

*Estudios de la Exactitud y Precisión de Estimadores en
Marcos Duales*

OSCAR FABIÁN MERCHÁN LÓPEZ
ESTADÍSTICO
CÓDIGO: 832426



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
NOVIEMBRE DE 2011

*Estudios de la Exactitud y Precisión de Estimadores en
Marcos Duales*

OSCAR FABIÁN MERCHÁN LÓPEZ
ESTADÍSTICO
CÓDIGO: 832426

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE
MAGÍSTER EN ESTADÍSTICA

DIRECTOR
LEONARDO TRUJILLO OYOLA, PH.D.
DOCTOR EN ESTADÍSTICA

CODIRECTOR
CRISTIANO FERRAZ, PH.D.
DOCTOR EN ESTADÍSTICA



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
NOVIEMBRE DE 2011

Título en español

Estudios de la exactitud y precisión de estimadores en marcos duales

Title in English

Accuracy and precision assessment of estimators for dual frame

Resumen: Las encuestas utilizando marcos duales son muy útiles en situaciones donde no es posible garantizar una cobertura completa de la población objetivo y se pueden disminuir considerablemente los costos de la encuesta comparativamente a un diseño con un solo marco muestral. Sin embargo, esta metodología no se utiliza con frecuencia en aplicaciones prácticas, ya sea por desconocimiento de los procedimientos, por la posible complejidad en su aplicación o por que sólo se cuenta con un marco muestral para llevar a cabo una investigación por muestreo. En el caso particular de Colombia son casi inexistentes las encuestas por muestreo que utilizan marcos duales para realizar estimaciones, en este documento se realiza un estudio mediante simulación de algunas medidas de exactitud y precisión utilizando la metodología de marcos duales tomando como referencia tres de los estimadores más citados en la literatura, también se realiza el análisis de estas medidas utilizando los datos de una encuesta muy importante para el sector agropecuario colombiano: la Encuesta Nacional Agropecuaria realizada por el DANE.

Abstract: Dual frame surveys are very useful when it is not possible to guarantee a complete coverage of the target population and may result in considerable cost savings over a single frame design with comparable precision. However, this technique is not very often applied from national statistical offices (NSO) due to its complexity and also because in most surveys only a single sample frame is available. Regarding the Colombian case, the implementation of dual frame surveys is very rare and therefore, this paper explore through a simulation study some measures of accuracy and precision using dual frame methodology considering the three most well known available estimators for the population total in the recent statistical literature, additionally a analysis of this measures with data from a very important survey for the Colombian agricultural sector: the National Agricultural Survey at DANE (National Statistical Administrative Department - Colombian NSO).

Palabras clave: Marcos duales, estimador de Hartley, estimador de Fuller-Burmeister, estimador de pseudo máxima verosimilitud, Encuesta Nacional Agropecuaria - ENA (Colombia)

Keywords: dual frame, Hartley estimator, Fuller-Burmeister estimator, pseudo maximum likelihood estimator, National Agricultural Survey - NAS (Colombia)

Nota de aceptación

Trabajo de tesis

Aprobado

“Mención Meritoria o Laureada”

Jurado
Jurado 1

Jurado
Jurado 2

Jurado
Jurado 3

Director
Leonardo Trujillo Oyola

Codirector
Cristiano Ferraz

Bogotá, D.C., Noviembre de 2011

Dedicado a

Este trabajo va dedicado en primera instancia a Dios que me dio la fortaleza para llevar a cabo la realización del mismo, a mis padres Martha y Carlos, que me han brindado todo el apoyo posible para cumplir mis metas más anheladas, a mi hermano Jhon y a Ingrid.

Agradecimientos

Este trabajo, si bien ha requerido de esfuerzo y mucha dedicación por parte del autor y su director y codirector de tesis, no hubiese sido posible su finalización sin la cooperación desinteresada de todas y cada una de las personas que a continuación cito y muchas de las cuales han sido un soporte muy fuerte en momentos en los cuales se requirió ayuda.

Primero y antes que nada, dar gracias a Dios, por estar con nosotros en cada paso que damos, por fortalecer nuestros corazones e iluminar nuestra mente y por haber puesto en nuestro camino a aquellas personas que han sido soporte y compañía durante todo el proceso de estudio y realización del mismo.

Un especial agradecimiento a mis directores de tesis Leonardo y Cristiano, por su paciencia, orientación y colaboración durante todo el proceso desarrollado y que tiene como fruto en el día de hoy este documento.

Gracias a cada uno de los maestros que participaron en el desarrollo profesional durante la carrera y en el posgrado, sin su ayuda y conocimientos no hubiera sido posible el desarrollo de esta tesis.

Finalmente, agradezco a los compañeros y amigos de la maestría, especialmente en el seminario donde se debatieron las propuestas de tesis y se dieron consejos, enseñanzas y ayuda para el cumplimiento de los objetivos de este trabajo.

"Si avanzo sígueme, si me detengo empújame, si retrocedo mátame.", Ernesto Che Guevara.

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	IV
Introducción	V
1. Marco Teórico	1
1.1. Conceptos Generales	1
1.1.1. Marco muestral	1
1.1.2. Comparación entre marcos de área y marcos de lista	1
1.1.3. Marcos Duales	1
1.2. Estimadores para marcos duales	4
1.2.1. Estimador de Hartley	4
1.2.2. Estimador de Fuller y Burmeister	6
1.2.3. Estimador Pseudo Máximo Verosímil	9
2. Metodología	14
2.1. Diseños muestrales actuales	14
2.1.1. Diseño muestral de la ENA	14
2.1.2. Diseño muestral para hortalizas	15
2.2. Estudio de simulación	16
2.2.1. Sesgo relativo	16
2.2.2. Error cuadrático medio empírico	16
2.2.3. Probabilidad de cobertura	16
2.2.4. Procedimiento de simulación	17

2.2.5. Modelo de generación de datos	18
3. Resultados	19
3.1. Modelo	19
3.2. Caso aplicado	27
Conclusiones	29
Trabajo futuro	32
Bibliografía	33
Programa en R	35

Índice de tablas

1.1. Notación que se utilizará en este documento.	3
3.1. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 40 % en la segunda etapa del marco A.	20
3.2. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 60 % en la segunda etapa del marco A.	22
3.3. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 80 % en la segunda etapa del marco A.	24
3.4. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 100 % (conglomerados) en la segunda etapa del marco A.	26
3.5. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 100 % (conglomerados) en la segunda etapa del marco A.	28

Índice de figuras

1.1. Esquema general de marcos duales.	2
1.2. Esquema general de muestras de marcos duales.	3

Introducción

Las encuestas utilizando marcos duales son muy útiles en situaciones donde no se puede proporcionar una cobertura completa de la población objetivo y cuando se quieren disminuir los costos de la encuesta. Sin embargo, esta metodología no se utiliza con frecuencia en aplicaciones prácticas, ya sea por desconocimiento de los procedimientos, por la posible complejidad en su aplicación o por que sólo se cuenta con un marco muestral para llevar a cabo una investigación por muestreo.

La metodología de marcos duales no es un concepto nuevo, su aparición se remonta a la década de los 60's, y desde aquella época hasta la fecha han sido mucho los autores que han contribuido al desarrollo de teoría en este campo del muestreo de poblaciones finitas. Hartley y Rao han sido los personajes más preponderates en la generación de nuevo conocimiento relacionado a la metodología de marcos duales (Hartley (1962); Hartley (1974); Lohr & Rao (2000); Lohr & Rao (2006); Skinner & Rao (1996); Rao & Changbao (2010)).

En el caso particular de Colombia son casi inexistentes las encuestas por muestreo que utilizan la estimación con marcos duales. Los problemas para la construcción de los marcos pueden ser una posible causa, no obstante es evidente que las teconologías de información y el avance de las comunicaciones de la actualidad, son una gran ventaja con la cual no se contaba hace algunas décadas. En este documento se presenta una propuesta para la implementación de la metodología de marcos duales basada en datos provenientes de una encuesta muy importante para el país, la Encuesta Nacional Agropecuaria - ENA (Departamento Administrativo Nacional de Estadística - DANE 2010).

En la literatura existen diversos estimadores propuestos para trabajar con marcos duales (marcos múltiples), por esta razón se pretende evaluar la precisión y eficiencia de los principales estimadores, mediante el estudio de simulaciones de Monte Carlo. Algunos autores han realizado estudios de simulación similares; sin embargo, los diseños muestrales usualmente utilizados son el diseño muestral MAS (muestreo aleatorio simple) o muestreo en dos etapas utilizando en cada etapa MAS. Se pretende en este trabajo evaluar tres estimadores para marcos duales, utilizando diseños muestrales distintos a los comúnmente presentados, como el diseño πPT (diseño proporcional al tamaño sin repetición), diseños muestrales bietápicos y combinaciones de los mismos, para los datos generados mediante simulación y datos directamente tomados de la ENA.

Los datos de la ENA se obtienen anualmente por el DANE; la ENA se viene realizando desde principios de los 90's y su manejo ha ido cambiando a través de los años entre la Corporación Colombia Internacional (CCI) y el DANE, en conjunto con el Ministerio de Agricultura y Desarrollo Rural (MADR). Para esta encuesta se estiman algunos indicado-

res asociados a variables como: uso de la tierra, producción y rendimiento, para cultivos transitorios y permanentes, área en pastos, producción de leche e inventario pecuario.

Paralelamente a la ENA, se realizan algunas encuestas agropecuarias denominadas por producto, como la de hortalizas, la encuesta nacional de arroz mecanizado o la encuesta de sacrificio de ganado; lo cual genera más de una estimación de indicadores para algunos productos agropecuarios. Actualmente, no existe una propuesta metodológica que integre más de una encuesta para obtener una única estimación de indicadores obtenidos para el mismo producto agropecuario, lo cual también podría ser beneficioso en cuantos a costos y precisión, y cuya implementación es el objetivo principal de esta investigación.

En este documento se evaluarán mediante medidas de precisión y eficiencia, los tres estimadores más mencionados en la literatura reciente para marcos duales, utilizando como referencia datos provenientes de la ENA y de la encuesta de hortalizas realizada por el DANE. Los principales objetivos que persigue esta tesis son, confrontar mediante simulación los estimadores propuestos para marcos duales aplicados a la ENA, comparar varias combinaciones de diseños muestrales con diferentes tamaños y fracciones muestrales a seleccionar en los dos marcos a usar, determinar la ganancia o pérdida en términos de eficiencia de los estimadores propuesto y evaluar la ganancia o pérdida en términos de precisión de los estimadores bajo las combinaciones muestrales planteados.

Marco Teórico

1.1. Conceptos Generales

1.1.1. Marco muestral

En el contexto de muestreo, un **marco** es el material o instrumento usado para identificar y ubicar todos los elementos de la población (Cochran 1977). Una unidad muestral puede ser un elemento o un conjunto de elementos. En este sentido se pueden definir en general dos tipos de marco:

- Marcos de lista de elementos.
- Marcos de grupos de elementos (por ejemplo, los marcos de áreas).

1.1.2. Comparación entre marcos de área y marcos de lista

- El costo de un marco de áreas generalmente es mucho más alto que el de un marco de lista, ya que generalmente los marcos de áreas requieren de material más caro para su consolidación como las fotografías aéreas y las imágenes satelitales.
- En algunos casos los marcos de lista no dan cobertura completa de población.
- Los marcos de lista son frecuentemente más robustos ante datos atípicos, puesto que las inconsistencias encontradas son más fáciles de identificar.
- Los marcos de áreas no se desactualizan tan rápido como los de lista, estos últimos se afectan debido a cambios en datos de propiedad, razón social o porque el elemento ya no pertenece al universo de estudio al realizar la actualización del marco.

1.1.3. Marcos Duales

Cuando se unen dos marcos muestrales que permiten una cobertura completa de la población se habla de "Marcos Duales" (Hartley 1962). Al trabajar con marcos duales existen dos supuestos:

1. Toda unidad en la población de interés pertenece como mínimo a uno de los marcos.
2. Es posible registrar para cada unidad en la muestra si ella pertenece o no al otro marco.

En la figura 1.1 se presenta la forma general al trabajar con marcos duales; donde al unir dos marcos, en este caso el marco A y el marco B, se da cobertura completa a la población objetivo. Al realizar esta unión se generan 3 dominios excluyentes: el dominio a que contiene los elementos o grupos de elementos (dependiendo del tipo de marco) que solamente están en el marco A, el dominio b que contiene los elementos o grupos de elementos que solamente están en el marco B y el dominio ab que contiene los elementos o grupos de elementos que pertenecen tanto al marco A como al marco B, a éste último dominio se le denomina usualmente como el traslape.

Un ejemplo clásico en el sector agropecuario de marcos duales, es tomar un marco de lista conformado por un listado de las fincas o de los productores de un determinado producto y unirlo con un marco de áreas de dicho producto conformado mediante cartografía y fotografía digital para dar cobertura completa a la población; en este caso la metodología es de gran utilidad, ya que al tomar por separado los marcos no se podría dar una estimación completa de los indicadores para un determinado producto, como condición esencial para la metodología los marcos se deben traslapar en alguno de sus niveles de desagregación, como por ejemplo a nivel de segmento de tierra (conglomerado), a nivel predial, nivel municipal o nivel departamental, entre otros.

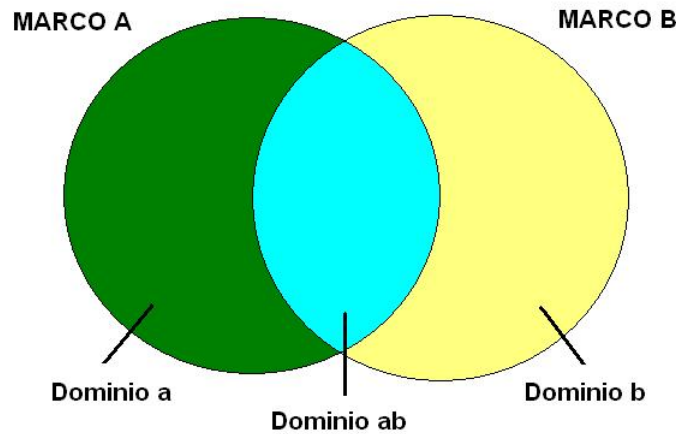


FIGURA 1.1. Esquema general de marcos duales.

Una de las condiciones al trabajar con marcos duales es la extracción de muestras independientes en cada uno de los marcos. La representación gráfica al extraer muestras de marcos duales se presenta en la figura 1.2; en ella se puede apreciar que se generan dos dominios adicionales: el dominio ab' que contiene los elementos de la muestra extraída del marco A que pertenecen al traslape de los marcos y el dominio ab'' que contiene los elementos de la muestra extraída del marco B que pertenecen al traslape de los marcos.

La notación que se usará en el documento se encuentra en la tabla 1.1. Para el dominio ab se puede apreciar que hay dos notaciones distintas, que indican la pertenencia a la muestra extraída del marco A o del marco B. Así n'_{ab} indica el número de elementos de

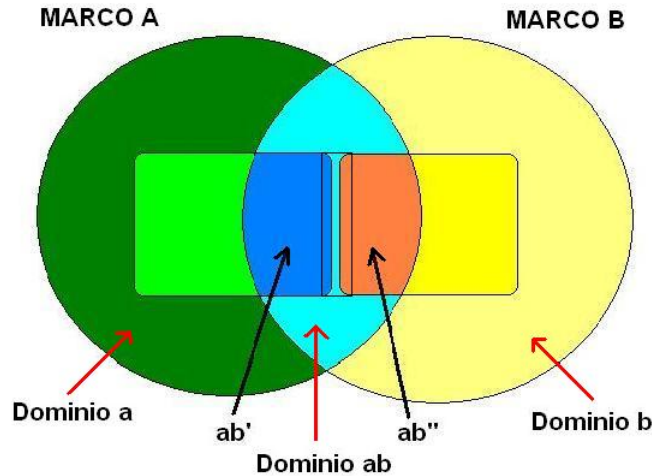


FIGURA 1.2. Esquema general de muestras de marcos duales.

la muestra del marco A que pertenecen al traslape, n''_{ab} indica el número de elementos de la muestra del marco B que pertenecen al traslape, \hat{t}'_{yab} es el total estimado del traslape generado por la muestra extraída del marco A, \hat{t}''_{yab} es el total estimado del traslape generado por la muestra extraída del marco B, \bar{y}'_{ab} es el promedio para el traslape generado por la muestra extraída del marco A y \bar{y}''_{ab} es el promedio para el traslape generado por la muestra extraída del marco B.

TABLA 1.1. Notación que se utilizará en este documento.

CANTIDAD DE INTERÉS	MARCO		DOMINIO		
	A	B	a	b	ab
Población	U_A	U_B	U_a	U_b	U_{ab}
Tamaño poblacional	N_A	N_B	N_a	N_b	N_{ab}
Total poblacional	t_{yA}	t_{yB}	t_{ya}	t_{yb}	t_{yab}
Media poblacional	μ_{yA}	μ_{yB}	μ_{ya}	μ_{yb}	μ_{yab}
Muestra	S_A	S_B	S_a	S_b	S_{ab}
Tamaño muestral	n_A	n_B	n_a	n_b	n'_{ab} n''_{ab}
Total muestral	\hat{t}'_{yA}	\hat{t}'_{yB}	\hat{t}'_{ya}	\hat{t}'_{yb}	\hat{t}'_{yab} \hat{t}''_{yab}
Media muestral	\bar{y}_A	\bar{y}_B	\bar{y}_a	\bar{y}_b	\bar{y}'_{ab} \bar{y}''_{ab}

Cuatro casos diferentes pueden considerarse dependiendo del conocimiento del número total de unidades en los marcos y en los dominios y de la posibilidad para asignar el número de elementos a seleccionar (tamaños muestrales) de cada dominio (Ospina 1981).

1. Todos los tamaños de los dominios se conocen. Es factible asignar tamaños de muestra a los dominios. Dominios = Estratos.
2. Todos los tamaños de los dominios se conocen. Los tamaños de muestra pueden asignarse únicamente a los marcos. Dominios = Post-estratos.
3. Los tamaños de los dominios no se conocen pero los tamaños de marco sí se conocen. Los tamaños de muestra pueden asignarse solamente a los marcos. Dominios = Dominios propios.

4. Tanto los tamaños de los dominios como los tamaños de los marcos no se conocen, pero las magnitudes relativas de los marcos se conocen. Los tamaños de muestra sólo pueden asignarse a los marcos. Dominios = Dominios en poblaciones de tamaño desconocido.

Bajo el escenario 1 se puede aplicar directamente la metodología estándar para muestreo estratificado y en el escenario 4 sólo será posible la estimación de medias poblacionales, por estas razones en este trabajo sólo se trabajará bajo los escenarios 2 y 3.

1.2. Estimadores para marcos duales

Los marcos múltiples han sido estudiados por diversos autores como Bankier (1986), Fuller & Burmeister (1972), Hartley (1962), Hartley (1974), Kalton & Anderson (1986), Skinner (1991) y Skinner & Rao (1996), entre otros. Lohr & Rao (2000) estudiaron la estimación de la varianza para marcos duales usando el método jackknife y examinaron el desempeño de los intervalos de confianza bajo la aproximación normal mediante simulaciones.

Lohr & Rao (2006) derivaron estimadores lineales óptimos y pseudo-verosímiles para el total poblacional cuando se toman muestras independientes de marcos múltiples bajo diseños muestrales probabilísticos. Gonzalez-Villalobos & Wallace (1991) proponen el uso de marcos múltiples en encuestas agrícolas alrededor del mundo.

A continuación se presentan los fundamentos de los tres estimadores más mencionados en la literatura actual de marcos duales: el estimador de Hartley, el estimador de Fuller-Burmeister y el estimador Pseudo Máximo Verosímil o de Skinner-Rao.

1.2.1. Estimador de Hartley

Uno de los objetivos al analizar datos provenientes de encuestas con marcos múltiples usualmente es estimar el total poblacional t_y , usando información de muestras independientes tomada de los marcos. En el caso de dos marcos, el total poblacional t_y se puede escribir como:

$$t_y = t_{ya} + t_{yab} + t_{yb}$$

donde t_{ya} es el total del dominio a , t_{yab} es el total del dominio ab , y t_{yb} es el total para el dominio b .

Hartley (1962) propuso un estimador al trabajar con marcos duales, el cual incluye un promedio ponderado para estimar el total del dominio ab :

$$\hat{t}_{yH}(p) = \hat{t}_{ya} + p\hat{t}'_{yab} + (1-p)\hat{t}''_{yab} + \hat{t}_{yb}$$

donde \hat{t}_{ya} es el total poblacional estimado para el dominio a , \hat{t}'_{yab} es el total poblacional estimado para el dominio ab usando la muestra del marco A, \hat{t}''_{yab} es el total poblacional estimado para el dominio ab usando la muestra del marco B, \hat{t}_{yb} es el total poblacional estimado para el dominio b y generalmente con $0 \leq p \leq 1$.

