### 5.3 Especificación del modelo empírico

Para este análisis se formuló un modelo econométrico de sección cruzada, porqu las observaciones corresponden a las características de las personas encuestada en un mismo período de tiempo. Se supone que las variables independientes está idénticamente distribuidas con media cero y varianza, es decir: $e_i \sim iid(0,\sigma^2)$. Perc como los modelos no reportan de forma exacta las relaciones (Caridad, 1998 la discrepancia o error entre los valores medidos reales de la variable explicada los estimados mediante el modelo, se recogerán en la variable que se denominar perturbación aleatoria $e$. En el modelo se quiere probar que la variable dependient producción está en función de las variables independientes antigüedad, edad, capita humano, apiarios y camioneta. La metodología econométrica parte de la función matemática siguiente $Y = f(X_1, X_2)$ (5.1)

Para su solución se añade el término de error o perturbación que denotaremos co $e_i$ para construir el modelo econométrico que se escribe a continuación:

$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \mu_i$ (5.2)

i= 1,2,3,4,……N

donde :

Y = la variable dependiente o endógena

X = la variable independiente o explicativa

α = la ordenada y la pendiente del modelo.

β = los coeficientes de la regresión.

$\mu_i$ = variable que recoge los errores

i = Subíndice porque la muestra contiene datos de sección cruzada, y

N = el número de observaciones de la muestra.

En esta expresión, la variable $Y_{1,}$ representa la producción, la variable $X_{i1}$, expresa la antigüedad; $X_{i2}$, expresa la edad que tiene cada apicultor; $X_{i3}$, el capital humano $X_{i4,}$ representa los apiarios que instalan; $X_{i5}$, recoge la variable que representa a los apicultores que poseen camioneta. El término $\mu_i$ es una perturbación aleatoria o componente de error. Si se considera que un modelo no recoge todas las variables que influyen sobre $Y_{1,}$ y, además, que hay errores de medición y un imprevisible comportamiento humano, se espera que $\mu_{i,}$ recoja los efectos de las variables omitidas

Para analizar el desarrollo de la actividad apícola se busca la relación de la producción, con la variable de la antigüedad, la edad, el capital humano, apiarios y camionetas. En la medida en que el modelo esté especificado de manera correcta, los coeficientes informan directamente sobre el sentido (positivo o negativo) de la relación

Se estimó un modelo de regresión lineal normal clásico (MRLNC) que según Pena Trapero et al (1999) teóricamente se parte del supuesto de que la variable Y es función de k factores explicativos de su comportamiento:

$$Y = f(x_1, x_2, \dots, x_k) \qquad (5.3)$$

Asimismo se presume que las *n* observaciones, son formadas por un mecanismo lógico que se basa en las siguientes hipótesis:

Hipótesis de linealidad: $Y = X\beta + e$

donde:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}; \quad X = \begin{pmatrix} X_{11} & X_{21} & X_{31} & \cdots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \cdots & X_{kn} \end{pmatrix}; \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}; \qquad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

(n x 1) (n x k) (k x 1) (n x 1)

es decir: $Y_i = \beta_{1i} X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i \quad \forall$ i = 1, 2, ..., n

Si se quiere que en el modelo exista término independiente, la variable $X_1$ tiene que ser igual a uno, o lo que es lo mismo, la primera columna de la matriz X tiene que ser un vector de unos (vector iota, i). A este regresor se le llama regreso ficticio. La esperanza del vector de la variable aleatoria es cero: E($e$) = 0

La matriz de varianzas y covarianzas del vector de variables aleatorias es: *E(ee´)* = $\sigma_e^2 \mathrm{I}$, o sea los componentes del vector "e" tienen idéntica varianza (homoscedasticidad) y además las covarianzas son 0, debido a que los elementos del vector *e* están incorrelacionados.

El rango de la matriz *X* es *k*, el número de regresores, y debe ser menor o igua a *n*, el número de observaciones. Esta condición es necesaria para que la matri *X'X* sea invertible. Además, las variables explicativas no pueden ser linealment dependientes.

La matriz *X* es una matriz aleatoria o no estocástica.

El vector de la variable aleatoria sigue una distribución normal multivariante d parámetros: $e \rightarrow N(0, \sigma_u^2 I)$, es decir, es un vector normal esférico.

