

---

# Una revisión de los estimadores óptimos de calibración usando métodos de cuasi-verosimilitud en muestreo bifásico

## A Revision about Optimal Calibration Estimators Using Quasi-Likelihood Methods under Two-Phase Sampling

Hugo Andrés Gutiérrez Rojas<sup>a</sup>  
hugogutierrez@usantotomas.edu.co

---

### Resumen

Los estimadores óptimos de calibración utilizan información auxiliar completa para producir estimaciones más eficientes. Cuando no se dispone de este recurso, una alternativa es realizar un muestreo en dos fases para recopilar la información auxiliar en una primera fase y después utilizarla en el diseño o estimación de la segunda fase. Se compara la eficiencia de los estimadores óptimos de calibración, cuando la relación entre la variable de estudio y las variables de información auxiliar es lineal y log-lineal. En este último caso se utilizan métodos de cuasi-verosimilitud para la estimación de los parámetros del modelo de super-población.

**Palabras clave:** estimadores de calibración, información auxiliar, muestreo bifásico.

### Abstract

Optimal calibration estimators use auxiliary information to produce more efficient estimates. When the auxiliary information is not available, we proceed to design a two-phase sampling where the auxiliary information is gather in the first phase and it is used in the design or estimation of the second phase. We compare the efficiency of the optimal calibration estimators when the relationship between the variable of interest and the auxiliary information is both linear and log-linear. In this last escenario, we make use of the quasi-likelihood method in order to estimate the super-population parameters.

**Key words:** auxiliary information, calibration estimators, two-phase sampling.

---

<sup>a</sup>Director. Centro de Investigaciones y Estudios Estadísticos (CIEES).

## 1. Introducción

Para obtener estimaciones eficientes es deseable conocer información de variables auxiliares, que estén altamente correlacionadas con la variable de interés. Por ejemplo, en la construcción del estimador clásico de calibración propuesto por Deville & Särndal (1992), es necesario conocer el total poblacional de las variables auxiliares. Sin embargo, no en todos los casos es posible tener acceso a información auxiliar en toda la población. Bajo el anterior escenario, es necesario recurrir al uso del muestreo en dos fases<sup>1</sup>. En la primera fase de muestreo se selecciona una muestra grande y se miden las variables auxiliares; con esto se obtiene una buena estimación del total de cada variable auxiliar en la población. En la segunda fase se toma una muestra más pequeña y se mide la variable de interés. De esta manera, y como lo afirma Wu & Luan (2003), la mayor ventaja de utilizar el muestreo en dos fases es la ganancia de precisión sin un incremento significativo en el coste.

Recientemente se han propuesto estimadores de calibración que no sólo son más eficientes que los estimadores clásicos sino que también son óptimos en su clase (Wu 2003). Sin embargo, esta optimalidad está supeditada al conocimiento del valor de las variables de información auxiliares para toda la población. El muestreo en dos fases provee, entonces, una solución ideal para el uso de los estimadores óptimos de calibración bajo tales circunstancias. No obstante, antes de abordar estos temas, es necesario explicar a grandes rasgos en qué consiste la técnica de calibración.

### Estimadores de calibración

Suponga el universo  $U = \{1, \dots, N\}$  el conjunto de elementos en una población finita y  $s$  el conjunto de los elementos que conforman una muestra seleccionada aleatoriamente mediante diseño de muestreo  $p(\cdot)$ . Sea  $y_k$ ,  $k \in s$ , el valor que toma la variable de interés en el individuo  $k$ -ésimo y  $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{Pk})'$ , el valor del vector de información auxiliar, compuesto por  $P$  variables auxiliares, para este mismo individuo.  $\pi_k$  se define como la probabilidad de inclusión de primer orden inducida por el diseño de muestreo  $p(s)$  y se asume que los totales poblacionales

$$\mathbf{t}_x = \sum_{k \in U} \mathbf{x}_k$$

son conocidos. El estimador de calibración para el total poblacional de  $y$ , definido como  $t_y = \sum_{k \in U} y_k$ , se define de la siguiente manera

$$\hat{t}_{y,cal} = \sum_{k \in s} w_k y_k,$$

---

<sup>1</sup>El uso de este diseño de muestreo está caracterizado porque la información auxiliar puede estar en dos niveles (Esteveo & Särndal 2001): alguna información puede estar al nivel de la población y otra puede estar al nivel de la primera fase de muestreo. En este trabajo suponemos que la información auxiliar está disponible sólo en la primera fase de muestreo.

donde las ponderaciones  $w_k$  minimizan una distancia  $\Phi_s$  entre las ponderaciones básicas<sup>2</sup> y las  $w_k$  sujetas a la siguiente restricción:

$$\sum_{k \in s} w_k \mathbf{x}'_k = \mathbf{t}_s' \quad (1.1)$$

En la construcción de un estimador de calibración hay dos componentes básicos: una distancia  $\Phi_s$  y un conjunto de restricciones tales dadas por (1.1). La distancia Ji-cuadrado dada por  $\Phi_s = \sum_{k \in s} (w_k - d_k)^2 / (d_k q_k)$ , es la más usada en la práctica. Nótese que los factores  $q_k$ , ponderaciones en la distancia, no están correlacionados con  $d_k$ . Existen otras distancias usadas para la construcción de estos estimadores, pero los resultados son asintóticamente equivalentes al usar la distancia Ji-cuadrado (Deville & Särndal 1992). De esta manera, al desarrollar la minimización, se llega fácilmente a que los pesos de calibración quedan definidos como

$$w_k = d_k + d_k (\mathbf{t}_s - \hat{\mathbf{t}}_{s\pi})' \left( \sum_{k \in s} q_k d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} q_k \mathbf{x}_k$$

donde  $\hat{\mathbf{t}}_{s\pi} = \sum_{k \in s} d_k \mathbf{x}_k$  es el estimador de Horvitz-Thompson para el total del vector de información auxiliar. Por tanto, el estimador de calibración para el total poblacional queda definido de la siguiente manera

$$\begin{aligned} \hat{t}_{y,cal} &= \sum_{k \in s} w_k y_k \\ &= \hat{t}_{y\pi} + (\mathbf{t}_s - \hat{\mathbf{t}}_{s\pi})' \hat{\mathbf{B}} \end{aligned}$$

donde  $\hat{t}_{y\pi} = \sum_{k \in s} d_k y_k$  es el estimador de Horvitz-Thompson para el total de la variable de interés y  $\hat{\mathbf{B}} = \left( \sum_{k \in s} q_k d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in s} q_k d_k \mathbf{x}_k y_k$  es una matriz de coeficientes de regresión.