Se puede apreciar que si los estimadores \hat{t}_{ya} , \hat{t}_{yb} , \hat{t}'_{yab} y \hat{t}''_{yab} son insesgados para t_{ya} , t_{yb} , t_{yab} y t_{yab} respectivamente, entonces $\hat{t}_{yH}(p)$ será insesgado para t_y .

Prueba:

$$\begin{aligned}
E[\hat{t}_{yH}(p)] &= E[\hat{t}_{ya} + p\hat{t}'_{yab} + (1-p)\hat{t}''_{yab} + \hat{t}_{yb}] \\
&= E[\hat{t}_{ya}] + pE[\hat{t}'_{yab}] + (1-p)E[\hat{t}''_{yab}] + E[\hat{t}_{yb}] \\
&= t_{ya} + pt_{yab} + (1-p)t_{yab} + t_{yb} \\
&= t_{ya} + t_{yab} + t_{yb} \\
&= t_y \square
\end{aligned}$$

Hartley (1974) propuso en la ecuación determinar el valor de p de tal forma que se minimice la varianza de $\hat{t}_{yH}(p)$, esto bajo el supuesto de que las muestras de los marcos son independientes. Puesto que las muestras son extraídas independientemente de cada uno de los marcos no hay covarianzas entre los totales de los dominios provenientes de marcos distintos, y se pueden separar los términos de la varianza del estimador de Hartley para los dos marcos:

$$V[\hat{t}_{yH}(p)] = V[\hat{t}_{ya} + p\hat{t}'_{yab}] + V[(1-p)\hat{t}''_{yab} + \hat{t}_{yb}]$$

y el valor de p que minimiza dicha varianza es:

$$p_{opt} = \frac{V(\hat{t}''_{yab}) + Cov(\hat{t}''_{yab}, \hat{t}_{yb}) - Cov(\hat{t}_{ya}, \hat{t}'_{yab})}{V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})}$$

Prueba:

$$\begin{aligned}
V[\hat{t}_{yH}(p)] &= V[\hat{t}_{ya} + p\hat{t}'_{yab}] + V[(1-p)\hat{t}''_{yab} + \hat{t}_{yb}] \\
&= V(\hat{t}_{ya}) + p^2V(\hat{t}'_{yab}) + 2pCov(\hat{t}_{ya}, \hat{t}'_{yab}) + (1-p)^2V(\hat{t}''_{yab}) \\
&\quad + V(\hat{t}_{yb}) + 2(1-p)Cov(\hat{t}''_{yab}, \hat{t}_{yb}) \\
&= V(\hat{t}_{ya}) + p^2V(\hat{t}'_{yab}) + 2pCov(\hat{t}_{ya}, \hat{t}'_{yab}) + (1-2p+p^2)V(\hat{t}''_{yab}) \\
&\quad + V(\hat{t}_{yb}) + (2-2p)Cov(\hat{t}''_{yab}, \hat{t}_{yb}) \\
&= V(\hat{t}_{ya}) + p^2V(\hat{t}'_{yab}) + 2pCov(\hat{t}_{ya}, \hat{t}'_{yab}) + V(\hat{t}''_{yab}) - 2pV(\hat{t}''_{yab}) + p^2V(\hat{t}''_{yab}) \\
&\quad + V(\hat{t}_{yb}) + 2Cov(\hat{t}''_{yab}, \hat{t}_{yb}) - 2pCov(\hat{t}''_{yab}, \hat{t}_{yb}) \\
&= p^2[V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})] + p[2Cov(\hat{t}_{ya}, \hat{t}'_{yab}) - 2V(\hat{t}''_{yab}) - 2Cov(\hat{t}''_{yab}, \hat{t}_{yb})] \\
&\quad + V(\hat{t}_{ya}) + V(\hat{t}''_{yab}) + V(\hat{t}_{yb}) + 2Cov(\hat{t}''_{yab}, \hat{t}_{yb})
\end{aligned}$$

Ahora se encuentra el valor de p que minimice $V[\hat{t}_{yH}(p)]$ es decir:

$$\begin{aligned}
\min_p \{V[\hat{t}_{yH}(p)]\} &= \min_p \{p^2[V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})] + p[2Cov(\hat{t}_{ya}, \hat{t}'_{yab}) - 2V(\hat{t}''_{yab}) - 2Cov(\hat{t}''_{yab}, \hat{t}_{yb})] \\
&\quad + V(\hat{t}''_{yab})\} + V(\hat{t}_{ya}) + V(\hat{t}''_{yab}) + V(\hat{t}_{yb}) + 2Cov(\hat{t}''_{yab}, \hat{t}_{yb})
\end{aligned}$$

derivando con respecto a p se tiene:

$$\frac{dV[\hat{t}_{yH}(p)]}{dp} = 2p[V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})] + 2Cov(\hat{t}_{ya}, \hat{t}'_{yab}) - 2V(\hat{t}''_{yab}) - 2Cov(\hat{t}''_{yab}, \hat{t}_{yb})$$

al igualar a 0 y despejar p se obtiene:

$$p_{opt} = \frac{V(\hat{t}''_{yab}) + Cov(\hat{t}''_{yab}, \hat{t}_{yb}) - Cov(\hat{t}_{ya}, \hat{t}'_{yab})}{V(\hat{t}'_{yab}) + V(\hat{t}''_{yab})} \quad \square$$

En la práctica las covarianzas necesarias para el cálculo del p óptimo son desconocidas, entonces se deben estimar con los datos muestrales.

Ahora supongamos que se extraen muestras MAS en los dos marcos y conocemos N_{ab} , entonces tenemos:

$$\hat{t}_{yH}(p) = N_a \bar{y}_a + N_{ab}(p\bar{y}'_{ab} + (1-p)\bar{y}''_{ab}) + N_b \bar{y}_b$$

En este caso la varianza del estimador viene dada por:

$$\begin{aligned} Var(\hat{t}_{yH}) &= \frac{N_A^2}{n_A} \left[\left(1 - \frac{N_{ab}}{N_A}\right) \sigma_a^2 + \frac{N_{ab}}{N_A} p^2 \sigma_{ab}^2 \right] \\ &+ \frac{N_B^2}{n_B} \left[\left(1 - \frac{N_{ab}}{N_B}\right) \sigma_b^2 + \frac{N_{ab}}{N_B} (1-p)^2 \sigma_{ab}^2 \right] \end{aligned}$$

1.2.2. Estimador de Fuller y Burmeister

Fuller & Burmeister (1972) propusieron una modificación al estimador de Hartley que incorpora el tamaño estimado del dominio ab con respecto a las muestras de ambos marcos:

$$\hat{t}_{yFB}(p_1, p_2) = \hat{t}_{ya} + \hat{t}_{yb} + p_1 \hat{t}'_{yab} + (1-p_1) \hat{t}''_{yab} + p_2 (\hat{N}'_{ab} - \hat{N}''_{ab})$$

donde \hat{N}'_{ab} y \hat{N}''_{ab} son los estimadores respectivos del tamaño del dominio ab bajo la muestra del marco A y B respectivamente .

Se puede apreciar que si los estimadores \hat{t}_{ya} , \hat{t}_{yb} , \hat{t}'_{yab} , \hat{t}''_{yab} , \hat{N}'_{ab} y \hat{N}''_{ab} son insesgados para t_{ya} , t_{yb} , t_{yab} , t_{yab} , N_{ab} y N_{ab} respectivamente, entonces $\hat{t}_{yFB}(p_1, p_2)$ será insesgado para t_y .

Prueba:

$$\begin{aligned}
E[\hat{t}_{yFB}(p_1, p_2)] &= E[\hat{t}_{ya} + \hat{t}_{yb} + p_1\hat{t}'_{yab} + (1-p_1)\hat{t}''_{yab} + p_2(\hat{N}'_{ab} - \hat{N}''_{ab})] \\
&= E[\hat{t}_{ya}] + E[\hat{t}_{yb}] + p_1E[\hat{t}'_{yab}] + (1-p_1)E[\hat{t}''_{yab}] + p_2E[\hat{N}'_{ab}] - p_2E[\hat{N}''_{ab}] \\
&= t_{ya} + t_{yb} + p_1t_{yab} + (1-p_1)t_{yab} + p_2(N_{ab} - N_{ab}) \\
&= t_{ya} + t_{yb} + t_{yab} \\
&= t_{y\Box}
\end{aligned}$$

Al igual que para el estimador de Hartley se encuentran los valores de p_1 y p_2 que minimicen la varianza de \hat{t}_{yFB} , obteniéndose:

$$\begin{bmatrix} p_{1opt} \\ p_{2opt} \end{bmatrix} = - \left(Cov \begin{bmatrix} \hat{t}'_{yab} - \hat{t}''_{yab} \\ \hat{N}'_{ab} - \hat{N}''_{ab} \end{bmatrix} \right)^{-1} \begin{bmatrix} Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}'_{yab}, \hat{t}'_{yab} - \hat{t}''_{yab}) \\ Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}'_{yab}, \hat{N}'_{ab} - \hat{N}''_{ab}) \end{bmatrix}$$

Prueba: Escribiendo en forma vectorial el estimador $\hat{t}_{yFB}(p_1, p_2)$ tenemos:

$$\hat{t}_{yFB} = \begin{bmatrix} (\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}'_{yab}) & (\hat{t}'_{yab} - \hat{t}''_{yab}) & (\hat{N}'_{ab} - \hat{N}''_{ab}) \end{bmatrix} \begin{bmatrix} 1 \\ p_1 \\ p_2 \end{bmatrix}$$

Notando los términos del vector fila como:

$$\begin{aligned}
X_1 &= \hat{t}_{ya} + \hat{t}_{yb} + \hat{t}'_{yab} \\
X_2 &= \hat{t}'_{yab} - \hat{t}''_{yab} \\
X_3 &= \hat{N}'_{ab} - \hat{N}''_{ab}
\end{aligned}$$

entonces se tiene:

$$Var[\hat{t}_{yFB}] = \begin{bmatrix} 1 & p_1 & p_2 \end{bmatrix} \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & Cov(X_1, X_3) \\ Cov(X_1, X_2) & Var(X_2) & Cov(X_2, X_3) \\ Cov(X_1, X_3) & Cov(X_2, X_3) & Var(X_3) \end{bmatrix} \begin{bmatrix} 1 \\ p_1 \\ p_2 \end{bmatrix}$$

$$\begin{aligned}
Var[\hat{t}_{yFB}] &= Var(X_1) + 2p_1Cov(X_1, X_2) + 2p_2Cov(X_1, X_3) + p_1^2Var(X_2) \\
&\quad + 2p_1p_2Cov(X_2, X_3) + p_2^2Var(X_3)
\end{aligned}$$

Ahora se encuentra los valores de p_1 y p_2 tal que se minimice $Var[\hat{t}_{yFB}]$ es decir:

$$\begin{aligned}
min_{p_1, p_2} \{Var[\hat{t}_{yFB}]\} &= min_{p_1, p_2} \{Var(X_1) + 2p_1Cov(X_1, X_2) + 2p_2Cov(X_1, X_3) \\
&\quad + p_1^2Var(X_2) + 2p_1p_2Cov(X_2, X_3) + p_2^2Var(X_3)\}
\end{aligned}$$

derivando parcialmente con respecto a p_1 se tiene:

$$\frac{\partial Var[\hat{t}_{yFB}]}{\partial p_1} = 2Cov(X_1, X_2) + 2p_1Var(X_2) + 2p_2Cov(X_2, X_3)$$

igualando a 0 y despejando p_1 se obtiene:

$$p_1 = \frac{-Cov(X_1, X_2) - p_2Cov(X_2, X_3)}{Var(X_2)}$$

Análogamente derivando parcialmente con respecto a p_2 se tiene:

$$\frac{\partial Var[\hat{t}_{yFB}]}{\partial p_2} = 2Cov(X_1, X_3) + 2p_1Cov(X_2, X_3) + 2p_2Var(X_3)$$

igualando a 0 y despejando p_2 se obtiene:

$$p_2 = \frac{-Cov(X_1, X_3) - p_1Cov(X_2, X_3)}{Var(X_3)}$$

reemplazando p_2 se tiene:

$$p_{1opt} = \frac{Cov(X_1, X_2)Var(X_3) - Cov(X_1, X_3)Cov(X_2, X_3)}{Cov^2(X_2, X_3) - Var(X_3)Var(X_2)}$$

luego reemplazando p_1 se tiene:

$$p_{2opt} = \frac{-Cov(X_1, X_2)Cov(X_2, X_3) + Var(X_2)Cov(X_1, X_3)}{Cov^2(X_2, X_3) - Var(X_3)Var(X_2)}$$

Con un poco de álgebra adicional se llega a una expresión más compacta:

$$\begin{bmatrix} p_{1opt} \\ p_{2opt} \end{bmatrix} = - \left(Cov \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \right)^{-1} \begin{bmatrix} Cov(X_1, X_2) \\ Cov(X_1, X_3) \end{bmatrix}$$

finalmente reemplazando los valores de X_1 , X_2 y X_3 resulta:

$$\begin{bmatrix} p_{1opt} \\ p_{2opt} \end{bmatrix} = - \left(Cov \begin{bmatrix} \hat{t}'_{yab} - \hat{t}''_{yab} \\ \hat{N}'_{ab} - \hat{N}''_{ab} \end{bmatrix} \right)^{-1} \begin{bmatrix} Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}''_{yab}, \hat{t}'_{yab} - \hat{t}''_{yab}) \\ Cov(\hat{t}_{ya} + \hat{t}_{yb} + \hat{t}''_{yab}, \hat{N}'_{ab} - \hat{N}''_{ab}) \end{bmatrix} \square$$

Al igual que para el p óptimo del estimador de Hartley, los valores p_1 y p_2 óptimos incluyen covarianzas poblacionales, así que para poder encontrar estos valores en la práctica, se estiman dichas covarianzas con los datos de las muestras seleccionadas.

Ahora supongamos que se extraen muestras mediante el diseño MAS en ambos marcos y desconocemos N_{ab} , entonces tenemos:

$$\hat{t}_{y_{FB}} = (N_A - \hat{N}_{ab})\bar{y}_a + \hat{N}_{ab}(\tilde{y}_{ab}) + (N_B - \hat{N}_{ab})\bar{y}_b$$

donde:

$$\begin{aligned}\tilde{y}_{ab} &= w\tilde{y}'_{ab} + (1-w)\tilde{y}''_{ab} \\ w &= \frac{n'_{ab}(1-f_B)}{n'_{ab}(1-f_B) + n''_{ab}(1-f_A)}\end{aligned}$$

En este caso la varianza del estimador viene dada por:

$$\begin{aligned}Var(\hat{t}_{y_{FB}}) &= N_a(f_A^{-1} - 1)\sigma_a^2 + \frac{(1-f_A)(1-f_B)N_{ab}\sigma_{ab}^2}{(1-f_B)f_A + (1-f_A)f_B} \\ &+ N_b(f_B^{-1} - 1)\sigma_b^2 + (\mu_{y_{ab}} - \mu_{y_a} - \mu_{y_b})\delta\end{aligned}$$

donde:

$$\begin{aligned}\delta &= \frac{N_{ab}N_aN_bg_Ag_B}{n_A N_b g_B + n_B N_a g_A} \\ g_A &= \frac{N_A - n_A}{N_A - 1} \\ g_B &= \frac{N_B - n_B}{N_B - 1} \\ f_A &= \frac{n_A}{N_A} \\ f_B &= \frac{n_B}{N_B}\end{aligned}$$

y \hat{N}_{ab} es el valor más pequeño de la solución de la ecuación cuadrática:

$$\begin{aligned}(n_{AgB} + n_{BgA})\hat{N}_{ab}^2 - (n_A N_B g_B + n_B N_A g_A + n'_{ab} N_A g_B \\ + n''_{ab} N_B g_A)\hat{N}_{ab} + (n'_{ab} g_B + n''_{ab} g_A)N_A N_B = 0\end{aligned}$$

1.2.3. Estimador Pseudo Máximo Verosímil

Skinner & Rao (1996) propusieron un nuevo estimador llamado el estimador Pseudo-Máximo-Verosímil (PMV), el cual para diseños complejos como diseños de conglomerados o en más de una etapa con estratificación permite ajustar las estimaciones de los totales en los subgrupos al total general, el estimador viene dado por:

$$\hat{t}_{yPMV}(p) = \frac{N_A - \hat{N}_{ab}^{PMV}}{\hat{N}_a} \hat{t}_{ya} + \frac{\hat{N}_{ab}^{PMV}}{\theta \hat{N}'_{ab} + (1 - \theta) \hat{N}''_{ab}} [\theta \hat{t}'_{yab} + (1 - \theta) \hat{t}''_{yab}] \\ + \frac{N_B - \hat{N}_{ab}^{PMV}}{\hat{N}_b} \hat{t}_{yb}$$

donde \hat{N}_{ab}^{PMV} es la raíz más pequeña de la ecuación cuadrática:

$$(n_A + n_B) \hat{N}_{ab}^{2PMV} - (n_A N_B + n_B N_A + n_A \hat{N}'_{ab} + n_B \hat{N}''_{ab}) \hat{N}_{ab}^{PMV} \\ + n_A \hat{N}'_{ab} N_B + n_B \hat{N}''_{ab} N_A = 0$$

A diferencia del estimador de Hartley y el estimador de Fuller-Burmeister el cálculo de la esperanza y la varianza no es directo, ya que el estimador PMV no es lineal porque incluye el cociente entre estimadores, por esto se debe recurrir a otros métodos para su cálculo. Teniendo en cuenta lo anterior se busca el p óptimo que minimice la varianza asintótica de \hat{N}_{ab}^{PMV} , ya que si realiza la minimización con respecto a la varianza asintótica de $\hat{t}_{yPMV}(p)$ el estimador PMV tendrá el mismo problema que los estimadores anteriores y es que los pesos muestrales dependerán de la variable de interés y que se midió.

El valor de θ que minimiza la varianza asintótica de \hat{N}_{ab}^{PMV} viene dado por (Lohr & Rao 2000):

$$\theta_{opt} = \frac{N_a N_B V(\hat{N}''_{ab})}{N_a N_B V(\hat{N}''_{ab}) + N_b N_A V(\hat{N}'_{ab})}$$

Prueba: Como se mencionó, la estimación de la varianza para el estimador PMV es más complicada que para los otros dos estimadores mencionados, puesto que al no ser una función lineal de estimadores se debe calcular mediante métodos de linealización para su cálculo. Skinner & Rao (1996) realizan el cálculo de la varianza aproximada del estimador PMV mediante el método delta.

