# ALTERNATIVAS AL ESTIMADOR DE REGRESIÓN EN POBLACIONES FINITAS. APLICACIÓN A UN COLECTIVO DE EMPRESAS


**Santiago Murgui Izquierdo**

Departamento de Economía Aplicada

Universidad de Valencia

e-mail: Santiago.Murgui@uv.es

**Mª Cruz Molés Machí**

Departamento de Economía Aplicada

Universidad de Valencia

e-mail: Cruz.Moles@uv.es

**Mª Consuelo Colom Andrés**

Departamento de Economía Aplicada

Universidad de Valencia

e-mail: Consuelo.Colom@uv.es



## Resumen

La estimación de agregados en colectivos finitos con la metodología propia de un modelo de superpoblación, en general, ofrece algunos aspectos interesantes. Uno de ellos es la posibilidad de proponer estimadores diferentes a los que habitualmente se plantean en el contexto clásico de los diseños aleatorios. Otro aspecto de interés es la recomendación de seleccionar la muestra por un procedimiento intencionado, siempre con el objetivo de aumentar la precisión de los resultados y disminuir el error de muestreo. Estas cuestiones son analizadas en el caso particular de un colectivo de empresas de la comunidad valenciana, recurriendo a técnicas específicas de replicación para resolver alguno de los problemas que se plantean. Las propuestas y consideraciones finales son evaluadas empíricamente mediante la aplicación retroactiva de la metodología a datos censales del pasado.



*Palabras clave*: Modelo de superpoblación, población finita.

*Area temática*: Métodos cuantitativos.

## 1. Introducción.

Con cierta frecuencia se plantea la necesidad de efectuar estimaciones en poblaciones finitas, partiendo de muestras relativamente pequeñas. Las limitaciones al tamaño muestral suelen ser debidas, la mayor parte de las veces, a restricciones presupuestarias y a características específicas del universo que en la práctica impiden desarrollar el trabajo de campo en el tiempo deseado. En estos casos, si la única estructura aleatoria que se considera para realizar la inferencia es la que se deriva de la adopción de un diseño muestral aleatorio, es normal que la precisión de los resultados sea relativamente pequeña.

Cuando se dispone de información auxiliar de partida, una manera de disminuir los errores de estimación sin incrementar el tamaño de la muestra, consiste en incorporar esta información al proceso predictivo, proponiendo para ello la adopción de un modelo de superpoblación.

Las perspectivas que se abren cuando la estructura aleatoria que soporta la inferencia emanan de un modelo, se orientan en varias direcciones. Por un lado, surgen criterios de construcción de estimadores que conducen a soluciones, en ocasiones poco intuitivas, pero con mayor atractivo que los estimadores habitualmente propuestos en la literatura clásica del muestreo en poblaciones finitas. Por otro lado, la no necesidad de recurrir a diseños aleatorios, plantea el interés en buscar diseños intencionados que permitan intervenir sobre la capacidad predictiva de las estrategias, aunque manteniendo su carácter no informativo o ignorable en el sentido que definen Sugden y Smith (1984).

Murgui, Colom y Molés (2005) proponen un modelo de superpoblación que permite formalizar una estrategia de muestreo, apta para estimar agregados poblacionales en el universo particular de las cooperativas agrarias valencianas. El citado modelo y la correspondiente estrategia se manifiestan adecuados para la estimación de algunas variables de las que recoge el balance y las cuentas económicas de las empresas, sin embargo, conduce a resultados poco precisos en otros casos, como por ejemplo al estimar los agregados poblacionales del inmovilizado y de los fondos propios.

Este trabajo tiene como objetivo determinar una estrategia de muestreo para la estimación de las últimas variables señaladas. Para ello se propone un proceso basado en la integración de dos modelos, uno para la variable activo y otro para las respectivas variables inmovilizado y fondos propios. Los estimadores se eligen básicamente bajo el criterio de minimizar el error cuadrático medio y se propone un diseño intencionado que en términos generales sugiere elegir las cooperativas más grandes.

La predicción de los agregados poblacionales de interés se efectúa en dos pasos, en el primero se estiman separadamente el activo total y el ratio global de cada variable con respecto al activo. En el segundo paso, se construye un estimador compuesto para cada uno de los agregados buscados, siendo necesario recurrir a métodos aproximativos para evaluar sus propiedades.