En la sección 2 de este artículo se examinan los resultados encontrados por Wu (2003), en la sección 3 se estudian los resultados del estimador clásico del muestreo en dos fases, el estimador  $\pi^*$ . En la sección 4 se profundiza en algunas funciones que optimizan el estimador de calibración bajo un enfoque de inferencia basada en un modelo de super-población y se revisan las expresiones para el estimador del total poblacional en un diseño bifásico MAS-MAS<sup>3</sup>. Finalmente, mediante una simulación de Monte Carlo, se simula una población y se compara el sesgo relativo y la eficiencia relativa de los estimadores resultantes contra el estimador  $\pi^*$ .

## 2. Estimadores óptimos de calibración

Como lo afirma Wu (2003), existen dos variantes en la construcción de un estimador de calibración: una está dada por la escogencia de la distancia y la otra

<sup>2</sup>Las ponderaciones básicas se definen como el inverso de la probabilidad de inclusión para cada elemento incluido en la muestra y dadas por  $d_k = 1/\pi_k$

<sup>3</sup>El diseño bifásico MAS-MAS aplica un diseño de muestreo aleatorio simple sin reemplazo en la primera y en la segunda fase.

está dada por el conjunto de ecuaciones de calibración<sup>4</sup>. En áreas como la demografía existe la costumbre de calibrar sobre muchas variables, para que se logre estimar con varianza nula los totales conocidos de las variables auxiliares, sin importar que el estimador resultante pueda perder eficiencia. En estos términos, sería mejor utilizar la menor cantidad de ecuaciones de calibración para no estropear el buen comportamiento del estimador. La pregunta que debe plantearse el investigador es ¿cuál es la mejor ecuación de calibración que se debe usar en la construcción de un estimador de este tipo?

Si  $u_k = u(\mathbf{x}_k)$ , donde  $u(\cdot)$  es una función de valor real, entonces una nueva forma de construir un estimador de calibración estaría dada por la consecución de unos pesos  $w_k$  restringidos<sup>5</sup> a

$$\sum_{k \in s} w_k u(\mathbf{x}_k) = \sum_{k \in U} u(\mathbf{x}_k)$$

Por tanto, la pregunta se torna más diáfana y se convierte en ¿cuál función  $u(\cdot)$  hace al estimador  $\hat{t}_{y_{cal}}$  más eficiente?. Ahora, es bien sabido que bajo la inferencia basada en el diseño de muestreo, no existe un estimador insesgado de mínima varianza uniformemente (Cassel et al. 1976). Sin embargo, es posible obtener un estimador óptimo bajo la inferencia asistida por modelos de super-población. La respuesta a estas preguntas está dada por la propuesta de Wu (2003), que construyó un estimador óptimo de calibración suponiendo que las respuestas de  $y_k$  pueden ser vistas como realizaciones del sigui-ente modelo de super-población semi-paramétrica

$$E_{\xi}(y_k | \mathbf{x}_k) = \mu(\mathbf{x}_k, \boldsymbol{\theta}), \quad V_{\xi}(y_k | \mathbf{x}_k) = [\nu(\mathbf{x}_k)]^2 \sigma^2, \quad (2.1)$$

donde  $\mu(\cdot, \cdot)$  y  $\nu(\cdot)$  son funciones conocidas,  $\boldsymbol{\theta}$  y  $\sigma^2$  son parámetros desconocidos del modelo. Se asume que los  $y_k$ ,  $k \in U$ , son condicionalmente independientes dadas las  $\mathbf{x}_k$ . Nótese que  $\nu$  puede ser una función conocida de  $\mu$  como en los modelos lineales generalizados.

Los estimadores óptimos, asistidos por un modelo de super-población  $\xi$ , que minimizan el valor esperado de la varianza basada en un diseño de muestreo,  $E_{\xi}(V_p(\hat{Y}))$ , han sido discutidos<sup>6</sup> por muchos autores. Por ejemplo, en Isaki & Fuller (1982) esta varianza esperada tomó el nombre de varianza anticipada.

**Resultado 2.1** (Teorema 1 de Wu (2003)). *Sea  $t_{y, C_u}$  un estimador de calibración del total poblacional de la característica de interés, construido utilizando la restricción (2), donde  $C_u = \{u(\mathbf{x}_1), u(\mathbf{x}_2), \dots, u(\mathbf{x}_N)\}$  es la familia de vectores de todas las posibles funciones de valor real aplicadas a la información auxiliar. Dentro de la clase de estimadores de calibración  $t_{y, C_u}$ , la escogencia de*

$$C_{\mu} = \{\mu(\mathbf{x}_1, \boldsymbol{\theta}), \mu(\mathbf{x}_2, \boldsymbol{\theta}), \dots, \mu(\mathbf{x}_N, \boldsymbol{\theta})\}$$

<sup>4</sup>Nótese que si el vector de información auxiliar tiene  $P$  variables auxiliares, entonces habrán  $P$  ecuaciones de calibración.

<sup>5</sup>Bajo este marco de referencia aparece una reducción en la cantidad de restricciones que se utilizan en la calibración.

<sup>6</sup>Los términos  $E_p$  y  $V_p$  se refieren a la esperanza y varianza bajo un diseño muestral  $p(\cdot)$ , y  $E_{\xi}$  y  $V_{\xi}$  denotan la esperanza y varianza bajo un modelo de super-población  $\xi$ .

minimiza  $E_{\xi}(V_p(\hat{Y}))$  bajo el modelo de super-población dado por (2.1) y suponiendo condiciones de regularidad en el diseño de muestreo.