Al aplicar el método delta se encuentra que el estimador PMV se puede expresar como:

$$\hat{t}_{yPMV} \approx \hat{t}_y + N \nabla^T (\hat{\eta} - \eta)$$

donde:

$$\begin{aligned}
N &= N_a + N_b + N_{ab} \\
\eta &= \begin{bmatrix} \mu_{ya} \\ \mu_{yab} \\ \frac{N_{ab}}{N} \\ \mu_{yb} \\ \mu_{yab} \\ \frac{N_{ab}}{N} \end{bmatrix} \\
\nabla &= \begin{bmatrix} \frac{N_a}{N} \\ \frac{\theta N_{ab}}{N} \\ \phi(\mu_{yab} - \mu_{ya} - \mu_{yb}) \\ \frac{N_b}{N} \\ \frac{(1-\theta)N_{ab}}{N} \\ (1-\phi)(\mu_{yab} - \mu_{ya} - \mu_{yb}) \end{bmatrix}
\end{aligned}$$

luego la varianza aproximada del estimador PMV viene dada por:

$$\begin{aligned}
aVar[\hat{t}_{yPMV}] &= Var[t_y + N\nabla^T(\hat{\eta} - \eta)] \\
&= Var[N\nabla^T\hat{\eta}] \\
&= Var(N_a\hat{\mu}_{ya} + \theta N_{ab}\hat{\mu}'_{yab} + \lambda\phi\hat{N}'_{ab}) \\
&+ Var(N_b\hat{\mu}_{yb} + N_{ab}(1-\theta)\hat{\mu}''_{yab} + \lambda(1-\phi)\hat{N}''_{ab}) \quad (1.1) \\
&= V_A + V_B
\end{aligned}$$

con:

$$\begin{aligned}
\lambda &= \mu_{yab} - \mu_{ya} - \mu_{yb} \\
\phi &= \frac{n_A N_b}{n_A N_b + n_B N_a} \quad (1.2)
\end{aligned}$$

$$\theta = \frac{n_A N_B}{n_A N_B + n_B N_A} \quad (1.3)$$

de estas ecuaciones se aprecia que hay una expresión de θ que se puede utilizar en el estimador PMV, pero al igual que para los estimadores de Hartley y de Fuller-Burmeister se busca el valor óptimo de θ que minimice en este caso la varianza de \hat{N}_{ab}^{PMV} . Utilizando la siguiente variable indicadora $y_i = 1$ si $i \in ab$ y $y_i = 0$ si $i \notin ab$ se tiene que $\hat{t}_{yPMV} = \hat{N}_{ab}^{PMV}$ y reemplazando en la ecuación 1.1 se obtiene:

$$\begin{aligned}
aVar[\hat{N}_{ab}^{PMV}] &= Var[\theta N_{ab} + \lambda\phi\hat{N}'_{ab}] + Var[(1-\theta)N_{ab} + \lambda(1-\phi)\hat{N}''_{ab}] \\
&= \lambda^2\phi^2 Var(\hat{N}'_{ab}) + \lambda^2(1-\phi)^2 Var(\hat{N}''_{ab})
\end{aligned}$$

Ahora minimizando con respecto a ϕ se tiene que:

$$\min_{\phi}\{a\text{Var}[\hat{N}_{ab}^{PMV}]\} = \min_{\phi}\{\lambda^2\phi^2\text{Var}(\hat{N}'_{ab}) + \lambda^2(1-\phi)^2\text{Var}(\hat{N}''_{ab})\}$$

derivando con respecto a ϕ se tiene:

$$\frac{d[\lambda^2\phi^2\text{Var}(\hat{N}'_{ab}) + \lambda^2(1-\phi)^2\text{Var}(\hat{N}''_{ab})]}{d\phi} = 2\lambda^2\phi\text{Var}(\hat{N}'_{ab}) - 2\lambda^2(1-\phi)\text{Var}(\hat{N}''_{ab})$$

al igualar a 0 y despejar ϕ se obtiene:

$$\phi_{opt} = \frac{\text{Var}(\hat{N}''_{ab})}{\text{Var}(\hat{N}'_{ab}) + \text{Var}(\hat{N}''_{ab})} \quad (1.4)$$

ahora, utilizando las expresiones para ϕ y θ de las ecuaciones 1.2 y 1.3 se obtiene la siguiente relación:

$$\phi = \frac{\theta N_b}{\theta N_b + (1-\theta)\frac{N_B N_a}{N_A}} \quad (1.5)$$

y al igualar las ecuaciones 1.4 y 1.5 se tiene:

$$\begin{aligned} \frac{\text{Var}(\hat{N}''_{ab})}{\text{Var}(\hat{N}'_{ab}) + \text{Var}(\hat{N}''_{ab})} &= \frac{\theta N_b}{\theta N_b + (1-\theta)\frac{N_B N_a}{N_A}} \\ \frac{-N_b N_a \text{Var}(\hat{N}''_{ab})}{N_A} &= \theta [N_b \text{Var}(\hat{N}''_{ab}) - \frac{N_B N_a}{N_A} \text{Var}(\hat{N}''_{ab}) - N_b \text{Var}(\hat{N}'_{ab}) - N_b \text{Var}(\hat{N}''_{ab})] \end{aligned}$$

y despejando θ se obtiene el valor óptimo:

$$\begin{aligned} \theta_{opt} &= \frac{-N_B N_a \text{Var}(\hat{N}''_{ab})}{N_A N_b \text{Var}(\hat{N}''_{ab}) - N_B N_a \text{Var}(\hat{N}''_{ab}) - N_A N_b \text{Var}(\hat{N}'_{ab}) - N_A N_b \text{Var}(\hat{N}''_{ab})} \\ \theta_{opt} &= \frac{N_B N_a \text{Var}(\hat{N}''_{ab})}{N_B N_a \text{Var}(\hat{N}''_{ab}) + N_A N_b \text{Var}(\hat{N}'_{ab})} \square \end{aligned}$$

En la práctica las varianzas necesarias para el cálculo del θ óptimo son desconocidas, entonces se deben estimar con los datos muestrales. En la ecuación 1.1 se puede observar que no es posible calcular directamente la varianza aproximada para el estimador PMV, por eso se realiza una reparametrización para el marco A, haciendo $z_{Ai} = y_i - \mu_{ya}$ si i pertenece al dominio S_a y $z_{Ai} = \theta(y_i - \mu_{yab}) + \lambda\phi$ si i pertenece al dominio S_{ab} . Análogamente para el marco B se hace $z_{Bi} = y_i - \mu_{yb}$ si i pertenece al dominio S_b y $z_{Bi} = (1-\theta)(y_i - \mu_{yab}) + \lambda(1-\phi)$ si i pertenece al dominio S_{ab} .

Al remplazar en la varianza aproximada del estimador PMV se obtiene:

$$\begin{aligned} a\text{Var}[\hat{t}_{yPMV}] &= V_A + V_B \\ &= \text{Var}(z_{Ai}) + \text{Var}(z_{Bi}) \end{aligned}$$

donde $\text{Var}(z_{Ai})$ y $\text{Var}(z_{Bi})$ son las varianzas de las nuevas variables z_{Ai} y z_{Bi} bajo los diseños muestrales aplicados en los marcos A y B respectivamente; de esta expresión podemos apreciar que a diferencia de los estimadores de Hartley y Fuller-Burmeister, la aproximación de la varianza para el estimador PMV tiene la ventaja de no involucrar las covarianzas entre dominios.

Ahora supongamos que se extraen muestras MAS en ambos marcos y desconocemos N_{ab} . Entonces tenemos:

$$\begin{aligned} \hat{t}_{yPMV} &= (N_A - \hat{N}_{ab}^{PMV})\bar{y}_a + \frac{\hat{N}_{ab}^{PMV}}{p\hat{N}'_{ab} + (1-p)\hat{N}''_{ab}} [p\hat{N}'_{ab}\bar{y}'_{ab} + (1-p)\hat{N}''_{ab}\bar{y}''_{ab}] \\ &\quad + (N_B - \hat{N}_{ab}^{PMV})\bar{y}_b \end{aligned}$$

Metodología

Con el fin de comparar y evaluar los estimadores para marcos duales mencionados anteriormente, se tomarán como referencia los datos de la Encuesta Nacional Agropecuaria - ENA y la encuesta de hortalizas realizadas por el DANE, y se tendrán en cuenta los diseños muestrales que existen en la actualidad para dichas encuestas. A continuación se presentan los diseños muestrales de la ENA y de la encuesta de hortalizas realizados en la actualidad.

2.1. Diseños muestrales actuales

2.1.1. Diseño muestral de la ENA

El diseño muestral de la ENA se basa en el documento denominado "Modelo de costos y diseño muestral del sistema de información agropecuaria para la estimación de áreas producción y rendimiento" elaborado por Bautista (1994), donde se muestra de forma clara y concisa la metodología de estimación con base en muestreo de áreas. Este diseño es en dos etapas; el marco muestral es un marco de áreas con una cobertura de 37.900.546 hectáreas (169.587 segmentos en 31.588 UPMS).

Las unidades de la primera etapa son segmentos de tierra denominadas UPM (Unidades Primarias de Muestreo) de distinto tamaño; a su vez, cada UPM seleccionada en la muestra es dividida en segmentos, los cuales constituyen las unidades de la segunda etapa. El método de selección es πPT en la primera etapa y πPT en la segunda etapa, donde la variable auxiliar para la primera etapa es el área planimetrada de la UPM. Existen varios tipos de estratificación, los más importantes son por tipo de cobertura vegetal y piso térmico.

La ENA se realiza dos veces durante el año. Para el año 2010 el tamaño de muestra para el primer semestre fue de 2.537 segmentos que equivale a 293.252 hectáreas, y 5.894 segmentos que equivalen a 959.767 hectáreas para el segundo semestre (Corporación Colombia Internacional - CCI 2010). Los factores de expansión se calculan de la forma tradicional para un diseño $\pi PT - \pi PT$ estratificado (Särndal et al. 1992):

$$f_{hi\alpha} = \left(\frac{x_{hi}A_h}{X_h} \right) \left(\frac{x_{hi\alpha}a_{hi}}{x_{hi}} \right)$$

donde:

X_h : es el área planimetrada del estrato h.

x_{hi} : es el área planimetrada de la UPM i en el estrato h.

$x_{hi\alpha}$: es el área planimetrada del segmento α en la UPM i en el estrato h.

A_h : es la cantidad de UPMS seleccionadas en el estrato h.

a_{hi} : es la cantidad de segmentos seleccionados en la UPM i en el estrato h.

2.1.2. Diseño muestral para hortalizas

Paralelamente a la ENA se realizan las encuestas denominadas por productos dentro de los que se encuentran el maíz, la piscicultura, las hortalizas, entre otras. Aunque inicialmente el diseño muestral para la ENA se había planificado para el uso de técnicas para marcos duales, es decir usando el marco de áreas y su traslape con los marcos por productos; en este momento lo que se hace es extraer lo que está en los marcos por productos del marco de áreas y trabajar como si fueran estratos (escenario 1).

Teniendo en cuenta lo anterior la propuesta que se plantea es usar los datos de la encuesta de hortalizas para realizar simulaciones de Monte Carlo y evaluar el comportamiento de los 3 estimadores bajo los escenarios 2 y 3. El marco muestral de hortalizas tiene una cobertura de 10.874 hectáreas (distribuidas en 588 segmentos), el diseño muestral para la encuesta de hortalizas es *MAS* de segmentos y los criterios de estratificación son los mismos que para la ENA. El tamaño de muestra para la encuesta de hortalizas de 2010 es de 95 segmentos que equivalen a 1.735 hectáreas.

Adicionalmente se considerarán otros diseños muestrales correspondientes a todas las combinaciones entre *MAS* y πPT para cada una de las etapas de la ENA y para hortalizas:

- ENA *MAS* – *MAS* y hortalizas *MAS*.
- ENA *MAS* – *MAS* y hortalizas πPT .
- ENA πPT – πPT y hortalizas *MAS*.
- ENA πPT – πPT y hortalizas πPT .
- ENA *MAS* – πPT y hortalizas *MAS*.
- ENA *MAS* – πPT y hortalizas πPT .
- ENA πPT – *MAS* y hortalizas *MAS*.
- ENA πPT – *MAS* y hortalizas πPT .

Al utilizar πPT en la primera etapa (ENA) se toma como variable auxiliar el área planimetrada de la UPM, y para πPT en la segunda etapa (ENA u hortalizas) se toma como variable auxiliar el área planimetrada del segmento.

2.2. Estudio de simulación

Para evaluar los resultados los estimadores de Hartley, Fuller-Burmeister y Skinner-Rao (PMV) se realizó un estudio de simulación por el método de Monte Carlo con $M=1000$ repeticiones, y se consideraron como medidas de precisión y eficiencia los valores del sesgo relativo (SR), el error cuadrático medio empírico (ECME), las respectivas varianzas estimadas (s^2), el sesgo relativo de la varianza (SRV) y la probabilidad de cobertura (PC).

2.2.1. Sesgo relativo

El sesgo relativo se calculará como:

$$SR = \frac{1}{M} \sum_{m=1}^M \frac{\hat{t}_{y,m} - t_y}{t_y}$$

donde t_y es el total poblacional para la variable de interés, $\hat{t}_{y,m}$ es el total estimado \hat{t}_y para la m -ésima repetición. Igualmente se hizo para el valor de sesgo relativo de la varianza SRV, en ese caso las varianzas estimadas (s^2) se calcularon según lo expuesto en el capítulo anterior.

2.2.2. Error cuadrático medio empírico

El error cuadrático medio empírico (ECME) se calculará como:

$$ECME = \frac{1}{M} \sum_{m=1}^M (\hat{t}_{y,m} - t_y)^2$$

2.2.3. Probabilidad de cobertura

Definiendo un intervalo de confianza para el total de la variable de interés y bajo la aproximación Normal como:

$$IC(m) : \hat{t}_{y,m} \pm z_{1-\alpha/2} \sqrt{Var(\hat{t}_{y,m})}$$

donde $z_{1-\alpha/2}$ es el percentil $1 - \alpha/2$ de la distribución Normal estándar.

Notando la variable indicadora $\gamma(m)$ como, 1 si t_y pertenece a $IC(m)$ y 0 en caso contrario, entonces la probabilidad de cobertura se calculará como:

$$PC = 100\% \frac{1}{M} \sum_{m=1}^M \gamma(m)$$

2.2.4. Procedimiento de simulación

El procedimiento de simulación para llevar a cabo el cálculo del sesgo relativo, el error cuadrático medio empírico y la probabilidad de cobertura de los estimadores se realizó en los siguientes pasos:

1. De la información disponible de los marcos de la ENA (marco A) y la encuesta de hortalizas (marco B), se utilizó como variables para el análisis el área sembrada. Con dicha información se planteó en primera instancia un modelo para la generación de los datos, siguiendo la misma línea de trabajo de Skinner & Rao (1996); en segunda instancia se utilizó la información como provenía directamente de las encuestas, sin ninguna transformación, a manera de caso aplicado.
2. Determinar los tamaños de muestra a seleccionar de cada uno de los marcos. Para los datos provenientes del modelo propuesto y para los datos sin transformar se tomaron para el marco A dos tamaños de muestra: en la primera etapa 1000 (2.5 %) y 1750 (5 %) UPMS. En la segunda etapa del marco A se plantearon fracciones muestrales de 40 %, 60 %, 80 % y 100 % (conglomerados) para los datos del modelo y del 100 % para el caso aplicado. Para el marco B se tomaron tamaños de muestra de 59 (10 %), 160 (20 %) y 117 (30 %) tanto para los datos simulados como para los originales.
3. Extraer las muestras de cada uno de los marcos teniendo en cuenta los 8 diseños muestrales propuestos. Para esto se usaron algunas funciones de las librerías *sampling* (Tillé, Y. & Matie, A. 2009) y *TeachingSampling* (Gutiérrez, A. 2010) del software estadístico R (R Development Core Team 2007), en donde se encuentran rutinas para selección de muestra en una o dos etapas, para los diseños *MAS* y πPT .
4. Calcular los totales por dominios (escenario 3) o postestratos (escenario 2) con sus respectivas varianzas estimadas, que son necesarios a la hora de implementar los estimadores propuestos y sus varianzas estimadas, para ello se utilizaron algunas funciones de la librería *survey* (Lumley, T. 2010) de R.
5. Calcular las covarianzas por dominios (escenario 3) o postestratos (escenario 2) necesarias para estimar las varianzas de los estimadores de Hartley y Fuller-Burmeister.
6. Calcular el p óptimo para el estimador de Hartley, los valores p_1 y p_2 óptimos para el estimador de Fuller-Burmeister y el θ óptimo para el estimador PMV, a utilizar en la estimación de totales y varianzas estimadas de totales.
7. Calcular los totales con sus respectivas varianzas estimadas bajo cada una de las alternativas los estimadores.
8. Repetir los pasos 2 a 7 en 1000 ocasiones.
9. Estimar los correspondientes sesgos relativos, errores cuadráticos medios (con las respectivas varianzas estimadas) y las probabilidades de cobertura para cada uno de los estimadores, bajo las 1000 repeticiones.

2.2.5. Modelo de generación de datos

Siguiendo la misma línea de trabajo propuesta por Lohr & Rao (2006) para la simulación de datos, se genera un modelo de efectos aleatorios con la siguiente estructura:

$$y_{ij} = \mu_k + \alpha_i + \epsilon_{ij}$$

donde μ_k es la media de cada uno de los dominios, α_i es una variable aleatoria normal estándar, ϵ_{ij} es una variable aleatoria normal con media cero y la varianza de los datos de área sembrada, k denota el número de dominios ($k = 1, 2, 3$), i denota el número de marcos ($i = 1, 2$) y j denota el número de elementos en cada dominio ($j = 1, \dots, N_k$), y_{ij} es el valor simulado para la variable de interés y_k para el marco i con un número de elementos j para cada dominio.

Resultados

De acuerdo a la metodología planteada se obtuvieron los siguientes resultados para los datos bajo el modelo planteado y para el caso aplicado.

3.1. Modelo

Bajo el modelo establecido se calculó el sesgo relativo (SR) de los tres estimadores, el error cuadrático medio empírico (ECME) y la varianza (s^2), el sesgo relativo del estimador de la varianza (SRV) y la probabilidad de cobertura (PC) de los tres estimadores. En las tablas 3.1, 3.2, 3.3 y 3.4 se presentan los resultados de la simulación bajo los escenarios 2 y 3, con fracciones muestrales en la segunda etapa del marco A del 40 %, 60 %, 80 % y 100 % respectivamente (en este último caso el número de combinaciones de diseños muestrales se reduce a 4 por hacerse censo en la segunda etapa del marco A).

Con una fracción muestral del 40 % para el marco A (tabla 3.1) los valores de SR para los tres estimadores bajo el escenario 2 son inferiores al 1 % confirmando el insesgamiento de los estimadores de Hartley y Fuller-Burmeister, y el insesgamiento asintótico del estimador Pseudo Máximo Verosímil; para esta misma fracción muestral bajo el escenario 3, se encuentra que los SR para todos los diseños muestrales no superan el 1 %. Bajo el escenario 2 los SR más cercanos a cero se presentan para los diseños muestrales $MAS - MAS - MAS$ y $\pi PT - \pi PT - MAS$, para el escenario 3 los SR más pequeños se obtuvieron para los diseños $MAS - MAS - MAS$ y $MAS - \pi PT - \pi PT$.

Para los dos escenarios, los menores valores de ECME y s^2 con una fracción muestral del 40 %, se presentan para la combinación de diseños con MAS en la primera etapa del marco A; es de destacar que al aumentarse el tamaño de muestra en la primera etapa del marco A disminuye el ECME y s^2 , sin embargo, esto no sucede al aumentar el tamaño bajo el diseño muestral para el marco B. Al comparar los tres estimadores bajo el escenario 2, en un número importante de casos el estimador que tiene valores más altos de ECME es el de Fuller-Burmeister, y los estimadores de Hartley y Pseudo Máximo Verosímil tienen valores de ECME en su mayoría cercanos entre si, comparativamente con respecto al escenario 3 el comportamiento es distinto y en muchas de las combinaciones los ECME para los estimadores Pseudo Máximo Verosímil y Fuller-Burmeister son menores que para el estimador de Hartley.

TABLA 3.1. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico (ECME $\times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza (SRV $\times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 40% en la segunda etapa del marco A.

