

Discusión, seminario de desagregación de estimaciones utilizando técnicas de estimación de áreas pequeñas en Chile

Carolina Franco

10 de noviembre, 2022

Muchas gracias por la interesantes charlas

Las charlas proveen una excelente introducción al SAE y dan a luz a muchísimos temas interesantes técnicos y prácticos

Chile: Pionero en Latinoamérica en la aplicación de SAE!

CEPAL esta haciendo trabajo importantísimo en Latinoamérica en desarrollar y difundir SAE

Estimación de áreas pequeñas/ Small Area Estimation (SAE)

- Las encuestas a menudo no pueden estimar todas las cantidades de interés a través de métodos “directos” con precisión aceptable
- **Área pequeña:** dominio donde el tamaño de la muestra es demasiado pequeño para una estimación directa confiable
- **Estimador directo:** basado en datos de muestra solo para el dominio de interés
- **SAE:** a través de modelamiento, se incorpora información de otros dominios y de fuentes de datos auxiliares



Ejemplos de fuentes de datos auxiliares

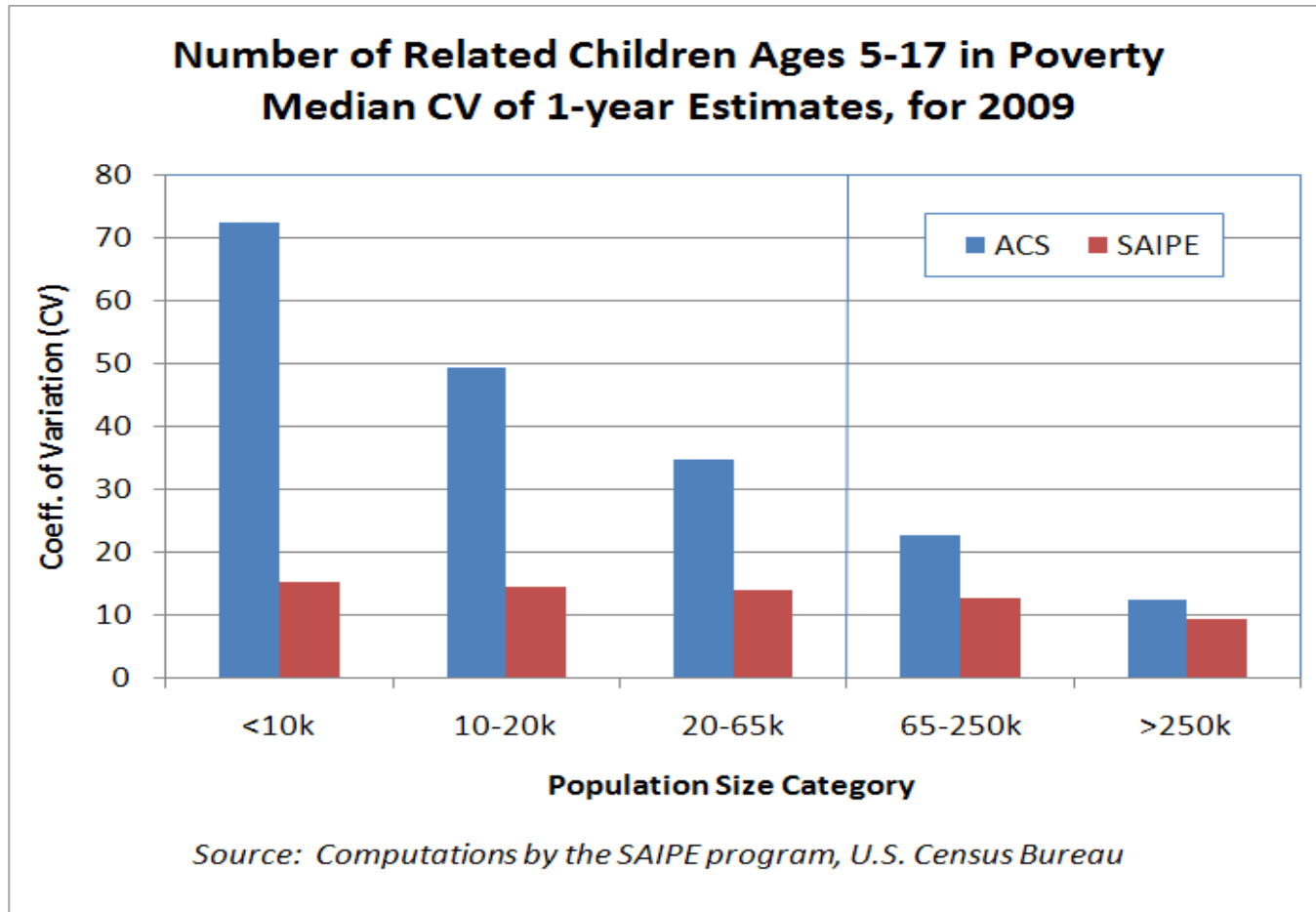
- **Censos**
- **Registros administrativos** : por ejemplo, ver Erciulescu, Franco y Lahiri (2021). Uso de registros administrativos en la estimación de áreas pequeñas
- **Misma encuesta, diferente año** (uso de modelos temporales, por ejemplo, Rao and You 1994, Franco y Bell, 2015, Lopez-Vizcaino et al. 2015)
- **Otras encuestas** , por ejemplo, Franco y Bell (2022), Franco y Maitra (2022, en revisión)
- **Información espacial** (por ejemplo, cuales dominios son vecinos?)
- **Datos comerciales, “Big Data”, datos satelitales, datos de telefonía celular, etc.**