Con este resultado podemos proseguir a la construcción del estimador óptimo de calibración resultante de minimizar Ji-cuadrado sujeta a la siguiente restricción

$$\sum_{k \in s} w_k \hat{\mu}_k = \sum_{k \in U} \hat{\mu}_k$$

Donde  $\hat{\mu}_k = \mu(\mathbf{x}_k, \boldsymbol{\theta})$ . La razón para esto se debe a que los valores del vector  $\boldsymbol{\theta}$  son desconocidos y se deben reemplazar por un estimador basado en la muestra seleccionada dado por  $\hat{\boldsymbol{\theta}}$ . La minimización se realiza usando un multiplicador de Lagrange como en Deville & Särndal (1992). De esta manera, es muy fácil conseguir la expresión del estimador óptimo de calibración, el cual está dado por (Wu & Sitter 2001)

$$\begin{aligned} \hat{t}_{y,opt} &= \sum_{k \in s} w_k y_k \\ &= \hat{t}_{y\pi} + (t_{\hat{\mu}} - \hat{t}_{\hat{\mu}\pi}) \hat{B}_y \end{aligned}$$

en donde  $t_{\hat{\mu}} = \sum_{k \in U} \hat{\mu}_k$  es el total poblacional de las funciones  $\hat{\mu}$ ,  $\hat{t}_{\hat{\mu}\pi}$  su correspondiente estimador de Horvitz-Thompson y

$$\hat{B}_y = \frac{\sum_{k \in s} d_k q_k \hat{\mu}_k y_k}{\sum_{k \in s} d_k q_k \hat{\mu}_k^2}$$

En resumen, los estimadores óptimos de calibración se han estudiado y profundizado en Wu & Sitter (2001) y Wu (2003), y su fundamento se encuentra en la inferencia asistida por modelos. Para motivar las condiciones de optimalidad se utilizó un modelo de super-población semi-paramétrica general dado por (2.1). Estos estimadores de calibración para el total poblacional de la característica de interés tienen las siguientes características:

1. Una distancia Ji-cuadrado cuyos factores de peso satisfacen

$$q_k > 0 \quad \text{y} \quad N^{-1} \sum_{k=1}^N q_k^2 = O(1).$$

2. Una sola restricción, dada por una reducción de dimensión  $u_k = \mu(\mathbf{x}_k, \boldsymbol{\theta})$ , donde la forma funcional  $\mu(\cdot, \cdot)$  puede ser arbitraria.

Algunos de los resultados más importantes de este método pueden ser resumidos de la siguiente manera (Wu 2003):

- Sea  $\hat{\boldsymbol{\theta}} = (\sum_{k \in S} d_k q_k \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_{k \in S} d_k q_k \mathbf{x}_k y_k$ . Si se usa  $u_k = \mathbf{x}_k' \boldsymbol{\theta}$  como variable de calibración, el estimador de calibración resultante es idéntico al estimador convencional de calibración dado por  $\hat{t}_{y_{cal}}$ . Por tanto, la clase de estimadores resultantes de este método es muy general, pues incluye al estimador original como un caso particular.
- Para cualquier estimador consistente de  $\boldsymbol{\theta}$  tal que  $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + o_p(1)$ , si se reemplaza  $\boldsymbol{\theta}$  por  $\hat{\boldsymbol{\theta}}$ , en las ecuaciones de calibración, el estimador de calibración resultante no cambia asintóticamente.
- Los estimadores óptimos de calibración obtenidos usando  $u_k = E_{\xi}(y_k | \mathbf{x}_k) = \mu(\mathbf{x}_k, \boldsymbol{\theta})$ , son óptimos bajo el criterio del mínima varianza esperada.
- Los estimadores óptimos de calibración son óptimos bajo el modelo de superpoblación  $\xi$ , pero aún si el modelo considerado es incorrectamente especificado, estos estimadores permanecen consistentes.

Dado que no existe un estimador insesgado con varianza mínima uniforme, la única escogencia de  $u(\cdot)$  que hace a  $\hat{t}_{y_{opt}}$  un estimador con las anteriores características es  $u(\mathbf{x}_k) = y_k$ , y por supuesto esto es prácticamente inútil. Por tanto se debe hacer  $u(\mathbf{x}_k) \approx y_k$ .

El lector debe notar que la estructura del modelo  $\xi$  dado por (2.1) es muy general e incluye dos importantes casos: el primero, el modelo de regresión lineal o no lineal dado por

$$y_k = \mu(\mathbf{x}_k, \boldsymbol{\theta}) + \nu_k \varepsilon_k \quad (2.2)$$

donde los  $\varepsilon_k$  son variables aleatorias independientes e idénticamente distribuidas con  $E_{\xi}(\varepsilon_k) = 0$ ,  $V_{\xi}(\varepsilon_k) = \sigma^2$  y  $\nu_k = \nu(\mathbf{x}_k)$  es una función conocida y estrictamente positiva.

El segundo caso se refiere al modelo lineal generalizado dado por

$$g(\mu_k) = \mathbf{x}_k' \boldsymbol{\theta}, \quad V_{\xi}(y_k | \mathbf{x}_k) = \nu(\mu_k) \quad (2.3)$$

donde  $\mu_k = E_{\xi}(y_k | \mathbf{x}_k)$ ,  $g(\cdot)$  es una función de vínculo y  $\nu(\cdot)$  es una función de varianza.

A continuación se describe el comportamiento de los estimadores óptimos de calibración bajo un modelo lineal y un modelo log-lineal.

## 2.1. $u(\mathbf{x})$ Vía mínimos cuadrados

Si la información auxiliar explica a la característica de interés de forma lineal, como se observa en la Figura 10.5, entonces tendría sentido el argumento que se expresa en Deville & Särndal (1992), en donde motivados por el estimador de razón, se argumenta que «...las ponderaciones [de calibración] que se ajustan bien

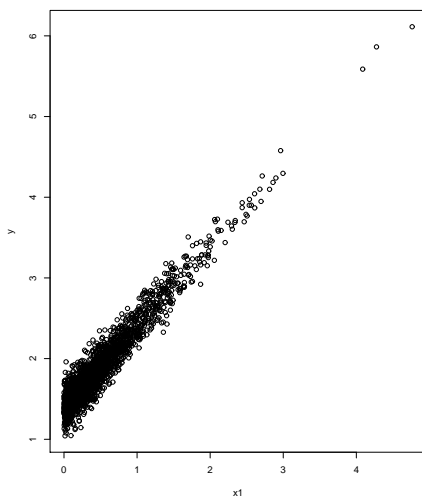


Figura 1: *Comportamiento lineal de la característica de interés explicada por la información auxiliar*

