Estimación en áreas pequeñas

Métodos indirectos basados en modelos: Método jerárquico Bayes y método basado en modelos generalizados lineales mixtos

Andrés Gutiérrez
Comisión Económica para América Latina y el Caribe

2018

References

References

(2018) Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.

References

(2018) Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.

References

- *(2018)* Molina, Isabel. *Estudio de los límites de desagregación de dates en encuestas de hogares para subgrupos de población y áreas geográficas y los requerimientos para superarlos: Fase II*. CEPAL.
- *(2015)* Rao, J.N.K y Isabel Molina. *Small Area Estimation*. Second ed. Wiley Series in Survey Methodology.

Introducción

- De nuevo, estimadores para áreas basados en modelos se consideran modelos indirectos porque usan información de otras áreas

Introducción

- De nuevo, estimadores para áreas basados en modelos se consideran modelos indirectos porque usan información de otras áreas
- Estimadores basados en modelos incorporan la heterogeneidad que no puede ser explicada por las variables auxiliares coleccionadas

Introducción

- De nuevo, estimadores para áreas basados en modelos se consideran modelos indirectos porque usan información de otras áreas
- Estimadores basados en modelos incorporan la heterogeneidad que no puede ser explicada por las variables auxiliares coleccionadas
- Esto se realiza incorporando efectos aleatorios de las áreas en los modelos de interés

Método jerárquico Bayes bajo el modelo BHF

Método jeráquico Bayes bajo el modelo BHF

- Molina, Nandrum, y Rao (2014) propusieron el método jerárquico Bayes (*hierarchical Bayes*, HB), que no requiere el uso del bootstrap

Método jeráquico Bayes bajo el modelo BHF

- Molina, Nandrum, y Rao (2014) propusieron el método jerárquico Bayes (*hierarchical Bayes*, HB), que no requiere el uso del bootstrap
- El método reparametriza el modelo de errores anidados en términos del coeficiente de correlación intraclase $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$

Método jeráquico Bayes bajo el modelo BHF

- Molina, Nandrum, y Rao (2014) propusieron el método jerárquico Bayes (*hierarchical Bayes*, HB), que no requiere el uso del bootstrap
- El método reparametriza el modelo de errores anidados en términos del coeficiente de correlación intraclase $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$
- Considera distribuciones *a priori* para los parámetros $(\boldsymbol{\beta}, \rho, \sigma_e^2)$

Método jeráquico Bayes bajo el modelo BHF

El modelo HB viene dado por:

1)

$$Y_{di}|u_d, \beta, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}'_{di}\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,$$

Método jeráquico Bayes bajo el modelo BHF

El modelo HB viene dado por:

1)

$$Y_{di}|u_d, \beta, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}'_{di}\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,$$

2)

$$u_d|\rho, \sigma_e^2 \overset{iid}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma_e^2\right), \quad d = 1, \ldots, D,$$

Método jeráquico Bayes bajo el modelo BHF

El modelo HB viene dado por:

1)

$$Y_{di}|u_d, \boldsymbol{\beta}, \sigma_e^2 \overset{ind}{\sim} N(\mathbf{x}'_{di}\boldsymbol{\beta} + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,$$

2)

$$u_d|\rho, \sigma_e^2 \overset{iid}{\sim} N\left(0, \frac{\rho}{1-\rho}\sigma_e^2\right), \quad d = 1, \ldots, D,$$

3)

$$\pi(\boldsymbol{\beta}, \rho, \sigma_e^2) \propto \frac{1}{\sigma_e^2}, \quad \epsilon \leq \rho \leq 1-\epsilon,\ \sigma_e^2 > 0, \boldsymbol{\beta} \in \mathbb{R}^p$$

donde $\epsilon > 0$ refleja la falta de información

Método jeráquico Bayes bajo el modelo BHF

- La distribución a priori de los parámetros del modelo se puede calcular en función de las distribuciones condicionadas usando la regla de cadena

Método jeráquico Bayes bajo el modelo BHF

- La distribución a priori de los parámetros del modelo se puede calcular en función de las distribuciones condicionadas usando la regla de cadena
- La densidad conjunta del parámetros $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ viene dada por