Finalmente se analiza la validez empírica de la estrategia propuesta, aplicando retroactivamente la metodología a los datos censales disponibles para los ejercicios comprendidos entre 1996 y 2000.

**2. Modelo.**

Considérese el universo $U = \left(u_1, u_2, ..., u_N\right)$ integrado por las cooperativas agrarias censadas en la Comunidad Valenciana. Por $A_{it}$, $I_{it}$ y $F_{it}$ respectivamente se denota al valor declarado por la cooperativa i en el ejercicio t para las variables activo, inmovilizado y fondos propios. La experiencia adquirida en oficinas administrativas como el Servicio de Cooperativas Agrarias de la Conselleria d'Agricultura (2002) permite concluir que, transcurrido un corto espacio de tiempo desde el cierre de un ejercicio t, únicamente puede disponerse de las cuentas presentadas por una muestra s de n cooperativas, resultando prácticamente imposible acceder al censo completo. En este contexto, se pretende estimar a partir de los datos disponibles los agregados poblacionales $A_t = \sum_{i \in U} A_{it}$, $I_t = \sum_{i \in U} I_{it}$ y $F_t = \sum_{i \in U} F_{it}$.

Teniendo en cuenta que el tamaño del universo de referencia para cada ejercicio se sitúa entorno a 500 entidades y la necesidad de acceder cuanto antes a las referidas estimaciones, no resulta apropiado intentar la observación de las variables de interés en un número elevado de entidades. Consecuentemente, una inferencia basada en la selección aleatoria de la muestra adolecería de niveles de error excesivamente grandes para los resultados.

Por otra parte, la experiencia también indica que pasado cierto tiempo desde el cierre de cada ejercicio, todas las cooperativas acaban proporcionando sus cuentas económicas, de manera que es muy posible que en el momento de efectuar la estimación de los agregados para el ejercicio t, podrá disponerse del censo completo de valores $A_{i(t-1)}$, $I_{i(t-1)}$ y $F_{i(t-1)}$, para $i = 1,2,....,N$ .

La forma habitual con que las empresas elaboran sus presupuestos, efectuando una revisión de los del ejercicio precedente, sugiere la existencia de cierta vinculación entre los valores observados para una misma variable entre dos ejercicios consecutivos. En base a ello, se procede a continuación a formular una modelización que recoja las relaciones interanuales de las distintas variables.

Murgui, Colom y Molés (2005) proponen para el activo total el modelo que definen las siguientes hipótesis:

$$H_1 : E[A_{it}] = \varphi_t A_{i(t-1)} \quad \text{para } i = 1,2,....,N$$

$$H_2 : V[A_{it}] = \eta_t^2 \tau_i \quad \text{para } i = 1,2,....,N$$

$$H_3 : C[A_{it}, A_{jt}] = 0 \quad \text{para } i \neq j$$

Por otra parte, en la práctica, es más frecuente elaborar los presupuestos anuales manejando ratios entre variables que sus propios valores absolutos, lo que sugiere proponer una modelización basada principalmente en los primeros. En particular, para la variable inmovilizado se van a considerar ratios sobre el activo correspondiente a cada ejercicio $R_{it} = \dfrac{I_{it}}{A_{it}}$ y $R_{i(t-1)} = \dfrac{I_{i(t-1)}}{A_{i(t-1)}}$ postulándose las siguientes hipótesis:

$$H_4 : E\left[R_{it}\right] = \alpha_t + \beta_t R_{i(t-1)} \text{ para } i = 1,2,....,N$$

$$H_5 : V\left[R_{it}\right] = \sigma_t^2 \ \nu_i \text{ para } i = 1,2,....,N$$

$$H_6 : C\left[R_{it}, R_{jt}\right] = 0 \text{ para } i \neq j$$

En ambos bloques de hipótesis, el del activo y el del inmovilizado, se consideran varias posibilidades para las varianzas, tanto la homoscedasticidad asociada con $\tau_i = 1$ y $\nu_i = 1$ , como las alternativas $\tau_i = A_{i(t-1)}$, $\tau_i = A^2_{i(t-1)}$ , $\nu_i = R_{i(t-1)}$ y $\nu_i = R^2_{i(t-1)}$.