Beneficios potenciales de SAE

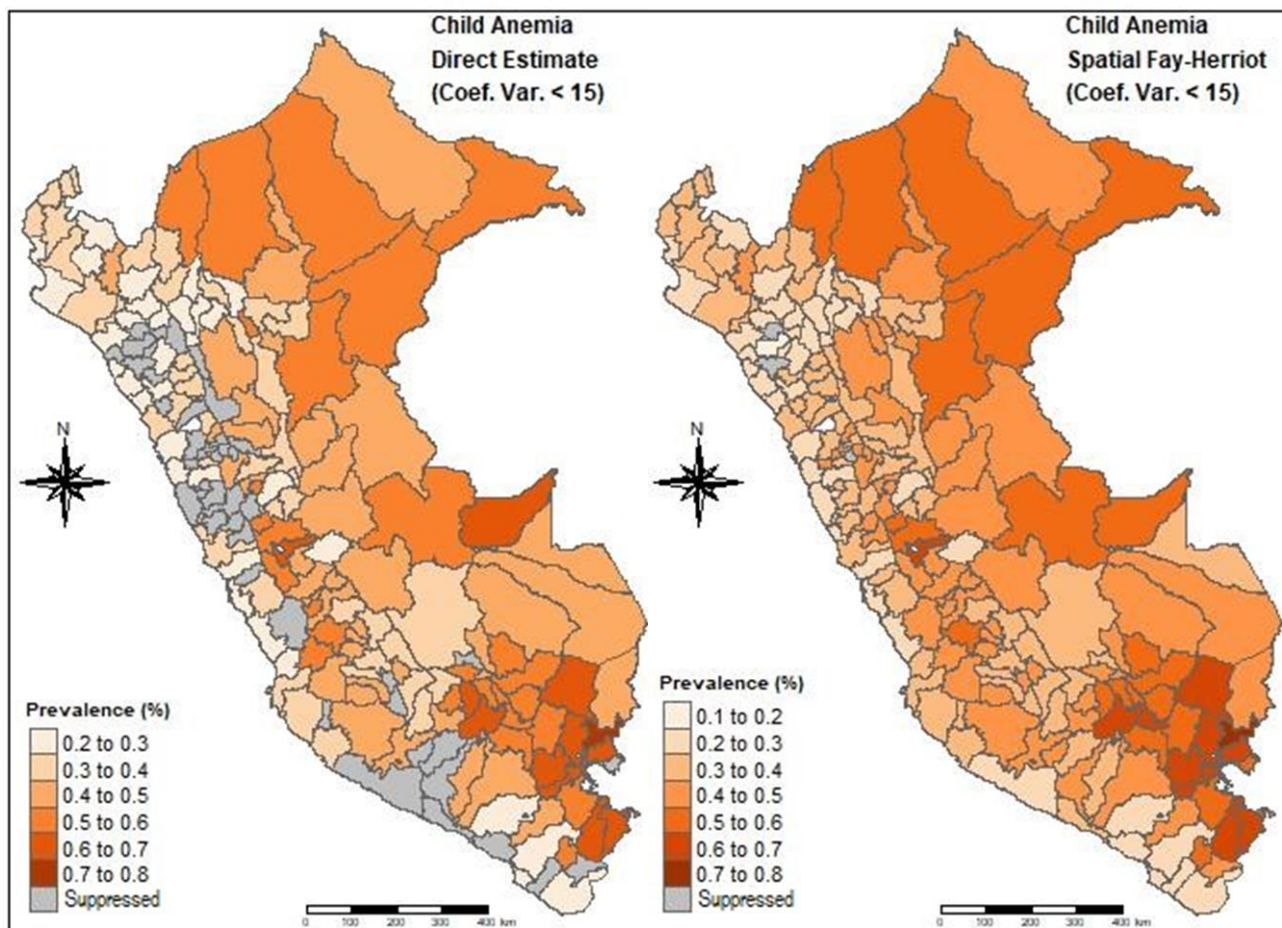
EN LAS CHARLAS, hemos visto que cuando se **implementa correctamente** y cuando se puede encontrar **buena información auxiliar**, SAE puede...