Método jeráquico Bayes bajo el modelo BHF

- La distribución a priori de los parámetros del modelo se puede calcular en función de las distribuciones condicionadas usando la regla de cadena
- La densidad conjunta del parámetros $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ viene dada por
- $\pi(\mathbf{u}, \boldsymbol{\beta}, \sigma_e^2, \rho|\mathbf{y}_s) =$
$\pi_1(\mathbf{u}|\boldsymbol{\beta}, \sigma_e^2, \rho, \mathbf{y}_s)\pi_2(\boldsymbol{\beta}|\sigma_e^2, \rho, \mathbf{y}_s)\pi_3(\sigma_e^2|\rho, \mathbf{y}_s)\pi_4(\rho|\mathbf{y}_s)$

Método jeráquico Bayes bajo el modelo BHF

- La distribución a priori de los parámetros del modelo se puede calcular en función de las distribuciones condicionadas usando la regla de cadena
- La densidad conjunta del parámetros $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ viene dada por
- $\pi(\mathbf{u}, \boldsymbol{\beta}, \sigma_e^2, \rho|\mathbf{y}_s) =$
  $\pi_1(\mathbf{u}|\boldsymbol{\beta}, \sigma_e^2, \rho, \mathbf{y}_s)\pi_2(\boldsymbol{\beta}|\sigma_e^2, \rho, \mathbf{y}_s)\pi_3(\sigma_e^2|\rho, \mathbf{y}_s)\pi_4(\rho|\mathbf{y}_s)$
- Todas las distribuciones tienen forma conocida excepto π_4, y generamos esos valores usando un método de rejilla

Método jeráquico Bayes bajo el modelo BHF

- Así, se pueden generar muestras de $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ directamente de la distribución a posteriori

Método jeráquico Bayes bajo el modelo BHF

- Así, se pueden generar muestras de $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ directamente de la distribución a posteriori
- Dado $\boldsymbol{\theta}$, las variables Y_{di} para todos los individuos verifican

$$Y_{di}|\boldsymbol{\theta} \overset{ind}{\sim} N(\mathbf{x}_{di}'\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D$$

Método jeráquico Bayes bajo el modelo BHF

- Así, se pueden generar muestras de $\boldsymbol{\theta} = (\mathbf{u}', \beta', \sigma_e^2, \rho)'$ directamente de la distribución a posteriori
- Dado $\boldsymbol{\theta}$, las variables Y_{di} para todos los individuos verifican

$$Y_{di}|\boldsymbol{\theta} \stackrel{ind}{\sim} N(\mathbf{x}_{di}'\beta + u_d, \sigma_e^2 k_{di}^2), \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D$$

- La densidad predictiva de $\mathbf{y}_d r$ viene dada por

$$f(\mathbf{y}_{dr}|\mathbf{y}_s) = \int \prod_{i \in r_d} f(Y_{di}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{y}_s) d\boldsymbol{\theta},$$

Método jeráquico Bayes bajo el modelo BHF

- Finalmente, el estimador HB viene dado por

$$\hat{\delta}_d^{HB} = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_s) = \int \delta_d(\mathbf{y}_d) f(\mathbf{y}_{dr}|\mathbf{y}_s) d\mathbf{y}_{dr}$$

lo que estimamos usando una simulación Monte Carlo

Método jeráquico Bayes bajo el modelo BHF

- Finalmente, el estimador HB viene dado por

$$\hat{\delta}_d^{HB} = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_s) = \int \delta_d(\mathbf{y}_d) f(\mathbf{y}_{dr}|\mathbf{y}_s) d\mathbf{y}_{dr}$$

  lo que estimamos usando una simulación Monte Carlo
- Generamos muestras de la distribución a posteriori $\pi(\boldsymbol{\theta}|\mathbf{y}_s)$

Método jeráquico Bayes bajo el modelo BHF: proceso Monte Carlo

- Primero, generamos un valor $\rho^{(a)}$ de $\pi_4(\rho|\mathbf{y}_s)$ con un método de Rejilla

Método jeráquico Bayes bajo el modelo BHF: proceso Monte Carlo

- Primero, generamos un valor $\rho^{(a)}$ de $\pi_4(\rho|\mathbf{y}_s)$ con un método de Rejilla
- Después, generamos $\sigma_e^{2(a)}$ de $\pi_3(\sigma_e^2|\rho^{(a)}, \mathbf{y}_s)$, $\boldsymbol{\beta}^{(a)}$ de $\pi_2(\boldsymbol{\beta}|\sigma_e^{2(a)}, \rho^{(a)}, \mathbf{y}_s)$ y $\mathbf{u}^{(a)}$ de $\pi_1(\mathbf{u}|\boldsymbol{\beta}^{(a)}, \sigma_e^{2(a)}, \rho^{(a)}, \mathbf{y}_s)$

