The role of adjusted maximum likelihood estimation in small area estimation

Partha Lahiri

Joint Program in Survey Methodology
University of Maryland, College Park

(Based on joint work with Huilin Li, New York University, and Masayo Yoshimori, Osaka University)

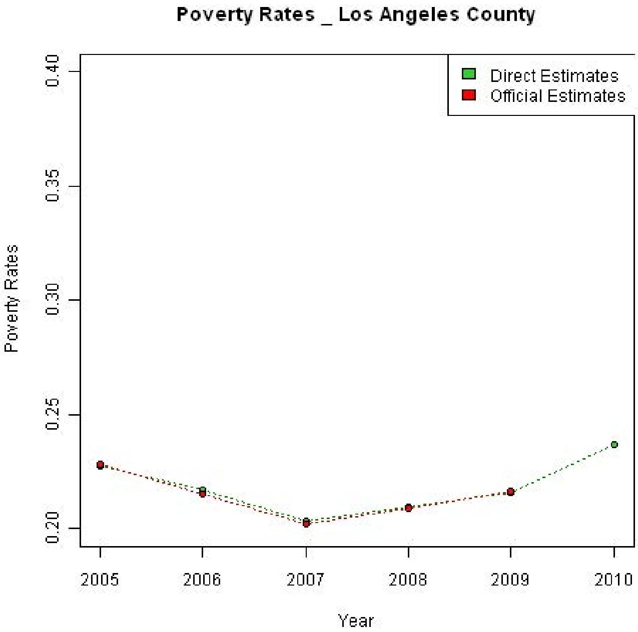
Sept 1, 2015

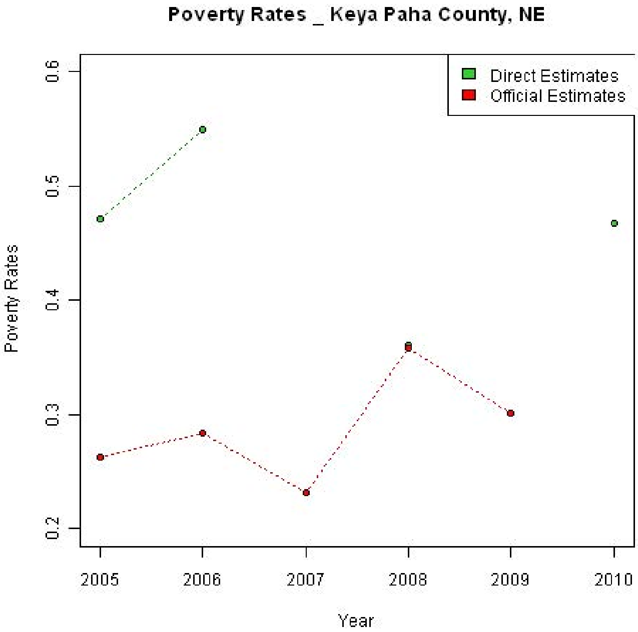
What is a Small Area or Domain?

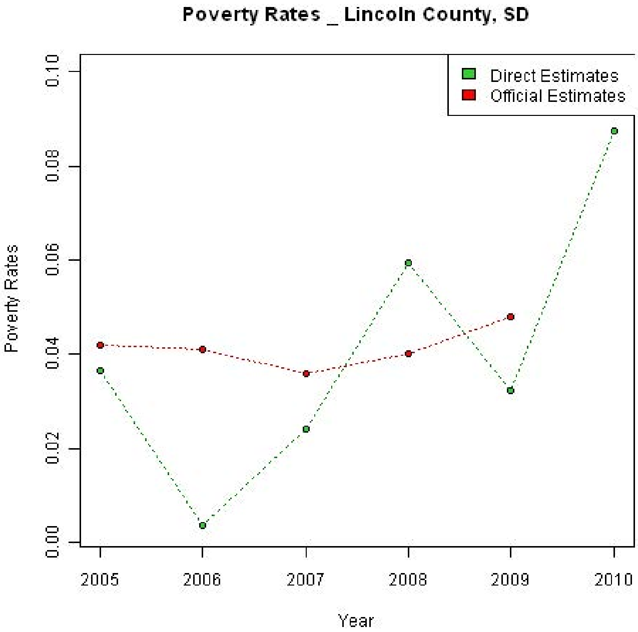
A subpopulation of interest with meager or no survey data.

Examples:

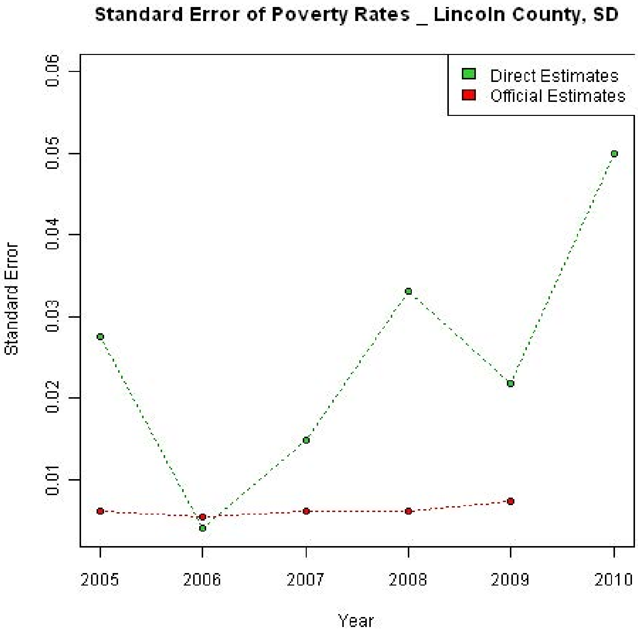
- In a nationwide survey, cells obtained by finer classification of age-group, race, gender even at the national level (small domains).
- In NHANSE III, a majority of US states do not have sample (small area).
- Even for a very large scale sample survey (e.g., American Community Survey), we can easily cite examples of small domains or areas (e.g., small counties or school districts).

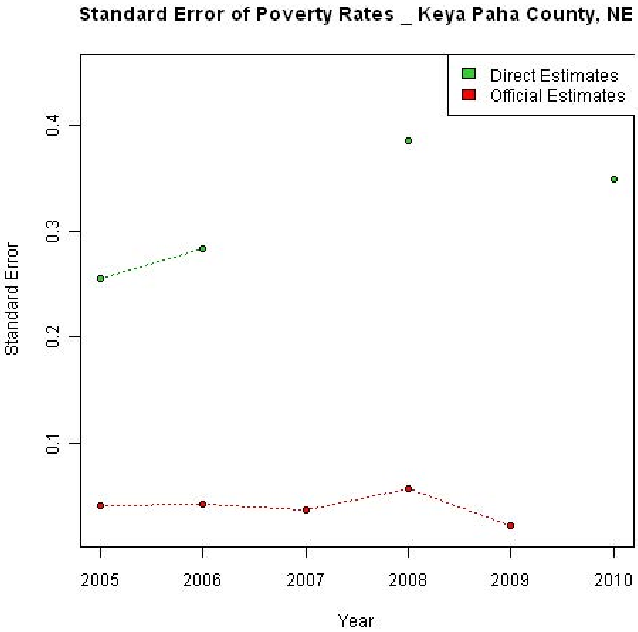

Poverty Rates _ Los Angeles County
Direct Estimates
Official Estimates
Poverty Rates
0.20
0.25
0.30
0.35
0.40
2005
2006
2007
2008
2009
2010
Year


Poverty Rates _ Keya Paha County, NE
Direct Estimates
Official Estimates
Poverty Rates
0.2
0.3
0.4
0.5
0.6
2005
2006
2007
2008
2009
2010
Year



Poverty Rates _ Lincoln County, SD
Direct Estimates
Official Estimates
Poverty Rates
0.00
0.02
0.04
0.06
0.08
0.10
2005
2006
2007
2008
2009
2010
Year


Standard Error of Poverty Rates _ Lincoln County, SD
Direct Estimates
Official Estimates
Standard Error
0.01
0.02
0.03
0.04
0.05
0.06
2005
2006
2007
2008
2009
2010
Year


Standard Error of Poverty Rates _ Keya Paha County, NE
Direct Estimates
Official Estimates
Standard Error
0.1
0.2
0.3
0.4
2005
2006
2007
2008
2009
2010
Year

The Fay Herriot Bayesian Model

Ref: Fay and Herriot (JASA 1979)

For $i = 1, \cdots, m$,

$$\textit{Level 1: (Sampling Distribution)}: \; y_i|\theta_i \sim N(\theta_i, D_i);$$
$$\textit{Level 2: (Prior Distribution)}: \; \theta_i \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}, A)$$

where

- m : number of small area;
- y_i : direct survey estimate of θ_i;
- θ_i : true mean for area i;
- $\boldsymbol{x}_i$: $p \times 1$ vector of known auxiliary variables;
- D_i: known sampling variance of the direct estimate;
- The $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}$ and model variance A are unknown.