TAMAÑOS DE MUESTRA			ESCENARIO 2									ESCENARIO 3																				
MUESTRA			$SR \times 10^{-3}$			ECME $\times 10^8$ ($s^2 \times 10^8$)			SRV $\times 10^{-2}$ (PC)			$SR \times 10^{-3}$			ECME $\times 10^8$ ($s^2 \times 10^8$)			SRV $\times 10^{-2}$ (PC)														
n_{IA}	$n_{IIA}(\%)$	n_B	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML												
MAS	MAS	MAS																														
1000	40	59	0.01	0.01	-0.11	12.06	(12.06)	12.06	(12.06)	12.04	(12.04)	0.41	(0.96)	0.41	(0.96)	0.65	(0.96)	0.19	-0.46	-0.50	11.41	(11.41)	11.41	(11.41)	11.39	(11.39)	6.55	(0.96)	6.57	(0.96)	6.07	(0.96)
1000	40	117	1.42	1.42	1.28	11.99	(11.99)	11.99	(11.99)	11.99	(11.99)	1.01	(0.95)	1.00	(0.95)	1.03	(0.95)	-1.20	-1.96	-1.97	12.25	(12.25)	12.25	(12.24)	12.20	(12.20)	-0.77	(0.94)	-0.72	(0.94)	-1.02	(0.95)
1000	40	160	-1.83	-1.83	-1.92	12.22	(12.21)	12.22	(12.21)	12.22	(12.22)	-0.94	(0.94)	-0.94	(0.94)	-1.04	(0.94)	0.82	0.02	0.18	11.51	(11.51)	11.48	(11.48)	11.46	(11.46)	5.45	(0.95)	5.65	(0.95)	5.23	(0.95)
1750	40	59	0.30	0.30	0.34	6.61	(6.61)	6.61	(6.61)	6.61	(6.61)	4.80	(0.96)	4.78	(0.96)	4.77	(0.96)	2.18	1.80	1.74	6.83	(6.83)	6.83	(6.83)	6.81	(6.81)	2.01	(0.95)	1.92	(0.95)	1.55	(0.96)
1750	40	117	-1.97	-1.97	-1.75	6.56	(6.55)	6.56	(6.55)	6.55	(6.55)	5.36	(0.96)	5.35	(0.96)	5.46	(0.96)	3.45	3.02	2.96	6.58	(6.56)	6.57	(6.56)	6.57	(6.56)	5.97	(0.96)	6.08	(0.96)	5.39	(0.97)
1750	40	160	-1.38	-1.38	-1.86	6.55	(6.55)	6.55	(6.55)	6.56	(6.56)	5.75	(0.96)	5.75	(0.96)	5.52	(0.96)	-0.97	-1.42	-1.33	6.31	(6.31)	6.30	(6.30)	6.31	(6.31)	10.01	(0.96)	10.18	(0.96)	9.30	(0.97)
MAS	MAS	π PT																														
1000	40	59	4.60	3.67	4.20	12.59	(12.56)	13.42	(13.40)	12.58	(12.56)	-3.50	(0.94)	-8.03	(0.94)	-0.87	(0.94)	3.52	2.96	2.90	11.92	(11.91)	11.92	(11.91)	11.94	(11.93)	1.80	(0.96)	1.81	(0.96)	0.94	(0.96)
1000	40	117	-4.28	-0.64	-4.69	12.12	(12.10)	21.01	(21.01)	12.12	(12.09)	0.30	(0.96)	-41.66	(0.96)	12.26	(0.96)	-3.79	-4.40	-4.50	11.86	(11.85)	11.84	(11.82)	11.84	(11.81)	2.77	(0.96)	3.01	(0.96)	2.38	(0.96)
1000	40	160	-0.63	0.03	-1.15	11.62	(11.62)	12.13	(12.13)	11.62	(11.61)	4.16	(0.96)	0.76	(0.96)	6.39	(0.96)	2.34	1.64	1.65	12.63	(12.63)	12.63	(12.62)	12.60	(12.60)	-3.37	(0.95)	-3.39	(0.95)	-3.85	(0.95)
1750	40	59	-1.63	-2.83	-1.68	7.19	(7.18)	8.13	(8.12)	7.17	(7.17)	-3.54	(0.94)	2.90	(0.94)	18.35	(0.95)	0.70	0.37	0.18	7.18	(7.18)	7.17	(7.17)	7.15	(7.15)	-3.29	(0.94)	-3.29	(0.95)	-3.60	(0.94)
1750	40	117	-1.64	-1.06	-1.80	6.70	(6.69)	7.13	(7.13)	6.69	(6.68)	3.56	(0.95)	1.75	(0.95)	17.45	(0.95)	0.55	0.17	0.14	7.26	(7.26)	7.26	(7.26)	7.25	(7.25)	-4.32	(0.93)	-4.31	(0.93)	-4.79	(0.93)
1750	40	160	1.35	-7.16	1.31	6.46	(6.46)	9.52	(9.46)	6.47	(6.47)	7.86	(0.95)	-10.06	(0.95)	23.74	(0.95)	-1.90	-2.33	-2.37	6.78	(6.78)	6.78	(6.77)	6.78	(6.77)	2.45	(0.96)	2.52	(0.96)	1.84	(0.96)
MAS	π PT	MAS																														
1000	40	59	3.81	3.81	3.30	10.73	(10.71)	10.73	(10.71)	10.71	(10.70)	12.86	(0.97)	12.85	(0.97)	12.95	(0.97)	8.20	7.51	7.56	11.97	(11.89)	11.94	(11.88)	11.94	(11.88)	2.36	(0.96)	2.44	(0.96)	1.71	(0.96)
1000	40	117	5.16	5.16	4.57	12.36	(12.33)	12.36	(12.33)	12.35	(12.33)	-1.80	(0.95)	-1.80	(0.95)	-1.77	(0.95)	1.79	1.06	1.03	12.05	(12.05)	12.03	(12.03)	12.05	(12.04)	0.76	(0.95)	0.86	(0.96)	0.08	(0.95)
1000	40	160	1.60	1.60	1.05	12.58	(12.58)	12.58	(12.58)	12.58	(12.58)	-3.70	(0.94)	-3.70	(0.94)	-3.72	(0.94)	2.70	1.83	1.95	12.15	(12.14)	12.12	(12.12)	12.13	(12.13)	0.48	(0.96)	0.66	(0.96)	-0.12	(0.95)
1750	40	59	-2.96	-2.96	-3.10	6.70	(6.69)	6.70	(6.69)	6.70	(6.69)	3.53	(0.96)	3.51	(0.96)	3.58	(0.96)	2.71	2.33	2.30	6.64	(6.63)	6.64	(6.63)	6.63	(6.63)	4.72	(0.95)	4.75	(0.95)	4.12	(0.95)
1750	40	117	3.21	3.21	3.06	6.80	(6.79)	6.80	(6.79)	6.80	(6.79)	2.06	(0.95)	2.05	(0.95)	2.01	(0.96)	1.15	0.70	0.71	6.68	(6.67)	6.67	(6.67)	6.66	(6.66)	4.05	(0.95)	4.07	(0.95)	3.56	(0.95)
1750	40	160	-3.65	-3.65	-3.71	6.32	(6.30)	6.32	(6.30)	6.32	(6.30)	9.78	(0.96)	9.77	(0.96)	9.74	(0.96)	1.61	1.15	1.11	7.18	(7.18)	7.18	(7.18)	7.17	(7.17)	-3.33	(0.94)	-3.33	(0.94)	-3.84	(0.94)
MAS	π PT	π PT																														
1000	40	59	-4.59	-3.13	-4.87	11.84	(11.81)	13.33	(13.32)	11.84	(11.81)	2.44	(0.95)	-8.44	(0.95)	5.51	(0.95)	0.14	-0.45	-0.64	11.83	(11.83)	11.81	(11.81)	11.80	(11.80)	2.58	(0.95)	2.74	(0.95)	2.15	(0.95)
1000	40	117	3.56	4.15	3.27	12.68	(12.66)	12.89	(12.87)	12.66	(12.65)	-4.23	(0.94)	-4.23	(0.94)	-1.33	(0.94)	-0.28	-0.95	-1.02	11.10	(11.10)	11.08	(11.08)	11.07	(11.07)	9.67	(0.96)	9.79	(0.96)	9.24	(0.96)
1000	40	160	1.89	1.88	1.55	11.42	(11.42)	11.41	(11.41)	11.41	(11.41)	6.29	(0.97)	6.50	(0.97)	6.56	(0.97)	1.40	0.73	0.69	11.83	(11.83)	11.81	(11.81)	11.81	(11.81)	3.13	(0.95)	3.24	(0.95)	2.65	(0.95)
1750	40	59	0.47	5.21	0.36	7.15	(7.15)	16.99	(16.96)	7.14	(7.14)	-3.15	(0.95)	-23.81	(0.95)	90.38	(0.95)	1.45	1.14	1.05	6.32	(6.32)	6.31	(6.31)	6.31	(6.30)	9.94	(0.96)	9.94	(0.96)	9.37	(0.96)
1750	40	117	-3.72	-3.74	-3.81	6.67	(6.66)	6.66	(6.65)	6.66	(6.64)	3.89	(0.95)	4.08	(0.95)	4.14	(0.95)	2.39	2.00	2.03	7.11	(7.11)	7.10	(7.10)	7.11	(7.11)	-2.35	(0.95)	-2.29	(0.95)	-3.05	(0.95)
1750	40	160	-1.96	-1.80	-2.06	6.87	(6.86)	6.90	(6.89)	6.86	(6.86)	0.89	(0.94)	0.53	(0.94)	0.96	(0.94)	1.38	0.95	0.89	7.17	(7.17)	7.16	(7.16)	7.15	(7.15)	-2.84	(0.95)	-2.75	(0.95)	-3.17	(0.95)
π PT	MAS	MAS																														
1000	40	59	-3.93	-3.93	-4.38	17.26	(17.24)	17.26	(17.24)	17.28	(17.26)	1.90	(0.95)	1.89	(0.95)	2.18	(0.95)	-3.07	-3.77	-3.68	17.68	(17.67)	17.66	(17.64)	17.63	(17.62)	3.44	(0.95)	3.57	(0.95)	-0.10	(0.95)
1000	40	117	4.72	4.72	4.22	16.80	(16.77)	16.80	(16.77)	16.83	(16.81)	4.98	(0.96)	4.98	(0.96)	4.71	(0.96)	-2.27	-3.04	-2.95	18.22	(18.21)	18.18	(18.16)	18.18	(18.17)	0.55	(0.96)	0.79	(0.96)	-3.01	(0.95)
1000	40	160	1.93	1.93	1.36	17.97	(17.97)	17.97	(17.97)	17.96	(17.96)	-1.71	(0.95)	-1.72	(0.95)	-1.64	(0.95)	-1.64	-2.53	-2.33	18.10	(18.09)	18.07	(18.06)	18.05	(18.04)	0.41	(0.95)	0.58	(0.95)	-3.16	(0.94)
1750	40	59	4.44	4.43	4.17	9.77	(9.74)	9.77	(9.74)	9.75	(9.73)	3.44	(0.96)	3.43	(0.96)	3.51	(0.96)	7.73	7.34	7.36	10.00	(9.93)	9.99	(9.93)	9.99	(9.92)	5.15	(0.96)	5.17	(0.96)	1.18	(0.95)
1750	40	117	0.27	0.27	-0.22	10.20	(10.20)	10.20	(10.20)	10.19	(10.19)	-1.43	(0.95)	-1.44	(0.95)	-1.58	(0.95)	2.14	1.72	1.78	10.38	(10.37)	10.37	(10.36)	10.35	(10.34)	0.60	(0.96)	0.65	(0.96)	-3.03	(0.95)
1750	40	160	-0.61	-0.62	-1.12	9.83	(9.83)	9.83	(9.83)	9.84	(9.84)	1.92	(0.95)	1.91	(0.95)	2.01	(0.96)	-3.74	-4.20	-4.12	9.15	(9.13)	9.14	(9.12)	9.15	(9.13)	14.62	(0.96)	14.79	(0.96)	10.49	(0.95)
π PT	MAS	π PT																														
1000	40	59	-2.15	-1.33	-2.22	17.90	(17.89)	18.46	(18.46)	17.90	(17.90)	-1.20	(0.95)	-0.21	(0.95)	2.37	(0.95)	1.87	1.26	1.36	17.47	(17.47)	17.46	(17.46)	17.42	(17.42)	4.58	(0.96)	4.61	(0.96)	0.75	(0.95)
1000	40	117	-0.63	-0.17	-0.89	17.23	(17.23)	17.80	(17.80)	17.22	(17.22)	2.24	(0.95)	-0.95	(0.95)	3.08	(0.95)	-1.46	-2.14	-2.16	19.70	(19.69)	19.65	(19.65)	19.67	(19.67)	-7.75	(0.94)	-7.55	(0.94)	-11.22	(0.94)
1000	40	160	-4.22	-4.46	-4.54	16.96	(16.93)	17.02	(17.00)	16.97	(16.94)	4.29	(0.95)	4.06	(0.95)	4.72	(0.95)	2.27	1.50	1.49	16.90	(16.89)	16.87	(16.87)	16.84	(16.84)	7.97	(0.96)	8.09	(0.96)	4.10	(0.95)
1750	40	59	3.47	3.57	3.39	8.89	(8.88)	8.98	(8.96)	8.91	(8.90)	13.19	(0.96)	14.28	(0.																	

Uno de los resultados más importantes de este documento es el cálculo del SRV y las respectivas PC para los tres estimadores, ya que son valores importantes para evaluar la precisión y eficiencia de los estimadores. Este tema no ha sido tratado a profundidad, puesto que existen análisis bajo escenarios que incluye información auxiliar; algunos trabajos importantes a destacar sobre este tema son los presentados por Rao & Changbao (2010) en el cual se propone un estimador denominado Pseudo Máximo Verosímil Empírico (PMVE) el cual es una modificación de la propuesta de Skinner-Rao y la propuesta presentada en los trabajos de Ferraz & Coelho (2009) y Coelho (2007) donde se plantean Estimadores de Regresión Generalizada - GREG para marcos duales.

En el trabajo de Rao & Changbao (2010) se evaluaron los intervalos de confianza para el estimador PMVE y el de Skinner y Rao, pero no se compararon los de Hartley y Fuller-Burmeister, algo que evalúa en este documento. Algunas de las causas de que estas comparaciones no se hayan hecho en el pasado pudieron haber sido ocasionadas por la dificultad que hasta hace algunos años representaba el cálculo de las covarianzas entre totales de dominios necesarias en las varianzas estimadas, y la capacidad y velocidad de los equipos de computo necesarios para tal fin.

Siguiendo con la tabla 3.1, se observa que en cuanto a SRV, en la mayoría de los casos bajo el escenario 2 los valores de SRV no exceden el 10 %, pero para algunos casos se presentan valores grandes que superan incluso el 50 %, además algo importante de notar es que en muchas de las combinaciones los SRV son positivos, es decir que para dichos casos se sobreestima la varianza. Comparando los tres estimadores no se observa que alguno destaque con mayor cantidad de valores de SRV cercanos a cero, pero si es de resaltar que el estimador de Hartley es el más estable en sus valores, puesto que en ninguno de los casos los SRV superan el 15 %; En cuanto a los diseños muestrales, las combinaciones que presentan valores más cercanos a cero son $MAS - MAS - MAS$ y $\pi PT - MAS - MAS$.

En el escenario 3 en ninguno de los casos los SRV exceden el 15 % y al igual que para el escenario 2 se tiende a sobreestimar la varianza, en general los tres estimadores se comportan similar en este aspecto con una ligera ventaja del estimador de Fuller-Burmeister, el cual en una cantidad importante de casos tiene menores valores que los demás, bajo este escenario cabe destacar los diseños muestrales $MAS - MAS - \pi PT$ y $MAS - \pi PT - MAS$ para los que los SVR no sobrepasan el 5 %. Para los dos escenarios no se aprecia una relación directa entre el aumento o disminución de los tamaños de muestra tanto del marco A como el B con respecto a los valores de SRV.

En cuanto a los valores de PC, para el escenario 2 los valores oscilan entre el 94 % y el 97 %; para el escenario 3 entre el 93 % y el 97 % . Bajo el escenario 2 la mayoría de los valores son del 96 % seguido de valores del 95 %; para el escenario 3 es similar la cantidad de casos con valores de PC de 95 % y 96 %. Al comparar los estimadores bajo el escenario 2, los estimadores de Hartley y Fuller-Burmeister siempre tienen valores iguales o menores de PC que el estimador PMV con excepción de la combinación $\pi PT - \pi PT - MAS$; caso contrario al escenario 3 en el cual a excepción de la combinación muestral $MAS - MAS - MAS$ el estimador PMV siempre tiene valores de PC a lo más iguales a los de Hartley y Fuller-Burmeister. Sin embargo, es importante resaltar que a grandes rasgos bajo los dos escenarios los tres estimadores se comportan de la misma forma en cuanto a PC. Al igual que para los valores de SRV, no se aprecia una tendencia que relacione los tamaños de muestra seleccionados en cada uno de los marcos con los valores de PC.

TABLA 3.2. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 60% en la segunda etapa del marco A.