Para este caso el modelo propuesto a estimar sería:

$Y = X\beta + e;\ \hat{Y} = Xb$ (5.4)

Donde $b$ es el vector de estimadores de los correspondientes parámetros Siguiendo MCO, se eligen aquellos estimadores que hacen mínima la suma de la diferencias cuadráticas entre los valores observados y los valores estimados de l variable dependiente, es decir, que minimizan la suma de los errores al cuadrado:

Mín $\Sigma (Y_i - \hat{Y}_i)^2$ = mín $\Sigma e_i^2$ (5.5)

Aplicando en el modelo propuesto el método de los mínimos cuadrados se obtiene estimadores lineales insesgados y óptimos (ELIO). Una vez aplicado el método s obtiene el vector de estimadores, $b$, a través de la siguiente expresión:

$b = (X'X)^{-1} X'Y$ (4.7.6)

$$\text{donde:}\quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix};\ X'X = \begin{pmatrix} n & \sum x_{2i} & \cdots & \sum x_{ki} \\ \sum x_{2i} & \sum x_{2i}^2 & \cdots & \sum x_{ki}x_{ki} \\ \vdots & \vdots & & \vdots \\ \sum x_{ki} & \sum x_{ki}x_{2i} & \cdots & \sum x_{ki}^2 \end{pmatrix}\ X'Y = \begin{pmatrix} \sum Y_i \\ \sum Y_i X_{2i} \\ \vdots \\ \sum Y_i X_{ki} \end{pmatrix}$$

(k $x$ 1) (k $x$ k) (k $x$ 1)

por lo que el modelo estimado se expresa:

$\hat{Y} = Xb$

$\hat{Y}_i = b_1 + b_2X_{2i} + b_3X_{3i} + \ldots + b_kX_{ki}$

La suma de cuadrados se presenta a partir de la variación total de Y, que pued
enunciarse como la suma de dos componentes: un componente que explica a l
regresión lineal y otro componente residual que no explica a la regresión lineal. S
sabemos que: Y = $\hat{Y}$ + *e*

Premultiplicando por la transpuesta: Y'Y = $\hat{Y}'\hat{Y}$ + $e'e$

Y expresándolo en forma de desviaciones: *SCT* = *SCE* + *SCR :* es decir, la suma
de los cuadrados totales es igual a la suma de cuadrados explicada por la regresión
más la suma de cuadrados de residuos, donde:

$$SCT = Y'Y - \frac{(i'Y)^2}{n} = \sum Y_i^2 - n\overline{Y}^2$$

$$SCE = \hat{Y}'\hat{Y} - \frac{(i'Y)^2}{n} = b'X'Y' - n\overline{Y}^2$$

$$SCR = e'e = SCT - SCE$$

Donde *i* es el vector cuyos elementos son todos iguales a uno.

Los estimadores mínimos cuadrados, son insesgados porque la esperanza de
estimador coincide con el parámetro a estimar: *E(b)* = β; 2. La matriz de varianzas
covarianzas es:

$$\Sigma_{bb} = \sigma_u^2 (X'X)^{-1} = \begin{pmatrix} \sigma_{b1}^2 & \sigma_{b1b2} & & \sigma_{b1bk} \\ \sigma_{b2b1} & \sigma_{b2}^2 & \cdots & \sigma_{b2bk} \\ \vdots & \vdots & & \vdots \\ \sigma_{bkb1} & \sigma_{bkb2} & \cdots & \sigma_{bk}^2 \end{pmatrix}$$

Si se define el estimador de la varianza de la variable aleatoria como:

$$S_u^2 = \frac{e'e}{n-k} = \frac{SCR}{N-K}$$

Se dice que el estimador mínimo cuadrático es insesgado porque su esperanza coincide con el parámetro a estimar: $E(S_u^2) = \sigma_u^2$

puesto que: *e* = Y— Xb = Y — X(X'X)-1X'Y =[I- X(X'X)-1X' ]Y= MY

donde la matriz M es una matriz simétrica e idempotente, $e'e = \Sigma e_i^2 = u'Mu$

y, si se calcula la esperanza, *E(u'Mu)* = $\boldsymbol{E(u'Mu) = \sigma_u^2 tr(M) = \sigma_u^2 (n-k) = \sigma_u^2}$

se obtiene $E(S_u^2) = E(\frac{e'e}{n-k}) = \frac{1}{n-k}\sigma_u^2(n-k) = \sigma_u^2$ (5.6)

Sabiendo esto, la estimación insesgada de la matriz de varianzas y covarianzas de los estimadores sería:

$$\boldsymbol{S_{bb} = S_u^2 (X'X)^{-1}} = \begin{pmatrix} S_{b1}^2 & \boldsymbol{S_{b1b2}} & & S_{b1bk} \\ S_{b2b1} & \boldsymbol{S_{b2}^2} & \cdots & S_{b2bk} \\ \vdots & \vdots & & \vdots \\ S_{bkb1} & S_{bkb2} & \cdots & S_{bk}^2 \end{pmatrix}$$