- Disminuir las medidas de incertidumbre (MSE/varianzas/CV) relativas a los estimadores directos
- Permitir que más áreas superen los filtros de calidad establecidos por los institutos nacionales de estadísticas
- Permitir la publicación de estimaciones a menores niveles de agregación
- Permitir la producción de estimaciones para áreas sin muestra en la encuesta

Ejemplo: Programa de Estimaciones de Pobreza e Ingresos en Áreas Pequeñas (SAIPE)



Ejemplo: Anemia en las provincias de Perú; niños de edad 0-4 (gráfico: Angelo Cozzubo)

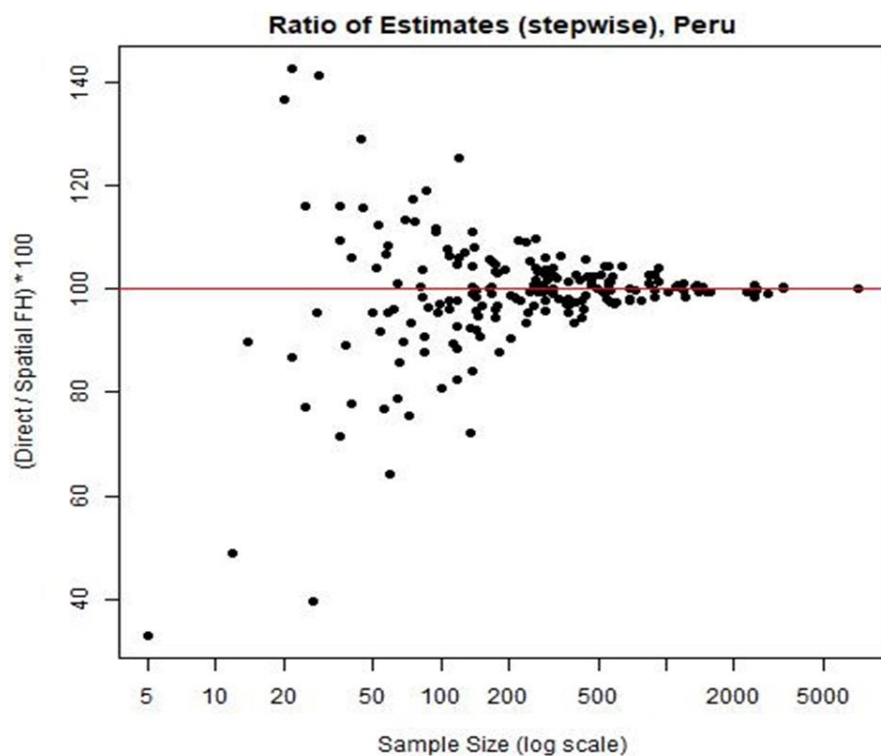


Solo se suprimen 3 predicciones de provincia con modelo espacial de Fay-Herriot

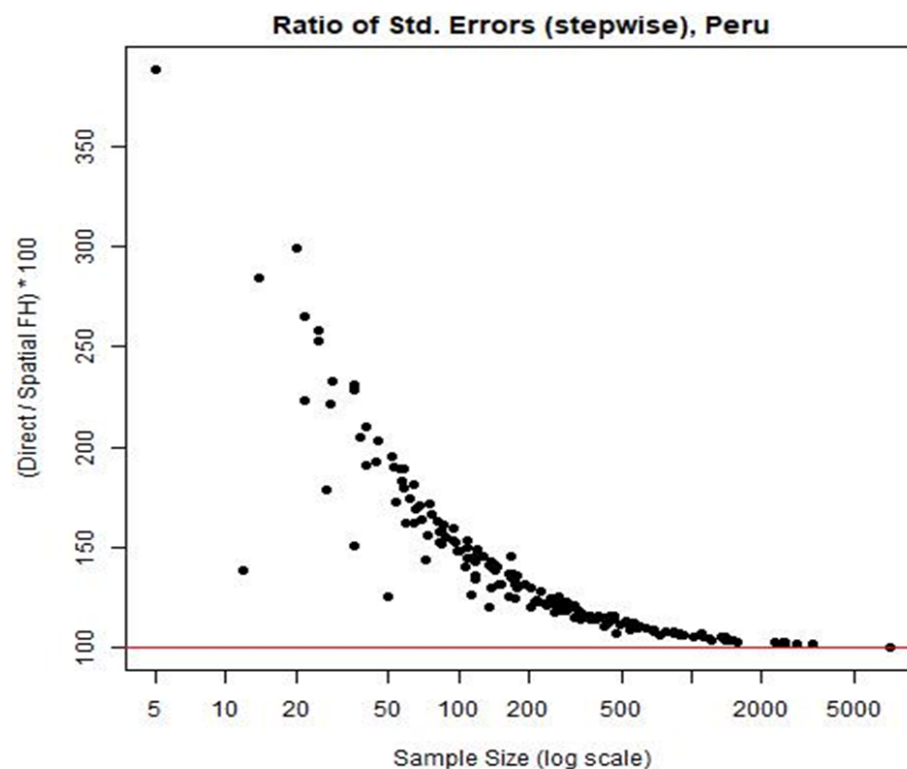
Se suprimen 42 estimados directos

Más detalles, anemia, niños 0-4, provincias Perú

Relación entre estimados directos y estimados del modelo



Note: FH = Fay-Harriot model. X-axis in logarithmic scale. Compiled by authors.



Note: FH = Fay-Harriot model. X-axis in logarithmic scale. Compiled by authors.

Algunos ejemplos de aplicaciones SAE en estadísticas oficiales en Estados Unidos

Ley de Derechos Electorales, Sección 203, determinaciones: <https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/>

Programa de Estimaciones de Seguro Médico para Áreas Pequeñas (SAHIE): <https://www.census.gov/programs-surveys/sahie.html>

Programa de Estimaciones de Pobreza e Ingresos en Áreas Pequeñas (SAIPE): <https://www.census.gov/programs-surveys/saipe.html>

Oficina de Estadísticas Laborales de EE UU.: Estadísticas de desempleo del área <https://www.bls.gov/lau/>

Elementos de un programa SAE exitoso, como las aplicaciones que vimos aquí

- Por lo general, una encuesta que mide la cantidad de interés sin sesgo (aproximadamente) pero con un tamaño de muestra limitada
- Buena información auxiliar, y experiencia en ella.
- Experiencia en estimación de áreas pequeñas, muestreo
- Experiencia en el tema de la encuesta
- ¡Mucho software para implementar SAE está disponible gratuitamente! A veces, el desarrollo de programas nuevos puede ser necesario/más apropiado
- “Temáticos, técnicos, información auxiliar”—INE

Ejemplos de aplicaciones SAE

- **Ejemplos de las presentaciones de hoy:** Desigualdad de ingresos, tasa de ocupación, planificación familiar, inseguridad alimentaria, acceso a la justicia, educación, medidas de pobreza, desempleo, victimización, seguridad ciudadana, etc.
- **Ejemplos de mi propia experiencia:** Pobreza de los niños en Estados Unidos, proporciones de ciudadanos estadounidenses, proporciones de personas que no hablan bien el inglés, proporciones de personas analfabetas, proporciones de gente sin cobertura de seguro de salud, proporciones de personas con discapacidades, desigualdad en los negocios, anemia en Perú, violencia en Perú