Las hipótesis $H_4$, $H_5$ y $H_6$ inicialmente definidas para el ratio del inmovilizado sobre el activo, se consideran igualmente válidas para el ratio de la variable fondos propios sobre el activo. De esta forma, la estrategia que se establezca para estimar el inmovilizado agregado será extensible a la estimación de los fondos globales.

**3- Elección de la estrategia de muestreo.**

Numerosos autores han estudiado la mejor estrategia muestral a asociar con un modelo como el que definen las hipótesis $H_1$, $H_2$ y $H_3$, cabe citar a Royall y Herson (1973) entre los primeros y más destacados. Posteriormente se han efectuado generalizaciones de este modelo particular introduciendo hipótesis que incorporan más de una variable auxiliar, entre otras la desarrollada por Valliant et al (2000).

El estimador lineal insesgado con mínimo error cuadrático para el activo global agregado $A_t$, es el que determina la expresión $\hat{A}_t = \sum_s A_{it} + \dfrac{\sum_s \frac{A_{i(t-1)} A_{it}}{\tau_i}}{\sum_s \frac{A^2_{i(t-1)}}{\tau_i}} \sum_{s'} A_{i(t-1)}$

y un estimador insesgado para su error cuadrático medio es $E\left[\left(\hat{A}_t - A_t\right)^2\right] = \eta_t^2 \left( \sum_{s'} \tau_i + \dfrac{1}{\sum_s \frac{A^2_{i(t-1)}}{\tau_i}} \left( \sum_{s'} A_{i(t-1)} \right)^2 \right)$, donde $\eta_t^2$ debe ser estimado a partir de los datos, siendo $\hat{\eta}_t^2 = \frac{1}{n-1}\sum_s \frac{\left(A_{it} - \hat{\varphi}_t A_{i(t-1)}\right)^2}{A_{i(t-1)}}$ un estimador insesgado del mismo.

Con respecto al inmovilizado y como paso previo a la estimación de su agregado global, va a determinarse en primer lugar una estrategia para la estimación del ratio $R_t = \frac{I_t}{A_t}$. Para ello se parte de la descomposición $R_t = \sum_{i \in U} R_{it}\gamma_{it}$ , siendo $\gamma_{it} = \frac{A_{it}}{A_t}$ una medida del peso relativo de cada entidad en el universo global del ejercicio t determinada a través del activo. En la práctica, el tamaño relativo de las cooperativas agrarias varía poco entre dos ejercicios consecutivos, por lo que los pesos de cada ejercicio $\gamma_{it}$ pueden aproximarse por los del ejercicio anterior $\gamma_{i(t-1)}$ y en adelante se denotarán por $\gamma_i$ .

A partir del modelo que definen las hipótesis $H_4$, $H_5$ y $H_6$, pueden considerarse distintos criterios de construcción de un estimador para el ratio global $R_t$ .

Criterio 1

Estimador lineal $\sum_{i \in s} R_{it} c_i$ , insesgado y con mínimo error cuadrático medio.

Criterio 2

Estimador determinado por la expresión, $\sum_{i \in s} R_{it}\gamma_i + \sum_{i \in s'} \left(\hat{\alpha}_t + \hat{\beta}_t R_{i(t-1)}\right)\gamma_i$ , siendo $s' = U - s$, y $\hat{\alpha}_t$ y $\hat{\beta}_t$ estimadores lineales insesgados de mínima varianza para los parámetros $\alpha_t$ y $\beta_t$ .

Criterio 3

Estimador determinado por la expresión $\sum_{i \in s} R_{it}\gamma_i + \sum_{i \in s'} \left(\hat{\alpha}_t + \hat{\beta}_t R_{i(t-1)}\right)\gamma_i$, siendo $\hat{\alpha}_t$ y $\hat{\beta}_t$ los valores que minimizan la suma de los residuos cuadráticos ponderados por el factor $\frac{1}{\nu_i}$, definida como $\sum_{i \in s} \left(R_{it} - \alpha_t - \beta_t R_{i(t-1)}\right)^2 \frac{1}{\nu_i}$.