TAMAÑOS DE MUESTRA			ESCENARIO 2									ESCENARIO 3									
MUESTRA			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			
n_{IA}	$n_{IIA} (%)$	n_B	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	
MAS	MAS	MAS																			
1000	60	59	0.24	0.23	-0.21	8.27 (8.27)	8.27 (8.27)	8.26 (8.26)	-5.02 (0.95)	-5.03 (0.95)	-4.95 (0.95)	0.48	-0.19	-0.10	7.83 (7.83)	7.84 (7.84)	7.80 (7.80)	0.80 (0.96)	0.67 (0.96)	0.17 (0.96)	
1000	60	117	-1.82	-1.82	-2.15	7.79 (7.79)	7.80 (7.79)	7.80 (7.79)	0.61 (0.95)	0.60 (0.95)	0.61 (0.95)	-1.80	-2.45	-2.48	7.84 (7.84)	7.83 (7.82)	7.85 (7.84)	1.10 (0.95)	1.25 (0.95)	-0.03 (0.95)	
1000	60	160	-1.49	-1.49	-1.86	7.93 (7.92)	7.93 (7.92)	7.94 (7.93)	-1.06 (0.95)	-1.06 (0.95)	-1.11 (0.95)	0.45	-0.28	-0.23	7.98 (7.98)	7.97 (7.97)	8.00 (8.00)	-1.02 (0.95)	-0.88 (0.95)	-2.15 (0.95)	
1750	60	59	0.42	0.42	0.28	4.20 (4.20)	4.20 (4.20)	4.20 (4.20)	6.87 (0.96)	6.85 (0.96)	6.71 (0.96)	0.54	0.19	0.18	4.45 (4.45)	4.45 (4.45)	4.45 (4.45)	1.61 (0.95)	1.58 (0.95)	0.60 (0.95)	
1750	60	117	-0.81	-0.81	-0.95	4.44 (4.44)	4.44 (4.44)	4.45 (4.45)	0.72 (0.95)	0.71 (0.95)	0.70 (0.95)	0.81	0.40	0.40	4.50 (4.50)	4.49 (4.49)	4.49 (4.49)	0.26 (0.95)	0.38 (0.96)	-0.49 (0.95)	
1750	60	160	-0.64	-0.64	-0.85	4.07 (4.07)	4.07 (4.07)	4.07 (4.07)	9.96 (0.97)	9.95 (0.97)	9.93 (0.96)	-1.05	-1.47	-1.45	4.22 (4.21)	4.21 (4.21)	4.21 (4.21)	7.58 (0.95)	7.70 (0.95)	6.53 (0.94)	
MAS	MAS	πPT																			
1000	60	59	0.50	0.57	0.18	8.07 (8.07)	8.08 (8.08)	8.07 (8.07)	-2.73 (0.94)	-2.70 (0.94)	-2.29 (0.94)	0.29	-0.22	-0.22	7.63 (7.63)	7.62 (7.62)	7.61 (7.61)	3.93 (0.95)	4.05 (0.95)	3.20 (0.95)	
1000	60	117	1.88	1.84	1.56	7.93 (7.92)	8.02 (8.02)	7.93 (7.93)	-1.13 (0.95)	-2.09 (0.95)	-0.51 (0.95)	2.39	1.81	1.70	7.64 (7.63)	7.64 (7.63)	7.61 (7.61)	3.76 (0.96)	3.73 (0.96)	3.03 (0.96)	
1000	60	160	4.14	4.18	3.86	8.20 (8.18)	8.21 (8.19)	8.20 (8.18)	-4.10 (0.95)	-4.19 (0.95)	-4.06 (0.95)	-3.08	-3.69	-3.70	7.92 (7.91)	7.92 (7.90)	7.92 (7.90)	0.25 (0.95)	0.32 (0.95)	-0.72 (0.95)	
1750	60	59	0.33	0.39	0.25	4.66 (4.66)	4.65 (4.65)	4.66 (4.66)	-3.96 (0.95)	-3.54 (0.95)	-3.52 (0.95)	2.34	2.05	1.85	4.24 (4.23)	4.23 (4.22)	4.23 (4.22)	7.17 (0.95)	7.23 (0.95)	6.19 (0.95)	
1750	60	117	-4.02	-4.54	-4.15	4.48 (4.46)	4.77 (4.74)	4.48 (4.47)	0.49 (0.95)	-1.56 (0.95)	4.99 (0.95)	-1.73	-2.05	-2.07	4.53 (4.53)	4.53 (4.52)	4.53 (4.53)	-0.03 (0.95)	0.02 (0.95)	-0.97 (0.95)	
1750	60	160	1.16	1.08	1.05	4.35 (4.34)	4.37 (4.37)	4.34 (4.34)	3.16 (0.96)	2.63 (0.96)	3.26 (0.96)	-1.40	-1.73	-1.66	4.27 (4.26)	4.27 (4.26)	4.26 (4.26)	6.07 (0.96)	6.04 (0.96)	5.13 (0.96)	
MAS	πPT	MAS																			
1000	60	59	-2.04	-2.04	-2.31	7.76 (7.75)	7.76 (7.75)	7.76 (7.75)	1.28 (0.95)	1.27 (0.95)	1.33 (0.95)	3.57	2.91	2.90	7.81 (7.80)	7.81 (7.80)	7.78 (7.77)	1.40 (0.96)	1.39 (0.96)	0.64 (0.96)	
1000	60	117	2.54	2.54	2.22	7.94 (7.93)	7.94 (7.93)	7.93 (7.93)	-1.14 (0.95)	-1.15 (0.95)	-1.05 (0.95)	-2.58	-3.33	-3.22	7.48 (7.47)	7.48 (7.47)	7.47 (7.46)	6.17 (0.95)	6.24 (0.95)	5.36 (0.95)	
1000	60	160	5.73	5.73	5.36	7.90 (7.86)	7.90 (7.86)	7.89 (7.85)	-0.13 (0.95)	-0.13 (0.95)	-0.06 (0.95)	7.24	6.43	6.57	7.60 (7.54)	7.58 (7.53)	7.56 (7.51)	5.18 (0.95)	5.29 (0.95)	4.44 (0.95)	
1750	60	59	0.45	0.45	0.27	4.21 (4.21)	4.21 (4.21)	4.21 (4.21)	6.38 (0.96)	6.36 (0.96)	6.29 (0.96)	0.82	0.48	0.29	4.52 (4.52)	4.52 (4.52)	4.52 (4.52)	0.24 (0.96)	0.19 (0.96)	-0.75 (0.96)	
1750	60	117	-1.15	-1.15	-1.33	4.20 (4.20)	4.20 (4.20)	4.20 (4.20)	6.72 (0.96)	6.71 (0.96)	6.84 (0.96)	-0.02	-0.46	-0.34	4.51 (4.51)	4.50 (4.50)	4.50 (4.50)	0.39 (0.95)	0.59 (0.95)	-0.39 (0.95)	
1750	60	160	0.90	0.90	0.65	4.07 (4.07)	4.07 (4.07)	4.07 (4.07)	10.15 (0.96)	10.15 (0.96)	10.15 (0.96)	-0.50	-0.94	-0.92	4.11 (4.11)	4.11 (4.11)	4.09 (4.09)	10.33 (0.96)	10.30 (0.96)	9.70 (0.96)	
MAS	πPT	πPT																			
1000	60	59	-6.37	-6.31	-6.66	7.76 (7.71)	7.80 (7.75)	7.78 (7.73)	1.45 (0.95)	3.35 (0.95)	5.82 (0.95)	5.51	4.98	4.92	7.87 (7.84)	7.85 (7.82)	7.86 (7.83)	1.37 (0.94)	1.49 (0.94)	0.42 (0.95)	
1000	60	117	2.17	2.50	1.88	7.18 (7.18)	7.16 (7.15)	7.19 (7.18)	9.28 (0.96)	11.09 (0.96)	10.63 (0.96)	0.65	0.04	-0.03	8.27 (8.27)	8.26 (8.26)	8.25 (8.25)	-4.37 (0.95)	-4.30 (0.95)	-4.99 (0.94)	
1000	60	160	-3.32	-3.28	-3.56	7.90 (7.89)	7.88 (7.87)	7.88 (7.87)	-0.79 (0.95)	-0.51 (0.95)	-0.53 (0.95)	3.31	2.64	2.69	7.88 (7.86)	7.87 (7.86)	7.88 (7.87)	0.81 (0.95)	0.84 (0.95)	-0.25 (0.95)	
1750	60	59	-1.21	-1.27	-1.27	4.32 (4.32)	4.33 (4.33)	4.32 (4.32)	3.70 (0.95)	3.57 (0.95)	3.86 (0.95)	6.02	5.75	5.62	4.42 (4.37)	4.41 (4.38)	4.40 (4.37)	3.61 (0.96)	3.52 (0.96)	2.80 (0.96)	
1750	60	117	4.02	3.85	3.97	4.38 (4.36)	4.39 (4.37)	4.38 (4.36)	2.98 (0.95)	3.10 (0.95)	3.38 (0.95)	-0.46	-0.82	-0.83	4.30 (4.30)	4.30 (4.30)	4.30 (4.30)	5.27 (0.96)	5.18 (0.96)	4.25 (0.95)	
1750	60	160	0.75	0.77	0.65	4.19 (4.19)	4.18 (4.18)	4.18 (4.18)	7.03 (0.96)	7.55 (0.96)	7.95 (0.96)	1.40	1.02	1.04	4.62 (4.62)	4.62 (4.62)	4.62 (4.62)	-1.90 (0.95)	-1.95 (0.95)	-2.88 (0.95)	
πPT	MAS	MAS																			
1000	60	59	2.13	2.13	1.86	10.74 (10.73)	10.74 (10.73)	10.72 (10.72)	3.37 (0.95)	3.36 (0.95)	3.72 (0.95)	4.00	3.30	3.46	11.34 (11.33)	11.33 (11.32)	11.33 (11.31)	4.26 (0.96)	4.29 (0.96)	-1.87 (0.95)	
1000	60	117	2.61	2.61	2.35	11.00 (11.00)	11.00 (11.00)	11.00 (10.99)	0.58 (0.95)	0.57 (0.95)	0.59 (0.95)	-0.87	-1.59	-1.44	11.15 (11.15)	11.15 (11.14)	11.13 (11.12)	5.41 (0.96)	5.44 (0.96)	-0.58 (0.95)	
1000	60	160	-1.58	-1.58	-1.85	11.30 (11.30)	11.30 (11.30)	11.29 (11.29)	-1.98 (0.94)	-1.98 (0.94)	-1.67 (0.95)	-1.29	-2.11	-2.01	10.78 (10.77)	10.77 (10.76)	10.77 (10.76)	9.56 (0.97)	9.67 (0.97)	3.20 (0.96)	
1750	60	59	3.29	3.29	3.18	5.82 (5.81)	5.82 (5.81)	5.82 (5.81)	8.93 (0.96)	8.91 (0.96)	8.97 (0.96)	1.01	0.67	0.54	6.31 (6.31)	6.31 (6.31)	6.31 (6.31)	6.94 (0.95)	6.91 (0.95)	0.66 (0.95)	
1750	60	117	-1.63	-1.63	-1.82	6.25 (6.25)	6.25 (6.25)	6.26 (6.25)	1.56 (0.96)	1.55 (0.96)	1.36 (0.96)	5.54	5.12	5.20	6.62 (6.58)	6.60 (6.57)	6.59 (6.55)	2.52 (0.95)	2.68 (0.95)	-3.17 (0.94)	
1750	60	160	0.52	0.52	0.40	5.43 (5.43)	5.43 (5.43)	5.43 (5.43)	16.76 (0.96)	16.75 (0.96)	16.78 (0.96)	-1.33	-1.77	-1.63	6.67 (6.67)	6.67 (6.66)	6.67 (6.67)	0.79 (0.94)	0.88 (0.94)	-5.18 (0.93)	
πPT	MAS	πPT																			
1000	60	59	-0.40	-0.94	-0.71	10.93 (10.93)	10.88 (10.88)	10.92 (10.92)	1.84 (0.97)	3.58 (0.97)	5.97 (0.97)	0.48	-0.08	-0.28	12.20 (12.20)	12.19 (12.19)	12.16 (12.16)	-3.70 (0.94)	-3.65 (0.94)	-9.02 (0.94)	
1000	60	117	-1.18	-0.94	-1.34	11.26 (11.26)	11.44 (11.44)	11.28 (11.28)	-1.39 (0.95)	-2.47 (0.95)	-0.90 (0.95)	1.83	1.27	1.24	11.31 (11.31)	11.29 (11.29)	11.28 (11.28)	4.20 (0.95)	4.38 (0.95)	-1.65 (0.94)	
1000	60	160	0.94	0.96	0.76	10.86 (10.86)	10.86 (10.86)	10.85 (10.85)	2.42 (0.95)	2.39 (0.95)	2.80 (0.95)	2.12	1.43	1.51	11.93 (11.93)	11.91 (11.91)	11.92 (11.91)	-1.56 (0.94)	-1.45 (0.94)	-7.27 (0.94)	
1750	60	59	-1.05	-0.71	-1.06	6.23 (6.23)	6.31 (6.31)	6.23 (6.23)	2.37 (0.96)	1.44 (0.96)	3.45 (0.95)	-1.03	-1.33	-1.53	6.75 (6.75)	6.75 (6.75)	6.74 (6.73)	-0.31 (0.95)	-0.34 (0.95)	-5.89 (0.94)	
1750	60	117	0.69	0.64	0.62	5.87 (5.87)	5.90 (5.90)	5.87 (5.87)	8.03 (0.95)	8.09 (0.95)	8.35 (0.95)	4.02	3.67	3.56	6.31 (6.29)	6.30 (6.29)	6.29 (6.28)	6.91 (0.96)	6.96 (0.96)	8.81 (0.94)	
1750	60	160	-0.59	-0.57	-0.66	6.10 (6.10)	6.10 (6.10)	6.10 (6.10)	3.83 (0.96)	3.79 (0.96)	3.96 (0.96)	3.33	2.95	3.07	6.53 (6.52)	6.52 (6.51)	6.53 (6.52)	3.28 (0.95)	3.38 (0.96)	-2.92 (0.94)	
πPT	πPT	MAS																			
1000	60	59	0.33	0.33	0.07	11.18 (11.18)	11.18 (11.18)	11.17 (11.17)	-0.80 (0.96)	-0.80 (0.96)	-0.71 (0.96)	-4.69	-5.34	-5.21	11.88 (11.85)	11.87 (11.84)	11.84 (11.81)	-0.81 (0.95)	-0.72 (0.95)	-6.33 (0.94)	
1000	60	117	1.60	1.60	1.35	10.34 (10.33)	10.34 (10.33)	10.33 (10.33)	7.76 (0.96)	7.76 (0.96)	7.82 (0.96)	3.94	3.20	3.23	12.23 (12.21)	12.20 (12.19)	12.21 (12.20)	-3.82 (0.95)	-3.64 (0.95)	-9.32 (0.94)	
1000	60	160	-1.47	-1.47	-1.68	10.58 (10.58)	10.58 (10.58)	10.59 (10.58)	4.90 (0.95)	4.89 (0.95)	4.84 (0.95)	-5.44	-6.19	-6.05	11.91 (11.87)	11.89 (11.84)	11.92 (11.88)	-1.25 (0.94)	-1.03 (0.95)	-6.92 (0.94)	
1750	60	59	1.83	1.83	1.75	6.35 (6.35)	6.35 (6.35)	6.35 (6.35)	0.32 (0.94)	0.30 (0.94)	0.45 (0.95)	-1.41	-1.75	-1.87	7.19 (7.18)	7.18 (7.18)	7.17 (7.17)	-6.22 (0.94)	-6.19 (0.94)	-11.50 (0.93)	
1750	60	117	-1.70	-1.70	-1.83	6.42 (6.42)	6.42 (6.42)	6.42 (6.42)	-1.11 (0.96)	-1.12 (0.96)	-1.07 (0.96)	-1.17	-1.56	-1.60	6.72 (6.71)	6.71 (6.71)	6.71 (6.71)	0.29 (0.96)	0.31 (0.96)	-5.52 (0.95)	
1750	60	160	1.62	1.61	1.51	6.37 (6.37)	6.37 (6.37)	6.38 (6.37)	-0.37 (0.95)	-0.38 (0.95)	-0.50 (0.96)	0.38	-0.07	-0.08	6.29 (6.29)	6.29 (6.29)	6.28 (6.28)	7.10 (0.96)	7.09 (0.96)	1.09 (0.96)	
πPT	πPT	πPT																			
1000	60	59	1.73	1.56	1.58	11.04 (11.04)	11.07 (11.07)	11.02 (11.02)	0.19 (0.96)	0.11 (0.96)	0.80 (0.95)	0.58	-0.04	-0.04	10.80 (10.80)	10.					

Los datos registrados de la tabla 3.2 hacen referencia a los resultados de la simulación con una fracción muestral del 60 % para el marco A, en ellos se encuentra que los valores de SR para los tres estimadores bajo los dos escenarios son inferiores al 1 %, al igual que para la fracción del 40 %. Bajo el escenario 2 los SR más cercanos a cero se presentan para los diseños muestrales $MAS - MAS - MAS, \pi PT - MAS - \pi PT$ y $\pi PT - \pi PT - MAS$; para el escenario 3 los SR más pequeños se obtuvieron para los diseños $MAS - MAS - MAS$ y $MAS - MAS - \pi PT$, también se aprecia que al comparar los dos escenarios en cuanto a SR los valores del escenario 3 en general son mayores que los del escenario 2.

Al observar los valores de ECME y s^2 con una fracción muestral de 60 % en la tabla 3.2, los menores valores se presentan para la combinación de diseños con MAS en la primera etapa del marco A tanto para el escenario 2 como el 3, al igual que para la fracción muestral del 40 %; en este aspecto las combinaciones de diseños muestrales $MAS - MAS - \pi PT$ y $MAS - \pi PT - MAS$ presentan los menores valores de ECME bajo los 2 escenarios.

También es de resaltar que, al aumentarse el tamaño de muestra en la primera etapa del marco A se disminuye el ECME y s^2 y la varianza a casi la mitad, pero esto no sucede al aumentar el tamaño de muestra a extraer del marco B; al comparar los tres estimadores para el escenario 2, los valores de ECME y s^2 fueron similares, no obstante, bajo el escenario 3 el estimador de Hartley nunca tuvo el valor más pequeño bajo las combinaciones de diseños muestrales propuestos.

Siguiendo con la fracción del 60 % se observa que en cuanto a SRV bajo el escenario 2, sólo para una combinación muestral se excede el 15 %, y para el escenario 3 estos valores no exceden en ningún caso el 12 %. A diferencia de la fracción muestral de 40 %, para el escenario 2 los SRV se reparten equitativamente entre valores positivos y negativos (sobrestimación y subestimación de la varianza respectivamente) y son en gran parte valores positivos bajo el escenario 3. Al comparar los tres estimadores en el escenario 2 no se observa que alguno de ellos en sobresalga con los valores más cercanos a cero, pero para el escenario 3 sí se observa en todos los casos que el estimador Pseudo Máximo Verosímil siempre es el menor en cuanto a SRV. Por otra parte las combinaciones de diseños muestrales que presentan valores más cercanos a cero son $MAS - MAS - \pi PT$ y $\pi PT - \pi PT - MAS$ para el escenario 2 y $MAS - MAS - MAS$ y $MAS - \pi PT - \pi PT$ para el escenario 3, además para los dos escenarios no se aprecia una relación directa de aumento o disminución de SRV respecto a los tamaños de muestra tanto del marco A como el B.

Los valores de PC para los escenarios 2 y 3 oscilan entre el 94 % y el 97 % y en su mayoría son del 95 % y 96 %. Adicionalmente al comparar los tres estimadores en cuanto a PC no se aprecian diferencias importantes en este aspecto. Igualmente que para los valores de SRV, en los valores de PC no se aprecia una tendencia o relación de los tamaños de muestra seleccionados en cada uno de los marcos con los valores de PC.

TABLA 3.3. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 80% en la segunda etapa del marco A.