Método jeráquico Bayes bajo el modelo BHF: proceso Monte Carlo

- Primero, generamos un valor $\rho^{(a)}$ de $\pi_4(\rho|\mathbf{y}_s)$ con un método de Rejilla
- Después, generamos $\sigma_e^{2(a)}$ de $\pi_3(\sigma_e^2|\rho^{(a)}, \mathbf{y}_s)$, $\boldsymbol{\beta}^{(a)}$ de $\pi_2(\boldsymbol{\beta}|\sigma_e^{2(a)}, \rho^{(a)}, \mathbf{y}_s)$ y $\mathbf{u}^{(a)}$ de $\pi_1(\mathbf{u}|\boldsymbol{\beta}^{(a)}, \sigma_e^{2(a)}, \rho^{(a)}, \mathbf{y}_s)$
- Para cada de los A valores del vector $\boldsymbol{\theta}$, generamos los valores de los individuos afuera de la muestra $\mathbf{y}_{dr}^{(a)}$, y creamos el vector censal $\mathbf{y}_d^{(a)} = (\mathbf{y}'_{ds}, (\mathbf{y}_{dr}^{(a)})')'$

Método jeráquico Bayes bajo el modelo BHF:

- Para cada vector censal, producimos $\delta_d^{(a)} = \delta_d(\mathbf{y}_d^{(a)})$

Método jeráquico Bayes bajo el modelo BHF:

- Para cada vector censal, producimos $\delta_d^{(a)} = \delta_d(\mathbf{y}_d^{(a)})$
- Se aproxima El estimador HB por

$$\hat{\delta}_d^{HB} = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_s) \approx \frac{1}{A}\sum_{a=1}^{A}\delta_d^{(a)}$$

Método jeráquico Bayes bajo el modelo BHF:

- Para cada vector censal, producimos $\delta_d^{(a)} = \delta_d(\mathbf{y}_d^{(a)})$
- Se aproxima El estimador HB por

$$\hat{\delta}_d^{HB} = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_s) \approx \frac{1}{A}\sum_{a=1}^{A}\delta_d^{(a)}$$

- La varianza se aproxima por

$$V(\delta_d|\mathbf{y}_s) \approx \frac{1}{A}\sum_{a=1}^{A}\left(\delta_d^{(a)} - \hat{\delta}_d^{HB}\right)^2$$

Método jeráquico Bayes bajo el modelo BHF:

- Para cada vector censal, producimos $\delta_d^{(a)} = \delta_d(\mathbf{y}_d^{(a)})$
- Se aproxima El estimador HB por

$$\hat{\delta}_d^{HB} = E_{\mathbf{y}_{dr}}(\delta_d|\mathbf{y}_s) \approx \frac{1}{A}\sum_{a=1}^{A}\delta_d^{(a)}$$

- La varianza se aproxima por

$$V(\delta_d|\mathbf{y}_s) \approx \frac{1}{A}\sum_{a=1}^{A}\left(\delta_d^{(a)} - \hat{\delta}_d^{HB}\right)^2$$

- Para un indicador FGT, el estimador HB se aproxima en la forma:

$$\hat{F}_{\alpha d}^{HB} \approx \frac{1}{A}\sum_{a=1}^{A}F_{\alpha d}^{(a)}$$

Resumen del estimador del método HB:

- Indicadores objetivos: Parámetros generales

Resumen del estimador del método HB:

- Indicadores objetivos: Parámetros generales
- Requerimientos de datos:

Resumen del estimador del método HB:

- Indicadores objetivos: Parámetros generales
- Requerimientos de datos:
  - Microdatos de las p variables auxiliares de la misma de la variable de interés

Resumen del estimador del método HB:

- Indicadores objetivos: Parámetros generales
- Requerimientos de datos:
  - Microdatos de las p variables auxiliares de la misma de la variable de interés
  - Área de interés obtenida de la misma encuesta

Resumen del estimador del método HB:

- Indicadores objetivos: Parámetros generales
- Requerimientos de datos:
    - Microdatos de las p variables auxiliares de la misma de la variable de interés
    - Área de interés obtenida de la misma encuesta
    - Microdatos de las p covariables a partir de un censo o registro administrativo