Criterio 4

Estimador que determina la expresión $\sum_{i \in s} R_{it}\gamma_i + \sum_{i \in s'} \left(\hat{\alpha}_t + \hat{\beta}_t R_{i(t-1)}\right)\gamma_i$, siendo $\hat{\alpha}_t$ y $\hat{\beta}_t$ los valores que minimizan la suma de los residuos cuadráticos ponderados por el peso relativo $\gamma_i$, definida como $\sum_{i \in s} \left(R_{it} - \alpha_t - \beta_t R_{i(t-1)}\right)^2 \gamma_i$ .

Otros criterios

Igualmente podrían proponerse criterios similares al 3 y 4 modificando la ponderación de los residuos cuadráticos. Por ejemplo, suprimiendo la ponderación, o bien, introduciendo los pesos $\frac{\gamma_i}{\nu_i}$.

De entre los criterios propuestos para la construcción de un estimador de $R_t$, el que ofrece mayor atractivo es el primero, ya que asegura una precisión máxima para las estimaciones. No obstante, se comprueba que los tres primeros conducen al mismo estimador determinado por la expresión:

$$\hat{R}_t = \ddot{R}_t + \left(\sum_{i \in s} \gamma_i\right)\left(\overline{R}_t - \ddot{R}_t\right) + \frac{\ddot{C}}{\ddot{V}}\left(R_{t-1} - \ddot{R}_{t-1} - \left(\sum_{i \in s} \gamma_i\right)\left(\overline{R}_{t-1} - \ddot{R}_{t-1}\right)\right) \qquad (1)$$

donde $\overline{R}_t = \frac{1}{\left(\sum_{i \in s} \gamma_i\right)} \sum_{i \in s} R_{it}\gamma_i$ es la media muestral ponderada con los pesos relativos de las cooperativas para el ejercicio t; $\overline{R}_{t-1}$ la misma media pero referida al ejercicio t-1; $\ddot{R}_t = \dfrac{1}{\left(\sum_{i\in s} \dfrac{1}{\nu_i}\right)} \sum_{i\in s} R_{it} \dfrac{1}{\nu_i}$ la media muestral ponderada con la inversa del factor conocido de la varianza para el ejercicio t; $\ddot{R}_{t-1}$ la misma media referida al ejercicio t-1; $\ddot{V} = \dfrac{1}{\left(\sum_{i\in s} \dfrac{1}{\nu_i}\right)} \sum_{i\in s} \left(R_{i(t-1)} - \ddot{R}_{t-1}\right)^2 \dfrac{1}{\nu_i}$ la varianza muestral ponderada con la inversa del factor conocido de la varianza para el ejercicio t-1; y finalmente $\ddot{C} = \dfrac{1}{\left(\sum_{i\in s} \dfrac{1}{\nu_i}\right)} \sum_{i\in s} \left(R_{it} - \ddot{R}_t\right)\left(R_{i(t-1)} - \ddot{R}_{t-1}\right) \dfrac{1}{\nu_i}$ la covarianza ponderada entre las observaciones muestrales del ejercicio t y t-1.

El criterio 4 conduce a la expresión:

$$\tilde{R}_t = \overline{R}_t + \frac{C}{V}\left(R_{t-1} - \overline{R}_{t-1}\right) \tag{2}$$

en la que $C$ y $V$ denotan respectivamente covarianza y varianza sobre las observaciones muestrales ponderadas con el peso relativo de las entidades.

El estimador (2) coincide con el habitualmente denominado estimador de regresión, mientras que el (1), aunque mantiene una estructura similar al de regresión, presenta algunas singularidades. La primera es que está basado en estadísticos muestrales ponderados con dos esquemas diferentes, por un lado los pesos relativos de las entidades y por otro la inversa de los factores conocidos de las varianzas. La segunda es la corrección que introduce en las respectivas medias muestrales, basada en la diferencia observada entre las medias ponderadas con ambos pesos.

Por construcción, el estimador (1) tendrá asociado un menor error cuadrático medio que el (2) y por tanto, una mayor capacidad predictiva. Esto permite concluir que la adopción de un modelo facilita la obtención de buenos estimadores, sin la restricción que supondría ajustarse a una estructura algebraica más o menos intuitiva.