*a las variables auxiliares [reproducen exactamente su total poblacional], también se ajustan bien a la variable de estudio...»*

En el caso multivariado, la función que hace óptimo al estimador de calibración está dada por

$$u(\mathbf{x}_k, \boldsymbol{\theta}) = \mathbf{x}'_k \boldsymbol{\theta} = \theta_0 + \theta_1 x_{k1} + \dots + \theta_P x_{kP} \tag{2.4}$$

En donde  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_P)$  es estimado a través de mínimos cuadrados ponderados, como en una regresión múltiple. Por lo tanto, la característica de interés sigue el siguiente modelo de super-población

$$y_k = \mathbf{x}'_k \boldsymbol{\theta} + \nu_k \varepsilon_k \tag{2.5}$$

donde los  $\varepsilon_k$  son independientes e idénticamente distribuidos con  $E_\xi(\varepsilon_k) = 0$  y  $V_\xi(\varepsilon_k) = \sigma^2$ , y  $\nu_k = \nu(\mathbf{x}_k) = 1$ . Por tanto, al estimar  $\boldsymbol{\theta}$  usando la técnica de mínimos cuadrados se tiene que

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \left( \sum_{k \in s} q_k d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} \sum_{k \in s} q_k d_k \mathbf{x}_k y_k \\ &= (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \end{aligned}$$

donde  $\mathbf{V} = \text{diag}(d_1 q_1, \dots, d_n q_n) = \frac{1}{\sigma^2} \text{diag}(d_1, \dots, d_n)$ .

**Resultado 2.2.** De esta forma, el estimador de calibración del total poblacional resultante del anterior modelo de super-población está dado por

$$\hat{t}_{y,opt} = t_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})' \hat{\boldsymbol{\theta}} \quad (2.6)$$

*Demostración.*

$$\begin{aligned} \hat{t}_{y,opt} &= \hat{t}_{y\pi} + (t_{\hat{\mu}} - \hat{t}_{\hat{\mu}\pi}) \hat{B}_y \\ &= \hat{t}_{y\pi} + \left( \sum_{k \in U} \hat{\mu}_k - \sum_{k \in U} d_k \hat{\mu}_k \right) \hat{B}_y \\ &= \hat{t}_{y\pi} + \left( \sum_{k \in U} \mathbf{x}'_k \hat{\boldsymbol{\theta}} - \sum_{k \in U} d_k \mathbf{x}'_k \hat{\boldsymbol{\theta}} \right) \hat{B}_y \\ &= \hat{t}_{y\pi} + \left( \sum_{k \in U} \mathbf{x}'_k - \sum_{k \in U} d_k \mathbf{x}'_k \right) \hat{\boldsymbol{\theta}} \hat{B}_y \\ &= \hat{t}_{y\pi} + \left( \sum_{k \in U} \mathbf{x}_k - \sum_{k \in U} d_k \mathbf{x}_k \right)' \hat{\boldsymbol{\theta}} \hat{B}_y \\ &= \hat{t}_{y\pi} + \left( \sum_{k \in U} \mathbf{x}_k - \sum_{k \in U} d_k \mathbf{x}_k \right)' \hat{\boldsymbol{\theta}} \end{aligned}$$

puesto que  $\hat{B}_y = 1$ . Lo anterior se tiene de la definición de  $\hat{B}_y$  teniendo en cuenta que

$$\hat{\mu}_k = \mathbf{x}'_k \boldsymbol{\theta} = \mathbf{x}'_k (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}$$

Por tanto,

$$\begin{aligned} \sum_{k \in s} d_k q_k \hat{\mu}_k^2 &= \mathbf{y}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \\ &= \mathbf{y}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \\ &= \sum_{k \in s} d_k q_k \hat{\mu}_k y_k \end{aligned}$$

□

Nótese que el termino  $\hat{B}_Y$  es igual a uno y por tanto desaparece, lo que hace que el estimador óptimo de calibración sea idéntico al estimador de calibración clásico dado por (10.4.5).

## 2.2. $u(\mathbf{x})$ Vía modelo lineal generalizado

¿Qué sucede si la información auxiliar no describe a la característica de interés con un comportamiento lineal, como se observa en la Figura 2.?

Es ésta la parte más importante del desarrollo práctico en los estimadores óptimos de calibración. Al respecto, el usuario puede pensar por un instante en los siguientes cuestionamientos:

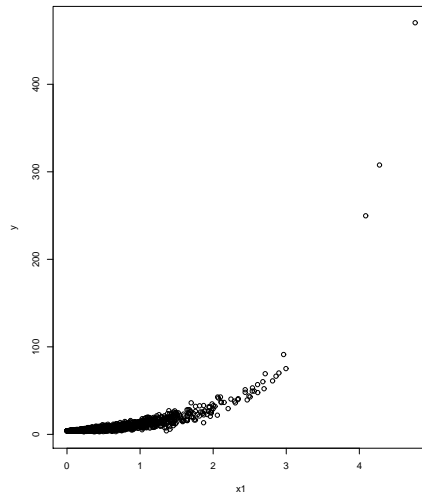


Figura 2: *Comportamiento no lineal de la característica de interés explicada por la información auxiliar*

- Si una característica de información auxiliar explica muy bien a la característica de interés, entonces calibrar con respecto a esta información auxiliar sería muy conveniente. Sin embargo, esta relación no siempre será lineal.
- Si queremos estimaciones perfectas deberíamos utilizar a la misma característica de interés para calibrar, pero como ésto es un absurdo se debe utilizar  $u(\mathbf{x})$  semejante a  $y$ .

