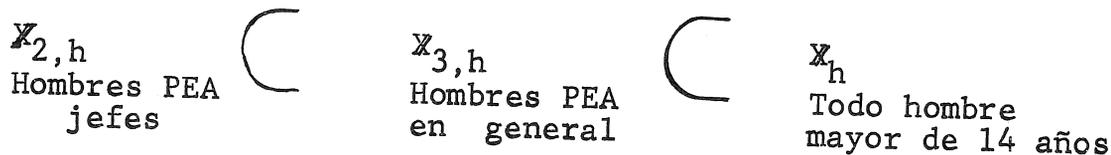
La notación X indica conjuntos de variables exógenas (vector de variables exógenas para un caso en el corte transversal); sus subíndices denotan más precisamente el ámbito y los subconjuntos de los datos que se tratan.

Ejemplo:

$X_{2,h}$ Indica el conjunto de las variables exógenas correspondientes a cabeza de hogar - PEA (2) y además hombres (h).

Es de hacer notar que los diferentes conjuntos exógenos cumplen con relaciones de contención estrictas.

Ejemplo:



ESTIMACION DEL MODELO RECURRENTE

Problema de las variables endógenas que aparecen como variables explicativas.

Distinguiremos dos tiempos:

tiempo de estimación

tiempo de pronóstico

En tiempo de estimación se utilizaron las variables endógenas-explicativas con sus valores reales y no con los valores estimados en ecuaciones anteriores.

En cambio en tiempo de pronóstico se hará lo único posible: uti-

lizar los valores estimados.

JUSTIFICACION DEL METODO DE ESTIMACION UTILIZANDO VALORES REALES PARA LAS VARIABLES ENDOGENAS EXPLICATIVAS

Esta bibliografía (*) y el método de máxima verosimilitud con información completa (del que es un caso particular el modelo recurrente), permiten afirmar que si las diferentes perturbaciones u_i correspondientes a cada ecuación son independientes, entonces es perfectamente lícito realizar la estimación con las variables $\ln w_i$ en lugar de las variables $\ln \hat{w}_i$

No hemos realizado pruebas de hipótesis que nos aseguren la independencia de las perturbaciones u_i de cada ecuación, fundamentalmente porque para una muestra grande como la utilizada, es probable que su costo fuese bastante alto.

La decisión, se basó en consideraciones mucho más empíricas; sencillamente hemos realizado estimaciones de \hat{p} y \hat{E} con ambas posibilidades (utilizando las variables reales $\ln w$ y las variables estimadas $\ln \hat{w}$), obteniendo resultados notablemente mejores con las variables reales $\ln w$. (En los experimentos realizados la correlación múltiple de las estimaciones \hat{p} y \hat{E} , aumentó sistemáticamente entre 20 y 30 puntos cuando se utilizó la variable real $\ln w$ en lugar de las estimaciones $\ln \hat{w}$).

EL PROBLEMA DE LAS ESTIMACIONES GENERALES SIN RECURRIR A ESTRATIFICACIONES POR EDAD

Hemos optado por dejar para una segunda etapa la modelización estratificada por edades por varias razones:

- 1) Los experimentos generales en que no se recurre a estratificaciones afinadas, si bien pagan un precio en bondad de ajuste y significación debido a lo heterogéneo de la población, en cambio brindan panoramas de gran interés para el investigador.

No significa lo anterior negar la importancia de recursos que como la estratificación desean lograr poblaciones homogéneas, sencillamente pensamos que el estudio general es un paso previo, habitualmente imprescindible, si se pretende que los modelos hagan el mayor aporte a la comprensión de la realidad.

- 2) Imposibilitados de recurrir a la sofisticada metodología estadística aconsejada para los modelos PROBIT (ver p/ejemplo,

(*) WILLIAM MERRIL/KARL FOX. Introducción a la Estadística Económica/J. JOHNSTON.