TAMAÑOS DE MUESTRA			ESCENARIO 2									ESCENARIO 3																				
MUESTRA			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$														
n_{IA}	$n_{IIA} (%)$	n_B	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML												
MAS	MAS	MAS	-0.09	-0.09	-0.57	5.82	(5.82)	5.82	(5.82)	5.84	(5.84)	0.95	(0.95)	0.94	(0.95)	0.61	(0.95)	-0.69	-1.19	-1.22	5.93	(5.93)	5.92	(5.92)	5.92	(5.92)	-0.03	(0.94)	0.05	(0.94)	-1.47	(0.94)
1000	80	59	1.66	1.66	1.34	5.41	(5.41)	5.41	(5.41)	5.41	(5.41)	8.55	(0.96)	8.54	(0.96)	8.68	(0.96)	-0.49	-1.14	-1.22	5.99	(5.98)	5.98	(5.98)	5.98	(5.98)	-0.94	(0.95)	-0.91	(0.95)	-2.51	(0.95)
1000	80	160	-0.56	-0.56	-0.84	5.38	(5.38)	5.38	(5.38)	5.38	(5.37)	9.37	(0.95)	9.36	(0.95)	9.41	(0.95)	-1.59	-2.18	-2.14	5.94	(5.93)	5.93	(5.92)	5.93	(5.92)	-0.25	(0.95)	-0.06	(0.95)	-1.71	(0.95)
1750	80	59	-3.24	-3.24	-3.32	3.23	(3.22)	3.23	(3.22)	3.23	(3.21)	4.25	(0.96)	4.22	(0.96)	4.44	(0.96)	-0.14	-0.37	-0.75	3.16	(3.16)	3.16	(3.16)	3.17	(3.17)	7.23	(0.95)	7.24	(0.95)	5.19	(0.95)
1750	80	117	-2.24	-2.24	-2.30	3.13	(3.12)	3.13	(3.12)	3.13	(3.12)	7.53	(0.96)	7.52	(0.96)	7.60	(0.96)	0.00	-0.30	-0.33	3.06	(3.06)	3.05	(3.05)	3.07	(3.07)	10.57	(0.96)	10.61	(0.96)	8.27	(0.96)
1750	80	160	-1.33	-1.33	-1.37	3.16	(3.16)	3.16	(3.16)	3.16	(3.16)	6.37	(0.96)	6.36	(0.96)	6.50	(0.96)	-1.33	-1.65	-1.68	3.67	(3.67)	3.67	(3.66)	3.67	(3.67)	-7.61	(0.94)	-7.53	(0.94)	-9.14	(0.94)
MAS	MAS	π PT	0.51	1.93	0.07	5.42	(5.42)	7.58	(7.58)	5.40	(5.40)	8.53	(0.96)	15.12	(0.96)	70.25	(0.96)	5.21	4.81	4.50	6.45	(6.42)	6.44	(6.42)	6.41	(6.39)	-7.90	(0.94)	-7.89	(0.94)	-9.06	(0.94)
1000	80	117	-0.36	-0.13	-0.59	5.97	(5.97)	5.98	(5.98)	5.99	(5.98)	-1.70	(0.94)	-0.72	(0.94)	-1.07	(0.95)	-2.92	-3.41	-3.55	5.80	(5.79)	5.79	(5.78)	5.80	(5.78)	2.42	(0.95)	2.48	(0.95)	0.74	(0.95)
1000	80	160	1.77	1.80	1.54	5.56	(5.56)	5.57	(5.57)	5.56	(5.55)	5.40	(0.96)	5.25	(0.96)	5.47	(0.96)	-2.84	-3.35	-3.43	6.02	(6.01)	6.01	(5.99)	5.98	(5.97)	-1.51	(0.95)	-1.32	(0.95)	-2.55	(0.95)
1750	80	59	-1.92	-2.00	-1.93	3.12	(3.12)	3.11	(3.11)	3.12	(3.11)	7.66	(0.96)	8.79	(0.96)	8.92	(0.96)	-0.26	-0.43	-0.81	3.38	(3.38)	3.38	(3.38)	3.40	(3.40)	-0.11	(0.95)	-0.24	(0.95)	-2.18	(0.95)
1750	80	117	0.48	0.52	0.39	3.34	(3.34)	3.33	(3.33)	3.34	(3.34)	0.73	(0.96)	1.39	(0.96)	0.83	(0.96)	0.89	0.62	0.46	3.16	(3.16)	3.16	(3.16)	3.16	(3.16)	6.81	(0.96)	6.89	(0.96)	5.18	(0.96)
1750	80	160	0.95	3.43	0.95	3.03	(3.03)	10.12	(10.11)	3.03	(3.03)	10.82	(0.95)	-36.22	(0.95)	101.22	(0.95)	1.13	0.84	0.74	2.90	(2.90)	2.90	(2.90)	2.90	(2.89)	16.43	(0.96)	16.48	(0.96)	14.73	(0.96)
MAS	π PT	MAS	-0.47	-0.47	-0.89	5.54	(5.54)	5.54	(5.54)	5.56	(5.56)	4.15	(0.95)	4.13	(0.95)	3.84	(0.95)	1.11	0.63	0.56	6.06	(6.05)	6.04	(6.04)	6.04	(6.04)	-2.26	(0.95)	-2.09	(0.95)	-3.65	(0.95)
1000	80	117	0.66	0.66	0.39	5.18	(5.18)	5.18	(5.18)	5.19	(5.19)	11.73	(0.96)	11.71	(0.96)	11.72	(0.96)	3.70	3.13	3.14	6.53	(6.51)	6.51	(6.50)	6.53	(6.52)	-9.04	(0.94)	-8.87	(0.93)	-10.71	(0.93)
1000	80	160	-4.80	-4.81	-5.09	5.87	(5.84)	5.87	(5.84)	5.87	(5.84)	-0.65	(0.95)	-0.66	(0.95)	-0.59	(0.95)	-0.96	-1.61	-1.59	5.45	(5.45)	5.44	(5.44)	5.44	(5.44)	8.71	(0.95)	8.86	(0.95)	7.11	(0.95)
1750	80	59	-2.24	-2.24	-2.29	3.39	(3.38)	3.39	(3.38)	3.39	(3.38)	-2.20	(0.95)	-2.23	(0.95)	-2.17	(0.95)	-0.73	-0.97	-1.21	3.33	(3.33)	3.33	(3.33)	3.33	(3.33)	1.87	(0.95)	1.89	(0.95)	0.19	(0.95)
1750	80	117	2.16	2.16	2.12	3.06	(3.06)	3.06	(3.06)	3.06	(3.06)	8.88	(0.95)	8.86	(0.95)	8.85	(0.95)	-1.50	-1.83	-1.90	3.41	(3.41)	3.41	(3.40)	3.39	(3.39)	-0.80	(0.95)	-0.66	(0.95)	-1.87	(0.94)
1750	80	160	-2.29	-2.29	-2.28	3.11	(3.10)	3.11	(3.10)	3.11	(3.10)	7.23	(0.97)	7.22	(0.97)	7.14	(0.97)	-1.59	-1.93	-1.95	3.28	(3.28)	3.28	(3.27)	3.28	(3.28)	3.15	(0.95)	3.24	(0.95)	1.45	(0.95)
MAS	π PT	π PT	-2.00	-2.35	-2.31	5.48	(5.47)	5.57	(5.56)	5.47	(5.46)	6.03	(0.96)	6.76	(0.96)	8.59	(0.96)	-1.95	-2.34	-2.67	5.54	(5.54)	5.54	(5.53)	5.53	(5.52)	6.95	(0.96)	6.97	(0.96)	5.36	(0.96)
1000	80	117	0.52	1.09	0.26	5.60	(5.60)	5.83	(5.82)	5.60	(5.60)	4.29	(0.96)	1.64	(0.96)	6.04	(0.96)	0.45	-0.05	-0.18	6.01	(6.01)	6.00	(6.00)	6.00	(6.00)	-1.69	(0.94)	-1.56	(0.94)	-3.15	(0.94)
1000	80	160	-0.80	-0.73	-0.90	5.92	(5.92)	5.93	(5.93)	5.91	(5.91)	-1.55	(0.94)	-1.77	(0.94)	-1.44	(0.95)	0.47	-0.07	-0.22	5.90	(5.90)	5.89	(5.89)	5.87	(5.87)	0.60	(0.95)	0.59	(0.96)	-0.74	(0.95)
1750	80	59	0.06	0.31	-0.01	3.10	(3.10)	3.18	(3.18)	3.10	(3.10)	7.13	(0.96)	8.54	(0.96)	12.35	(0.96)	4.21	4.03	3.61	3.28	(3.26)	3.28	(3.26)	3.27	(3.25)	3.96	(0.96)	3.80	(0.96)	2.35	(0.96)
1750	80	117	-1.56	-0.13	-1.59	3.23	(3.22)	5.76	(5.76)	3.23	(3.23)	3.26	(0.96)	-24.89	(0.96)	28.68	(0.96)	1.54	1.26	1.07	3.35	(3.34)	3.34	(3.34)	3.34	(3.34)	1.04	(0.95)	1.02	(0.95)	-0.56	(0.94)
1750	80	160	0.93	1.00	0.82	2.99	(2.99)	2.98	(2.97)	2.99	(2.99)	11.08	(0.96)	11.95	(0.96)	11.46	(0.96)	0.03	-0.29	-0.41	3.02	(3.02)	3.02	(3.02)	3.01	(3.01)	11.98	(0.96)	11.94	(0.96)	10.35	(0.96)
π PT	MAS	MAS	-2.38	-2.39	-2.71	8.39	(8.39)	8.39	(8.39)	8.40	(8.39)	-0.38	(0.93)	-0.39	(0.93)	-0.47	(0.93)	1.99	1.49	1.40	8.60	(8.59)	8.58	(8.58)	8.63	(8.63)	4.84	(0.95)	4.94	(0.95)	-4.00	(0.94)
1000	80	117	0.89	0.89	0.71	8.14	(8.14)	8.14	(8.14)	8.14	(8.14)	2.51	(0.96)	2.50	(0.96)	2.38	(0.95)	1.83	1.17	1.20	8.32	(8.32)	8.31	(8.31)	8.32	(8.32)	8.52	(0.96)	8.61	(0.96)	-0.26	(0.95)
1000	80	160	-1.76	-1.76	-1.93	8.74	(8.73)	8.74	(8.73)	8.75	(8.74)	-3.95	(0.95)	-3.95	(0.95)	-3.88	(0.95)	1.03	0.39	0.53	8.50	(8.50)	8.49	(8.49)	8.49	(8.49)	6.26	(0.96)	6.40	(0.96)	-2.21	(0.95)
1750	80	59	1.08	1.08	1.08	4.72	(4.72)	4.72	(4.72)	4.73	(4.72)	0.83	(0.95)	0.81	(0.95)	0.84	(0.95)	-2.07	-2.34	-2.58	4.68	(4.68)	4.67	(4.67)	4.68	(4.67)	10.18	(0.96)	10.33	(0.96)	1.63	(0.95)
1750	80	117	-0.29	-0.30	-0.30	5.01	(5.01)	5.01	(5.01)	5.01	(5.01)	-4.92	(0.94)	-4.93	(0.94)	-5.06	(0.94)	0.76	0.45	0.41	5.02	(5.02)	5.01	(5.01)	5.03	(5.03)	2.76	(0.95)	2.86	(0.95)	-5.85	(0.94)
1750	80	160	-0.90	-0.90	-0.96	4.30	(4.30)	4.30	(4.30)	4.30	(4.30)	11.25	(0.97)	11.24	(0.97)	11.18	(0.97)	0.75	0.38	0.46	4.98	(4.98)	4.97	(4.97)	4.97	(4.97)	3.39	(0.95)	3.44	(0.95)	-4.96	(0.94)
π PT	MAS	π PT	-1.93	-2.58	-2.17	8.86	(8.85)	9.26	(9.25)	8.87	(8.86)	-5.23	(0.94)	-9.10	(0.94)	-4.58	(0.94)	2.80	2.38	2.27	8.75	(8.74)	8.74	(8.73)	8.75	(8.74)	3.59	(0.95)	3.71	(0.95)	-4.95	(0.94)
1000	80	117	0.80	1.26	0.69	8.66	(8.66)	8.84	(8.84)	8.66	(8.66)	-3.32	(0.94)	-5.12	(0.94)	-3.22	(0.94)	-1.24	-1.76	-1.80	8.12	(8.12)	8.12	(8.11)	8.13	(8.13)	11.14	(0.96)	11.18	(0.96)	1.95	(0.96)
1000	80	160	-2.05	-2.07	-2.24	8.99	(8.99)	8.99	(8.98)	8.99	(8.99)	-6.82	(0.95)	-4.68	(0.95)	-4.36	(0.94)	7.80	7.29	7.18	8.44	(8.37)	8.42	(8.35)	8.42	(8.36)	7.90	(0.95)	8.07	(0.95)	-0.81	(0.94)
1750	80	59	-4.68	-4.28	-4.69	4.63	(4.61)	5.33	(5.31)	4.62	(4.60)	3.73	(0.95)	1.87	(0.95)	27.94	(0.95)	2.44	2.23	1.98	4.96	(4.96)	4.96	(4.95)	4.97	(4.96)	4.31	(0.95)	4.31	(0.95)	-4.34	(0.94)
1750	80	117	0.64	1.43	0.59	4.53	(4.53)	4.53	(4.53)	4.53	(4.53)	5.19	(0.96)	14.85	(0.95)	28.27	(0.95)	0.25	0.01	-0.11	4.65	(4.65)	4.64	(4.64)	4.65	(4.65)	10.80	(0.96)	10.84			

Pasando a los resultados con una fracción muestral del 80 % en la segunda etapa del marco A, (ver tabla 3.3) los valores de SR para los tres estimadores son inferiores al 0.6 % y al 0.8 % bajo los escenarios 2 y 3 respectivamente. En general bajo los dos escenarios, todas las combinaciones de diseños muestrales se comportan similar en cuanto a SR.

Analógo a las fracciones del 40 % y 60 %, los menores valores de ECME y s^2 con una fracción muestral del 80 % se presentan para la combinación de diseños con *MAS* en la primera etapa del marco A, y a medida que aumenta el tamaño de muestra del marco A disminuye el ECME. Al cotejar los tres estimadores, el estimador de Fuller-Burmeister es el que presenta mayor cantidad de valores altos en todas las combinaciones muestrales sobre el escenario 2, distinto al escenario 3 donde el estimador de Hartley es el que tiene esta característica.

En la tabla 3.3 también se aprecia que en cuanto a SRV, bajo el escenario 2 en algunos casos se alcanzan porcentajes superiores al 25 % (valores absolutos) mayoritariamente en el estimador PMV caso contrario a los valores absolutos de SRV del escenario 3 en donde ninguna de las combinaciones excede el 20 %. Una característica que se evidencia con esta fracción muestral, es que el estimador de Fuller-Burmeister tiene los menores valores de SRV comparado con los otros dos estimadores bajo los dos escenarios. Algo también de resaltar es que no se marca una clara diferencia en cuanto a SRV entre las combinaciones de diseños muestrales.

En cuanto a los valores de PC, para el escenario 2 los valores oscilan entre el 94 % y el 96 % a excepción de tres casos donde se obtiene 93 % y 97 %; bajo el escenario 3 sólo hay un caso en donde se alcanza el 93 %. Comparando los estimadores se aprecia que los estimadores de Hartley y Fuller-Burmeister se comportan similar, pero el Pseudo Máximo Verosímil presenta en varios casos valores de PC menores. Con esta fracción muestral las combinaciones de diseños muestrales tienen valores similares de PC.

Finalmente, trabajando con una fracción muestral del 100 % (ver tabla 3.4) en la segunda etapa del marco A (conglomerados), se observa que, bajo los dos escenarios propuestos el valor máximo de SR en valor absoluto es de 0.5 % y que el estimador PMV en general presenta los menores valores en cuanto a SR. Bajo el escenario 2 los valores de ECME son un poco más elevados para el estimador de Fuller-Burmeister que para los otros dos y dichos valores son muy parecidos entre estimadores en el escenario 3.

En cuanto a SRV se presentan valores que alcanzan hasta el 40 % en valor abosulto para el escenario 2; principalmente en los estimadores PMV y Fuller-Burmeister, pero bajo el escenario 3 los SRV alcanzan un valor máximo de 17 % . El estimador PMV presenta en general menores valores de SRV que los otros dos estimadores, esta misma característica se presenta para los valores de PC bajo los dos escenarios, en donde se alcanzan valores entre el 94 % y el 97 %.

TABLA 3.4. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico ($ECME \times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza ($SRV \times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 100 % (conglomerados) en la segunda etapa del marco A.

TAMAÑOS DE MUESTRA		ESCENARIO 2									ESCENARIO 3									
MUESTRA		$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			
n_A	n_B	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	
MAS	MAS																			
1000	59	1.29	1.29	0.72	4.47 (4.47)	4.47 (4.47)	4.47 (4.47)	5.32 (0.95)	5.30 (0.95)	5.44 (0.95)	1.36	0.66	0.75	4.51 (4.51)	4.50 (4.50)	4.51 (4.51)	4.72 (0.96)	4.92 (0.96)	3.14 (0.96)	
1000	117	0.16	0.16	-0.26	4.72 (4.72)	4.72 (4.72)	4.72 (4.72)	-0.14 (0.96)	-0.15 (0.96)	-0.18 (0.96)	1.58	0.97	0.87	4.50 (4.50)	4.49 (4.48)	4.50 (4.50)	5.49 (0.96)	5.72 (0.96)	3.70 (0.95)	
1000	160	0.67	0.67	0.38	4.68 (4.68)	4.68 (4.68)	4.67 (4.67)	0.99 (0.95)	0.98 (0.95)	1.16 (0.95)	-4.52	-5.21	-5.11	4.92 (4.90)	4.92 (4.89)	4.92 (4.89)	-3.53 (0.95)	-3.39 (0.95)	-4.87 (0.95)	
1750	59	2.10	2.10	2.04	2.47 (2.47)	2.47 (2.47)	2.47 (2.47)	9.13 (0.96)	9.09 (0.96)	9.08 (0.96)	-1.89	-2.24	-2.07	2.49 (2.48)	2.49 (2.48)	2.50 (2.50)	8.67 (0.96)	8.62 (0.96)	6.69 (0.96)	
1750	117	-0.74	-0.74	-0.85	2.40 (2.40)	2.40 (2.40)	2.40 (2.40)	12.14 (0.96)	12.12 (0.96)	12.30 (0.96)	-0.78	-1.13	-1.12	2.68 (2.68)	2.68 (2.68)	2.67 (2.67)	0.92 (0.96)	0.90 (0.96)	-0.32 (0.96)	
1750	160	0.36	0.36	0.28	2.58 (2.58)	2.58 (2.58)	2.58 (2.58)	4.68 (0.96)	4.67 (0.96)	4.53 (0.95)	-1.31	-1.81	-1.54	2.66 (2.66)	2.66 (2.65)	2.65 (2.65)	1.72 (0.95)	1.88 (0.95)	0.67 (0.95)	
MAS	πPT																			
1000	59	0.45	3.28	0.01	4.66 (4.66)	11.64 (11.63)	4.65 (4.65)	1.43 (0.95)	-39.62 (0.95)	55.00 (0.95)	-1.43	-1.88	-2.61	4.71 (4.70)	4.71 (4.70)	4.70 (4.69)	0.76 (0.95)	0.74 (0.95)	-0.19 (0.95)	
1000	117	-2.47	-4.13	-2.86	4.75 (4.75)	8.07 (8.05)	4.76 (4.75)	-0.53 (0.94)	-40.17 (0.94)	8.60 (0.94)	1.05	0.45	0.31	4.40 (4.39)	4.39 (4.39)	4.40 (4.40)	7.99 (0.97)	8.13 (0.96)	6.42 (0.96)	
1000	160	0.79	0.80	0.44	4.44 (4.44)	4.43 (4.43)	4.44 (4.44)	6.50 (0.96)	6.86 (0.96)	6.82 (0.96)	0.19	-0.42	-0.58	4.31 (4.31)	4.31 (4.31)	4.29 (4.29)	10.18 (0.96)	10.14 (0.95)	9.20 (0.96)	
1750	59	0.69	0.93	0.52	2.43 (2.43)	2.49 (2.49)	2.43 (2.43)	10.96 (0.97)	8.22 (0.96)	10.99 (0.96)	1.48	1.26	0.98	2.64 (2.64)	2.64 (2.64)	2.64 (2.64)	2.71 (0.95)	2.72 (0.95)	1.43 (0.95)	
1750	117	-1.41	-1.50	-1.54	2.43 (2.43)	2.44 (2.44)	2.43 (2.43)	11.06 (0.96)	11.37 (0.96)	11.82 (0.96)	1.89	1.60	1.36	2.31 (2.31)	2.31 (2.31)	2.31 (2.31)	17.44 (0.96)	17.42 (0.96)	16.05 (0.96)	
1750	160	3.72	3.71	3.64	2.46 (2.45)	2.46 (2.45)	2.46 (2.44)	10.16 (0.97)	9.95 (0.97)	10.45 (0.96)	-1.19	-1.52	-1.75	2.53 (2.53)	2.53 (2.53)	2.54 (2.53)	7.18 (0.96)	7.15 (0.96)	5.80 (0.95)	
πPT	MAS																			
1000	59	-0.58	-0.58	-1.05	6.08 (6.08)	6.08 (6.08)	6.06 (6.06)	12.06 (0.96)	12.05 (0.96)	12.60 (0.96)	0.89	0.40	0.14	7.40 (7.40)	7.38 (7.38)	7.37 (7.37)	-0.23 (0.95)	-0.13 (0.95)	-9.13 (0.94)	
1000	117	-1.07	-1.06	-1.30	6.68 (6.68)	6.68 (6.68)	6.67 (6.67)	1.56 (0.95)	1.55 (0.95)	1.73 (0.95)	-0.28	-1.11	-0.84	7.06 (7.06)	7.05 (7.05)	7.05 (7.05)	3.94 (0.95)	4.10 (0.95)	-4.93 (0.94)	
1000	160	-1.55	-1.55	-1.86	6.02 (6.01)	6.02 (6.01)	6.02 (6.02)	13.14 (0.96)	13.14 (0.96)	13.18 (0.96)	1.43	0.65	0.74	6.94 (6.94)	6.93 (6.93)	6.91 (6.91)	7.14 (0.96)	7.23 (0.96)	-1.45 (0.95)	
1750	59	0.61	0.61	0.47	3.71 (3.71)	3.71 (3.71)	3.71 (3.70)	5.10 (0.95)	5.07 (0.95)	5.19 (0.95)	0.79	0.39	0.52	4.12 (4.12)	4.12 (4.12)	4.12 (4.12)	1.94 (0.95)	1.98 (0.95)	-6.86 (0.94)	
1750	117	0.46	0.46	0.39	3.86 (3.86)	3.86 (3.86)	3.86 (3.86)	0.83 (0.94)	0.82 (0.94)	0.93 (0.94)	-2.23	-2.60	-2.65	4.08 (4.07)	4.08 (4.07)	4.07 (4.06)	3.64 (0.96)	3.66 (0.96)	-4.72 (0.95)	
1750	160	-0.37	-0.37	-0.42	3.70 (3.70)	3.70 (3.70)	3.70 (3.70)	5.07 (0.95)	5.06 (0.95)	5.20 (0.95)	3.34	2.96	2.97	4.03 (4.02)	4.03 (4.02)	4.03 (4.02)	5.26 (0.96)	5.26 (0.96)	-3.58 (0.95)	
πPT	πPT																			
1000	59	-1.42	-1.69	-1.80	6.92 (6.92)	6.92 (6.92)	6.89 (6.89)	-1.63 (0.96)	-1.02 (0.96)	-0.68 (0.96)	1.46	0.98	0.54	7.01 (7.01)	6.99 (6.99)	6.99 (6.99)	5.00 (0.95)	5.14 (0.95)	-3.74 (0.94)	
1000	117	-1.01	1.72	-1.32	6.65 (6.65)	12.77 (12.77)	6.65 (6.65)	2.18 (0.95)	-39.27 (0.95)	27.89 (0.95)	0.83	0.25	0.02	7.32 (7.32)	7.31 (7.31)	7.28 (7.28)	0.84 (0.94)	0.91 (0.94)	-6.88 (0.94)	
1000	160	0.68	0.73	0.42	6.86 (6.86)	6.85 (6.85)	6.86 (6.86)	-0.83 (0.94)	-0.73 (0.94)	-0.75 (0.95)	-1.93	-2.57	-2.66	7.20 (7.20)	7.20 (7.19)	7.18 (7.18)	2.24 (0.95)	2.32 (0.95)	-6.09 (0.94)	
1750	59	3.08	3.08	3.01	3.25 (3.24)	3.25 (3.24)	3.24 (3.23)	20.06 (0.97)	19.90 (0.97)	20.14 (0.97)	-2.39	-2.61	-2.77	3.66 (3.66)	3.66 (3.65)	3.67 (3.66)	15.78 (0.97)	15.80 (0.97)	6.13 (0.96)	
1750	117	-2.82	-2.85	-2.95	3.65 (3.64)	3.64 (3.64)	3.65 (3.64)	6.83 (0.96)	6.90 (0.96)	6.75 (0.96)	0.52	0.20	-0.21	3.89 (3.89)	3.89 (3.89)	3.90 (3.90)	8.48 (0.96)	8.47 (0.96)	-0.64 (0.95)	
1750	160	0.18	0.10	0.03	3.91 (3.91)	3.91 (3.91)	3.92 (3.92)	-0.72 (0.94)	12.04 (0.94)	9.46 (0.94)	2.35	2.03	1.82	4.13 (4.12)	4.13 (4.12)	4.11 (4.11)	2.50 (0.95)	2.48 (0.95)	-5.60 (0.95)	

3.2. Caso aplicado

En la tabla 3.5 se presentan los resultados para el caso práctico, en donde se utiliza una fracción muestral para la segunda etapa del marco A del 100 %, dicha fracción es la que se utiliza actualmente en la ENA.

Bajo el escenario 2 tenemos que los valores de SR para los tres estimadores son inferiores al 1 % . Bajo el escenario 3 en dos combinaciones en las cuales se extrajo la muestra mediante πPT para el marco B se superó el 1 %. Los SR más cercanos a cero bajo el escenario 2, se presentan para los diseños muestrales con *MAS* en el marco A; esto también sucede para el escenario 3.