El error cuadrático medio del estimador (1) está determinado por la expresión:

$$E\left[\left(\hat{R}_t - R_t\right)^2\right] = \sigma_t^2 \left( \sum_{i \in s'} \gamma_i^2 \nu_i + \frac{\left(\sum_{i \in s'} \gamma_i\right)^2}{\left(\sum_{i \in s} \frac{1}{\nu_i}\right)} \left( 1 + \frac{\left(\ddot{R}_{t-1} - \overline{R'_{t-1}}\right)^2}{\ddot{V}} \right) \right) \qquad (3)$$

donde $\overline{R'_{t-1}}$ denota la media de las observaciones no seleccionadas en la muestra en el ejercicio t-1, ponderadas con los pesos relativos de las entidades.

Se demuestra que un estimador insesgado para el parámetro $\sigma^2$ es el que define $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i \in s} \left( R_{it} - \hat{\alpha}_t - \hat{\beta}_t R_{i(t-1)} \right)^2 \frac{1}{\nu_i}$, donde $\hat{\beta}_t = \frac{\ddot{C}}{\ddot{V}}$ y $\hat{\alpha}_t = \ddot{R}_t - \hat{\beta}_t \ddot{R}_{t-1}$. A partir de él, para los valores habituales del factor $\nu_i$ es inmediato obtener un estimador insesgado para el error cuadrático medio de $\hat{R}_t$.

A partir de los estimadores $\hat{A}_t$ y $\hat{R}_t$, teniendo en cuenta la relación $I_t = A_t R_t$, puede proponerse el siguiente estimador para el inmovilizado agregado $\hat{I}_t = \hat{A}_t \hat{R}_t$.

Ante la imposibilidad de realizar cálculos exactos, para analizar las propiedades de este último estimador se ha procedido a aproximar su expresión efectuando el correspondiente desarrollo de Taylor:

$$\hat{I}_t \approx A_t R_t + R_t \left( \hat{A}_t - A_t \right) + A_t \left( \hat{R}_t - R_t \right) \qquad (4)$$

A partir de (4) es fácil comprobar la insesgadez aproximada del estimador $\hat{I}_t$, así como determinar una expresión aproximada para su error cuadrático medio:

$$\hat{E}\left[\left(\hat{I}_t - I_t\right)^2\right] = \hat{E}\left[\left(\hat{R}_t - R_t\right)^2\right] \left( \sum_{i \in U} A_{i(t-1)} \right) \left( \hat{\eta}_t^2 + \hat{\varphi}_t^2 \left( \sum_{i \in U} A_{i(t-1)} \right) \right) +$$

$$+\hat{E}\left[\left(\hat{A}_t - A_t\right)^2\right]\left(\hat{\sigma}_t^2\left(\sum_{i\in U}\gamma_i^2\right) + \left(\hat{\alpha}_t + \hat{\beta}_t R_{t-1}\right)^2\right) \qquad (5)$$

En el siguiente epígrafe se efectúa una valoración empírica de la bondad de esta aproximación utilizando técnicas de simulación para estimar el error cuadrático medio.

Una vez obtenido un estimador de la característica poblacional de interés y una aproximación a su error asociado, procede determinar el diseño o forma de seleccionar las unidades muestrales. Al resolver la estimación de $A_t$ bajo el modelo que definen las hipótesis $H_1$, $H_2$ y $H_3$, es fácil comprobar que el error cuadrático del estimador es mínimo si se selecciona una muestra intencionada, integrada por las entidades que en el ejercicio t-1 presentaron un mayor valor de la variable de referencia. Murgui, Colom y Molés (2005) ofrecen resultados empíricos sobre esta estrategia de estimación del activo global agregado.

Cuando el objetivo es estimar el ratio $R_t$, se observa que el error de estimación determinado por (3) depende fundamentalmente de tres factores: la suma de los pesos de las entidades no seleccionadas en la muestra, la varianza muestral de los datos correspondientes al ejercicio t-1 y la diferencia entre las medias muestrales ponderadas con los pesos ya señalados. El último factor, inicialmente no parece que pueda adoptar un valor elevado, además su influencia se verá reducida en la medida en que la varianza muestral adopte valores relativamente grandes. Consecuentemente, parece recomendable elegir entidades con un peso relativo elevado para la variable activo y a ser posible que determinen una varianza muestral elevada para el ratio en el ejercicio t-1.