Resumen del estimador del método HB:

- Ventajas:

Resumen del estimador del método HB:

- Ventajas:
  - Basado en datos a nivel de individuo, lo que provee información más detallada

Resumen del estimador del método HB:

- Ventajas:
  - Basado en datos a nivel de individuo, lo que provee información más detallada
  - Se puede estimar cualquier indicador que es una función de la variable Y_{di}

Resumen del estimador del método HB:

- Ventajas:
  - Basado en datos a nivel de individuo, lo que provee información más detallada
  - Se puede estimar cualquier indicador que es una función de la variable Y_{di}
  - Es insesgado bajo el modelo si los parámetros son conocidos

Resumen del estimador del método HB:

- Ventajas:
  - Basado en datos a nivel de individuo, lo que provee información más detallada
  - Se puede estimar cualquier indicador que es una función de la variable Y_{di}
  - Es insesgado bajo el modelo si los parámetros son conocidos
  - Minimiza la varianza a posteriori

Resumen del estimador del método HB:

- Ventajas:

Resumen del estimador del método HB:

- Ventajas:
  - Resulta prácticamente igual al estimador EB

Resumen del estimador del método HB:

- Ventajas:
  - Resulta prácticamente igual al estimador EB
  - Una vez se ajusta el modelo, se puede estimar en subáreas sin reajustar el modelo

Resumen del estimador del método HB:

- Ventajas:
  - Resulta prácticamente igual al estimador EB
  - Una vez se ajusta el modelo, se puede estimar en subáreas sin reajustar el modelo
  - Una vez se ajusta el modelo, se puede estimar cualquier indicador sin reajustar el modelo

Resumen del estimador del método HB:

- Ventajas:

Resumen del estimador del método HB:

- Ventajas:
    - No se usa el procedimiento *Markov Chain Monte Carlo*, MCMC, al contrario de muchos procesos bayesianos

Resumen del estimador del método HB:

- Ventajas:
  - No se usa el procedimiento *Markov Chain Monte Carlo*, MCMC, al contrario de muchos procesos bayesianos
  - No requiere el uso de métodos bootstrap para estimar ECM, lo que disminuye el tiempo computacional

Resumen del estimador del método HB:

- Ventajas:
  - No se usa el procedimiento *Markov Chain Monte Carlo*, MCMC, al contrario de muchos procesos bayesianos
  - No requiere el uso de métodos bootstrap para estimar ECM, lo que disminuye el tiempo computacional
  - El cálculo de intervalos creíbles o cualquier otro resumen de la distribución es automático

Resumen del estimador del método HB:

- Desventajas:

Resumen del estimador del método HB:

- Desventajas:
  - Es basado en un modelo, por tanto es necesario comprobar dicho modelo

Resumen del estimador del método HB:

- Desventajas:
  - Es basado en un modelo, por tanto es necesario comprobar dicho modelo
  - No tiene en cuenta el diseño muestral

Resumen del estimador del método HB:

- Desventajas:
  - Es basado en un modelo, por tanto es necesario comprobar dicho modelo
  - No tiene en cuenta el diseño muestral
  - Puede ser seriamente afectado por atípicos aislados

Resumen del estimador del método HB:

- Desventajas:
  - Es basado en un modelo, por tanto es necesario comprobar dicho modelo
  - No tiene en cuenta el diseño muestral
  - Puede ser seriamente afectado por atípicos aislados
  - El método no se puede extender a un modelo más complejos sin perder algunas ventajas mencionadas, como el evitar los métodos MCMC por ejemplo

Métodos basados en modelos lineales generalizados mixtos

Métodos basados en modelos lineales generalizados mixtos

- Los modelos mixtos hasta ahora no dan predicciones entre $[0, 1]$

Métodos basados en modelos lineales generalizados mixtos

- Los modelos mixtos hasta ahora no dan predicciones entre $[0, 1]$
- Para estimar proporciones, sería útil usar un modelo que proporciona valores en ese rango

Métodos basados en modelos lineales generalizados mixtos

- Los modelos mixtos hasta ahora no dan predicciones entre $[0, 1]$
- Para estimar proporciones, sería útil usar un modelo que proporciona valores en ese rango
- Esto incluye la incidencia de pobreza F_{0d}, pero no la brecha de pobreza, F_{1d}