Si se conoce que la información auxiliar disponible no describe a la característica de interés de forma lineal, se ponen en tela de juicio la aplicación de los estimadores clásicos de calibración motivadas por Deville & Särndal (1992). Por tanto, si los valores de la característica de interés son considerados como realizaciones de un modelo de super-población  $\xi$  como en (2.1) que puede ser descrito a través de su primer y segundo momento, entonces claramente el modelo lineal generalizado (MLG), descrito detalladamente en McCullagh & Nelder (1989) y dado por (2.3). La mayor particularidad del MLG es que la varianza de la característica de interés depende de la media  $\mu_k$ . Además, en el MLG se considera que la característica de interés se relaciona con las variables de información auxiliar mediante la media  $\mu_k$  y una función de vínculo  $g(\cdot)$  tal que

$$g(\mu_k) = \theta_0 + \theta_1 x_{k1} + \dots + \theta_P x_{kP}$$

Nótese que el modelo clásico de regresión lineal es un caso particular del MLG en donde  $g(\mu_k) = \mu_k$  y  $V(\mu_k) = 1$ . Por supuesto, existen otras formas de la función de varianza y vínculos no lineales también son permitidos. Por ejemplo, entre las

funciones de vínculo y de varianza más populares están el vínculo logarítmico dado por  $g(\mu_k) = \log(\mu_k)$  y las funciones de varianza de Poisson dada por  $V(\mu_k) = \mu_k$  y la varianza Gamma dada por  $V(\mu_k) = \mu_k^2$ .

El MLG es un método semi-paramétrico y requiere especificaciones solamente en el primer y segundo momento. La función de vínculo  $\mu_k$  está relacionada a las variables independientes y la función de varianza describe cómo la variación en la característica de interés está relacionada con la media.

Los coeficientes  $(\theta_0, \theta_1, \dots, \theta_k)$  pueden ser estimados, como en nuestro caso, usando el método de máxima cuasi-verosimilitud. Para el caso más general, el estimador del vector de parámetros poblacionales  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_P)'$ , es la solución de la siguiente ecuación

$$\mathbf{D}'\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = 0 \quad (2.7)$$

La anterior, no es más que una generalización de las ecuaciones normales en un modelo de regresión múltiple. Donde  $\mathbf{y} = (y_1, \dots, y_n)'$  y  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ ,  $\mathbf{V} = \text{diag}(V(\mu_1), \dots, V(\mu_n))$  son las estructuras de media y varianza del modelo respectivamente, y  $\mathbf{D} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}$ . Los parámetros  $\theta_p$ ,  $p = 1, \dots, P$ , se encuentran implícitos en (2.7). En el caso más simple, el modelo lineal clásico, se tiene que  $\mu_k = \theta_0 + \theta_1 x_{k1} + \dots + \theta_P x_{kP}$ ,  $\boldsymbol{\mu} = \mathbf{X}'\boldsymbol{\theta}$  y  $\mathbf{D} = \mathbf{X}'$ . Luego, (2.7) queda convertida en  $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\theta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ , las cuales corresponden a las ecuaciones normales de la regresión múltiple.

Por otro lado, en cualquier otro modelo, en donde la función de vínculo sea distinta de la identidad, la mayor dificultad para encontrar el estimador máximo cuasi-verosímil de  $\boldsymbol{\theta}$  es que para resolver (2.7) se necesita utilizar procedimientos iterativos.

**Resultado 2.3.** *Bajo un modelo de super-población MLG, el estimador óptimo de calibración está dado por*

$$\hat{t}_{y,opt} = \hat{t}_{y\pi} + (t_{\hat{\mu}} - \hat{t}_{\hat{\mu}\pi})\hat{B}_y \quad (2.8)$$

con

$$\hat{B}_y = \frac{\sum_{k \in s} d_k q_k \hat{\mu}_k y_k}{\sum_{k \in s} d_k q_k \hat{\mu}_k^2}$$

donde  $\hat{\mu}_k = g^{-1}(\mathbf{x}'_k \boldsymbol{\theta})$  y  $g^{-1}(\cdot)$  es la inversa de la función de vínculo.

El software estadístico **R** (R Development Core Team 2008) tiene implementada la función `glm`, la cual permite estimar los parámetros del MLG. Suponga que se desea encontrar el estimador de máxima cuasi-verosimilitud de  $\boldsymbol{\theta} = (\theta_0, \theta_1)'$  para el modelo

$$\mu_k = \exp(\theta_0 + \theta_1 x_{k1}), \quad V_{\xi}(y_k | \mathbf{x}_k) = \nu(\mu_k)^2 = \mu_k^2 \quad (2.9)$$

Por supuesto, desde (10.9.3), se tiene que la función de vínculo es el logaritmo. Las siguientes líneas de código muestran cómo obtener  $\hat{\boldsymbol{\theta}}$

```
theta0 <- lm(Y ~ X)
theta1 <- glm(Y ~ X, start=theta0,quasi(var="mu^2",link="log"))
```

Nótese que `theta0` es el estimador de mínimos cuadrados y sirve como estimador inicial para el proceso iterativo. Análogamente, es posible crear un código propio para computar las estimaciones del vector de parámetros basado en McCullagh & Nelder (1989, p. 327).

```
tol<-0.000000001
theta0 <- solve(t(X)%*%X,t(X)%*%(Y)) ## valores iniciales
dif<-1
while(dif>=tol) ## condición de iteración
{
mu <- exp(as.vector(X)%*%theta0))
V <- diag(1/mu)
theta1 <- theta0+solve(t(X)%*%X,t(X)%*%V%*%(Y-mu))
dif <- max(abs(theta1-theta0))
theta0 <- theta1
}
```

Por supuesto, el resultado del anterior código debe coincidir con la salida que arroje el procedimiento `glm` de **R**. Claramente, la optimalidad de los estimadores propuestos por Wu (2003) depende del modelo que relacione a la característica de interés con la información auxiliar. Directamente, esta relación está supeditada a la información auxiliar *per se*; es decir, si ésta no está disponible, no es posible construir un estimador óptimo. La estrategia sugerida en las secciones anteriores está enfocada en la implementación de un diseño de muestreo en dos fases.

### 3. Estimadores óptimos de calibración bajo muestreo bifásico (MAS-MAS)

#### 3.1. Muestreo en dos fases

El muestreo bifásico se utiliza cuando hay poco o nulo conocimiento sobre la población. Por ejemplo, el estimador de razón combinada requiere que todos los elementos puedan ser estratificados y que  $\sum_{k \in U} \mathbf{x}_k$  sea conocido. Sin embargo, en muchos estudios esta información no está disponible. Cuando el marco de muestreo contiene poca o deficiente información se tienen dos opciones (Särndal et al. 1992):

1. Usar un diseño simple como MAS o MASC y al combinarlo con el estimador de Horvitz-Thompson se gana precisión al aumentar el tamaño de la muestra.
2. Obtener información acerca de la población y construir un nuevo marco muestral. Usando el estimador de regresión se logra la precisión deseada con un tamaño de muestra moderado.