The Bayesian Method

Parameter of Interest: θ_i

Inferences based on the posterior distribution of θ_i:

$$\theta_i | y; \beta, A \overset{ind}{\sim} N(\hat{\theta}_i^B,\ \sigma_i^2(A)),$$

where

- $\hat{\theta}_i^B = (1 - B_i)y_i + B_i \boldsymbol{x}_i' \boldsymbol{\beta}$
- $B_i = \frac{D_i}{A + D_i}$
- $\sigma_i^2(A) = (1 - B_i)D_i$

Adjusted Maximum Likelihood Method

Adjusted Likelihood:

$$h_i(A)L(A)$$

where

- $L(A)$: REML likelihood
- $h_i(A)$: an adjustment factor

Different Choices for the adjustment factor:

- Approximation of the posterior distribution under flat prior on β and A: $h_i(A) = A$ (Morris and Tang, Stat Sci. 2011)
- Construction of empirical Bayes parametric bootstrap confidence interval: $h_i(A) = A$ (Li and Lahiri, JMVA 2010)
- Construction of the second-order efficient empirical Bayes confidence interval: Yoshimori and Lahiri (AS 2014)
- Strictly positive estimator of A with the identical higher-order asymptotic properties of standard MLE of A: Yoshimori and Lahiri (JMVA 2014)

Confidence Interval for θ_i

An interval, denoted by I_i, is called a $100(1-\alpha)\%$ interval for θ_i if

$$P(\theta_i \in I_i|\beta, A) = 1 - \alpha,\ \forall\beta \in R^p, A \in R^+,$$

where

- the probability P is with respect to the joint distribution of $\{(y_i, \theta_i), i = 1, \cdots, m\}$ under the Fay-Herriot model;
- R^+ is the positive part of the real line.

A General Form of Confidence Interval for θ_i

Most of the intervals proposed in the literature can be written as:

$$\left(\hat{\theta}_i + q_1(\alpha)\hat{\tau}_i(\hat{\theta}_i),\ \hat{\theta}_i + q_2(\alpha)\hat{\tau}_i(\hat{\theta}_i)\right)$$

where

- $\hat{\theta}_i$ is an estimator of θ_i;
- $\hat{\tau}_i(\hat{\theta}_i)$ is an estimate of the measure of uncertainty of $\hat{\theta}_i$;
- $q_1(\alpha)$ and $q_2(\alpha)$ are chosen suitably in an effort to attain coverage probability close to the nominal level $1 - \alpha$.

Direct Confidence Interval

The choice $\hat{\theta}_i = y_i$ leads to the direct interval I_i^D given by

$$I_i^D : y_i \pm z_{\alpha/2}\sqrt{D_i},$$

where $z_{\alpha/2} \equiv z$ is the upper $100(1-\alpha/2)\%$ point of $N(0,1)$.

Remarks:

- The coverage probability is $1-\alpha$;
- When D_i is large, the length is too large to make any reasonable conclusion.

Synthetic Confidence Interval

Ref: Hall and Maiti (JRSS, 2006)

$$(x_i^T \hat{\beta}_{(-i)} + q_1(\alpha)\sqrt{\hat{A}_{(-i)}},\ x_i^T \hat{\beta}_{(-i)} + q_2(\alpha)\sqrt{\hat{A}_{(-i)}})$$

where

- $\hat{\beta}_{(-i)}$ and $\hat{A}_{(-i)}$ are consistent estimators of β and A, respectively, based on all but the ith area data.
- $L_i^*[q_2(\alpha)] - L_i^*[q_1(\alpha)] = 1 - \alpha$ where L_i^* is a parametric bootstrap approximation of the distribution L_i of $\frac{\theta_i - x_i' \hat{\beta}_{(-i)}}{\hat{A}_{(-i)}}$.

Remarks:

- The coverage is $1 - \alpha + O(m^{-1.5})$.
- This approach could be useful in situations especially when y_i is missing for the ith area.
- The method is synthetic (Rao 2005; Chatterjee, Lahiri, Li 2008).
- $\hat{A}_{(-i)}$ could be zero.

Bayesian Credible Interval

Assume β and A are known.

$$I_i^B(A) : \hat{\theta}_i^B(A) \pm z_{\alpha/2}\sigma_i(A),$$

where $\sigma_i(A) = \sqrt{(1 - B_i)D_i}$.

Remarks:

- The Bayesian credible interval cuts down the length of the direct confidence interval by $100 \times (1 - \sqrt{1 - B_i})\%$
- The maximum benefit from the Bayesian methodology is achieved when B_i is large.

Empirical Bayes Confidence Interval: Balanced Case ($x_i'\beta = \mu,\ D_i = D$)

Ref: Cox (1975)

$$I_i^{Cox}(\hat{A}) : \hat{\theta}_i^{EB}(\hat{A}) \pm z_{\alpha/2}\sigma(\hat{A}),$$

where

- $x_i^T\beta = \mu$ is estimated by the sample mean $\bar{y} = m^{-1}\sum_{i=1}^m y_i$ and
- A by the ANOVA estimator:
  $\hat{A}_{ANOVA} = \max\left\{(m-1)^{-1}\sum_{i=1}^m (y_i - \bar{y})^2 - D,\ 0\right\}$.

Remarks:

- Like the Bayesian credible interval, the length of the Cox interval is smaller than that of the direct interval.
- The distribution of $\frac{\theta_i - \hat{\theta}_i^{\mathrm{EB}}}{\sigma(\hat{A})}$ is not a standard Normal. Thus, it is not appropriate to use the Normal quantile $z_{\alpha/2}$ as the cut-off points.
- The Cox empirical Bayes confidence interval introduces a coverage error of the order $O(m^{-1})$, not accurate enough in most small area applications.
- length of the interval is zero when $\hat{A}_{ANOVA} = 0$

Other EB Confidence Intervals

- Replace $z_{\alpha/2}$ by $z_{\alpha'/2}$ to reduce coverage error (Cox 1975).
- Replace $\sigma(\hat{A})$ by a measure of uncertainty that captures uncertainty due to estimation of the hyperparameters β and A Ref: Morris (1983) Prasad and Rao (1990)
- Replace $z_{\alpha/2}$ by $z_{\alpha/2}c_i(\hat{A})$ to reduce the coverage error to $O(m^{-1.5})$ (Datta et al. 2002; Basu et al. 2003; Yoshimori 2013)
- Parametric bootstrap (Laird and Louis 1987; Carlin and Louis 1996; Chatterjee et al. 2008)

Use of adjusted MLE: Parametric Bootstrap EB Confidence Interval

Ref: Chatterjee, Lahiri and Li (2008, AS), Li and Lahiri (2010, JMVA)

- Draw bootstrap sample from the following bootstrap model:
  (i) $y_i^*|\theta_i^* \overset{ind}{\sim} N(\theta_i^*, D_i)$
  (ii)$\theta_i^* \overset{ind}{\sim} N(x_i'\hat{\beta}, \hat{A})$
- Compute $\hat{\beta}^*$ and $\hat{A}^*$ from y^* using adjusted REML with $h_i(A) = A$. Then we have $\hat{\theta}_i^{\mathrm{EB}*} = (1 - \hat{B}^*)y_i^* + \hat{B}^* x_i^{'}\hat{\beta}^*$, and $\sigma_i^2(\hat{A}^*) = (1 - \hat{B}^*)D_i$;
- Compute $(\theta_i^* - \hat{\theta}_i^{\mathrm{EB}*})/\sigma_i(\hat{A}^*)$.
- Use the distribution $\mathcal{L}_i^*$ of $\frac{\theta_i^* - \hat{\theta}_i^{\mathrm{EB}*}}{\sigma_i(\hat{A}^*)}$ to approximate the distribution $\mathcal{L}_i$ of $\frac{\theta_i - \hat{\theta}_i^{\mathrm{EB}}}{\sigma_i(\hat{A})}$.

Let q_1 and q_2 be real numbers such that

$$\mathcal{L}_i^*(q_2) - \mathcal{L}_i^*(q_1) = 1 - \alpha.$$

Parametric Bootstrap Confidence Interval

$$\mathrm{CI}_i^{\mathrm{PB}} = \left(\hat{\theta}_i^{\mathrm{EB}} + q_1\sigma_i(\hat{A}),\ \hat{\theta}_i^{\mathrm{EB}} + q_2\sigma_i(\hat{A})\right).$$

Theorem

Under reg. cond. $\Pr(\theta_i \in \mathrm{CI}_i^{\mathrm{PB}}) = 1 - \alpha + O(m^{-3/2})$.

Use of adjusted MLE: second-order efficient EB confidence interval

$$I_i^{Cox}(\hat{A}_{h_i}) : \hat{\theta}_i^{EB}(\hat{A}_{h_i}) \pm z_{\alpha/2}\sigma_i(\hat{A}_{h_i}),$$

where

- $\hat{A}_{h_i}$ is obtained by maximizing the following adjusted residual likelihood:

  $$L_{i;ad}(A) \propto h_i(A) \times L(A),$$

  with respect to A over $(0, \infty)$;

- $h_i(A)$ is a general area specific adjustment factor;

- $L(A)$ is the standard residual likelihood function.