Métodos basados en modelos lineales generalizados mixtos

- Los modelos mixtos hasta ahora no dan predicciones entre $[0, 1]$
- Para estimar proporciones, sería útil usar un modelo que proporciona valores en ese rango
- Esto incluye la incidencia de pobreza F_{0d}, pero no la brecha de pobreza, F_{1d}
- Para hacer eso, es habitual usar un modelo lineal generalizado mixto (*generalized linear mixed models*, GLMM)

Métodos basados en modelos lineales generalizados mixtos

- Asumimos que

$$Y_{di}|v_d \sim \text{Bern}(p_{di}),\ g(p_{di}) = \mathbf{x}'_{di}\boldsymbol{\alpha} + v_d$$

y

$$v_d \overset{iid}{\sim} N(0, \sigma^2_v), \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D$$

Métodos basados en modelos lineales generalizados mixtos

- Asumimos que

$$Y_{di}|v_d \sim \text{Bern}(p_{di}),\ g(p_{di}) = \mathbf{x}'_{di}\boldsymbol{\alpha} + v_d$$

y

$$v_d \overset{iid}{\sim} N(0, \sigma_v^2), \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D$$

- v_d es el efecto de área d y $\boldsymbol{\alpha}$ es el vector de coeficientes de la regresión

Métodos basados en modelos lineales generalizados mixtos

- Asumimos que

$$Y_{di}|v_d \sim \text{Bern}(p_{di}),\ g(p_{di}) = \mathbf{x}'_{di}\boldsymbol{\alpha} + v_d$$

y

$$v_d \overset{iid}{\sim} N(0, \sigma_v^2), \quad i = 1, \ldots, N_d,\ d = 1, \ldots, D$$

- v_d es el efecto de área d y $\boldsymbol{\alpha}$ es el vector de coeficientes de la regresión
- $g : (0, 1) \to \mathbb{R}$ es la función link (biyectiva, con derivada continua)

Métodos basados en modelos lineales generalizados mixtos

- Asumimos que

$$Y_{di}|v_d \sim \mathrm{Bern}(p_{di}),\ g(p_{di}) = \mathbf{x}'_{di}\boldsymbol{\alpha} + v_d$$

y

$$v_d \overset{iid}{\sim} N(0,\sigma_v^2), \quad i = 1,\ldots,N_d,\ d = 1,\ldots,D$$

- v_d es el efecto de área d y $\boldsymbol{\alpha}$ es el vector de coeficientes de la regresión
- $g:(0,1) \to \mathbb{R}$ es la función link (biyectiva, con derivada continua)
- El link logístico, $g(p) = \log(p/(1-p))$, el más utilizado

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- El mejor predictor, el que minimiza el ECM bajo el modelo, viene dado por

$$\tilde{P}_d^B(\boldsymbol{\theta}) = E(P_d|\mathbf{y}_{ds};\boldsymbol{\theta}) = \frac{1}{N_d}\left\{\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})\right\}$$

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- El mejor predictor, el que minimiza el ECM bajo el modelo, viene dado por

$$\tilde{P}_d^B(\boldsymbol{\theta}) = E(P_d|\mathbf{y}_{ds};\boldsymbol{\theta}) = \frac{1}{N_d}\left\{\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})\right\}$$

- En la práctica, obtenemos el predictor EB reemplazando θ por estimaciones consistentes, es decir

$$\hat{P}_d^{EB} = \tilde{P}_d^B(\hat{\boldsymbol{\theta}})$$

donde $\hat{\boldsymbol{\theta}}$ se encuentra ajustando el modelo GLMM a los datos muestrales $\mathbf{y}_s = (\mathbf{y}'_{1s}, \ldots, \mathbf{y}'_{Ds})'$

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- Una manera de estimar $E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})$ sería utilizar el Teorema de Bayes y que las variables Y_{di} son independientes dado v_d