- En resumen, SAE trae posibilidades de mejorar la producción de estadísticas de muchísimos temas!

La importancia de SAE.

Hemos visto a través de estas charlas que SAE....

- Puede “incrementar la oferta estadística”
- Puede ayudar en la asignación eficiente de recursos (ejemplo: SAE para pobreza comunal en Chile)
- Puede ayudar con la implementación y evaluación de programas
- Puede ayudar a informar la política basada en evidencia (por ejemplo, políticas públicas para corregir desigualdades)
- Aprovecha de la disponibilidad de varios tipos de datos

El futuro de SAE

- La necesidad de SAE está creciendo
- Interés en entender las desigualdades en varios países del mundo
- Descubrimiento del potencial que tienen los métodos de SAE
- El estudio de SAE está mas activo que nunca
- Todos los años hay conferencias sobre SAE que atraen gente de varias parte del mundo (¡¡La próxima, en Perú!!)
- SAE permite utilizar técnicas diversas de estadística, y tiene inmensa utilidad práctica
- Estamos en un época de “proliferación de información”—una oportunidad para SAE

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

Encontrar buenas fuentes de información

- Estas charlas han resaltado sobre todo el uso de registros y censos. También se han mencionado los modelos **temporales y espaciales**.
- Los **modelos temporales son prometedores** no solo para mejorar las estimaciones del año de interés, sino también **para estimar cambios a través de los años**
- Datos satelitales, BIG DATA (ya hay algunos trabajos en esa dirección, como los mencionados por Andres en CEPAL, y otros).
- **La calidad y poder predictivo de los datos auxiliares es importante (como ha sido resaltado por todas las charlas de hoy)**

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

Encontrar buenas fuentes de información

- Si las fuentes de información auxiliaría tienen diferencias sistemáticas entre las áreas (como entre distintos grupos de razas, edad, etc.) eso puede tener consecuencias indeseables si los modelos no las capturan
- Discusión sobre que características buscar en datos auxiliares en Erciulescu, Franco, Lahiri, 2021
- Incorporar todos los errores conocidos al modelo. Por ejemplo, cuando se usan covariables que vienen de otras encuestas, incorporar sus errores de muestreo (ver Bell, Chung, Franco, Datta, 2019)
- **Cuidado con el uso de covariables de la misma encuesta!** Violan los supuestos de modelos FH

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

¿Modelos de nivel de unidad o de nivel de área?