ELEMENTS OF ECONOMETRICS-JAN KMENTA), y limitados también en cuanto a recurrir a la validación a través del pronóstico con frontado con la realidad. Decidimos, como ya dijimos, poner a prueba la simplificación de utilizar el modelo lineal clásico para la estimación de E y \hat{P} , realizando experimentos y análisis, con variables y fenómenos de efecto muy conocido para los especialistas, en particular: las variables edad y experiencia laboral. Estos experimentos presumiblemente hubieran sido poco claros y de difícil interpretación de haber trabajado con poblaciones altamente homogéneas en relación a la edad, y por la alta correlación edad-experiencia, también en relación a la experiencia laboral.

- 3) La estratificación habría aumentado bastante los costos de computador, no sólo porque se multiplica la cantidad de modelos, sino además porque la metodología de experimentación e incluso de ensayo y error sería notablemente más compleja sin el conocimiento general que actualmente poseemos.

PROBLEMA DE LA HETEROSCEDASTICIDAD EN LA ESTIMACION DE LAS VARIABLES \hat{P} y \hat{E} .

Las variables P y E son endógenas y dicotómicas. Toda la bibliografía a nuestro alcance y que trata el tema, coincide en señalar que en estos casos se produce heterocedasticidad.

Nos limitaremos a citar a J. JOHNSTON (Métodos de econometría) 2da. edición en castellano -pág. 227-

"Goldberger ha demostrado que una dificultad que concurre con la aplicación de los mínimos cuadrados clásicos cuando Y es dicotómica es que no puede mantenerse la hipótesis de que las perturbaciones son homoscedásticas".

En esta situación las estimaciones serán lineales e insesgadas pero no óptimas (de mínima variancia). La condición de mínima variancia se lograría con el procedimiento en dos etapas sugerido por Goldberger - J. JOHNSTON (pág. 228).

DEMOSTRACION DE LA EXISTENCIA DE HETEROSCEDASTICIDAD

Sea nuestro modelo para el caso i :

$$1.1) P_i = b_0 + b_1 x_{1_i} + b_2 x_{2_i} + \dots + b_{13} x_{13_i} + u_i$$

llamando $X_i = (1, x_{1_i}, x_{2_i}, \dots, x_{13_i})$

$$\beta = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{13} \end{bmatrix}$$

$$P_i = X_i' \beta + u_i$$

$$u_i = P_i - X_i' \beta$$

Como P_i vale 0 ó 1

u_i vale:

$$1 - X_i' \beta \text{ para } P_i = 1$$

$$\text{ó } - X_i' \beta \text{ para } P_i = 0$$

En estas condiciones aceptando $E(u_i) = 0$, la función de distribución de u_i es:

u_i	$P(u_i)$
$- X_i' \beta$	$1 - X_i' \beta$
$1 - X_i' \beta$	$X_i' \beta$

$$\begin{aligned} \text{Var}(u_i) &= E(u_i^2) = (-X_i' \beta)^2 (1 - X_i' \beta) + (1 - X_i' \beta)^2 X_i' \beta = \\ &= (X_i' \beta) (1 - X_i' \beta) = E(P_i) (1 - E(P_i)) \end{aligned}$$

Siendo obvia la heteroscedasticidad ya que esta variancia depende de P_i y que además es fácil deducir que la matriz de variancias y covariancias es:

$$E(u_i u_i') = \begin{bmatrix} X_1' \beta (1 - X_1' \beta) & 0 \dots 0 \\ 0 & X_2' \beta (1 - X_2' \beta) & 0 \dots 0 \\ \vdots & & \vdots \\ 0 \dots & & X_{13}' \beta (1 - X_{13}' \beta) \end{bmatrix}$$

LA VALIDACION DE LOS MODELOS A TRAVES DEL EFECTO DE LAS
VARIABLES EDAD (A.A²) Y EXPERIENCIA LABORAL (E.LAB.E.LAB²).

El ejemplo más espectacular para fundamentar la utilidad de los experimentos generales que hemos realizado, ocurrió con la "Experiencia laboral" en la estimación de \hat{P}_h (Probabilidad de pertenecer a la población económicamente activa). En esta estimación, en los 2 primeros pasos de ensayo el sistema "STEP-WISE" selecciona espontáneamente E.LAB² y E.LAB, logrando muy buenos indicadores estadísticos tanto en bondad de ajuste como en significación.