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- Una manera de estimar $E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})$ sería utilizar el Teorema de Bayes y que las variables Y_{di} son independientes dado v_d
- En este caso,

$$E(Y_{di}|\mathbf{y}_{ds};\hat{\boldsymbol{\theta}}) = \frac{E\{h(\mathbf{x}'_{di}\boldsymbol{\alpha} + v_d)f(\mathbf{y}_{ds}|v_d);\hat{\boldsymbol{\theta}}\}}{E\{f(\mathbf{y}_{ds}|v_d);\hat{\boldsymbol{\theta}}\}}, \quad i \in r_d,$$

donde $h = g^{-1}$ es el link inverso
$h(\mathbf{x}'_{di}\boldsymbol{\alpha} + v_d) = \exp(\mathbf{x}'_{di}\boldsymbol{\alpha} + v_d)/\left\{1 + \exp(\mathbf{x}'_{di}\boldsymbol{\alpha} + v_d)\right\}$ y

$$f(\mathbf{y}_{ds}|v_d) = \prod_{i\in s_d} p_{di}^{Y_{di}}(1 - p_{di})^{(1-Y_{di})}$$

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- Podemos usar un proceso Monte Carlo para generar $v_d^{(r)} \sim N(0, \hat{\sigma}_v^2)$, $r = 1, \ldots, R$

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- Podemos usar un proceso Monte Carlo para generar $v_d^{(r)} \sim N(0, \hat{\sigma}_v^2)$, $r = 1, \ldots, R$
- Después, calculamos

$$E(Y_{di}|\mathbf{y}_{ds}; \hat{\boldsymbol{\theta}}) \approx \frac{R^{-1} \sum_{r=1}^{R} h(\mathbf{x}'_{di}\hat{\boldsymbol{\alpha}} + v_d^{(r)}) \hat{f}(\mathbf{y}_{ds}|v_d^{(r)})}{R^{-1} \sum_{r=1}^{R} \hat{f}(\mathbf{y}_{ds}|v_d^{(r)})}, \quad i \in r_d,$$

donde $\hat{f}$ es la distribución condicionada $f(\mathbf{y}_{ds}|v_d)$ en $\hat{\boldsymbol{\theta}}$

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- $$\tilde{P}_d^B(\boldsymbol{\theta}) = E(P_d|\mathbf{y}_{ds};\boldsymbol{\theta}) = \frac{1}{N_d}\left\{\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})\right\}$$

  tiene ECM mínimo y es insesgado bajo el modelo linear generalizado mixto

Métodos basados en modelos lineales generalizados mixtos: Método mejor predictor

- $$\tilde{P}_d^B(\boldsymbol{\theta}) = E(P_d|\mathbf{y}_{ds};\boldsymbol{\theta}) = \frac{1}{N_d}\left\{\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})\right\}$$

  tiene ECM mínimo y es insesgado bajo el modelo linear generalizado mixto
- Sin embargo, el proceso que se ha descrito es computacionalmente intensivo debido a las réplicas Monte Carlo

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Existen estimadores que se obtienen directamente de la salida del software que estima el GLMM

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Existen estimadores que se obtienen directamente de la salida del software que estima el GLMM
- Cuando se hace la regresión, el software estima $\hat{\alpha}$ y $\hat{v}_d$

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Existen estimadores que se obtienen directamente de la salida del software que estima el GLMM
- Cuando se hace la regresión, el software estima $\hat{\boldsymbol{\alpha}}$ y $\hat{v}_d$
- Se puede crear un estimador por el método de la analogía (*plug-in estimator*) que viene dado por:

$$\hat{P}_d^{PI} = \frac{1}{N_d}\left(\sum_{i \in s_d} Y_{di} + \sum_{i \in r_d} \hat{p}_{di}\right)$$

donde

$$\hat{p}_{di} = h(\mathbf{x}'_{di}\hat{\boldsymbol{\alpha}} + \hat{v}_d)$$

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Existen estimadores que se obtienen directamente de la salida del software que estima el GLMM
- Cuando se hace la regresión, el software estima $\hat{\alpha}$ y $\hat{v}_d$
- Se puede crear un estimador por el método de la analogía (*plug-in estimator*) que viene dado por:

$$\hat{P}_d^{PI} = \frac{1}{N_d}\left(\sum_{i\in s_d} Y_{di} + \sum_{i\in r_d} \hat{p}_{di}\right)$$

donde

$$\hat{p}_{di} = h(\mathbf{x}'_{di}\hat{\alpha} + \hat{v}_d)$$

- El estimador plug-in no es insesgado a menos que la función link es lineal

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Aunque es más fácil calcular, el estimador plug-in no es insesgado a menos que la función link es lineal