Dado que un tamaño de muestra grande o la construcción de un nuevo marco de muestreo para toda la población resultaría muy costoso, una tercera opción es usar un diseño de muestreo en dos fases, así:

- a) En la primera fase seleccionar una muestra de tamaño moderado de elementos  $s_a$ , de tamaño  $n_a$ , mediante un diseño  $p_a(\cdot)$ . Para los elementos en  $s_a$  obtener información sobre una o más variables auxiliares puesto que esta opción resulta menos costosa que obtenerla en la población.
- b) En la segunda fase seleccionar una muestra  $s$ , de tamaño  $n$ , de  $s_a$ , con la ayuda de la información obtenida en la primera fase, mediante un diseño de muestreo  $p(\cdot | s_a)$ , donde  $s$  es la submuestra seleccionada de  $s_a$ . A continuación se observa la variable de estudio  $y$  para los elementos seleccionados en  $s$ .

### 3.1.1. El estimador $\pi^*$

Nótese que el estimador de Horvitz-Thompson  $\hat{t}_{y\pi} = \sum_{s_a} y_k / \pi_{ak}$  se podría calcular si  $y_k$  y  $\pi_{ak}$  fueran conocidos para todo  $k \in s_a$ . Sin embargo,  $y_k$  sólo es conocido para  $k \in s$ . Ahora, dados  $s_a$ , y  $\sum_{s_a} y_{ak}$ , el total poblacional de la característica de interés se estima mediante el estimador  $\pi^*$  definido como

$$\hat{t}_{y\pi^*} = \sum_s \frac{y_k}{\pi_k^*} = \sum_s \frac{y_k}{\pi_{ak}\pi_{k|s_a}} = \sum_s \check{y}_k \quad (3.1)$$

donde  $\pi_{ak}$  es la probabilidad de inclusión del  $k$ -ésimo elemento de la población  $U$  en la muestra  $s_a$ ,  $\pi_{k|s_a}$  es la probabilidad de inclusión del  $k$ -ésimo elemento de la muestra  $s_a$  en la submuestra final  $s$ , condicionada a la selección de  $s_a$  en la primera fase.

**Resultado 3.1.** *En muestreo bifásico el total poblacional es estimado insesgadamente por el estimador  $\pi^*$ . Cuando el diseño de muestreo corresponde a un MAS en la primera y segunda fase, éste estimador toma la siguiente forma*

$$\hat{t}_{y\pi^*} = \frac{N}{n} \sum_s y_k \quad (3.2)$$

La prueba de este resultado y otras propiedades del estimador  $\pi^*$  pueden ser consultadas en Särndal et al. (1992).

## 3.2. Estrategia de muestreo óptima

Para aprovechar las bondades del diseño de muestreo en dos fases y las bondades de los estimadores óptimos de calibración, se considera el caso en donde una muestra  $s_a$  de tamaño fijo  $n_a$  es tomada del universo  $U$  usando MAS y  $P$  variables auxiliares, se miden para todo  $k \in s_a$ . En una segunda fase se selecciona una submuestra

$s$  de tamaño fijo  $n$  usando el mismo diseño MAS y se mide la característica de interés  $y_k$  para todo  $k \in s$  y se utiliza un estimador óptimo de calibración para estimar el total poblacional  $t_y$ .

El siguiente resultado da cuenta del estimador resultante al implementar la anterior estrategia de muestreo. Éste es válido para cualquier diseño de muestreo en dos fases. Luego, se considera el caso particular que le da el nombre a este artículo.

**Resultado 3.2.** *Para un caso de muestreo bifásico arbitrario y utilizando el modelo de superpoblación (2.1), el estimador óptimo de calibración para el total poblacional se define como  $\hat{t}_{y,opt} = \sum_{k \in s} w_k y_k$ , donde los pesos calibrados  $w_k$  minimizan  $\Phi_s = \sum_{k \in s} (w_k - d_k)^2 / (d_k q_k)$ , sujetos<sup>7</sup> a:*

$$\sum_{k \in s} w_k \mu(\mathbf{x}_k, \hat{\theta}) = \frac{1}{\pi_{ak}} \sum_{k \in s_a} \mu(\mathbf{x}_k, \hat{\theta}) \quad (3.3)$$

Por tanto, la familia de estimadores resultantes para el total poblacional, que surge teniendo en cuenta los diseños de muestreo en la primera y segunda fase, está dada por

$$\hat{t}_{y,opt} = \sum_{k \in s} d_k y_k + \left\{ \sum_{k \in s_a} \frac{1}{\pi_{ak}} \mu(\mathbf{x}_k, \hat{\theta}) - \sum_{k \in s} d_k \mu(\mathbf{x}_k, \hat{\theta}) \right\} \hat{B}_Y, \quad (3.4)$$

donde

$$\hat{B}_Y = \frac{\sum_{k \in s} d_k q_k \hat{\mu}_k y_k}{\sum_{k \in s} d_k q_k \hat{\mu}_k^2} \quad (3.5)$$

con  $d_k = \frac{1}{\pi_k^*}$ ,  $q_k > 0$  y  $\hat{\mu}_k = \mu(\mathbf{x}_k, \hat{\theta})$ .

Es claro que del anterior resultado nace una familia de estimadores que se diferencian en la forma en que el parámetro poblacional  $\theta$  es estimado; a esta familia de estimadores se les llama **estimadores óptimos de calibración** porque minimizan la varianza esperada cuando el modelo asumido es correctamente especificado.