- Los **modelos de nivel de unidad** tienen requisitos de datos más estrictos
- Por ejemplo, muchos requieren información de covariables para todas las unidades en la población
- Otro reto es incorporar el efecto del muestreo cuando este es “informativo” (Skinner, 1989, Binder, 1983, You and Rao, 2002, Pfefferman and Scherkov, 2007, Savitsky and Toth, 2016)
- Si se ignora eso, puede resultar en un sesgo en los estimadores

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

¿Modelos de nivel de unidad o de nivel de área?

- Los **modelos de nivel de área** (como FH) pueden ser más fáciles de aplicar
- También tienen retos
- Estimación de la varianza de muestreo: estimadores directos son muy variables. Es buena practica aplicar algún método para reducir su incertidumbre, como Funciones Generales de Varianza (FGV), mencionado también en el contexto ENUSC 2018
- Además, si se desean estimados de varios niveles geográficos, y se hacen modelos separados (ejemplo: SAIPE), estos no necesariamente son consistentes entre ellos. La costumbre es usar calibración, pero es un arreglo “ad-hoc”

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

Modelos de nivel de unidad o de nivel de área

¿Cuáles son mejores?

- Algunos trabajos han hecho esta comparación, pero la literatura es limitada en este tema (Hidiroglou y You, 2016, Newhouse, Merfeld, Rakrishnan, Swartz, Lahiri, 2022)
- Los dos tipos de modelos se han usado en aplicaciones importantes (por ejemplo, la pobreza!)
- Depende de la disponibilidad de datos, de la situación

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

¿Qué tipo de modelo usar?

- Muchas veces, se usan modelos con suposiciones de normalidad, como el Fay-Herriot
- A veces las proporciones son extremas y hay una gran cantidad de observaciones con valor cero
- Normalidad: simetría, cero probabilidad de observación de cero, posible obtener resultados para proporciones que no están en $[0,1]$
- La asimetría se puede también capturar con transformaciones de los datos, pero estas pueden tener sus propios problemas (ejemplo: SAIPE)
- Notar también que los intervalos de confianza simétricos no siempre son apropiados (ver Franco, Little, Louis, Slud, 2019)

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

¿Qué tipo de modelo usar?

- Otras distribuciones permiten la asimetría, masa de probabilidad en cero, y se pueden definir modelos que automáticamente dan estimadores en $[0,1]$
- Por ejemplo, modelos binomiales como los BLN ilustrados en Franco y Bell (2015, 2022)
- Otros retos surgen cuando se quiere estudiar varios niveles/severidad de pobreza
- Cuando los datos son categóricos u ordinales, es importante respetar las relaciones entre las categorías
- En ese caso, se pueden considerar, por ejemplo, modelos multinomiales (por ejemplo, Molina et al. 2007).

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

Comparación y diagnósticos de modelos—vimos muy buenos ejemplos en las charlas de hoy