DETALLES DEL MODELO PARCIAL

Correlación múltiple 0.64

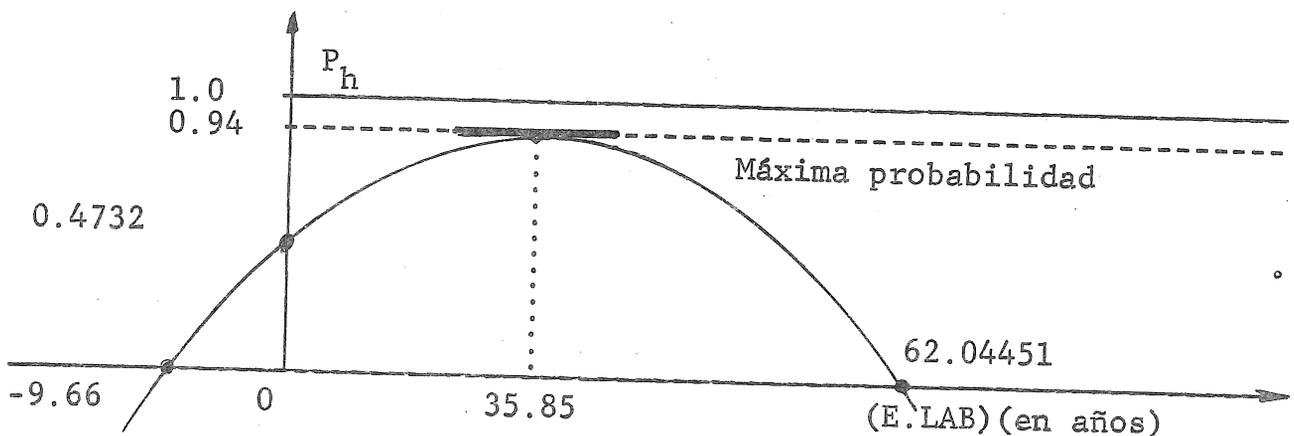
F-Fisher 1429.053

$$\hat{P} = - 0.00079 E.LAB^2 + 0.04138 E.LAB + 0.47372$$

t(-50.699) -31- t(43.350)-30-

(0.25) (0.25)

0.00002 0.00095



UNA INTERPRETACION DEL EXPERIMENTO SOBRE EXPERIENCIA LABORAL

Aproximadamente a los 36 años de experiencia laboral, la probabilidad de pertenecer a la PEA alcanza el máximo con un valor muy cercano al suceso seguro (0.94). Cuando se comienza la actividad laboral (experiencia laboral nula) la probabilidad de pertenecer a la PEA es 0.47372.

Estas conclusiones además de muy razonables en cuanto al conocimiento que se posee de la población a estudio, son muy coincidentes, con el efecto edad (ver Cuadro III) que ubica la máxima probabilidad de pertenecer a la PEA (Hombres mayores de 14 años) a los 50 años de edad.

Si además se observan los otros experimentos del Cuadro III todo resulta razonable y satisfactorio.

Estas apreciaciones tan razonables nos inducen a pensar que el problema de heteroscedasticidad en \hat{E} y \hat{P} no fue importante, que estimar con las variables reales $\ln w$, asumiendo que las perturbaciones u_i son independientes fue una decisión acertada y que en general los modelos multiecuacionales recurrentes son aceptables.

ESTAS FUERON LAS LUCES CON RESPECTO AL EXPERIMENTO CON EXPERIENCIA LABORAL, VEAMOS AHORA LAS SOMBRAS

El modelo que hemos expuesto son los dos primeros pasos del modelo 1.1. que se detallan más adelante (ver apéndice con detalle de modelos) y que incluye 13 variables de las cuales 10 son altamente significativas, la correlación múltiple alcanza el valor de 0.674 y se obtiene un valor F-Fisher de 259.305, siendo el límite al 1% 3.16 ($F_{0.99}$).

Como vemos este modelo completo sigue siendo muy aceptable.

Nuestras interrogantes son las siguientes:

¿En qué medida la existencia de años de experiencia laboral no implica casi completamente el hecho de pertenecer a la PEA, que es justamente lo que pretendemos explicar?