Al observar los valores de ECME y s^2 los tres estimadores tienen valores muy cercanos, sin embargo, estos valores son menores para el escenario 3 que para el escenario 2; al igual que para el SR, los valores de ECME y s^2 son menores bajo los dos escenarios para los diseños muestrales *MAS* en el marco A.

A diferencia de los datos obtenidos mediante el modelo, los valores de PC oscilan entre el 92 % y el 96 % para el escenario 2, y entre el 91 % y el 96 % para el escenario 3 con valores más cercanos al 95 % bajo el escenario 2. Al comparar los tres estimadores en cuanto a PC no se aprecia una diferencia importante en sus valores, pero sí es llamativo que en la combinación de diseños muestrales $\pi PT - \pi PT$ bajo el escenario 3, la mayoría de valores de PC son menores o iguales al 94 %.

TABLA 3.5. Sesgo relativo ($SR \times 10^{-3}$), error cuadrático medio empírico (ECME $\times 10^8$) con la respectiva varianza estimada en paréntesis ($s^2 \times 10^8$) y Sesgo relativo de la varianza (SRV $\times 10^{-2}$) con la probabilidad de cobertura (PC) en paréntesis para los tres estimadores y los diseños muestrales planteados bajo los escenarios 2 y 3 con fracción muestral del 100 % (conglomerados) en la segunda etapa del marco A.

TAMAÑOS DE MUESTRA		ESCENARIO 2									ESCENARIO 3									
n_A	n_B	$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			$SR \times 10^{-3}$			$ECME \times 10^8 (s^2 \times 10^8)$			$SRV \times 10^{-2} (PC)$			
		H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	H	FB	PML	
MAS	MAS																			
1000	59	-3.9	-3.9	-4.2	1.52 (1.51)	1.52 (1.51)	1.52 (1.51)	-0.35 (0.94)	-0.35 (0.94)	-0.18 (0.95)	-0.3	-0.2	-0.4	1.45 (1.44)	1.45 (1.44)	1.45 (1.44)	5.89 (0.94)	5.88 (0.94)	4.98 (0.94)	
1000	117	-1.0	-1.0	-1.1	1.52 (1.51)	1.52 (1.51)	1.52 (1.51)	-0.06 (0.93)	-0.06 (0.93)	0.15 (0.94)	-5.1	-5.0	-5.0	1.51 (1.50)	1.51 (1.50)	1.51 (1.50)	1.13 (0.93)	1.11 (0.93)	0.57 (0.93)	
1000	160	-3.9	-3.9	-4.0	1.44 (1.43)	1.44 (1.43)	1.44 (1.43)	4.47 (0.94)	4.48 (0.94)	4.34 (0.94)	5.3	5.4	5.4	1.57 (1.56)	1.57 (1.56)	1.57 (1.56)	-0.90 (0.94)	-0.87 (0.94)	-1.48 (0.93)	
1750	59	0.6	0.6	0.7	0.77 (0.76)	0.77 (0.76)	0.77 (0.76)	12.92 (0.95)	12.93 (0.95)	12.88 (0.95)	-2.3	-2.2	-2.4	0.83 (0.82)	0.83 (0.82)	0.84 (0.83)	4.31 (0.94)	4.29 (0.94)	3.82 (0.94)	
1750	117	4.3	4.3	4.5	0.83 (0.82)	0.83 (0.82)	0.83 (0.82)	5.67 (0.95)	5.67 (0.95)	5.61 (0.95)	1.4	1.5	1.4	0.79 (0.78)	0.80 (0.79)	0.80 (0.79)	10.62 (0.96)	10.56 (0.96)	9.85 (0.96)	
1750	160	-4.0	-4.0	-3.9	0.85 (0.84)	0.85 (0.84)	0.85 (0.84)	1.60 (0.94)	1.60 (0.94)	1.61 (0.94)	2.6	2.7	2.7	0.89 (0.88)	0.89 (0.88)	0.89 (0.88)	-1.22 (0.94)	-1.23 (0.94)	-1.66 (0.94)	
MAS	πPT																			
1000	59	0.8	0.8	0.5	1.41 (1.40)	1.41 (1.40)	1.41 (1.40)	8.79 (0.95)	8.80 (0.95)	9.05 (0.95)	12.6	12.7	12.3	1.55 (1.54)	1.55 (1.54)	1.55 (1.54)	1.73 (0.96)	1.76 (0.96)	1.11 (0.96)	
1000	117	-1.5	-1.5	-1.6	1.52 (1.51)	1.52 (1.51)	1.52 (1.51)	-0.15 (0.93)	-0.15 (0.93)	-0.12 (0.93)	4.0	4.1	4.0	1.48 (1.47)	1.49 (1.48)	1.48 (1.47)	3.38 (0.95)	3.32 (0.95)	3.01 (0.95)	
1000	160	0.0	0.0	-0.2	1.40 (1.39)	1.40 (1.39)	1.40 (1.39)	8.45 (0.95)	8.45 (0.95)	8.39 (0.95)	2.0	2.1	2.0	1.39 (1.38)	1.39 (1.38)	1.38 (1.37)	11.30 (0.96)	11.32 (0.96)	10.81 (0.96)	
1750	59	-2.6	-2.6	-2.5	0.86 (0.85)	0.86 (0.85)	0.86 (0.85)	0.58 (0.94)	0.58 (0.94)	0.51 (0.94)	-2.7	-2.6	-2.7	0.82 (0.81)	0.82 (0.81)	0.82 (0.81)	6.16 (0.95)	6.15 (0.95)	5.42 (0.95)	
1750	117	-1.2	-1.2	-1.2	0.85 (0.84)	0.85 (0.84)	0.85 (0.84)	1.76 (0.95)	1.76 (0.95)	1.78 (0.95)	0.4	0.5	0.3	0.81 (0.80)	0.81 (0.80)	0.81 (0.80)	7.55 (0.94)	7.56 (0.94)	7.13 (0.95)	
1750	160	0.4	0.4	0.5	0.80 (0.79)	0.80 (0.79)	0.80 (0.79)	8.95 (0.95)	8.95 (0.95)	8.85 (0.95)	0.3	0.3	0.2	0.83 (0.82)	0.83 (0.82)	0.83 (0.82)	5.57 (0.95)	5.56 (0.95)	4.82 (0.95)	
πPT	MAS																			
1000	59	5.6	5.6	5.4	2.26 (2.25)	2.26 (2.25)	2.26 (2.25)	-2.49 (0.94)	-2.48 (0.94)	-2.10 (0.93)	3.9	4.0	3.8	2.10 (2.09)	2.10 (2.09)	2.10 (2.09)	7.90 (0.94)	7.97 (0.94)	4.55 (0.94)	
1000	117	-1.4	-1.4	-1.5	2.15 (2.14)	2.15 (2.14)	2.15 (2.14)	2.40 (0.94)	2.40 (0.94)	2.62 (0.94)	-2.5	-2.4	-2.5	2.28 (2.27)	2.28 (2.27)	2.28 (2.27)	-0.36 (0.93)	-0.33 (0.93)	-3.53 (0.93)	
1000	160	-0.3	-0.3	-0.3	2.27 (2.26)	2.27 (2.26)	2.27 (2.26)	-2.84 (0.93)	-2.84 (0.93)	-2.59 (0.93)	5.1	5.2	5.0	2.32 (2.31)	2.33 (2.32)	2.32 (2.31)	-1.42 (0.93)	-1.46 (0.93)	-4.25 (0.93)	
1750	59	2.7	2.7	2.9	1.12 (1.11)	1.12 (1.11)	1.12 (1.11)	13.09 (0.95)	13.09 (0.95)	13.27 (0.95)	-0.8	-0.7	-0.9	1.32 (1.31)	1.32 (1.31)	1.32 (1.31)	-0.86 (0.93)	-0.88 (0.93)	-3.76 (0.93)	
1750	117	4.2	4.2	4.4	1.17 (1.16)	1.17 (1.16)	1.17 (1.16)	9.37 (0.96)	9.37 (0.96)	9.45 (0.96)	1.5	1.6	1.5	1.23 (1.22)	1.24 (1.23)	1.23 (1.22)	5.41 (0.95)	5.39 (0.95)	2.22 (0.94)	
1750	160	1.0	1.0	1.2	1.22 (1.21)	1.22 (1.21)	1.22 (1.21)	5.71 (0.94)	5.71 (0.94)	5.60 (0.94)	3.3	3.3	3.3	1.23 (1.22)	1.23 (1.22)	1.23 (1.22)	5.63 (0.94)	5.61 (0.94)	2.53 (0.94)	
πPT	πPT																			
1000	59	5.5	5.5	5.3	2.38 (2.37)	2.38 (2.37)	2.38 (2.37)	-4.24 (0.92)	-4.24 (0.92)	-4.31 (0.92)	-4.8	-4.7	-4.9	2.21 (2.20)	2.21 (2.20)	2.21 (2.20)	2.17 (0.91)	2.16 (0.91)	-1.02 (0.91)	
1000	117	-0.9	-0.9	-0.9	2.34 (2.33)	2.34 (2.33)	2.34 (2.33)	-5.05 (0.93)	-5.05 (0.93)	-4.81 (0.93)	-2.6	-2.4	-2.6	2.21 (2.20)	2.22 (2.21)	2.21 (2.20)	2.33 (0.93)	2.29 (0.93)	-0.68 (0.93)	
1000	160	0.4	0.4	0.3	2.07 (2.06)	2.07 (2.06)	2.06 (2.05)	7.94 (0.95)	7.95 (0.95)	8.32 (0.94)	10.2	10.3	10.2	2.18 (2.17)	2.18 (2.17)	2.18 (2.17)	8.29 (0.95)	8.27 (0.95)	4.50 (0.94)	
1750	59	4.9	4.9	5.0	1.17 (1.16)	1.17 (1.16)	1.17 (1.16)	9.81 (0.95)	9.82 (0.95)	9.90 (0.95)	1.9	2.0	1.8	1.23 (1.22)	1.23 (1.22)	1.23 (1.22)	6.17 (0.94)	6.13 (0.94)	2.98 (0.94)	
1750	117	3.4	3.4	3.5	1.16 (1.15)	1.16 (1.15)	1.16 (1.15)	9.70 (0.95)	9.70 (0.95)	9.65 (0.95)	6.5	6.6	6.3	1.38 (1.37)	1.38 (1.37)	1.38 (1.37)	-3.76 (0.92)	-3.75 (0.92)	-6.45 (0.92)	
1750	160	3.5	3.5	3.6	1.25 (1.24)	1.25 (1.24)	1.25 (1.24)	1.40 (0.95)	1.40 (0.95)	1.17 (0.95)	-0.3	-0.3	-0.5	1.28 (1.27)	1.28 (1.27)	1.28 (1.27)	2.62 (0.93)	2.61 (0.93)	-0.52 (0.93)	

Conclusiones

- Los marcos múltiples y particularmente lo tratado en este trabajo que fueron los marcos duales son útiles en diversas áreas del conocimiento, en las cuales al querer realizar una encuesta por medio de muestreo probabilístico en muchos casos no se cuenta con un marco muestral que de cobertura completa a la población objetivo, pero puede ser posible conseguir otras fuentes de información que resulten en un nuevo marco que unido al inicial completen la cobertura, dando en consecuencia un conjunto de unidades o grupos de unidades que estén presentes en ambos marcos (traslape) y cuya estimación para una variable determinada es el principal interés y tema de investigación en la metodología de marcos duales. En el caso de las encuestas del sector agropecuario del país, particularmente la Encuesta Nacional Agropecuaria - ENA y las encuestas focalizadas para algunos productos, dan la posibilidad de contar con marcos de lista, marcos de áreas y diversos registros administrativos para aplicar la metodología de marcos duales y estimar diversos indicadores asociados a productos de gran importancia como es el caso de las hortalizas, las cuales una participación importante en el Producto Interno Bruto (PIB) agropecuario y para la determinación del índice de precios al consumidor (IPC) de los mismos.
- Al evaluar los estimadores de Hartley, Fuller-Burmeister y Pseudo Máximo Verosímil (Skinner-Rao) por medio del sesgo relativo (SR) se encontró que, para las fracciones muestrales en la segunda etapa del marco A de 40 %, 60 %, 80 % y 100 % (conglomerados) dichos valores son inferiores al 1 %, corroborando empíricamente el insesgamiento de los estimadores de Hartley y Fuller-Burmeister, y el insesgamiento asintótico para el estimador de Pseudo Máxima Verosimilitud, una propiedad que se observa en los resultados obtenidos con los datos obtenidos mediante el modelo, es que a medida que se aumenta dicha fracción muestral van disminuyendo los SR, no obstante el aumento en el tamaño de muestra a seleccionar del marco B no produce el mismo efecto. A modo de comparación entre las combinaciones de diseños muestrales para los marcos duales presentados, destaca la combinación $MAS - MAS - MAS$ con menores valores de sesgo relativo, en este aspecto no se divisan marcadas diferencias en los valores de los sesgos relativos al cotejar los dos escenarios.
- En cuanto a eficiencia se refiere, validada por medio del error cuadrático medio empírico (ECME) y varianza estimada (s^2) los tres estimadores tienen diferentes comportamientos dependiendo de la fracción muestral de la segunda etapa del marco A que se utilizó; con las fracciones del 40 % y 60 % al contrastar los escenarios 2 y 3 los ECME tienen valores similares para los 3 estimadores y también dependiendo de la combinación de diseños muestrales. Al aumentarse las fracciones muestrales al

80 % y 100 % se tienen menores valores de ECME y s^2 . Los estimadores de Fuller-Burmeister y Hartley tienen valores muy parecidos en cuanto a ECME y s^2 en la gran mayoría de los casos y el estimador Pseudo Máxima Verosímil en un número importante de casos presenta cantidades de ECME por debajo de los otros dos, pero en el análisis conjunto de todos los resultados no se podría establecer un "ganador" en este sentido. Al realizar un análisis interno dentro de cada fracción muestral, el error cuadrático medio y la varianza estimada disminuyen cuando se aumenta el tamaño de muestra de la primera etapa del marco A, pero al igual que para el sesgo relativo no se encuentra esta relación con el tamaño de muestra para el marco B. También se aprecia que bajo la combinación de diseños muestrales con *MAS* en la primera etapa del marco A estos valores son más bajos que para los que tienen πPT en esta etapa.

- Los puntos más importantes e innovadores que se presentaron en esta tesis son el cálculo del sesgo relativo de la varianza estimada (SRV) y la probabilidad de cobertura (PC), propiedades que permiten complementar la bibliografía disponible en la actualidad sobre los tres estimadores presentados. Al aumentar la fracción muestral de la primera etapa del marco A va disminuyendo el SRV, específicamente se puede observar una disminución significativa al aumentar la fracción del 60 % al 80 %. Otra característica que se debe resaltar, es que a pesar de que en general los valores de SRV están por debajo del 15 %, para algunas combinaciones existen valores de más del 50 % lo que representaría una subestimación o sobreestimación importante en la varianza. Al cotejar los tres estimadores, el estimador de Fuller-Burmeister y Pseudo Máximo Verosímil tienen valores menores que el de Hartley, este hecho no implica que tengan los valores más cercanos a cero porque para algunos valores hay una subestimación importante de la varianza. A diferencia de los valores de SR y ECME los valores de SRV no tienen una relación directa con el tamaño muestral de la primera etapa del marco A, esto tampoco sucede al comparar los dos escenarios y las combinaciones de diseños muestrales. Estas características para los valores de SRV también se presentan para los valores de PC, donde se destaca que el estimador Pseudo Máximo Verosímil tiene valores generalmente menores o iguales (subcobertura) a Fuller-Burmeister y Hartley, no obstante los tres estimadores tienen valores de PC que en todos los casos no son del 95 % variando entre el 93 % y el 97 %, aspecto importante a tener en cuenta al realizar inferencia sobre los totales estimados bajo la metodología de marcos duales.
- Al revisar los resultados basados en los datos tomados directamente de las mediciones de la ENA, los menores valores en cuanto a SR bajo los dos escenarios se encuentran para la combinación de diseños muestrales *MAS* – *MAS* esta cualidad también se presenta para los valores de ECME y s^2 , pero no para los valores de SRV en donde no se destaca claramente alguna combinación de diseños muestrales con valores más pequeños. A nivel de los estimadores, no se destaca alguna mejores características en cuanto a ECME y s^2 , pero en cuanto a SRV el estimador Pseudo Máximo Verosímil presenta generalmente los menores valores. Al igual que para los resultados de los datos provenientes del modelo los valores de PC para todos los casos no alcanzan el 95 %, inclusive llegando en uno de los casos hasta el 92 % lo cual se presenta como una dificultad a la hora de plantear pruebas de hipótesis, sacar conclusiones con los totales estimados y contrastes que se pretendan evaluar. Finalmente, al comparar la facilidad en la que se llega a los resultados es evidente que el estimador de Skinner y Rao disminuye la cantidad de cálculos al no tener que estimar las covarianzas entre

totales de dominios, pero el desarrollo del programa computacional es un poco más complicado.

- A modo de síntesis de los principales resultados encontrados en este documento, y de acuerdo a las características encontradas bajo los escenarios, combinación de diseños muestrales, respectivos tamaños de muestra a extraer de cada marco y las medidas evaluadas de los estimadores; se recomienda usar para la implementación de marcos duales en la ENA el estimador de Skinner-Rao, ya que permite tener consistencia en las estimaciones de los subgrupos, además su cálculo no involucra las covarianzas entre dominios lo que limita los resultados a la variable que se este utilizando, con la desventaja adicional que se presentaría al tener 3 o más marcos, en cuanto a la combinación de diseños muestrales se recomienda el uso de *MAS* tanto para la primera etapa del marco A como para la muestra a extraer del marco B, y para la segunda etapa del marco se puede trabajar con *MAS* y πPT , finalmente los tamaños de muestra dependerán de la precisión que se requiera para las estimaciones y el escenario de trabajo dependerá del conocimiento de los tamaños de los dominios.

Trabajo futuro

Como continuación de este trabajo más adelante se presentan algunas ideas derivadas de la investigación realizada para el cumplimiento de los objetivos del mismo, por ejemplo el valor de p para el estimador de Hartley que podría interpretarse como la ponderación para la estimación que aporta cada marco, en la práctica no es tal en muchos de los casos, puesto que p no siempre se encuentra entre 0 y 1, por esta razón una opción puede ser la de replantear el cálculo de p colocando restricciones de tal forma que su valor caiga dentro del intervalo mencionado.

A pesar de que las estrategias muestrales evaluadas en este documento tiene un grado de dificultad importante comparado con las planteadas en otros trabajos, en la práctica la estimación de la varianza para algunos diseños muestrales y para parámetros diferentes a totales puede ser complicada en el caso de marcos duales, los métodos de estimación van desde métodos estándar para muestreos estratificados, hasta técnicas de linealización de Taylor, métodos de remuestreo como Jackknife y Bootstrap y otras propuestas como en Demnati et al. (2007) y Demnati & Rao (2004).

Otro de los posibles trabajos a futuro que se podrían plantear tienen que ver con la continuación de los programas de computo para los estimadores propuestos, ya que en esta tesis se trabajo para el caso de dos marcos y en algunas investigaciones se puede requerir de 3 o más marcos para dar cobertura completa a la población, esto implicaría un aumento en la cantidad de cálculos a efectuar debido a las covarianzas entre los nuevos dominios que se generan y sería un gran aporte una generalización de las rutinas planteadas para el caso de 2 o más marcos muestrales involucrados. Además de incorporar más de dos marcos en las rutinas, también se pueden complementar con las propuestas que involucran información auxiliar, ya no para el diseño muestral sino en los estimadores.