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Aunque es más fácil calcular, el estimador plug-in no es insesgado a menos que la función link es lineal
- Sin embargo, el link logístico $g(p) = log(p/(1-p))$ es aproximadamente lineal para $p \in (02, 08)$

Métodos basados en modelos lineales generalizados mixtos: método plug-in

- Aunque es más fácil calcular, el estimador plug-in no es insesgado a menos que la función link es lineal
- Sin embargo, el link logístico $g(p) = log(p/(1-p))$ es aproximadamente lineal para $p \in (02, 08)$
- Debido a esta propiedad, se puede comprobar que el EB y plug-in de la proporción $P_d = \overline{Y}_d$ se parecen al EBLUP, $\hat{P}_d^{EBLUP} = \hat{\bar{Y}}_d^{EBLUP}$, basado en el modelo con errores anidados (BHF)

Métodos basados en modelos lineales generalizados mixtos: ECM bootstrap

El ECM del estimador EB o plug-in se puede estimar con un procedimiento bootstrap 1) Ajustar el modelo GLMM $Y_{di}|v_d \sim \text{Bern}(p_{di}),\ g(p_{di}) = \mathbf{x}'_{di}\boldsymbol{\alpha} + v_d$ a los datos de la muestra para obtener los estimadores $\hat{\sigma}_v^2$ y $\hat{\boldsymbol{\alpha}}$ 2) Generar efectos aleatorios bootstrap

$$v_d^{*(b)} \overset{iid}{\sim} N(0, \hat{\sigma}_v^2), \quad d = 1, \ldots, D$$

Métodos basados en modelos lineales generalizados mixtos: ECM bootstrap

3) Generar el censo bootstrap $\mathbf{y}_d^{*(b)} = (Y_{d1}, \ldots, Y_{dN_d})'$ en la siguiente forma:

$$Y_{di}^{*(b)} \overset{ind}{\sim} \text{Bern}(p_{di}^{*(b)}),$$

y

$$p_{di}^{*(b)} = h(\mathbf{x}'_{di}\hat{\alpha} + v_d^{*(b)}), \quad i = 1, \ldots, N_d, \; d = 1, \ldots, D,$$

Calcular los verdaderos valores de los indicadores para el censo $P_d^{*(b)} = \bar{Y}_d^{*(b)}$, $d = 1, \ldots, D$.

Métodos basados en modelos lineales generalizados mixtos: ECM bootstrap

3) Generar el censo bootstrap $\mathbf{y}_d^{*(b)} = (Y_{d1}, \ldots, Y_{dN_d})'$ en la siguiente forma:

$$Y_{di}^{*(b)} \overset{ind}{\sim} \text{Bern}(p_{di}^{*(b)}),$$

y

$$p_{di}^{*(b)} = h(\mathbf{x}'_{di}\hat{\alpha} + v_d^{*(b)}), \quad i = 1, \ldots, N_d, \ d = 1, \ldots, D,$$

Calcular los verdaderos valores de los indicadores para el censo $P_d^{*(b)} = \bar{Y}_d^{*(b)}$, $d = 1, \ldots, D$.

4) Para cada área, extraer del censo los elementos de la muestra Y_{di}, $i \in s_d^{*(b)}$ para construir el vector $\mathbf{y}_{ds}^{*(b)}$ y, después, $\mathbf{y}_s^{*(b)} = ((\mathbf{y}_{1s}^{*(b)})', \ldots, (\mathbf{y}_{Ds}^{*(b)})')'$ el vector con los valores en la muestra de todas las áreas

Métodos basados en modelos lineales generalizados mixtos: ECM bootstrap

5) Ajustar el modelo GLMM a los datos bootstrap $\mathbf{y}_s^{*(b)}$ y calcular $\hat{P}_d^{EB*(b)}$, $d = 1, \ldots, D$

Métodos basados en modelos lineales generalizados mixtos: ECM bootstrap

5) Ajustar el modelo GLMM a los datos bootstrap $\mathbf{y}_s^{*(b)}$ y calcular $\hat{P}_d^{EB*(b)}$, $d = 1, \ldots, D$

6) Repetir pasos 2-5 B veces. El estimador bootstrap de ECM viene dado por

$$mse_B(\hat{P}_d^{EB}) = B^{-1} \sum_{b=1}^{B} (\hat{P}_d^{EB*(b)} - P_d^{*(b)})^2$$