A Higher-Order Expansion of Coverage

We obtain the following expansion under certain regularity conditions:

$$P(\theta_i \in I_i^{Cox}(\hat{A}_{h_i})) = 1 - \alpha + z\phi(z)\frac{a_i + b_i(h_i(A))}{m} + O(m^{-1.5}),$$

where

$$\begin{aligned}
z &= z_{\alpha/2} \\
a_i &= -\frac{m}{tr(V^{-2})}[\frac{4D_i}{A(A+D_i)^2} + \frac{(1+z^2)D_i^2}{2A^2(A+D_i)^2}] - \frac{mD_i}{A(A+D_i)}x_i' Var(\tilde{\beta})x_i \\
b_i &= \frac{2m}{tr(V^{-2})}\frac{D_i}{A(A+D_i)} \times \frac{\partial \log(h_i(A))}{\partial A} \\
\tilde{\beta} &= \hat{\beta}(A) = (X'V^{-1}X)^{-1}X'V^{-1}y
\end{aligned}$$

A Second-order Efficient Empirical Bayes Confidence Interval: Choice of $h_i(A)$

For small area i, we suggest an adjusted REML estimator of A where the adjustment factor satisfies the following differential equation:

$$a_i + b_i(h_i(A)) = 0.$$

Let $\hat{A}_i$ denote a solution to the above. Then our proposed empirical Bayes confidence interval for θ_i is given by

$$I_i^{YL}(\hat{A}_i) : \hat{\theta}_i^{EB}(\hat{A}_i) \pm z_{\alpha/2}\sigma_i(\hat{A}_i).$$

Since $\sigma_i(\hat{A}_i) < \sqrt{D_i}$, the length of this interval, like the original Cox interval $I_i^{Cox}(\hat{A}_{ANOVA})$, is always less than that of the direct interval I_i^D.

Choice of $h_i(A)$ when OLS of β is used

$$h_i(A) = A^{(1+z^2)/4}(A+D_i)^{(7-z^2)/4}\exp[-tr(V^{-1})x_i'(X'X)^{-1}X'VX(X'X)^{-1}x_i/2] \\ [\prod_{i=1}^{m}(A+D_i)]^{x_i'(X'X)^{-1}x_i/2} \times C.$$

where C is a generic constant free of A and $z = z_{\alpha/2}$.

For the balanced case $D_i = D$ $(i = 1, \cdots, m)$

$$h_i(A) = A^{(1+z^2)/4}(A+D)^{(7-z^2)/4+mx_i'(X'X)^{-1}x_i/2}C.$$

where C is a generic constant and free from A. In this balanced case, we show the uniqueness of the solution $\hat{A}_i$ if $m > \frac{4+p}{1-x_i'(X'X)^{-1}x_i}$.

Monte Carlo Simulation: The Fay-Herriot Model with $x_i^T\beta = 0$ and Unequal Sampling Variances

- $A = 1,\ m = 15$
- Five groups of small area $\{G_1,\ G_2,\ G_3,\ G_4,\ G_5\}$; within each group, D_i's are identical for the 3 small areas;
- Two D_i patterns:
  (a) $(0.7,\ 0.6,\ 0.5,\ 0.4,\ 0.3)$
  (b) $(4.0,\ 0.6,\ 0.5,\ 0.4,\ 0.1)$.
- $B = 6,000,\ R = 10^4$.

Simulation Results: The Fay-Herriot Model with $x_i^T \beta = 0$ and Unequal Sampling Variances

Table: Coverage Probability and Average Length

Pattern	G	MLE		Bootstrap		Adj. MLE		Direct	
a	1	89.8	(2.4)	94.5	(2.7)	95.3	(2.8)	95.1	(3.3)
	2	90.3	(2.3)	94.5	(2.5)	95.3	(2.6)	94.9	(3.0)
	3	90.6	(2.1)	94.6	(2.4)	95.2	(2.4)	95.2	(2.8)
	4	91.2	(2.0)	94.9	(2.2)	95.2	(2.2)	95.1	(2.5)
	5	91.1	(1.8)	94.3	(1.9)	95.0	(2.0)	94.7	(2.1)
b	1	88.3	(3.3)	94.5	(4.0)	95.8	(4.3)	94.9	(7.8)
	2	90.0	(2.3)	94.5	(2.5)	95.1	(2.6)	95.0	(3.0)
	3	90.4	(2.1)	94.6	(2.4)	95.3	(2.5)	94.9	(2.8)
	4	91.0	(2.0)	94.7	(2.2)	95.3	(2.2)	95.1	(2.5)
	5	93.1	(1.1)	94.7	(1.2)	95.0	(1.2)	95.0	(1.2)

Point Estimation of A and B_i: Likelihood-Based Methods

Profile Maximum Likelihood estimator (PML estimator)

$$\hat{A}_{PML} = \arg\max_{0<A<\infty} L_p(A, \boldsymbol{y}),$$

where

- $L_p(A, \boldsymbol{y}) = K|V|^{-1/2} \exp\{-\frac{1}{2} y' P y\}$, where K is a generic constant free from A;
- $P \equiv P(A) = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

Residual Maximum Likelihood estimator (REML estimator)

$$\tilde{A}_{RE} = \arg\max_{0<A<\infty} h_{RE}(A) L_p(A, \boldsymbol{y}),$$

where $h_{RE}(A) = |X'V^{-1}(A)X|^{-1/2}$.

Estimator of B_i: $\hat{B}_i = \frac{D_i}{\hat{A}+D_i}$

Remark: In some applications, $\hat{B}_i = 1$

Use of adjusted MLE: Point Estimation

Two Choices:

- Li and Lahiri (2010): $h(A) = A$
- Yoshimori and Lahiri (2014): $h(A) = \{\text{arctan}[tr(I - B(A))]\}^{1/m}$, where
    - $B \equiv B(A) = \text{diag}(B_1(A), \cdots, B_m(A))$;
    - I: an identity matrix of dimension m.

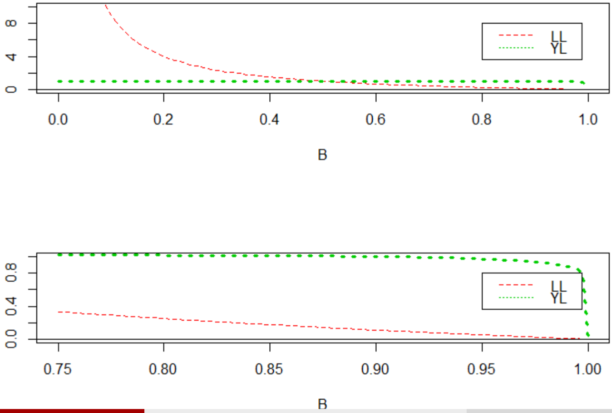
Adjusted MLE

$$\hat{A} = \arg \max_{0<A<\infty} h(A)L(A, \boldsymbol{y}),$$

Asymptotic Comparison

Result: For $A > 0$, $\log h_{YL}(A)$ is flatter than $\log h_{LL}(A)$ in the following sense

$$0 < \frac{\partial \log h_{YL}}{\partial A} < \frac{\partial \log h_{LL}}{\partial A}.$$



Higher Order Asymptotic Properties of Different Maximum Likelihood Methods

Let BV denote the bias to variance ratio for an estimator $\hat{B}_i$ of B_i.

Theorem 1

Under standard regularity conditions, we have, for large m,

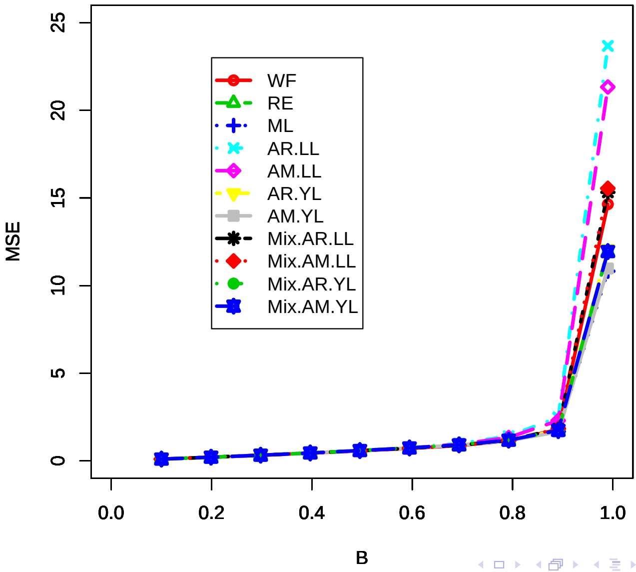
$$BV_{RE} = \frac{1}{B_i} + o(1),\ BV_{YL.RE} = \frac{1}{B_i} + o(1),\ BV_{LL.RE} = -\frac{1}{1-B_i} + o(1),$$

$$BV_{PML} = \frac{1}{B_i}[1 + (A + D_i)\frac{H}{2}] + o(1),\ BV_{YL.PML} = \frac{1}{B_i}[1 + (A + D_i)\frac{H}{2}] + o(1),$$

$$BV_{LL.PML} = -\frac{1}{1-B_i} + \frac{(A + D_i)^2}{D_i}\frac{H}{2} + o(1),$$

where $H = tr[V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}] > 0.$

MSE of EBLUP;p=1



Poverty Mapping for the Chilean Comunas

Partha Lahiri
JPSM, University of Maryland, College Park, USA

Livorno, Italy, June 16, 2015

[Based on joint work with Carolina Casas-Cordero and Jenny Encina]

Introduction

- The eradication of poverty has been at the center of various public policies in Chile and has guided public policy efforts.

- The nationwide survey estimate of the poverty rate has declined since the early 90´s suggesting some progress towards this goal. Erratic time series patterns, however, have emerged for small *comunas* - the smallest territorial entity in Chile.

- For a handful of extremely small comunas, survey estimates of poverty rates are unavailable for some or all time points simply because the survey design, which traditionally focuses on precise estimates for the nation and large geographical areas, excludes these comunas for some or all of the time points.