¿Experiencia laboral merecer aparecer en el modelo 1.1?

¿Estamos ante un modelo fuertemente auto-regresivo con algo de tautología?

¿Cómo afecta la auto-regresión a la heteroscedasticidad de \hat{P} ?

Por supuesto en estas reflexiones y como nota optimista, no podemos olvidar a los jubilados, las amas de casa inactivas, los rentistas, etc., que no perteneciendo a la PEA puedan perfectamente tener años de experiencia laboral siendo posible que las 10 variables significativas que aparecen en la ecuación 1.1 aisen precisamente esos efectos totalmente válidos.

Debemos confesar que no tenemos una opinión muy clara en torno a todas estas interrogantes, y que la inclusión de "Experiencia Laboral" en la explicación de la probabilidad \hat{P} de pertenecer a la PEA, debe ser un tema de profunda discusión.

CUADRO I - ESQUEMA INICIAL DE LA MODELIZACION

Probabilidad de participar en la fuerza de trabajo

(1.1) $\hat{P}_h = f_h (X_h)$

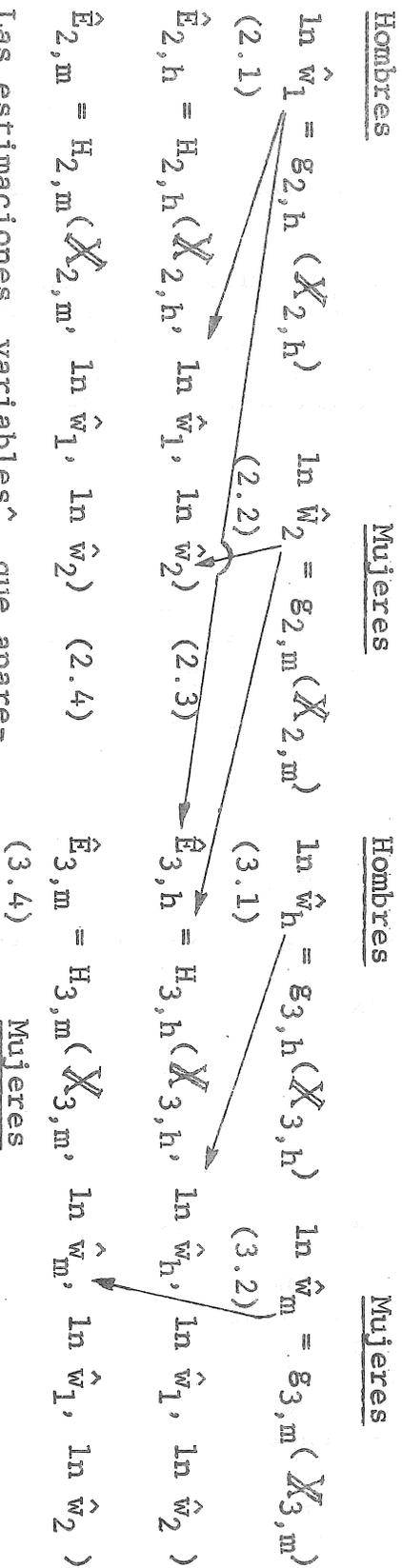
(1.2) $\hat{P}_m = f_m (X_m)$

(1.1.1) Población: Todo hombre mayor de 14 años

(1.2) Población: Toda mujer mayor de 14 años

2.) Población: Hombres y mujeres pertenecientes a la PEA

(3.) Población: Hombres y mujeres pertenecientes a la PEA en general (mayores de 14 años)



Las estimaciones, variables \hat{w} , que aparecen como exógenas son realmente estimaciones en tiempo de pronóstico o predicción.