Resumen del predictor EB/plug-in basado en GLMM

- Indicadores objetivos: Proporciones o totales de una variable binaria (e.g.carencia o no de determinado bien o servicio)

Resumen del predictor EB/plug-in basado en GLMM

- Indicadores objetivos: Proporciones o totales de una variable binaria (e.g.carencia o no de determinado bien o servicio)
- Requerimientos de datos:

Resumen del predictor EB/plug-in basado en GLMM

- Indicadores objetivos: Proporciones o totales de una variable binaria (e.g.carencia o no de determinado bien o servicio)
- Requerimientos de datos:
  - Microdatos de las p covariables obtenidas de la misma encuesta de la variable de interés

Resumen del predictor EB/plug-in basado en GLMM

- Indicadores objetivos: Proporciones o totales de una variable binaria (e.g.carencia o no de determinado bien o servicio)
- Requerimientos de datos:
  - Microdatos de las p covariables obtenidas de la misma encuesta de la variable de interés
  - Áreas de interés obtenidas de la misma encuesta

Resumen del predictor EB/plug-in basado en GLMM

- Indicadores objetivos: Proporciones o totales de una variable binaria (e.g.carencia o no de determinado bien o servicio)
- Requerimientos de datos:
  - Microdatos de las p covariables obtenidas de la misma encuesta de la variable de interés
  - Áreas de interés obtenidas de la misma encuesta
  - Microdatos de las p covariables de un censo o registro. Esto es necesario para calcular la esperanza $E(Y_{di}|\mathbf{y}_{ds};\boldsymbol{\theta})$

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
  - El número de observaciones usadas es el tamaño muestral, mucho mayor que el número de áreas

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
  - El número de observaciones usadas es el tamaño muestral, mucho mayor que el número de áreas
  - El modelo GLMM incorpora heterogeneidad no explicada entre las áreas

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
  - El número de observaciones usadas es el tamaño muestral, mucho mayor que el número de áreas
  - El modelo GLMM incorpora heterogeneidad no explicada entre las áreas
  - No se necesita conocer ninguna varianza, al contrario que para el modelo FH

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
  - El estimador del ECM bajo el modelo es un estimador estable y insesgado bajo el diseño (y bajo el diseño cuando el número de áreas es grande)

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
  - El estimador del ECM bajo el modelo es un estimador estable y insesgado bajo el diseño (y bajo el diseño cuando el número de áreas es grande)
  - Se puede estimar en cualquier subárea sin reajustar el modelo

Resumen del predictor EB/plug-in basado en GLMM

- Ventajas:
    - El estimador del ECM bajo el modelo es un estimador estable y insesgado bajo el diseño (y bajo el diseño cuando el número de áreas es grande)
    - Se puede estimar en cualquier subárea sin reajustar el modelo
    - Se puede estimar en áreas no muestreadas

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - Es basado en un modelo y por tanto es necesario analizar el modelo (a través de los residuos, por ejemplo)

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - Es basado en un modelo y por tanto es necesario analizar el modelo (a través de los residuos, por ejemplo)
  - No tiene en cuenta el diseño muestral y, por eso, no es insesgado bajo el diseño

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - Es basado en un modelo y por tanto es necesario analizar el modelo (a través de los residuos, por ejemplo)
  - No tiene en cuenta el diseño muestral y, por eso, no es insesgado bajo el diseño
  - El uso de microdatos de un censo puede conllevar problemas de confidencialidad

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - Es basado en un modelo y por tanto es necesario analizar el modelo (a través de los residuos, por ejemplo)
  - No tiene en cuenta el diseño muestral y, por eso, no es insesgado bajo el diseño
  - El uso de microdatos de un censo puede conllevar problemas de confidencialidad
  - El estimador ECM bootstrap *no* es insesgado bajo el modelo para el ECM bajo el modelo para un área concreta

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - El predictor EB (no el plug-in) es computacionalmente intensivo

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - El predictor EB (no el plug-in) es computacionalmente intensivo
  - El ECM del estimador EB usando un proceso bootstrap es excesivamente intensivo, pero se puede cortar usando el plug-in

Resumen del predictor EB/plug-in basado en GLMM

- Desventajas:
  - El predictor EB (no el plug-in) es computacionalmente intensivo
  - El ECM del estimador EB usando un proceso bootstrap es excesivamente intensivo, pero se puede cortar usando el plug-in
  - Requiere un reajuste para verificar la propiedad “benchmarking”