- Direct survey estimates of poverty rates typically do not meet the desired precision for small comunas and thus the assessment of implemented policies is not straightforward at the comuna level.

- In order to successfully monitor trends, identify influential factors, develop effective public policies and eradicate poverty at the comuna level, there is a growing need to improve on the methodology for estimating poverty rates at this level of geography.

- The need for socioeconomic data at lower levels of geography found its way into the Chilean legislation in 2007 when an amendment to the law of the *Fondo Común Municipal* (FCM) established a new set of indicators for its fund allocation algorithm among comunas. The regulation passed in 2009 required the Ministry to provide poverty rate estimates for all comunas in Chile.

- Regarding the production of comuna level estimates, an Expert Commission, appointed by the Ministry of Social Development (henceforth referred to as the *Ministry*) in 2010, raised concerns because of (1) the significant costs associated with sampling almost all comunas in the country, and (2) the relatively low precision for some comuna level estimates making the planned comparison among comunas and/or across time useless.

- The Commission recommended to (i) reduce the overall sample significantly, (ii) stop the production of comuna level direct estimates, and (iii) search for alternative data sources such as administrative records or develop a new data collection effort specifically designed for comuna level representation of social indicators of interest for various public policies.

- In 2010, the Ministry produced for the first time poverty rate estimates for all 345 comunas in Chile using both standard design-based and the Ministry-PNUD synthetic method.

The Poverty Measure Used in Chile

- In Chile, poverty is measured using the poverty rate, also known as Headcount Index, defined as the proportion of households with *income* below the *poverty threshold* or *poverty line*.

- The first ingredient of the poverty rate is the *poverty line*. For most Latin American countries, the poverty line is the cost of a basket of essential food and non-food items. This poverty line is expressed in per-capita terms. The methodology for estimating Chile´s poverty line was developed by the Comisión Económica para América Latina y el Caribe (CEPAL). Data from the Chilean expenditure survey *Encuesta de Presupuestos Familiares 1987-1988* was used to estimate the value of the food basket. Two different poverty lines were derived from the food basket ---- one for rural areas and the other for urban areas.

- The second ingredient of the poverty rate, the *per-capita income*, is the ratio of the *total household income* and the *household size*. Households whose per-capita income falls below the poverty line are considered in poverty. The poverty rate is then the percent of households in each region/comuna that are in poverty.

The Casen Survey

- Chile's official data source for poverty statistics is the National Socioeconomic Characterization Survey (Casen) - a survey sponsored every two or three years by the Ministry since 1987 with sample in most of the comunas.

- The Casen survey is a cross-sectional multipurpose household survey designed to understand the socioeconomic conditions of the population and the evaluation of social programs. The survey has been fielded regularly every two or three years since 1987.

- The 2009 Casen survey collected data from 246,924 persons in 71,460 households, representing a total of 16,607,007 persons living in private dwellings in Chile in November, 2009. The sampling design used was as follows:

  -The target population was defined to cover 334 out of the 345 comunas in the country.

  -Samples were drawn independently from 602 sampling strata formed by the comuna´s urban/rural subdivisions.

  -Using a two-stage sampling design, small geographic entities, known as *secciones*, were sampled at the first stage (Primary Sampling Units, PSUs) and housing units were sampled at the second stage (Secondary Sampling Units, SSUs) within each sampling strata.

  -The PSU´s were selected with probability proportional to size, measured in terms of the number of occupied housing units. A variable number of SSU´s were selected with equal probability using a systematic sampling algorithm with a random start within each selected PSU. Within each housing unit interviews were attempted with all households (*i.e.* no subsampling was implemented beyond the selection of the housing units).

Data Preparation

- Comuna level data derived from Casen 2009

  p_i : direct estimate of poverty rate for the *i*th comuna;

  $y_i = \sin^{-1}\sqrt{p_i}$; n_i : effective sample size

  $D_i = 1/(4n_i)$, an approximated sampling variance of y_i

- Comuna level administrative data

  -average wage for dependent workers

  -percentage of rural population

  -percentage of illiterate population

  -percentage of school attendance

  -the average of the comuna-level poverty rates from Casen 2000, 2003 and 2006

  -region-level indicators for the 7^{th}, 8^{th} and 9^{th} regions of the country

Description of SAE Method Implemented in Chile

Four Guidelines:

- method must use the Casen survey data directly to the extent possible since this is the largest data that collect information on most current poverty related variables
- poverty rate estimates should be close to the survey-weighted direct estimates for comunas with reasonably large samples
- method must not produce poverty rate estimates that considerably deviate from the corresponding direct survey estimates even for small comunas
- poverty count estimates, when aggregated over all the comunas in a given region, must produce the official survey-weighted count for that region.

Modeling

Level 1 (Sampling Model):

Given θ_i, y_i's are independent with $y_i \sim N(\theta_i, D_i)$;

Level 2 (Linking Model):

θ_i's are independent with $\theta_i \sim N(x_i{}'\beta, A)$,

- m is the number of comunas in Chile covered by Casen;
- $\theta_i = \sin^{-1}\sqrt{P_i}$; P_i is the true poverty rate;
- $x_i{}' = (x_{i0}, \cdots, x_{is-1})$ is a $s \times 1$ vector of s known fixed comuna specific auxiliary variables with $x_{i0} = 1$; $\beta = (\beta_0, \cdots, \beta_{s-1})$ is a $s \times 1$ column vector of unknown regression coefficients where β_0 denotes the intercept;
- A is the unknown model variance $(i = 1, \cdots, m)$.

Empirical Bayes Estimator of θ_i

Bayes estimator:

$$\hat{\theta}_i^B = (1 - B_i) y_i + B_i x_i{}' \beta, \text{ where } B_i = D_i \big/ (A + D_i).$$

An Empirical Bayes (EB) estimator of θ_i:

$$\hat{\theta}_i^{EB} = (1 - \hat{B}_i) y_i + \hat{B}_i x_i{}' \hat{\beta}, \text{ where } \hat{B}_i = D_i \Big/ \left(\hat{A} + D_i \right).$$

- The weight the EB estimator puts on the direct estimator y_i depends on the ratio $\hat{A} / D_i$.
- The choice of the adjusted maximum profile likelihood estimator of A over the usual residual maximum likelihood (REML) estimator was intentional and was used to assign more weight on the direct estimator since adjusted profile likelihood tends to have more upward bias than the REML.

- Since the adjusted maximum profile likelihood estimator is strictly positive, it avoids the common problem of the full shrinkage ($i.e., \hat{B}_i = 1$) that is often encountered with the REML-based empirical Bayes estimator of θ_i.
- In theory, EB estimates can go out of the admissible range$[0, \pi/2]$. Thus, $\hat{\theta}_i^{EB}$ is truncated to 0 if $\hat{\theta}_i^{EB}$ is negative and to $\pi/2$ if $\hat{\theta}_i^{EB}$ is greater than $\pi/2$.

Limited Translation Empirical Bayes Estimator of θ_i

$$\hat{\theta}_i^{LT} = \begin{cases} \hat{\theta}_i^{EB} & if \quad y_i - \sqrt{D_i} \le \hat{\theta}_i^{EB} \le y_i + \sqrt{D_i}, \\ y_i - \sqrt{D_i} & if \quad \hat{\theta}_i^{EB} \le y_i - \sqrt{D_i}, \\ y_i + \sqrt{D_i} & if \quad \hat{\theta}_i^{EB} \ge y_i + \sqrt{D_i}, \end{cases}$$

Back-transformation and raking

Back-transform: $\hat{P}_i = \sin^2 \hat{\theta}_i^{LT}$.

For a few comunas with no sample in the Casen 2009 survey, the estimates of the poverty rate were computed using the Ministry-PNUD synthetic method.

Whether a comuna is in the Casen sample or not, the final official raked SAE estimates of poverty rates for all the comunas that belong to the *r*th region are given by:

$$\hat{P}_i^{SAE} = \hat{P}_i \times R_r,$$

where

- $R_r = p_r^{regn} N_r^{regn} \Big/ \sum_{i=1}^{m_r^*} \hat{P}_i N_i$ is the raking factor common to all comunas in the region $r; m_r^*$ is the total number of comunas in region r; p_r^{regn} is the direct design-based estimate of the regional-level poverty rate using the original regional weights; N_i is an estimate of the population projection in comuna i belonging to region $r; N_r^{regn}$ is an estimate of the population projection in region $r;\ N_r^{regn} = \sum_{i=1}^{m_r^*} N_i$.

Confidence Intervals for the Poverty Rates

Step 1: Generate R independent parametric bootstrap samples$\{(y_i^{(r)},\theta_i^{(r)}),\ i=1,\cdots m\},\quad r=1,\cdots,R$ as follows: $\theta_i^{(r)} \sim N(x_i^T\hat{\beta},\hat{A}),\ y_i^{(r)} \mid \theta_i^{(r)} \sim N(\theta_i^{(r)},D_i),\ i=1,\ldots,m.$

Step 2: Produce estimates $\hat{A}^{(r)}$, $\hat{B}_i^{(r)}$ and $\hat{\beta}^{(r)}$ by replacing the original data with the parametric bootstrap samples generated in Step 1. We repeat this step R times.