CUADRO II - EFECTO EDAD PARA EL CASO HOMBRRES

Variable explicada	Población	Paso de ensayo	Función de comparación del efecto edad	Representación	Comentarios
$\ln \hat{w}$ (salario)	Hombres PEA en general	4	$Y = -0.00176 A^2 + 0.17940 A$		El mayor efecto se alcanza a los 51 años.
$\ln w_1$ (salario)	Hombres PEA jefes de hogar	12	$Y = 0.00065 A^2 + 0.06056 A$		El máximo efecto sobre el salario se logra a los 46 años
\hat{p}	Todo hombre mayor de 14 años.	11	$Y = -0.00023 A^2 + 0.02313 A$		El máximo efecto se logra a los 50 años
\hat{E}	Hombres PEA en general ocupado perteneciendo a la PEA	10	$Y = -0.00004A^2 + 0.0036 A$		La máxima probabilidad de estar ocupado en función sólo de la edad se observa a los 45 años.
E	Hombres	9	$Y = -0.00002 A^2 + 0.00117 A$		En los jefes de hogar el máximo efecto se logra 15 años antes.

CUADRO III - DESCRIPCION DE LAS VARIABLES

N°	DESCRIPCION	SIMBOLO
1	Años de Enseñanza Primaria	E.PRIM.
2	" " " Industrial	E.IND
3	" " " Secundaria	SECUND
4	" " " Preparatoria	PREP
5	" " " Normal	EM
6	" " " MILITAR	MI
7	" " " Universitaria	EF
8	Condición de profesional	PROF
9	Edad ₂	A ₂
10	Edad	A
11	Salario Propio	W _i
12	Salario del varón. Cabeza de hogar	W ₁
13	Salario de la cónyuge	W ₂
14	Logaritmo salario propio	ln w _i
15	" " varón. Cabeza de hogar	ln w ₁
16	" " de la cónyuge	ln w ₂
17	Total salarios del hogar	TSH
18	Ingresos adicionales del hogar	I.AD
19	Hijos 0-1 años	N ₀₁
20	Hijos 2-7 "	N ₂₇
21	Hijos 8-14 "	N ₈₁₄
22	Hijos 15 o más años	N ₁₅
23	Hijos total	NT
24	Probabilidad endógena de pertenecer a la PEA	P
25	Probabilidad de estar ocupado perteneciendo a la PEA	E
26	Estimación del salario	ln \hat{w}
27	Estimación de la P corregida	\hat{P}
28	Estimación de la E corregida	\hat{E}
29	Sólo enseñanza Primaria	SP
30	Experiencia Laboral ₂	E.LAB ₂
31	(Experiencia Laboral)	E.LAB
32	Años de escolaridad	AESC
33	Cabeza de hogar	CH
34	Tanto x 1 de hijos menores de 7 años	PN
35	Hombre o mujer	HM

Por limitaciones de espacio sólo presentaremos uno de los cuatro modelos recurrentes estimados, "JEFES DE HOGAR PERTENECIENTES A LA POBLACION ECONOMICAMENTE ACTIVA", restando los casos mujeres-cónyuges PEA, hombres en general y mujeres en general.

PRESENTACION DE LOS RESULTADOS DE LOS MODELOS

Las variables aparecen en el orden en que ingresaron según la elección en función del porcentaje de explicación, realizada por el sistema "STEP-WISE", lo que en definitiva es una medida excelente del aporte específico inicial de cada variable en la explicación de la variable endógena.

Los nombres mnemotécnicos son los detallados en páginas anteriores en (III); para mayor claridad del lector debajo del nombre mnemotécnico aparece el ordinal de la variable en el cuadro III, bajo la forma de -ii-.

Debajo del valor de cada coeficiente de regresión figuran:

- El valor del estadístico t-Student correspondiente, presentado como t(jj.jjj).
- Más abajo y entre paréntesis, el nivel aproximado de significación expresado en porcentaje (se consideran variables significativas las que no superan el nivel 5.0 %). En la misma línea, cuando sea necesario, se indica poca significación con * y sin significación con **.
- En tercer lugar, siempre debajo del coeficiente de regresión, se incluye el error standard correspondiente al coeficiente de regresión.

Debajo del encabezado aparece el N° del modelo asignado en el Cuadro (I) (Esquema inicial de modelización) y su formulación simbólica.

- Se especifica además en que modelo recurrente interviene la ecuación que se detalla.