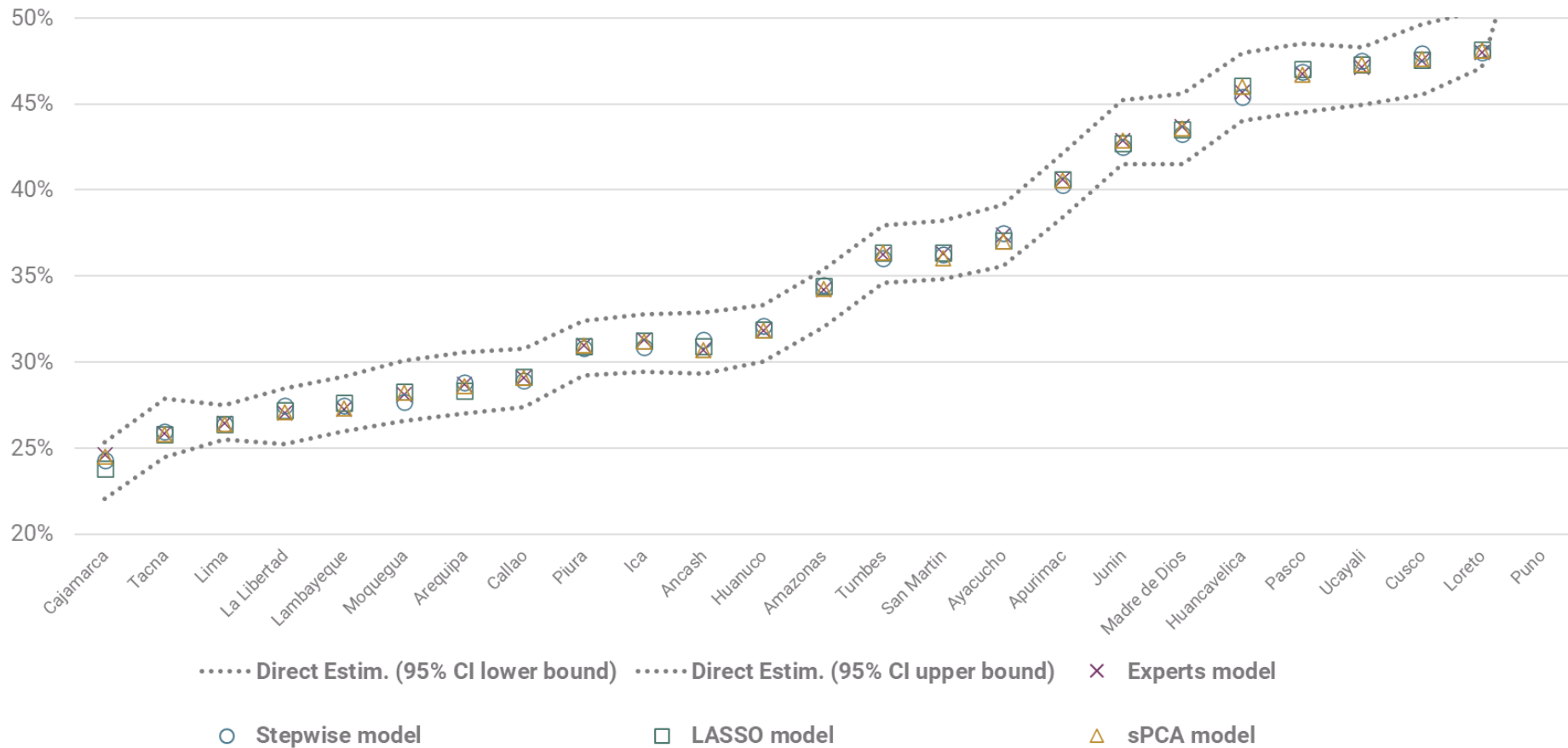
- Primero que todo, no olvidarse análisis de “sentido común”
- ¿Qué sabemos de la aplicación? ¿Son razonables las suposiciones del modelo? ¿Los estimadores? ¿Qué nivel de robustez hay?
- ¿Las suposiciones del modelo son también razonables para las áreas con pocos datos, para las áreas sin datos, para todo los tipos de dominios representados?
- Además, análisis formal
- Las comparaciones son más difíciles cuando uno compara modelos con muy distintas distribuciones (por ejemplo, binomiales vs normales)
- Hace falta más estudio del tema

Retos y desafíos más importantes que tienen los investigadores que hacen mapas de pobreza