Step 3: For each bootstrap simple, calculate the following pivotal quantity: $t_i^{(r)} = \left(\theta_i^{(r)} - \hat{\theta}_i^{EB(r)}\right) \Big/ \sqrt{D_i(1-\hat{B}_i^{(r)})}$, where $\hat{\theta}_i^{EB(r)} = (1-\hat{B}_i^{(r)})y_i^{(r)} + \hat{B}_i^{(r)}x_i{}'\hat{\beta}^{(r)}.$

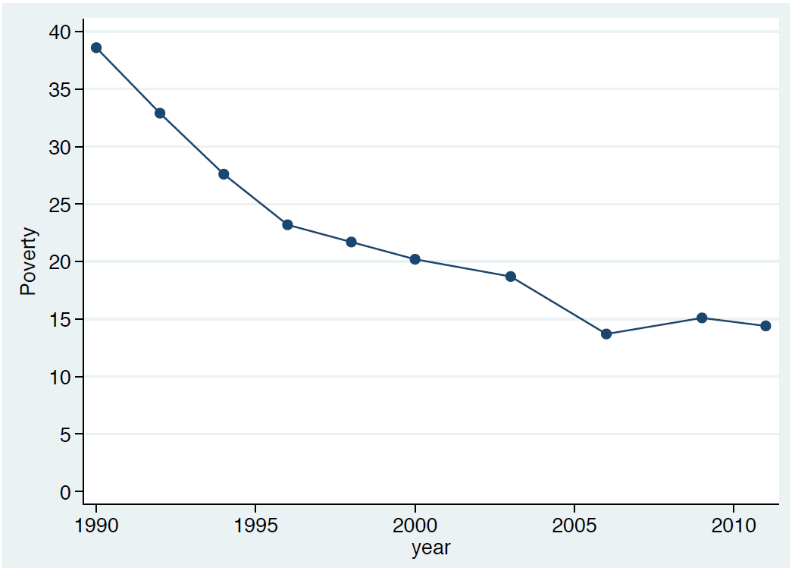
Step 4: For comuna i, obtain q_{1i} and q_{2i}, the $100\alpha/2$ and $100(1-\alpha/2)$ percentiles of $\{t_i^{(r)},\ r=1,\cdots,R\}$.

Step 5: For comuna i, an approximate $100(1-\alpha)\%$ confidence interval for θ_i is obtained as: (L_i, U_i), where $L_i = \hat{\theta}_i^{EB} + q_{1i}\sqrt{D_i(1-\hat{B}_i)}$ and $U_i = \hat{\theta}_i^{EB} + q_{2i}\sqrt{D_i(1-\hat{B}_i)}$. Note that the admissible range for θ_i is $[0, \pi/2]$. Thus, L_i is truncated to 0 if L_i is negative and U_i is truncated to $\pi/2$ if U_i is greater than $\pi/2$. The probability that (L_i, U_i) is not contained in $(0, \pi/2)$ is expected to be negligible unless $4n_i$ is very small. The truncated confidence interval for θ_i is denoted by (L_i^*, U_i^*).

Step 6: Finally, the lower and upper limits of the confidence interval (L_i^*, U_i^*) in Step 5 are back-transformed to yield the following approximate $100(1-\alpha)\%$ confidence interval of the poverty rate P_i : $\left(\sin^2 L_i^*,\ \sin^2 U_i^*\right)$. Note that the parametric bootstrap confidence interval for any one-to-one transformed parameter can be easily obtained using the simple back-transformation. In our case, the motivation for this back-transformed confidence interval comes from the fact that for any $0 < p < 1$ and $0 < \theta < \pi/2$, $\sin^{-1}\sqrt{p}$ and $\sin^2\theta$ are monotonically increasing functions of p and θ, respectively.

Appendix

Figure 21.1.
Estimates of national poverty rates in Chile, by year.



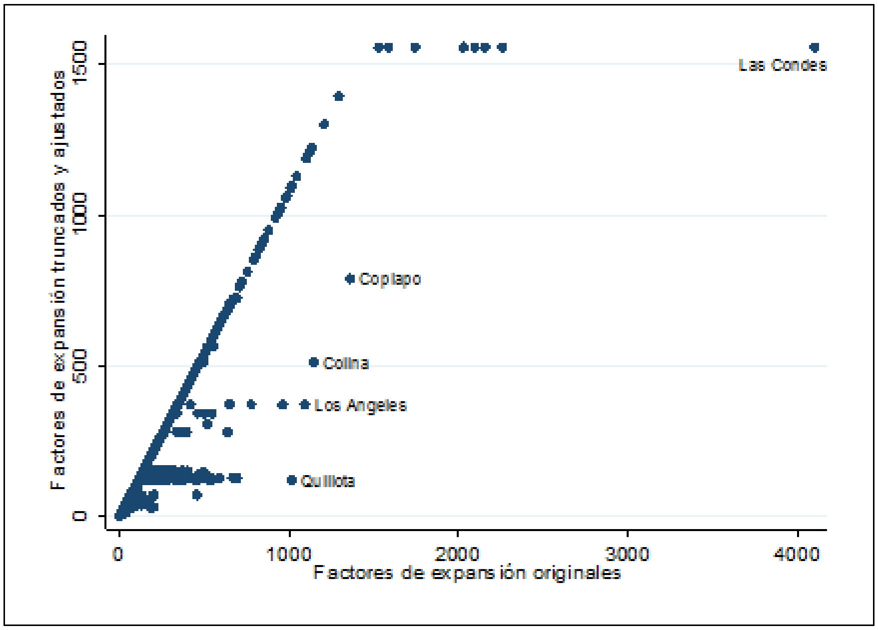
Source: Compiled by the authors based on Casen 1990, 1992, 1994, 1996, 1998, 2000, 2003, 2006, 2009 and 2011 data.

Figure 21.2.
Descriptive statistics for the original survey weights, cut-off point and total number of original comuna weights truncated, by region and zonal group. Casen 2009 data.

Truncation Groups	Descriptive statistics original comuna weights			Truncation point	Number of original comuna weights truncated
	Average	Minimum	Maximum		
1	87.4902	5	501	137.6	748
2	10.1405	2	34	63.0	0
3	94.3949	3	692	672.7	32
4	4.5907	1	27	32.5	0
5	59.8353	6	1.363	731.7	12
6	14.3449	4	47	83.4	0
7	95.7634	7	558	524.6	82
8	27.3082	4	134	127.6	58
9	74.0826	5	1.020	112.6	4,117
10	23.0577	3	100	75.7	105
11	53.7106	2	637	262.0	229
12	22.6640	2	110	87.7	171
13	69.0955	3	405	141.5	1,471
14	26.0215	4	160	49.8	1,760
15	66.7037	4	1,095	346.3	105
16	20.2520	4	203	30.6	2,929
17	60.0223	4	548	318.7	225
18	28.2342	6	183	37.2	2,152
19	70.5936	2	693	118.5	1,547
20	22.8171	3	461	67.9	673
21	37.9711	5	183	52.8	497
22	9.9334	3	34	11.8	328
23	85.7458	8	544	777.8	0
24	9.8642	2	41	14.1	67
25	147.5910	5	4,103	1,445.2	81
26	38.0404	2	1,147	476.6	18
27	53.8487	6	520	287.1	86
28	31.1182	6	237	135.4	54
29	106.1280	1	475	133.2	268
30	14.3313	1	57	103.1	0
Total	-	-	-	-	17,815

Source: Ministerio de Desarrollo Social [42].

Figure 21.3.
A plot of original weights (x-axis) and trimmed survey weights (y-axis) for all observations in Casen 2009.



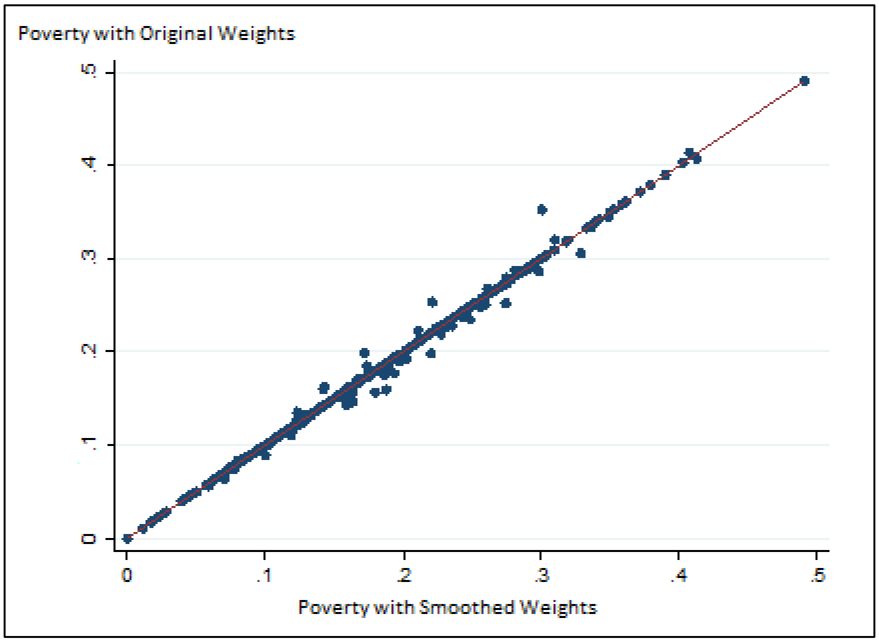
Source: Ministerio de Desarrollo Social [42].