En nuestro caso particular, al utilizar MAS como diseño de muestreo en la primera y segunda fase, el estimador  $\pi^*$  toma una forma conocida. Siendo  $\pi_{ak} = n_a/N$  y  $\pi_{k|s_a} = n/n_a$ , por tanto tenemos que

$$d_k = \frac{1}{\pi_k^*} = \frac{1}{\pi_{ak} \pi_{k|s_a}} = (N/n_a)(n_a/n) = N/n \quad (3.6)$$

Luego, bajo el modelo (2.1), el estimador óptimo de calibración para el total poblacional está dado por

$$\hat{t}_{y,opt} = \frac{N}{n} \sum_{k \in s} y_k + \left\{ \frac{N}{n_a} \sum_{k \in s_a} \mu(\mathbf{x}_k, \hat{\theta}) - \frac{N}{n} \sum_{k \in s} \mu(\mathbf{x}_k, \hat{\theta}) \right\} \hat{B}_Y \quad (3.7)$$

<sup>7</sup>Como no se conoce el total poblacional de las variables de información auxiliar, el lado derecho de la ecuación (3.3) no es más que una estimación insesgada de éste.

donde

$$\hat{B}_Y = \frac{\sum_{k \in s} q_k \hat{\mu}_k y_k}{\sum_{k \in s} q_k \hat{\mu}_k^2} \quad (3.8)$$

En el desarrollo de los estimadores de calibración, los  $q_k$  juegan un papel muy importante, dado que de su escogencia depende la forma final del estimador resultante. En la mayoría de los casos, los  $q_k$  se toman como iguales a uno aunque éstos también sirven para proponer la estructura de varianza que rige el modelo supuesto.

### 3.2.1. Relación lineal

Si la información auxiliar explica a la variable de respuesta de forma lineal, como por ejemplo se observa en la Figura 1, se sigue que el estimador óptimo de calibración del total poblacional bajo muestreo bifásico MAS-MAS es

$$\hat{t}_{y,opt} = \frac{N}{n} \sum_{k \in s} y_k + \left\{ \frac{N}{n_a} \sum_{k \in s_a} \mathbf{x}_k - \frac{N}{n} \sum_{k \in s} \mathbf{x}_k \right\}' \hat{\boldsymbol{\theta}} \quad (3.9)$$

donde

$$\hat{\boldsymbol{\theta}} = \left( \sum_{k \in s} q_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{k \in s} q_k \mathbf{x}_k y_k \quad (3.10)$$

### 3.2.2. Relación no lineal

Si la información auxiliar explica a la variable de respuesta de forma no lineal, como por ejemplo se observa en la Figura 2, se sigue que el estimador óptimo de calibración del total poblacional bajo muestreo bifásico MAS-MAS está dado por el Resultado 2.3. Por ejemplo, si la función de vínculo  $g = \log$ , entonces el estimador óptimo de calibración estaría dado por

$$\hat{t}_{y,opt} = \frac{N}{n} \sum_{k \in s} y_k + \left\{ \frac{N}{n_a} \sum_{k \in s_a} \exp(\mathbf{x}_k' \hat{\boldsymbol{\theta}}) - \frac{N}{n} \sum_{k \in s} \exp(\mathbf{x}_k' \hat{\boldsymbol{\theta}}) \right\} \hat{B}_Y \quad (3.11)$$

con

$$\hat{B}_y = \left\{ \frac{\sum_{k \in s} q_k \exp(\mathbf{x}_k' \hat{\boldsymbol{\theta}}) y_k}{\sum_{k \in s} q_k (\exp(\mathbf{x}_k' \hat{\boldsymbol{\theta}}))^2} \right\}$$

## 4. Simulación

En esta sección se realiza un estudio limitado por medio de una simulación de Monte Carlo. Éste se realiza con el fin de tener un acercamiento a la optimalidad

empírica de los estimadores que se revisaron en este artículo para un diseño bifásico MAS-MAS.

Se simuló un universo de tamaño  $N = 5000$  de un modelo de super-población,  $\xi$ . Se usaron dos modelos: uno de tipo lineal y otro no lineal descritos a continuación.

- El modelo lineal  $P_1$  es un modelo de regresión lineal simple con estructura de varianza no-homogénea,  $y_k = \theta_0 + \theta_1 x_k + x_k \varepsilon_k$ , para todo  $k \in U$ , con  $\varepsilon_k$  variables aleatorias independientes e idénticamente distribuidas de media cero y varianza constante  $\sigma^2$ . Nótese que las funciones de media y varianza para este modelo  $P_1$  son  $\mu_k = \mu(x_k, \boldsymbol{\theta}) = \theta_0 + \theta_1 x_k$  y  $\nu_k = v(x_k) = x_k$
- El modelo no lineal  $P_2$  es de tipo log-lineal,  $\log(y_k) = \theta_0 + \theta_1 x_k + \varepsilon_k$ . Las funciones de media y varianza para este modelo  $P_2$  son  $\mu_k = \exp(\theta_0 + \theta_1 x_k)$  y  $\nu_k = \mu_k$ , por tanto  $V_\xi(y_k|x_k) = \sigma^2 \mu_k^2$

En Wu (2003) se muestra que los estimadores resultan ser óptimos si el diseño muestral es regular<sup>8</sup>, por tanto distribuciones de colas pesadas como la distribución log-normal o la distribución gamma con parámetros de escala muy grandes, no podrán ser usadas para generar la información auxiliar. Por tanto, se generan los valores de  $x$  de una distribución gamma con parámetro de forma 1 y parámetro de escala 1/2. Esta variable toma valores no negativos y es sesgada a la derecha, esto es muy común en aplicaciones reales de encuestas por muestreo.

El valor del parámetro  $\theta_1$  se fijó en uno y el valor de  $\theta_0$  se fijó convenientemente tal que  $y_k > 0$  para ambos modelos. Se dispuso de cuatro valores para  $\sigma^2$  tal que el coeficiente de correlación, entre  $y$  y  $x$  para el modelo  $P_1$ , o entre  $\ln(y)$  y  $x$  para el modelo  $P_2$ , fuera 0.9, 0.8, 0.7, y 0.6, respectivamente.

En cada corrida de la simulación se seleccionó una primera muestra aleatoria simple  $s_a$ , de tamaño  $n_a = 2000$ , y de  $s_a$  se seleccionó una última muestra aleatoria simple de tamaño  $n = 100$ . Los parámetros  $\theta_0, \theta_1$  se estimaron usando mínimos cuadrados ponderados para el modelo  $P_1$ , y para el modelo  $P_2$  se utilizó el método de máxima cuasi-verosimilitud. Así, se calcularon los siguientes estimadores para el total poblacional de la característica de interés.

- $\hat{t}_{y\pi^*}$ : El estimador clásico al utilizar muestreo bifásico.
- $\hat{t}_{y,cal}$ : El estimador de calibración convencional.
- $\hat{t}_{y,opt}$ : El estimador óptimo de calibración.