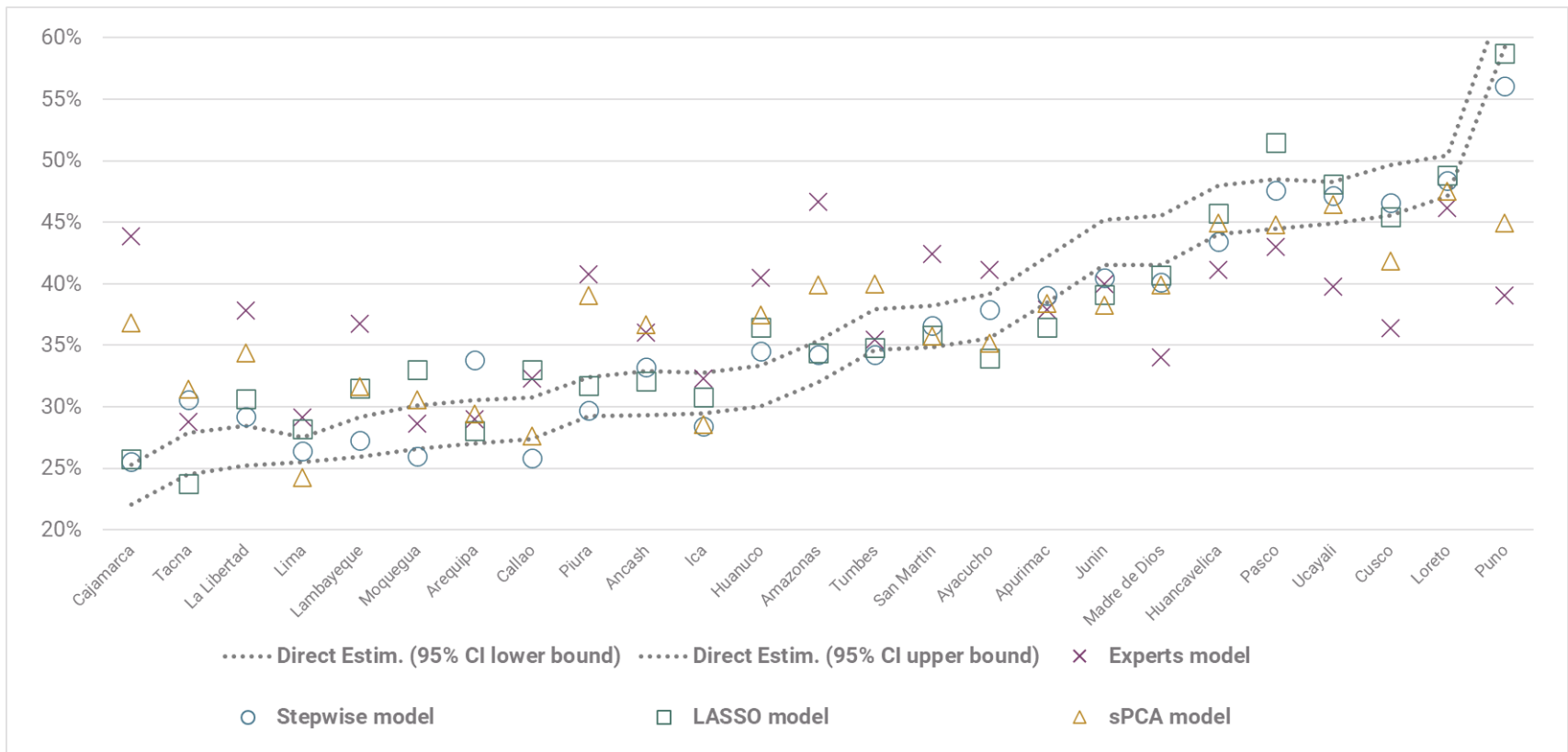
Comparación y diagnósticos de modelos-algunas alternativas

- AIC, BIC, WAIC, psis-LOO, métodos de ‘machine learning’, etc., útiles para elegir covariables, pero no necesariamente aplicables para comparar modelos muy distintos
- Análisis de residuos es útil, aunque a veces no da claridad para elegir entre modelos
- En algunos casos, es posible encontrar otra fuente para comparar con los resultados
- Agregar las estimaciones del modelo o las estimaciones sintéticas, comparar con estimadores directos (a un nivel en que los últimos sean confiables)
- Simulaciones de diseño
- Simulaciones de modelo
- Consultar con expertos

Ejemplo de comparación de estimadores de modelo al nivel de regiones en Perú



Los estimadores sintéticos, comparados con los estimadores directos



Metodologías más importantes para poder hacer SAE en esta década

- Es importante tener en cuenta tanto las propiedades técnicas de los modelos y las consideraciones prácticas de las aplicaciones
- SAE da lugar a toda clase de metodología de modelaje
- Aun así, es importante tener en cuenta las propiedades bajo el diseño de los estimadores de modelos
- Tanto los métodos sofisticados de modelaje como el conocimiento de los conceptos de muestreo serán importantes

Metodologías más importantes para poder hacer SAE en esta década

- Los métodos de MCMC y otras aproximaciones Bayesianas hacen posible desarrollar modelos más complejos que toman en cuenta varias relaciones importantes entre los datos y la información auxiliar (por ejemplo, usando STAN y JAGS)
- Muchas veces, los estimados son muy parecidos (Bayesianos y frecuentistas)
- La estimación de incertidumbre es más difícil para modelos muy complejos desde un punto de vista frecuentista, muchas veces se usa el Bootstrap

Conclusiones

- Los ejemplos que vimos hoy ilustran los beneficios de programas de SAE bien diseñados
- Chile es uno de los pioneros en Latinoamérica en el uso de SAE en estadísticas oficiales!
- Gracias a CEPAL por el trabajo excelente que están haciendo en difundir y desarrollar SAE!
- La utilidad e importancia de SAE seguirá creciendo globalmente
- La disciplina en si misma también va a continuar a crecer y a desarrollarse, pero ya hay un arsenal vasto de herramientas para los estadísticos que quieren aplicar los métodos de SAE

Un comentario

¡Próxima conferencia de SAE en Perú, Lima!
(**Andrés**)

SAE 2023: Small Area Estimation, Surveys and Data Science, 5-9 de junio, Pontificia Universidad Católica de Perú