Figure 21.4.
Descriptive statistics of number of cases at the Respondent level and the Household level. Casen 2009 data.

Quartiles of comunas respondent sample	Respondent level Sample			Household level Sample		
	Min	Mean	Max	Min	Mean	Max
1	53	491.0	610	20	152.7	198
2	612	654.9	692	155	195.7	239
3	693	752.2	864	177	214.1	265
4	873	1,064.3	1,608	211	294.4	409

Source: Compiled by authors based on Casen 2009 data.

Figure 21.5.
A plot of direct survey estimates with original survey weights (x-axis) and trimmed survey weights (y-axis) for comunas in Chile. Casen 2009 data.



Source: Ministerio de Desarrollo Social [42].

Figure 21.6.
Estimates of the design effect of the direct poverty rate in Chile using trimmed comuna weights, by region. Casen 2009 data.

N°	Region	Design Effect Estimates
1	Tarapacá	3.280
2	Antofagasta	5.750
3	Atacama	6.477
4	Coquimbo	4.665
5	Valparaíso	3.390
6	O´Higgins	4.307
7	Maule	4.870
8	Biobío	5.506
9	Araucanía	5.618
10	Los Lagos	6.095
11	Aysén	2.843
12	Magallanes	2.323
13	Metropolitana	3.290
14	Los Ríos	8.681
15	Arica y Parinacota	2.864

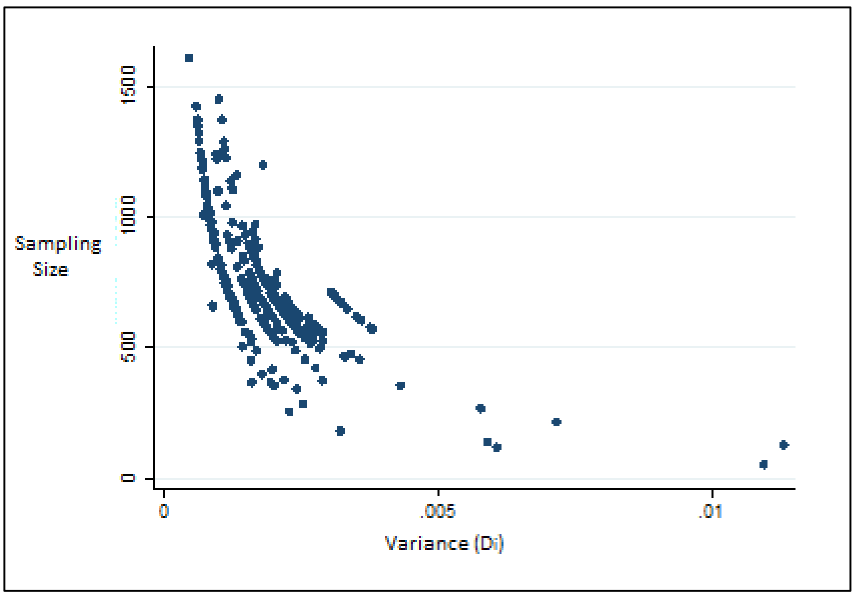
Source: Ministerio de Desarrollo Social [42].

Figure 21.7.
Descriptive statistics of D_i and B_i by groups of comunas formed using quartiles of the distribution of the sampling variances (D_i). Casen 2009 data.

Quartiles of D_i	Descriptive statistics of D_i			Descriptive statistics of B_i		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Group 1 (lowest 25% of D_i)	0.0009047	0.0004453	0.0012006	0.2765	0.1598	0.3390
Group 2	0.0014632	0.0012023	0.0016854	0.3836	0.3393	0.4186
Group 3	0.0019475	0.001686	0.0021987	0.4535	0.4186	0.4843
Group 4 (highest 25% of D_i)	0.0030616	0.0022082	0.0113181	0.5474	0.4854	0.8286
All	0.0018417	0.0004453	0.0113181	0.4150	0.1598	0.8286

Source: Ministerio de Desarrollo Social [42].

Figure 21.8.
Comuna-level sample size (y-axis) and comuna-level estimate of variance (x.-axis) of the direct estimate of the poverty rates. Casen 2009 data.



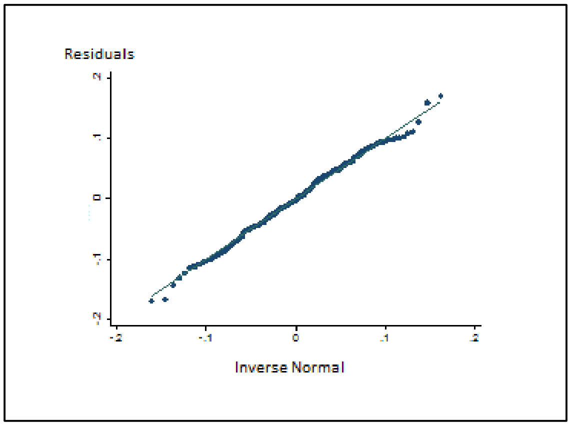
Source: Ministerio de Desarrollo Social [42].

Figure 21.9.
Initial set of auxiliary variables reviewed for their possible inclusion as comuna-level auxiliary variables in the area level model.

Number and Name of the auxiliary variable	Institution responsible for data collection	Frequency of publication of the data
#1. Subsidio Familiar	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#2. Subsidio al Pago del Consumo de Agua Potable y Servicio de Alcantarillado de Aguas Servidas	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#3. Bono Chile Solidario	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#4. Subsidio de Discapacidad Mental	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#5. Pensión Básica Solidaria (vejez e invalidez)	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	December
#6. Aporte Previsional Solidario (vejez e invalidez)	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	December
#7. Bonificación al Ingreso Ético Familiar	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#8. Beca de Apoyo a la Retención Escolar, BARE	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social.	monthly and yearly
#9. Afiliados Sistema de Capitalización Individual	Superintendencia de Pensiones	monthly and yearly
#10. Matrícula	Ministerio de Educación	Yearly
#11. Rendimiento	Ministerio de Educación	Yearly
#12. SIMCE	Ministerio de Educación	Yearly or every two years
#13. Titulados Educación Superior	Ministerio de Educación	Yearly
#14. Índice de Vulnerabilidad del Establecimiento (IVE-SINAE)	Junta Nacional Escolar y Becas (Junaeb)	Yearly
#15. Situación Nutricional estudiantes básica y media	Junta Nacional Escolar y Becas (Junaeb)	Yearly
#16. Población beneficiaria Fonasa	Ministerio de Salud	Yearly
#17. Atenciones sector privado	Ministerio de Salud	Yearly
#18. Razón de analfabetos respecto a la población de 10 y más años en la comuna	CENSO, INE	Every 10 years
#19. Porcentaje de Población Rural	CENSO, INE	Every 10 years
#20. Porcentaje de Asistencia Escolar Comunal	SINIM	monthly
#21. Tamaño promedio del hogar	CENSO, INE	Every 10 years
#22. Tasa de pobreza histórica	CASEN	Every 2 or 3 years
#23. Contribuciones de Vivienda	SII (http://www.sii.cl/avaluaciones/estadisticas/estadisticas_bbrr.htm#2)	Yearly
#24. Remuneraciones promedio de los trabajadores dependientes		Yearly

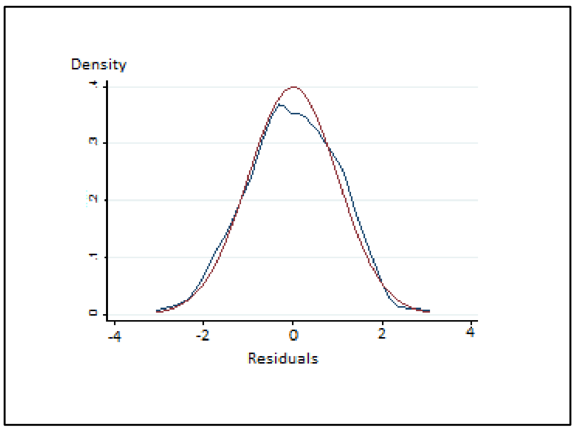
Source: Ministerio de Desarrollo Social [42].

Figure 21.10
QQ plot of the standardized residuals



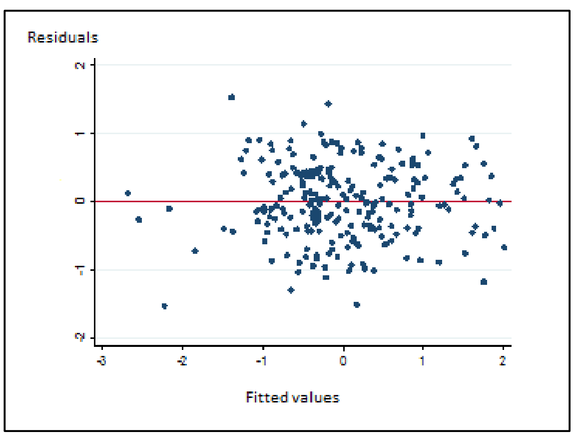
Source: Ministerio de Desarrollo Social [42].

Figure 21.11
Distribution of the standardized residuals (blue line)



Source: Ministerio de Desarrollo Social [42].

Figure 21.12
Plot of standardized residuals against fitted values



Source: Ministerio de Desarrollo Social [42].