Antes del desarrollo del modelo en coeficientes y variables, se detalla el coeficiente de correlación múltiple observado, el valor del estadístico F-Fisher y el límite al 1% de significación de la correspondiente prueba de hipótesis Fisher que valida el conjunto de las variables explicativas que participan en el modelo.

Debajo de la variable endógena o explicada, se especifica su media representada por $\overline{\text{xxx.xxxxx}}$, su desvío standard en la muestra y el error standard de la estimación.

MODELO GENERAL PARA HOMBRES MAYORES DE 14 AÑOS

Explicación de la probabilidad de participar en la PEA, para todo varón mayor de 14 años

$$(1.1) \hat{P}_h = f_h (X_h) \quad \text{Participa en los modelos recurrentes:} \quad (1.1, 3.1, 3.3) \\ (1.1, 2.1, 2.3)$$

13 variables

Correlación múltiple 0.674

F-Fisher 259.305 (límite al 1% 3.16 $F_{0,99}$)

$$\hat{P}_h = -0.01383 - 0.00050 E.LAB^2 + 0.02184 E.LAB + 0.00006 I.AD + 0.00003 TSH + 0.05303 E.IND$$

-24-

-31-

-30-

-18-

-17-

-2-

0.75049	t(-6.167)	t(3.654)	t(6.467)	t(6.888)	t(6.764)
0.43278	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)
0.32	0.00008	0.00598	0.00001	0.00000	0.00784

$$+ 0.00769 E.PRIM + 0.02438 SECUND + 0.04559 PREP + 0.07682 SP + 0.02511 A$$

-1-

-3-

-4-

-29-

-9-

t(0.00538)	t(3.857)	t(3.132)	t(2.220)	t(3.151)
(12.5) *	(0.25)	(0.25)	(1.25)	(0.25)
0.00538	0.00632	0.01456	0.03460	0.00797

$$- 0.00025 A^2 + 0.01666 EM - 0.00513 EF$$

-10-

-5-

-7-

t(-3.019)	t(0.942)	t(-0.585)
(0.25)	(25.0) **	**
0.00008	0.01768	0.00878

EXPLICACION DE LA PROBABILIDAD DE ESTAR OCUPADO PARA HOMBRES PEA JEFES DE HOGAR

$$(2.3) \hat{E}_{2,h} = H_{2,h} (X_{2,h} \ln \hat{w}_1, \ln \hat{w}_2) \text{ Participa en el modelo recurrente:}$$

(1.1, 2.1, 2.3)

7 variables Correlación múltiple 0.239

F-Fisher 18.683 (límite al 1% 5.65 $F_{0.99}$)

$\hat{E}_{2,h}$	=	0.56610	-	0.00048	ln \hat{w}_2	+	0.06108	ln \hat{w}_1	-	0.01064	PROF	-	0.00001	A ²
-25-					-16-			-15-		-8-			-10-	
<u>0.96722</u>		t(-0.368)		t(10.516)				t(-0.417)		t(-1.589)				
0.17810		**		(0.25)				**		(12.5) *				
0.173		0.00129		0.00581				0.02554		0.00000				
		- 0.01137		EM		- 0.00188		AESC		+ 0.00507				EF
		-5-				-32-				-7-				
t(-0.859)		t(-1.195)		t(0.765)										
(0.50) **		(12.5) *		(50.0) **										
0.01323		0.00157		0.00662										

BIBLIOGRAFIA

- Johnston, J. : Métodos de Econometría
1967- Universidad de Manchester - Inglaterra
- Kmenta, Jan : Elements of Econometrics
1971- EE.UU.
- Lebart, Ludovic; Fenelon, Jean Pierre, :
Estatistique et Informatique Appliquees.
1975 - Francia - Ed. Dunod
- Merril, William; Fox, Karl, : Introducción a la Estadística
Económica
1970 - Universidad de Iowa
- Pindyck, Robert S.; y Rubierfeld Daniel h.,:
Econometric Models An Economic Forecasts.
1976 - EE.UU.
- Wonnacott, Thomas H.; Wonnacott, Ronald J.:
Introductory Statistics for Business and
Economics.
1962 - N. York

.....