
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^a
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.

^aDirector. Centro de Investigaciones y Estudios Estadísticos (CIEES). Universidad Santo Tomás

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 fases¹. 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. π_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

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

¹El 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 Φ_s entre las ponderaciones básicas² y las w_k sujetas a la siguiente restricción:

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

En la construcción de un estimador de calibración hay dos componentes básicos: una distancia Φ_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

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

donde $\hat{\mathbf{t}}_{x\pi} = \sum_{k \in s} d_k \mathbf{x}_k$ 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

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

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 π^* . 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-MAS³. 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 π^* .

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

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

³El 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 calibración⁴. 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 restringidos⁵ 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 siguiente 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 σ^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 ν puede ser una función conocida de μ como en los modelos lineales generalizados.

Los estimadores óptimos, asistidos por un modelo de super-población ξ , que minimizan el valor esperado de la varianza basada en un diseño de muestreo, $E_{\xi}(V_p(\hat{Y}))$, han sido discutidos⁶ 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*

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

⁴Nótese que si el vector de información auxiliar tiene P variables auxiliares, entonces habrán P ecuaciones de calibración.

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

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

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 ξ , 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 ξ 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 ε_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

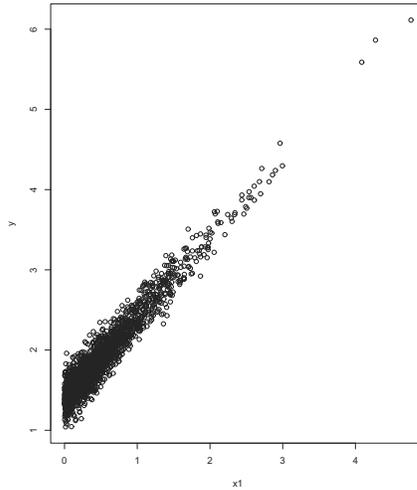


Figura 1: *Comportamiento lineal de la característica de interés explicada por la información auxiliar*

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

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

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

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

donde los ε_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:

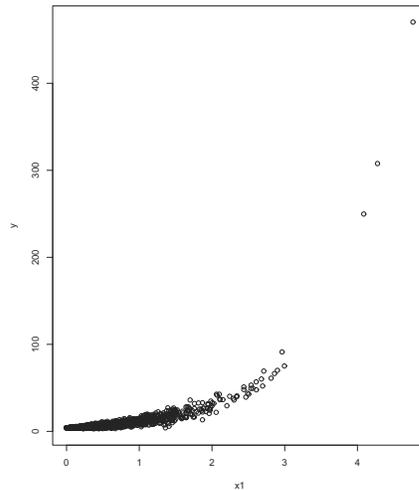


Figura 2: *Comportamiento no lineal de la característica de interés explicada por la información auxiliar*

- 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 ξ 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 μ_k . Además, en el MLG se considera que la característica de interés se relaciona con las variables de información auxiliar mediante la media μ_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