Figure 21.13.
Estimates of Spearman correlation coefficients and p-values for the squared standardized residuals of the OLS regression model in Figure 21.14.

Auxiliary Variable	Sperarman Correlation	P-values
Average wage of dependent workers	-0.0144	0.8264
Average of the poverty rate from Casen 2000, 2003 and 2006	-0.0148	0.8214
% of population in rural areas	-0.0065	0.9214
% of illiterate population	-0.0092	0.8882
% of population attending school	0.0072	0.9126
Dummy for region 7	0.0337	0.607
Dummy for region 8	-0.095	0.1467
Dummy for region 9	-0.0061	0.9256

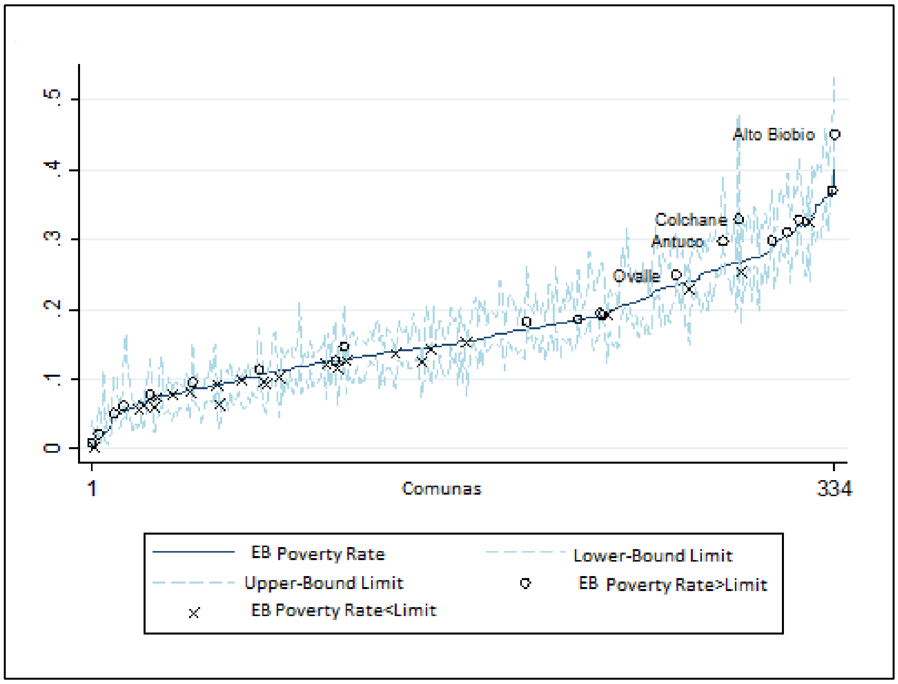
Source: Ministerio de Desarrollo Social [42].

Figure 21.14.
Output of regression analysis based on comunas with population more than 10.000 inhabitants (dependent variable: arcsine transformed direct survey estimate of the poverty rate with original and trimmed weights; independent variables: a set of variables used in the comuna level model with arcsine transformation for proportions and logarithmic transformation for the rest).

Independent variables	Regression coefficient estimate (t-statistics): original comuna weights	Regression coefficient estimate (t-statistics): trimmed comuna weights
Average wage of dependent workers (log)	-0.09575646 (3.52**)	-0.21927953 (3.52**)
Average of the poverty rate from Casen 2000, 2003 and 2006 (arcsin)	0.49548266 (7.92**)	0.48474029 (7.92**)
% of population in rural areas (arcsin)	-0.13409847 (4.96**)	-0.39252745 (4.96**)
% of illiterate population (arcsin)	0.40349163 (2.57*)	0.25176513 (2.57*)
% of population attending to school (arcsin)	-0.21883535 (2.23*)	-0.0938032 (2.23*)
Dummy for region 7 (=1)	0.03442978 (2.11*)	0.08671043 (2.11*)
Dummy for region 8 (=1)	0.03882056 (2.67**)	0.12474226 (2.67**)
Dummy for region 9 (=1)	0.105632 (6.04**)	0.28328927 (6.04**)
Constant	1.61477028 (4.24**)	-0.00203088 (0.06)
Number of observations	235	235
Adjusted R^2	0.67	0.67

Notes: * statistically significant at the 5% level; ** statistically significant at the 1% level.
Source: Ministerio de Desarrollo Social [42].

Figure 21.15
Limited translation empirical Bayes estimates of the comuna level poverty rates, and the upper and lower thresholds.



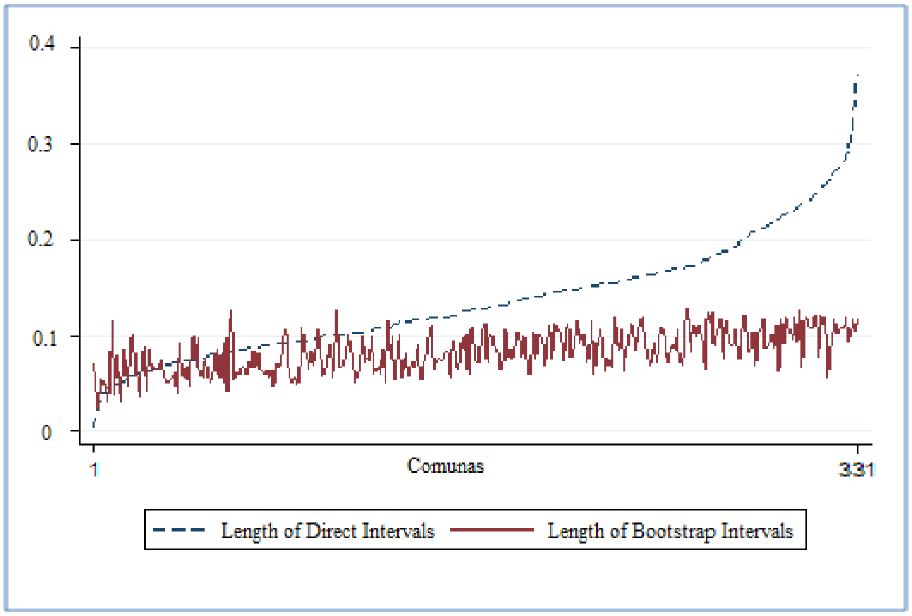
Source: Ministerio de Desarrollo Social [42].

Figure 21.16.
Raking factors used so that the total of the model-based estimates within each region matches the standard design-based official estimate for the region, by region.

N°	Region	R_r
1	Tarapacá	1.12172
2	Antofagasta	0.97455
3	Atacama	1.06685
4	Coquimbo	1.04309
5	Valparaíso	1.00387
6	O´Higgins	1.00430
7	Maule	1.05292
8	Biobío	0.99010
9	Araucanía	1.01628
10	Los Lagos	1.04088
11	Aysén	1.06255
12	Magallanes	0.97368
13	Metropolitana	0.97765
14	Los Ríos	1.08572
15	Arica y Parinacota	0.99486

Source: Ministerio de Desarrollo Social [42].

Figure 21.17.
Length of the direct and parametric bootstrap confidence intervals of the comuna-level poverty rates for comunas sorted by the limited translation empirical Bayes estimates of the poverty rate.



Note: The three comunas with the largest estimates of the length of the direct confidence interval were excluded from the graph. Source: Compiled by authors based on Casen 2009 data and Ministerio de Desarrollo Social [43].

Figure 21.18a.
Histograms of pivots in the parametric bootstrap method with 5,000 bootstrap samples.
Comuna of Puchucavi

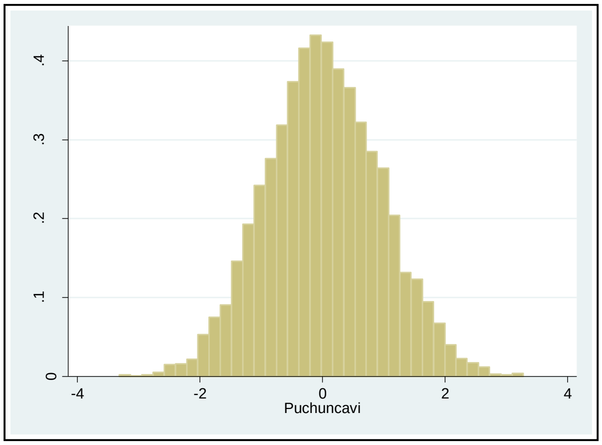


Figure 21.18b.
Histograms of pivots in the parametric bootstrap method with 5,000 bootstrap samples.
Comuna of Providencia

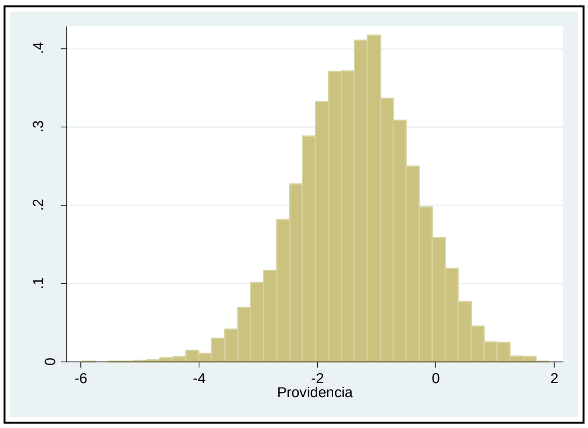
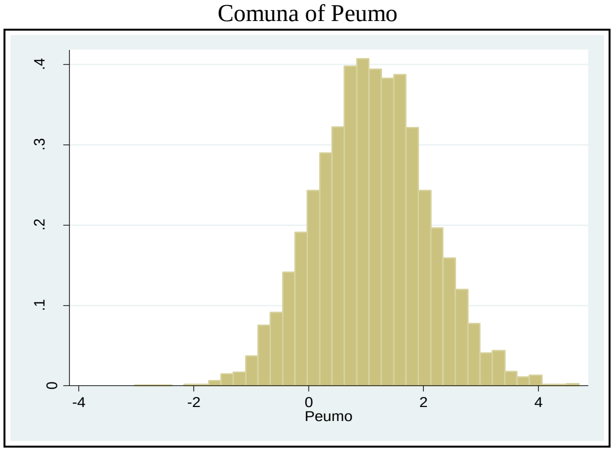


Figure 21.18c.
Histograms of pivots in the parametric bootstrap method with 5,000 bootstrap samples.
Comuna of Peumo



Source: Ministerio de Desarrollo Social [42].