Algunos ensayos numéricos realizados con la base de datos disponible, sugieren que la última condición no es especialmente relevante. De hecho, se ha comprobado que si para aumentar el valor de la varianza muestral debe sustituirse alguna entidad por otra que posea un menor peso relativo, el resultado es que acaba produciéndose un incremento del error del estimador.

La argumentación anterior permite concluir que, tanto para estimar el activo $A_t$ como el ratio $R_t$, el diseño muestral más adecuado es aquel que selecciona las entidades con mayor peso relativo con respecto a la variable activo. En consecuencia, si se fija como objetivo final la estimación del inmovilizado agregado $I_t$, a la vista de la aproximación (5) procede concluir que este mismo diseño debe considerarse como el más adecuado, puesto que en general conducirá a valores mínimos del error cuadrático medio.

En la medida en que las hipótesis $H_4$, $H_5$ y $H_6$ se consideran adecuadas, tanto para la variable inmovilizado como para la variable fondos propios, la estrategia óptima determinada para la estimación del inmovilizado agregado $I_t$, debe considerarse igualmente apropiada para la estimación del agregado de los fondos propios $F_t$.

**4- Valoración empírica.**

Para valorar la eficacia de la estrategia muestral propuesta, se ha procedido a su aplicación en una base de datos con información sobre las cooperativas agrarias valencianas, para los ejercicios comprendidos entre 1996 y 1999. Puesto que el modelo propuesto utiliza los datos censales de un año para obtener predicciones para el año siguiente, al efectuar una aplicación retroactiva sobre los datos disponibles ha podido ensayarse la estrategia propuesta en cuatro bienios, obteniéndose predicciones para los ejercicios 1997, 1998, 1999 y 2000.

Al margen de algunas depuraciones mínimas realizadas sobre los datos originales basadas en criterios de coherencia contable, la metodología exige la estabilidad del universo de referencia para cada bienio. En realidad, esta no es una condición muy restrictiva, puesto que la creación/desaparición de cooperativas tiene una incidencia realmente escasa en periodos tan cortos de tiempo. En todos los casos, la aplicación se ha efectuado sobre las entidades comunes a los dos años, detalle que deberá tenerse en cuenta cuando se pretenda interpretar los resultados en un contexto económico.

Para efectuar la predicción del año 2000 se ha seleccionado una muestra integrada por las 50 entidades que en el ejercicio 1999 tuvieron un mayor peso relativo de la variable activo. Este procedimiento de selección muestral debería haber sido seguido también en el resto de los procesos predictivos desarrollados, sin embargo, habida cuenta de los escasos cambios encontrados en los pesos relativos de las entidades en esos años, se ha optado por mantener fija la muestra para los cuatro ejercicios. En la práctica, este diseño tiene gran interés para la obtención de los datos muestrales, puesto que permite seleccionar un grupo de entidades colaboradoras y mantenerlo fijo durante varios años. De hecho, puesto que en el censo de los ejercicios 1997 y 1998 había desaparecido alguna de las cooperativas elegidas, la predicción para 1998 y 1999 se ha efectuado a partir de una muestra de 49 entidades y la de 1997 a partir de 48.

En lo que se refiere al tamaño muestral, de acuerdo con los resultados empíricos obtenidos por Murgui, Colom y Molés (2005) para la estimación del activo agregado, podría considerarse suficiente una muestra integrada por 20 entidades. No obstante, los errores asociados con la predicción de ratios en ensayos realizados sobre una muestra de este tamaño, se consideraron excesivamente altos. Posiblemente, en general, la precisión de los resultados podría ser aceptable con tamaños muestrales inferiores a 50. En cualquier caso, esta cifra parece asumible desde un punto vista práctico y siempre cabe la posibilidad de realizar nuevos ensayos que permitan ajustar más el tamaño final de la muestra.