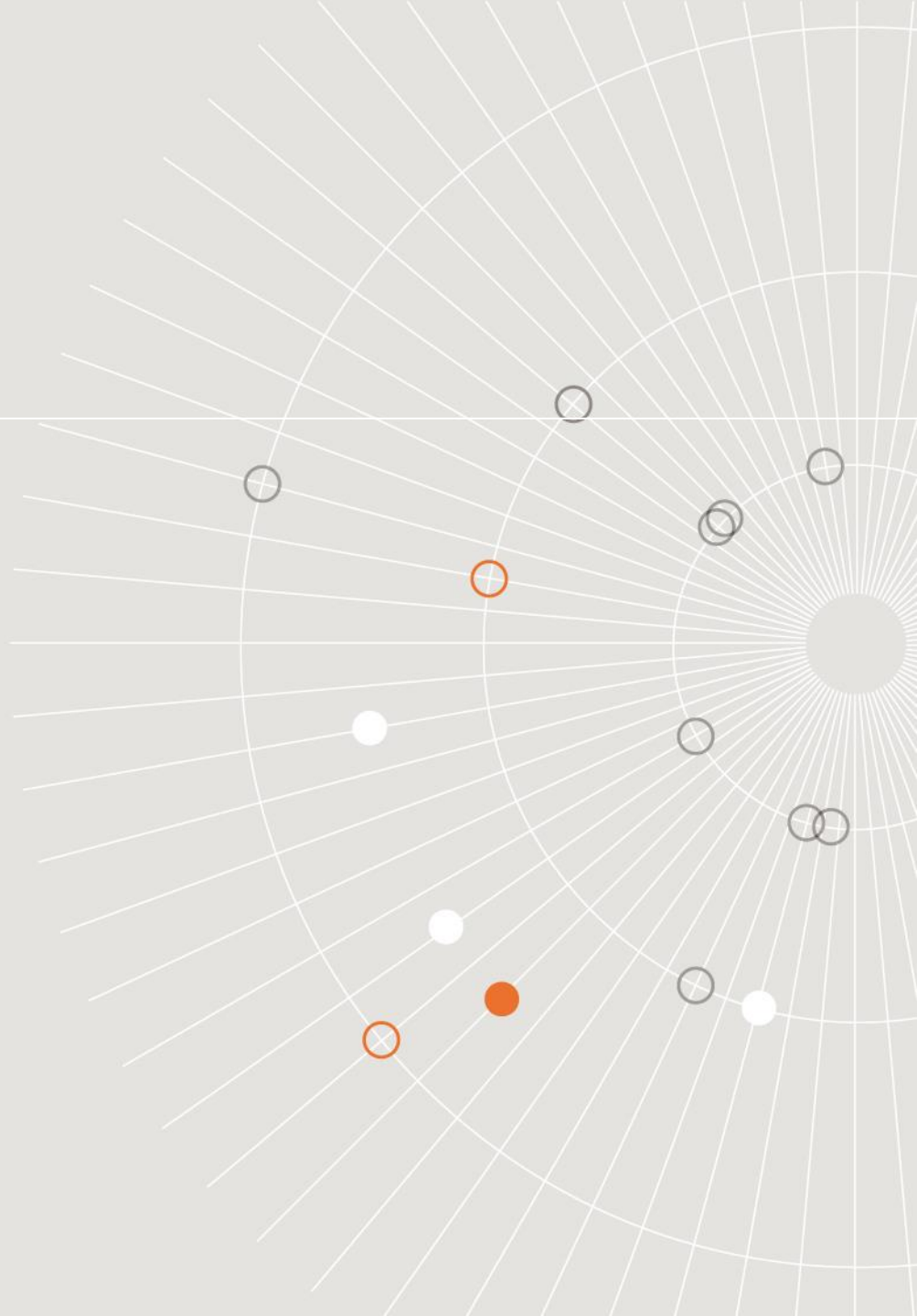
Sitio de web en construcción:

<https://sae2023.pucp.edu.pe/>

Recursos útiles

- Libro: Small Area Estimation, Molina and Rao (2015)
- ONU SAE Toolkit:
<https://unstats.un.org/wiki/display/SAE4SDG/>
- Mi introducción a SAE, junto con diapositivas y referencias:
<https://community.amstat.org/committeeoninternationalrelationsinstatistics/events2/webinars>
- Breve introducción a SAE: Erciulescu, A., Franco, C., and Lahiri, P. (2021). Use of administrative records in small area estimation. Chun, A. Y. and Larsen, M. (Eds.) Administrative records for survey methodology. New York: Wiley.

¿Preguntas?



Gracias

Carolina Franco
Franco-carolina@norc.org

 Research You Can Trust™

 **NORC** at the
University of
Chicago