References

1. Ley 18.695 Art.1. *Fija el Texto Refundido, Coordinado y Sistematizado de la Ley Nº 18.695, Orgánica Constitucional de Municipalidades*. Ministerio del Interior; Subsecretaría de Desarrollo Regional y Administrativo. Santiago, Chile, Publicación: 26-07-2006.
2. Ministerio de Desarrollo Social (2012). *Metodología del Diseño Muestral y Factores de Expansión Encuesta de Caracterización Socioeconómica Nacional*. Serie de Documentos Metodológicos N° 1, Observatorio Social.
3. United Nations Development Programme (UNDP) (1990). *Human Development Report 1990*. New York Oxford. Oxford University Press.
4. Programa de las Naciones Unidas para el Desarrollo (PNUD) Chile y Ministerio de Planificación de Chile (2000). *Desarrollo Humano en las comunas de Chile,* Santiago, Chile.
5. Ley 20.237 Art.1 (1). *Modifica el Decreto Ley Nº 3.063, de 1979, Sobre Rentas Municipales la Ley Nº 18.695, Orgánica Constitucional de Municipalidades, y Otros Cuerpos Legales, en Relación con el Fondo Común Municipal y Otras Materias Municipales*. Ministerio del Interior; Subsecretaría de Desarrollo Regional y Administrativo. Santiago, Chile, Publicación: 17-12-2007.
6. Decreto 1293 Art. 13(f). *Reglamento para la Aplicación Del Artículo 38 Del Decreto Ley N° 3.063, De 1979, Modificado por el Artículo 1° de la Ley N° 20.237*. Ministerio del Interior; Subsecretaría de Desarrollo Regional y Administrativo. Santiago, Chile, Publicación: 02-01-2009.
7. Comisión de Técnicos de la Encuesta Casen (2010). *Informe de la Comisión de Técnicos sobre la Encuesta Casen*. Santiago, Chile.
8. United Nations Expert Group on Poverty Statistics (Rio Group) (2006). *Compendium of Best Practices in Poverty Measurement*. United Nations Economic Commission for Latin America and the Caribbean (ECLAC) and Brazilian Institute for Geography and Statistics (IBGE). Rio de Janeiro, September 2006. ISBN 85-240-3908-6.
9. Comisión para la Medición de la Pobreza (2014). *Informe Final de la Comisión para la Medición de la Pobreza*. Santiago, Chile.
10. Comisión Económica para América Latina y el Caribe (CEPAL) (1990). *Una estimación de la medición de pobreza para Chile, 1987*, Santiago de Chile.
11. Ministerio de Planificación (2010). *Informe Metodológico Casen 2009*. División Social, Santiago, Chile.
12. Potter F.J. (1993). *The effect of weight trimming on nonlinear survey estimates*, San Francisco, CA: American Statistical Association.
13. Bell, W. (1997). *Models for county and state poverty estimates*. Preprint, Statistical Research Division, U. S. Census Bureau.
14. Bell, W., Basel, W., Cruse, C, Dalzell, L., Maples, J., OHara, B., and Powers, D. (2007). *Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties*, Census Report.
15. National Research Council (2000). *Small-area estimates of school-age children in poverty: Evaluation of Current Methodology*. Citro, C. and Kalton, G. (Eds.), National Academy Press, Washington DC.
16. Efron, B. and Morris, C. (1975). *Data analysis using Stein's estimator and its generalizations*, Journal of the American Statistical Association 70, 311-319.
17. Carter, G. and Rolph, J. (1974). *Empirical Bayes methods applied to estimating fire alarm probabilities*, Journal of the American Statistical Association 69, 880-885.
18. Jiang, J., Lahiri, P., Wan, S., and Wu, C. (2001). *Jackknifying in the Fay-Herriot Model with an example*, Proceedings of the Seminar on Funding Opportunity in Survey Research, Council of Professional Associations on Federal Statistics.
19. Raghunathan, T.E., Xie, D., Schenker, N., Parsons, V.L., Davis, W.W., Dodd, K.W., and Feuer, E.J. (2007). *Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening*. Journal of the American Statistical Association 102(478): 474.
20. Xie, D., Raghunathan, T.E., and Lepkowski, J.M. (2007). *Estimation of the proportion of overweight individuals in small areas - a robust extension of the Fay-Herriot model*. Statistics in Medicine 26(13): 2699-2715.
21. StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.
22. Box, G.E.P. (1979). *Robustness in the Strategy of Scientific Model Building* in Robustness in Statistics: Proceedings of a Workshop (1979) edited by RL Launer and GN Wilkinson.

23. Elbers, C., Lanjouw, J., and Lanjouw, P. (2003). *Micro-Level Estimation of Poverty and Inequality*, Econometrica 71:1 (2003), 355–364.
24. Elbers, C., Lanjouw, P., and Leite, P. (2008). *Brazil within Brazil: Testing the Poverty Map Methodology in Minas Gerais*, World Bank policy research working paper no. 4513.
25. Molina, I. and Rao, J.N.K. (2010). *Small area estimation of poverty indicators*, Canadian Journal of Statistics, 38, 369-385.
26. Agostini, C., Brown, P., and Gongora, D. (2008). *Distribución Espacial De La Pobreza En Chile*. Estudios De Economía, 35, N.1, 79-110.
27. Agostini, C., Brown, P., and Roman, A. (2010). *Estimando Indigencia y Pobreza Indígena Regional con Datos Censales y Encuestas de Hogares*. Cuadernos de Economía, 47(135), 125-150.
28. Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New York.
29. Pfeffermann, D. (2013). *New Important Developments in Small Area Estimation*, Statistical Science, 28, 40-68.
30. Fay, R.E. and Herriot, R.A. (1979). *Estimates of income for small places: An application of James_Stein procedure to census data*, Journal of the American Statistical Association 74, 269_277.
31. Li, H. and Lahiri, P. (2010). *An adjusted maximum likelihood method for solving small area estimation problems*. J. Multivariate Anal. 101 882–892. MR2584906
32. Efron, B. and Morris, C. (1972). *Limiting the risk of Bayes and empirical Bayes estimators-Part II: the empirical Bayes case.* Journal of the American Statistical Association 67: 130-139.
33. Chatterjee, S., Lahiri, P., and Li, H. (2008). *Parametric bootstrap approximation to the distribution of EBLUP, and related prediction intervals in linear mixed models*, The Annals of Statistics 36, 1221-1245.
34. Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval Estimation for a Binomial Proportion, Statistical Science, 16, 101-133.
35. Ha, N. and Lahiri, P. (2014a). *Hierarchical Bayesian Estimation of Small Areas Proportions Using Complex Survey Data*, unpublished manuscript.
36. Pfeffermann, D., Sikov, A., and Tiller, R. (2014). Single and two-stage cross-sectional and time series benchmarking procedures for small area estimation, *Test*, DOI: 10.1007/s11749-014-0400-8
37. Ha, N. and Lahiri, P. (2014b). *Comment: Single and two-stage cross-sectional and time series benchmarking procedures for small area estimation*, *Test*, DOI: 10.1007/s11749-014-0400-8
38. Chatterjee, S. and Lahiri, P. (2013), *A Simple Computational Method for Estimating Mean Squared Prediction Error in General Small-Area Model*, paper presented at the SAE 2013, Bangkok.
39. Alkire, S. and Foster, J. (2007). *Counting and Multidimensional Poverty Measurement*. OPHI Working Paper Series, N°7.5, OPHI.
40. Statistics Netherlands (2008). *Model-Based Estimation for Official Statistics*, Discussion paper (08002), The Hage, Netherlands.
41. Heady, P. and Ralphs, M. (2005). *EURAREA: an overview of the project and its findings*. Statistics in Transition 7, 557—570.
42. Ministerio de Desarrollo Social (2013a). *Procedimiento de cálculo de la Tasa de Pobreza a nivel Comunal mediante la aplicación de Metodología de Estimación para Áreas Pequeñas (SAE)*. Serie de Documentos Metodológicos N° 1, Observatorio Social.
43. Ministerio de Desarrollo Social (2013b). *Incidencia de la Pobreza a nivel Comunal, según Metodología de Estimación para Áreas Pequeñas. Chile 2009 y 2011*. Serie de Informes Comunales N° 1, Observatorio Social.

THANK YOU!