En el cuadro 1 se recogen los principales resultados obtenidos al estimar los agregados de las 3 variables y los 2 ratios que se han analizado en este trabajo: el activo total, el inmovilizado agregado, los fondos propios agregados, el ratio del inmovilizado sobre el activo y el ratio de los fondos propios sobre el activo. Los errores teóricos relativos se han calculado asumiendo la aproximación Normal para el estimador y una confianza del 95%, por lo que están calculados mediante la expresión $\dfrac{1{,}96\sqrt{\hat{E}\left[\left(\hat{X}_t - X_t\right)^2\right]}}{\hat{X}_t}$. Además, al disponer de los correspondientes datos censales, ha sido posible por otra parte calcular los errores reales determinados mediante el cociente $\frac{\hat{X}_t - X_t}{\hat{X}_t}$.

En un primer ensayo, el modelo considerado es el correspondiente a los valores particulares $\tau_i = A_{i(t-1)}$ y $\nu_i = 1$. En este caso, el estimador para el activo se reduce al conocido estimador de razón $\hat{A}_t = \frac{\sum_s A_{it}}{\sum_s A_{i(t-1)}} \sum_U A_{i(t-1)}$ y el de los ratios por $\hat{R}_t = \dot{R}_t + \left(\sum_{i \in s} \gamma_i\right)\left(\overline{R}_t - \dot{R}_t\right) + \frac{\dot{C}}{\dot{V}}\left(R_{t-1} - \dot{R}_{t-1} - \left(\sum_{i \in s} \gamma_i\right)\left(\overline{R}_{r-1} - \dot{R}_{t-1}\right)\right)$, donde $\dot{R}_t$, $\dot{V}$ y $\dot{C}$ denotan estadísticos muestrales sin ponderar. Igualmente se han particularizado las expresiones de los errores cuadráticos medios correspondientes a los valores indicados del factor conocido de las varianzas.

**Cuadro 1**
**Estimaciones y errores relativos (%)**

| | Activo | Ratio Inmovilizado | Ratio fondo | Inmovilizado | Fondo |
|---|---|---|---|---|---|
| **1999-2000** | | | | | |
| Estimación | 151905342 | 0,25 | 0,30 | 37251391 | 45784497 |
| Error relativo real | 0,71 | 2,01 | 2,34 | 3,83 | 2,84 |
| Error relativo teórico | 1,39 | 1,94 | 0,91 | 2,39 | 1,66 |
| **1998-1999** | | | | | |
| Estimación | 190843557 | 0,26 | 0,31 | 48982334 | 59540804 |
| Error relativo real | 1,00 | 6,14 | -2,15 | 8,24 | -1,18 |
| Error relativo teórico | 3,80 | 4,27 | 4,93 | 5,67 | 6,15 |
| **1997-1998** | | | | | |
| Estimación | 186971317 | 0,25 | 0,33 | 47071594 | 60784504 |
| Error relativo real | 2,63 | -0,46 | -0,27 | 0,98 | -9,13 |
| Error relativo teórico | 6,67 | 5,68 | 8,48 | 5,74 | 9,52 |
| **1996-1997** | | | | | |
| Estimación | 199321770 | 0,21 | 0,23 | 42263692 | 45977881 |
| Error relativo real | 0,86 | 0,07 | -8,02 | -1,79 | 6,59 |
| Error relativo teórico | 4,87 | 6,54 | 6,61 | 8,11 | 6,62 |

En general, los errores de estimación encontrados adoptan valores aceptables, únicamente en tres casos se incrementan hasta alcanzar cifras próximas al 8% y 9%. Inicialmente esto sugiere la posible existencia de valores anómalos entre los datos correspondientes, por lo que en otro momento se procederá a realizar el análisis oportuno.