El proceso se repitió  $B = 1000$  veces. La simulación fue programada en el paquete estadístico **R**. En la simulación, el desempeño de un estimador  $\hat{t}_y$  fue evaluado

<sup>8</sup>Un diseño muestral es regular si satisface: (i)  $\max_{k \in s} n d_k = O(1)$ , (ii). La distribución asintótica del estimador de Horvitz-Thompson es normal.

usando su sesgo relativo,  $SR$  y su eficiencia relativa,  $ER$ , definidas como

$$SR = B^{-1} \sum_{b=1}^B \frac{\hat{t}_{y,b} - t_y}{t_y} \quad (4.1)$$

$$ER = \frac{ECM(\hat{t}_{y\pi^*})}{ECM(\hat{t}_y)} \quad , \quad (4.2)$$

donde

$$ECM(\hat{t}_y) = B^{-1} \sum_{b=1}^B (\hat{t}_{y,b} - t_y)^2 \quad (4.3)$$

y  $\hat{t}_{y,b}$  se calculó en la  $b$ -ésima muestra simulada. Como se puede notar el estimador  $\hat{t}_{y\pi^*}$ , fue utilizado como línea base de comparación. Grandes valores para  $ER(> 1)$  representan alta eficiencia de un estimador en comparación al estimador  $\hat{t}_{y\pi^*}$ .

En las tablas 1 y 2 se puede observar cómo la optimalidad del estimador  $\hat{t}_{y,opt}$  está claramente soportada por los resultados de la simulación. En la Tabla 1 se nota que bajo el modelo lineal  $P1$  los estimadores  $\hat{t}_{y,cal}$  y  $\hat{t}_{y,opt}$  son idénticos. Se nota que se gana eficiencia si la correlación lineal crece. Para el modelo log-lineal  $P2$  el cambio en la eficiencia es bastante drástico, aunque el estimador  $\hat{t}_{y,cal}$  es más eficiente que  $\hat{t}_{y\pi^*}$ , el estimador  $\hat{t}_{y,opt}$  es mucho más eficiente que estos dos.

$\rho$	$\hat{t}_{y\pi^*}$	Modelo P1		Modelo P2	
		$\hat{t}_{y,cal}$	$\hat{t}_{y,opt}$	$\hat{t}_{y,cal}$	$\hat{t}_{y,opt}$
0.6	1.0	1.75	1.75	0.98	1.39
0.7	1.0	1.95	1.95	1.03	1.94
0.8	1.0	2.83	2.83	1.14	2.94
0.9	1.0	5.07	5.07	1.30	5.26

Tabla 1: *Eficiencia relativa de los estimadores para los modelos P1 y P2*

$\rho$	$Cal_{P1}/Opt_{P1}$	$Cal_{P1}/Cal_{P2}$	$Opt_{P1}/Opt_{P2}$	$Opt_{P2}/Cal_{P2}$
0.6	1.0	1.78	1.25	1.41
0.7	1.0	1.89	1.01	1.89
0.8	1.0	2.48	0.96	2.56
0.9	1.0	3.90	0.96	4.04

Tabla 2: *Proporción de la eficiencia relativa de los estimadores resultantes*

En términos de sesgo relativo, todos los valores correspondientes son menores del 0.5 por ciento. Aunque el estimador  $\hat{t}_{y,opt}$  tiene un sesgo aun mucho menor. En general, cuando la correlación aumenta, el sesgo decrece en ambos modelos. De los resultados se concluye que los estimadores óptimos de calibración funcionan bien bajo cualquier modelo, siendo el cociente entre los dos modelos muy cercano a uno, mientras que los estimadores convencionales pierden eficiencia cuando se utilizan bajo un modelo que no es lineal.

## 5. Comentarios finales

El desarrollo teórico revisado en este artículo, podría resultar muy útil cuando se desconoce el comportamiento de la población y cuando no se dispone de recursos suficientes para mejorar o construir un nuevo marco de muestreo o para aumentar el tamaño de muestra. En futuras investigaciones, se deben examinar los resultados para los casos en donde el diseño muestral es más complejo y cuando se tiene información auxiliar acerca de la inclusión de los individuos a grupos poblacionales en la etapa de estimación.

## Agradecimientos

Mi agradecimiento póstumo al profesor Leonardo Bautista, quien dirigió este trabajo incluso en los momentos más difíciles de su enfermedad. Él sembró muchas semillas en las mentes y corazones de cada uno de nosotros; hasta ahora sus frutos se están viendo y estoy seguro de que en un futuro no muy lejano, empezaremos a recoger lo que él sembró. También agradezco al Dr. Changbao Wu de la universidad de Waterloo, porque siempre estuvo atento para responder a nuestras múltiples inquietudes acerca de sus estimadores óptimos de calibración. Por último, agradezco a Dios que nos da la vida y los medios para poder investigar lo que él ya sabe.

**Recibido: 10 de noviembre de 2008**

**Aceptado: 6 de febrero de 2009**

## Referencias

- Cassel, C., Särndal, C. & Wretman, J. (1976), *Foundations of Inference in Survey Sampling*, Wiley.
- Deville, J. & Särndal, C. (1992), 'Calibration estimators in survey sampling.', *Journal of the American Statistical Association* **87**, 376–382.
- Estevao, V. M. & Särndal, C.-E. (2001), 'The ten cases of auxiliary information for calibration estimators in two-phase sampling.', *Journal of Official Statistics* **18**, 233–255.
- Isaki, C. T. & Fuller, W. A. (1982), 'Survey design under the regression superpopulation model.', *Journal of the American Statistical Association* **77**, 89–96.
- McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, Chapman and Hall.

- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Särndal, C., Swensson, B. & Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer, New York.
- Wu, C. (2003), ‘Optimal calibration estimators in survey sampling.’, *Biometrika* **90**, 937–951.
- Wu, C. & Luan, Y. (2003), ‘Optimal calibration estimators under two-phase sampling.’, *Journal of Official Statistics* **19**, 119–131.
- Wu, C. & Sitter, R. R. (2001), ‘A model-calibration approach to using complete auxiliary information from survey data.’, *Journal of the American Statistical Association* **96**, 185–193.