Los errores relativos teóricos que aparecen en las dos últimas filas de cada bienio, permiten apreciar cómo intervienen los errores separados de dos estimadores, en la determinación del error final de su producto. En general, este último siempre es superior al máximo de los errores por separado, aunque la relación no es simple, tal como refleja la aproximación (5). Precisamente el carácter aproximado del error (5), obtenido despreciando términos en el desarrollo de Taylor, ha inducido a considerar relevante efectuar un análisis empírico de la bondad de tal aproximación. El mecanismo de evaluación ensayado se basa en la replicación de todo el sistema asociado con la predicción de un ejercicio y consta de los siguientes pasos:

1º- Manteniendo fija la población auxiliar asociada con cada variable en el ejercicio t, determinada por los datos censales de la misma variable en t-1, se recurre al correspondiente modelo con hipótesis de Normalidad para la perturbación, para generar 500 poblaciones replicadas para el ejercicio t. El proceso se repite con cada una de las tres variables analizadas: el activo, el inmovilizado y los fondos propios.

2º- Manteniendo fija la muestra de 50 entidades con valores conocidos para el ejercicio t-1, en cada una de las poblaciones replicadas se determinan los valores correspondientes al ejercicio t, quedando especificada de este modo la muestra intencionada asociada a cada una de las replicaciones.

3º- Para cada una de las poblaciones replicadas se calculan los agregados de las variables inmovilizado y fondos propios. Asimismo, a partir de cada muestra, se obtiene la correspondiente estimación del agregado poblacional de cada una de las variables indicadas.

4º- Finalmente, para cada variable, se obtiene una aproximación empírica del error de la estimación calculando el promedio de las desviaciones cuadráticas entre los agregados estimados y los correspondientes valores obtenidos sobre las poblaciones generadas.

En el Cuadro 2 se recogen las aproximaciones al error de estimación obtenidas mediante el desarrollo (5) y mediante las replicaciones. La comparación se establece para las dos variables de interés: el inmovilizado y los fondos propios. Puede observarse que la aproximación es casi perfecta para fondo y muy aceptable para el inmovilizado. Dado el elevado esfuerzo de manipulación informática requerido, la comparación se ha limitado al bienio 99/00.

**Cuadro 2**
**Aproximaciones al error de estimación**

| | Inmovilizado | Fondo |
|---|---|---|
| **error relativo aprox. teórico** | 2,39 | 1,66 |
| **error relativo aprox. empírico** | 2,50 | 1,66 |

## 5- Conclusiones.

Tanto desde el punto de vista teórico, como práctico, una vez más la metodología de los modelos de superpoblación en poblaciones finitas se muestra eficaz para resolver algunos problemas de estimación a los que los diseños aleatorios no proporcionan una solución satisfactoria.

La elección del estimador óptimo es el resultado de aplicar un proceso de optimización planteado en el dominio que genera el modelo propuesto. Esto permite abordar la elección del estimador de una manera racional, conducente en ocasiones a expresiones que difícilmente podría haber sugerido la intuición científica.

La determinación de un diseño intencionado que en cierta medida permita controlar el error de estimación bajo un modelo, está perfectamente justificada teórica y empíricamente. Ni siquiera es necesario apelar a la solución óptima, basta con aproximarse a ella y procurar ventajas prácticas debidamente justificadas.

Finalmente, destacar la singularidad que presentan los procedimientos de replicación en poblaciones finitas. En lugar de generar submuestras aleatorias, procede replicar tanto las poblaciones como las muestras, aunque manteniendo siempre el carácter intencionado de las mismas. En este caso, la simulación ha permitido justificar aproximaciones algebraicas generadas por procedimientos relativamente simples.

**Bibliografía.**


1. Conselleria d’Agricultura, Peixca i Alimentació (2002): “Cuentas Económicas Agregadas del Cooperativismo Agrario”, Ed. Generalitat Valenciana.

2. Murgui, S.; Colom, M.C.; Molés, M.C. (2005): “Diseño y Evaluación Empírica de una Estrategia de Predicción por Muestreo en Cooperativas Agrarias”, *Estadística Española*, **159**.

3. Royall, R.M. and Herson, J. (1973): “Robust Estimation in Finite Populations I”, *Journal of American Statistical Association*, **68**, 880-889.

4. Sugden, R.A., and Smith, T.M.F. (1984): “Ignorable and Informative Designs in Survey Sampling Inference”, *Biometrika*, **71**,495-506.

5. Valliant, R.; Dorfman, A.H. and Royall, R.M. (2000): “Finite Population Sampling and Inference. A Prediction Approach”, New York: Wiley.