Bibliografía

- Bankier, M. . D. (1986), ‘Estimators based on severals stratified samples with applications to multiple frame surveys’, *Journal of the American Statistical Association* **81**, 1074–1079.
- Bautista, L. (1994), *Modelo de costos y diseño muestral del sistema de información agropecuaria para la estimaciñn de ñreas producciñn y rendimiento*, DANE, Bogotá.
- Cochran, W. G. (1977), *Sampling Techniques*, 3rd. edn, John Wiley and Sons, Inc., New York:.
- Coelho, H. (2007), A abordagem de cadastro duplo (dual frame): estimação assistida por modelos lineares con aplicação em pesquisas agropecuárias, Tesis de Maestría, Universidade Federal de Pernambuco, Estatística, Recife.
- Corporación Colombia Internacional - CCI (2010), *Metodología del diseño de la encuesta nacional agropecuaria: componente de muestreo*, Bogotá, Colombia.
- Demnati, A. & Rao, J. N. K. (2004), ‘Linearization Variance Estimators for Survey Data’, *Survey Methodology* **30**, 17–26.
- Demnati, A., Rao, J. N. K., Hidiroglou, M. A. & Tambay, J. L. (2007), *Linearization Variance Estimators for Dual Frame Survey Data*, Paper presented at the Joint Statistical Meetings, Salt Lake City.
- Departamento Administrativo Nacional de Estadística - DANE (2010), *Encuesta Nacional Agropecuaria - ENA*, Bogotá, Colombia.
URL: http://www.dane.gov.co/index.php?option=com_contentview = articleid = 240Itemid = 73
- Ferraz, C. & Coelho, H. (2009), ‘Inference for finite population under a dual frame approach’, *Federal University of Pernambuco* pp. 1–69.
- Fuller, W. A. & Burmeister, L. F. (1972), ‘Estimators for Samples Selected From Two Overlapping Frames’, in *ASA Proceedings of the Social Statistics Section* pp. 245–249.
- Gonzalez-Villalobos, A. & Wallace, M. (1991), ‘Multiple Frame Agriculture Surveys’, *Rome: Food and Agriculture Organization of the United Nations* **Vols. 1 and 2**.

-
- Gutiérrez, A. (2010), *Package 'TeachingSampling' Versión 1.4.9*, R Foundation for Statistical Computing.
URL: <http://www.R-project.org>
- Hartley, H. O. (1962), 'Multiple Frame Surveys', in *ASA Proceedings of the Social Statistics Section* pp. 203–206.
- Hartley, H. O. (1974), 'Multiple Frame Methodology and Selected Applications', *Sankhya, the Indian Journal of Statistics* **36**, 99–118.
- Kalton, G. & Anderson, D. W. (1986), 'Sampling Rare Populations', *Journal of the Royal Statistical Society Ser A.*, **149**, 65–82.
- Lohr, S. L. & Rao, J. N. K. (2000), 'Inference from Dual Frame Surveys', *Journal of the American Statistical Association* **95**, 271–280.
- Lohr, S. L. & Rao, J. N. K. (2006), 'Estimation in Multiple-Frame Surveys', *Journal of the American Statistical Association* **101**, 1019–1030.
- Lumley, T. (2010), *Package 'survey' Versión 3.22-2*, R Foundation for Statistical Computing.
URL: <http://www.R-project.org>
- Ospina, D. (1981), 'El problema de la estimación de parámetros cuando se trabaja con marcos de muestra superpuestos', *Revista Colombiana de Estadística* **4**, 1–20.
- R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>
- Rao, J. N. K. & Changbao, W. (2010), 'Pseudo Empirical Likelihood Inference for Multiple Frame Surveys', *Journal of the American Statistical Association* **105**, 1494–1503.
- Skinner, C. . J. (1991), 'On the Efficiency of Raking Ratio Estimation for Multiple Frame Surveys', *Journal of the American Statistical Association* **86**, 779–784.
- Skinner, C. J. & Rao, J. N. K. (1996), 'Estimation in Dual Frame Surveys With Complex Designs', *Journal of the American Statistical Association* **91**, 349–356.
- Särndal, C., Swensson, B. & Wretman, J. (1992), *Model-assisted Survey Sampling*, Springer-Verlag, New York.
- Tillé, Y. & Matie, A. (2009), *Package 'sampling' Versión 2.3*, R Foundation for Statistical Computing.
URL: <http://www.R-project.org>

APÉNDICE

Programa en R

```
##### CARGA DE LIBRERIAS #####
library(sampling)
library(survey)
library(TeachingSampling)
##### FUNCIONES #####
#Crea una población artificial de tamaño N basado en datos de una muestra
#de tamaño n, el campo "name" indica que nombre se quiere colocar a la variable
artipop<-function(N,y,name)
{
  y<-as.vector(unlist(y))
  n<-length(y)
  nv<-log(N/n)/(log(2))
  nvr<-floor(nv+1)
  for(k in 1:nvr){
    y<-append(y,y, after=n)}
  var<-data.frame(y[1:N])
  names(var)<-c(name)
  return(var)
}
#Extrae muestras aleatorias simples para diseños estratificados o cluster
#"x" dataframe con el marco, "st" variable de estratificación o cluster
#"f" fracción muestral dentro de cada estrato o cluster
Sterswor=function(x,st,f,method,aux)
{
  N <- apply(data.frame(x[st]), 2, table)
  n <- round(f*N)
  a=matrix(0,nrow=sum(n),ncol=1)
  b=matrix(0,nrow=sum(n),ncol=1)
  pik=matrix(0,nrow=sum(n),ncol=1)
  ns=matrix(0,nrow=sum(n),ncol=1)
  Ns=matrix(0,nrow=sum(n),ncol=1)
  cum=cumsum(n)
  cum2=cumsum(N)
  iden=c(1:sum(N))
  H <- length(n)
  if(method=="srswor"){
    for(k in 1:length(n)){
      if(k==1)
      {
        if(n[k]<1){n[k]<-1}
        a[1:n[k]]<-as.matrix(sample(iden[1:N[k]],n[k]))
        pik[1:n[k]]<- n[k]/N[k]
        Ns[1:n[k]]<- N[k]
        ns[1:n[k]]<- n[k]
      }
      if(k>1)
      {
        if(n[k]<1){n[k]<-1}
        a[(cum[k-1]+1):(cum[k])]<-as.matrix(sample(iden[(cum2[k-1]+1):(cum2[k])],n[k]))
        pik[(cum[k-1]+1):(cum[k])] <- n[k]/N[k]
        Ns[(cum[k-1]+1):(cum[k])] <- N[k]
        ns[(cum[k-1]+1):(cum[k])] <- n[k]
      }
    }
    sample <- x[a,]
    sam <- data.frame(sample,pik,Ns,ns)
    return(sam)
  }
  if(method=="pips"){
    sizes <-x[,aux]
    s <- S.piPS(n[1],sizes[1:N[1]])
    stratstart <- 1 + N[1]
    if (H > 1)
      for (h in 2:H) {
```

```

        stratend <- stratstart + N[h] - 1
        tmp <- S.piPS(n[h],sizes[stratstart:stratend])
tmp2 <- tmp[,1]+ stratstart - 1
tmp3 <- cbind(tmp2,tmp[,2])
        s <- rbind(s, tmp3)
        stratstart <- stratend + 1
    }
s2=x[s[,1],]
s2=data.frame(s2,pik=s[,2])
    return(s2)
}
}
lstage=function (x,st,f,method,aux)
{
    Nh <- apply(data.frame(x[st]), 2, table)
n <- round(f*Nh)
nm1 <- which(n<1)
n[nm1] <- 1
H <- length(n)
if(method=="pips"){
    sizes <-x[,aux]
        s <- S.piPS(n[1],sizes[1:Nh[1]])
        stratstart <- 1 + Nh[1]
        if (H > 1)
            for (h in 2:H) {
                stratend <- stratstart + Nh[h] - 1
                tmp <- S.piPS(n[h],sizes[stratstart:stratend])
tmp2 <- tmp[,1]+ stratstart - 1
tmp3 <- cbind(tmp2,tmp[,2])
                s <- rbind(s, tmp3)
                stratstart <- stratend + 1
            }
s2=x[s[,1],]
s2=data.frame(s2,pik=s[,2])
    return(s2)
}
}
#Extrae muestras en dos etapas bajo los diseños PIPT y MAS
twostage<-function(frame,st1,nI,f,design,PIK=NULL,auxI,auxII)
{
m1=design[1]
m2=design[2]
if(m1=="srswor"){
mI=cluster(frame,st1, nI, method=m1,pik=PIK,description=FALSE)
sI=getdata(frame,mI)
}
if(m1=="pips"){
ag=aggregate(marcoA[,auxI],list(marcoA[,st1]),max)
ag2=ag[,1]
res=S.piPS(nI,ag[,2])
sel=ag2[res[,1]]
sel2=data.frame(sel,res[,2])
names(sel2)<-c(st1,"Prob")
sI=merge(frame,sel2)
}
sII=Stsrswor(sI,st1,f,method=m2,auxII)
sII=data.frame(sII,pik2st=sII$Prob*sII$pik)
return(sII)
}
#Calcula la varianza estimada del estimador HT (pi-estimador) para 2 dominios
VarPi_est <- function(deltak1,Pik1,Pik,yk,dom,d1,d2)
{
d <- cbind(1,2)
label<- cbind(d1,d2)
if (dom[1]==d2) {
d <- cbind(2,1)
label<- cbind(d2,d1)
}
clas <- levels(as.factor(dom))
ndom <- length(clas)
zd <- Domains(dom)
yzd <- yk*zd
VARPI <- matrix(0,nrow=ndom,ncol=1)
Vi <- matrix(0,nrow=ndom,ncol=1)
ti <- matrix(0,nrow=ndom,ncol=1)
TPI <- matrix(0,nrow=ndom,ncol=1)
CVE <- matrix(0,nrow=ndom,ncol=1)
for (h in 1:ndom){
for (k in 1:length(Pik)){
ti[h] <- yzd[k,h]/Pik[k]
TPI[h] <- ti[h]+ TPI[h]
for (l in 1:length(Pik)){
Vi[h] <- -0.5*((deltak1[k,l])/Pik1[k,l])*(yzd[k,h]/Pik[k])-(yzd[l,h]/Pik[l])^2
VARPI[h] <- Vi[h] + VARPI[h]
}
}
}
CVE <- 100*sqrt(VARPI)/TPI
stats <- cbind(TPI,VARPI,CVE)
rownames(stats) <- paste("Dom",label,sep=" ")
colnames(stats) <- c("Total", "Variance", "cve")
}
if (d == c(2,1)) {
stats <- stats[d,]
}
}

```



```

return(stats)
}
#Función que calcula los totales por dominios (o postestratos) y las
#varianzas de los dominios (o postestratos) necesarias para el cálculo
#de las varianzas en los estimadores de Hartley, Fuller-Burmeister y PML
#"s" dataframe con la muestra, "y" vector de la variable de interés,
#"iden" identificador utilizado cuando hay varias etapas (si hay
#una etapa colocar "0 si hay dos o más etapas colocar el identificador de
#las "UPMS + USMS + ..."), "pik" vector con las probabilidades de
#inclusión, "dom" vector que identifica a que dominio pertenece la unidad
#si al dominio (a) o si está en el traslape, "dab" número o carácter
#indicando el nombre del dominio traslapado, "over" valor lógico
#indica si se conoce Nab o se desconoce, "size" tamaños de los dominios
#poblacionales Na y Nab (escenario 2)
DomainEstimate <- function(s,y,iden,pik,dm,dab,over=FALSE,size=NULL)
{
y<-data.frame(y)
one<-matrix(1,nrow(y))
vars<-data.frame(y,one)
colnames(vars)<- c("y","N")
s<-data.frame(s,N=1)
domain<-dm
lev<-levels(factor(domain))
des <-svydesign(id=iden,variables=vars,weights=1/pik, data=s)
if(over==FALSE){
  Estimation <- svyby(vars,~domain, data = s , design=des, svytotal)
  Estimation <- data.frame(Estimation)
  nres<-nrow(Estimation)
  if(nres==1 || is.na(Estimation[2,2])){Estimation<-rbind(Estimation,c(-999999,0,0,0,0))}
  if(Estimation[1,1]==dab){
    Estimation<-Estimation[c(2,1),]}
  if(over==TRUE){
    npos<-data.frame(size)
    names(npos)<-c("domain","Freq")
    clas <- Domains(dm)
    if(ncol(clas==1)){clas<-cbind(clas,0)}
    ya<- as.matrix(y*clas[,1])
    yab <- as.matrix(y*clas[,2])
    onea<- as.matrix(one*clas[,1])
    oneab<- as.matrix(one*clas[,2])
    vars<- data.frame(ya,yab,onea,oneab)
    pos <- postStratify(des, ~domain, npos,partial=TRUE)
    espos <-data.frame(svytotal(vars, pos))
    Estimation <- data.frame(lev,espos[1:2,1],espos[3:4,1],espos[1:2,2],espos[3:4,2])
    names(Estimation)<- c("domain","y","N","se.y","se.N")
    nres<-nrow(Estimation)
    if(nres==1 || is.na(Estimation[2,2])){Estimation<-rbind(Estimation,c(-999999,0,0,0,0))}
    if(Estimation[1,1]==dab){
      Estimation<-Estimation[c(2,1),]}
    }
  Estima<-list()
  Estima$estimation <- Estimation
  Estima$design <- des
  Estima$sample <- s
  Estima$data <- y
  Estima$weights <- weights(des)
  Estima$domains <- dm
  Estima$overlap <- dab
  Estima$iden <- iden
  Estima$nres <- nres
  Estima
}
#Estima los deltakl (necesarias para la varianza de HT) mediante el método
#de Hartley-Rao
deltaHR <- function(pik)
{
deltaest=matrix(nrow=length(pik),ncol=length(pik))
spi<-sum(pik)
n <-length(pik)
mpi<-spi/n
for (i in 1:length(pik)){
for (j in 1:length(pik)){
deltaest[i,j]<- 1-(n-pik[i]-pik[j]+mpi)/(n-1)}}
return(deltaest)
}
#Estima los deltakl (necesarias para la varianza de HT) mediante el método
#de Overton
delta0 <- function(pik)
{
deltaest=matrix(nrow=length(pik),ncol=length(pik))
n <-length(pik)
for (i in 1:length(pik)){
for (j in 1:length(pik)){
deltaest[i,j]<- 1-(n-(pik[i]+pik[j])/2)/(n-1)}}
return(deltaest)
}
#Función que calcula las covarianzas entre dominios(o postestratos)
#necesarias para el cálculo de las varianzas en los estimadores
#de Hartley y Fuller-Burmeister
#"method" es el campo del método a utilizar por defecto aplica el del paquete
#survey pero también se puede aplicar el método 1 (Rao) el método 2 (Overton)
# o el método 3 (exacto) el cual requiere de los deltakl y los pikl
#"y" vector de la variable de interés, "dom" vector que identifica a que
#dominio pertenece la unidad si al dominio (a) o si está en el traslape,

```

```

#"dab" número o carácter indicando el nombre del dominio traslapado
CovarDomain<- function (method=FALSE,deltakl,pikl,iden,pik,y,dom,dab,over=FALSE,size=NULL)
{
  y<-data.frame(y)
  clas <- Domains(dom)
  if(ncol(clas==1)){clas<-cbind(clas,0)}
  if(dom[1]==dab){
    clas <- clas[,c(2,1)]}
  ya<- as.matrix(y*clas[,1])
  yab <- as.matrix(y*clas[,2])
  oneab<-matrix(1,nrow(y))
  oneab<- as.matrix(oneab[,2])
  vars<-data.frame(ya,yab,oneab,pik,dom)
  names(vars)<-c("ya","yab","oneab","pik","clasA")
  vari<-names(vars[,1:3])
  des <-svydesign(id=iden,weights=1/pik, data=vars)
  if(over==TRUE){
    npos<-data.frame(size)
    names(npos)<-c("domain","Freq")
    domain<-dom
    des <- postStratify(des, ~domain, npos,partial=TRUE)}
    tdom<-svytotal(make.formula(vari),des,covmat=TRUE)
    covdom<-vcov(tdom)
    Covarpi<-c(covdom[1,2],covdom[1,3],covdom[2,3])
    if(method==1)
    {deltaexp <- as.matrix(delta0(pik))}
    if(method==2){
      deltaexp <- deltaHR(pik)}
    if(method==3) {deltaexp <- deltakl/pikl}
    if(method==1|method==2|method==3){
      yayabexp <- (ya/pik) %*% t(yab/pik)
      s1 <- (deltaexp) * (yayabexp)
      Cov1 <- sum(s1)
      yaNabexp <- (ya/pik) %*% t(oneab/pik)
      s2 <- (deltaexp) * (yaNabexp)
      Cov2 <- sum(s2)
      yabNabexp <- (yab/pik) %*% t(oneab/pik)
      s3 <- (deltaexp) * (yabNabexp)
      Cov3 <- sum(s3)
      Covarpi<-data.frame(Cov1,Cov2,Cov3)}
    names(Covarpi)<-c("Cov(tyd1,tyd2)","Cov(tyd1,Nd2)","Cov(tyd2,Nd2)")
    return(Covarpi)
  }
  #Función que calcula el p-óptimo para el estimador de Hartley
  #Requiere la función DomainEstimate aplicada al marco A y al marco B
  #Requiere la función CovarDomain aplicada al marco A y al marco B
  popth <- function(tA,tB,covA,covB)
  {
    vabA <- (tA$estimation[2,4])^2
    vabB <- (tB$estimation[2,4])^2
    covaabA <-covA[1]
    covabBb <-covB[1]
    pH <- (vabB+covabBb-covaabA)/(vabB+vabA)
    names(pH)<- c("poptH")
    return(pH)
  }
  #Función que calcula el total para marcos duales usando el estimador
  #de Hartley, Requiere la función DomainEstimate aplicada al marco A
  #y al marco B y el valor de p
  tyHest <- function(tA,tB,p)
  {
    tya <- tA$estimation[1,2]
    tyabA<- tA$estimation[2,2]
    tyb <- tB$estimation[1,2]
    tyabB<- tB$estimation[2,2]
    pH<-p
    if(tyabA==0){pH<-0}
    if(tyabB==0){pH<-1}
    if((tyabA==0 && tyabB==0) || tA$estimation[1,3]==0 || tB$estimation[1,3]==0){return(tyH<- -9999999)}
    tyH <- tya + pH*tyabA + (1-pH)*tyabB + tyb
    names(tyH)<-c("tyH")
    return(tyH)
  }
  #Función que estima la varianza para el estimador de Hartley
  #Requiere la función DomainEstimate aplicada al marco A y al marco B
  #Requiere la función CovarDomain aplicada al marco A y al marco B
  #y el valor de p
  VarH <- function(tA,tB,covA,covB,p){
    va<-(tA$estimation[1,4])^2
    vabA <- (tA$estimation[2,4])^2
    vb<-(tB$estimation[1,4])^2
    vabB <- (tB$estimation[2,4])^2
    covaabA <-covA[1]
    covabBb <-covB[1]
    VH <-((p^2)*(vabA+vabB))+p*(2*covaabA-2*vabB-2*covabBb)+(va+vb+vabB+2*covabBb)
    names(VH)<- c("VarH")
    return(VH)
  }
  #Función que estima el tamaño del traslape para el estimador de Fuller-Burmeister
  #Requiere la función DomainEstimate aplicada al marco A y al marco B
  NabFB_est <- function(tA,tB)
  {
    sA<-tA$sample
    sB<-tB$sample
  }

```

```

dA<-tA$domains
dB<-tB$domains
dab<-tA$overlap
domabA<-subset(sA,dA==dab)
domabB<-subset(sB,dB==dab)
nabA<-nrow(domabA)
nabB<-nrow(domabB)
N_A<-tA$estimation[1,3]+tA$estimation[2,3]
NB<- tB$estimation[1,3]+tB$estimation[2,3]
NabA<-tA$estimation[2,3]
NabB<-tB$estimation[2,3]
nA<-nrow(sA)
nB<-nrow(sB)
gA<-(N_A - nA)/(N_A - 1)
gB<-(NB - nB)/(NB - 1)
z1 <-(nabA*gB+nabB*gA)*N_A*NB
z2 <-nA*NB*gB+nB*N_A*gA+nabA*N_A*gB+nabB*NB*gA
z3 <-(nA*gB)+(nB*gA)
if(missing(z1) || is.null(z1) || is.na(z1) || !is.numeric(z1)){z1<-0}
if(missing(z2) || is.null(z2) || is.na(z2) || !is.numeric(z2)){z2<-0}
if(missing(z3) || is.null(z3) || is.na(z3) || !is.numeric(z3)){z3<-0}
roots<-polyroot(c(z1,-z2,z3))
NabFB <- min(Re(roots))
names(NabFB)<-c("NabFB")
return(NabFB)
}
#Función que calcula los p-óptimos (p1 y p2) para el estimador de
#Fuller y Burmeister
#Requiere la función DomainEstimate aplicada al marco A y al marco B
#Requiere la función CovarDomain aplicada al marco A y al marco B
poptFB <- function(tA,tB,covA,covB)
{
vabA <- (tA$estimation[2,4])^2
vabB <- (tB$estimation[2,4])^2
covaabA <-covA[1]
covabbB <-covB[2]
vNabA <- (tA$estimation[2,5])^2
vNabB <- (tB$estimation[2,5])^2
covaNabA <-covA[2]
covbNabB <-covB[2]
covabANabA <-covA[3]
covabBNabB <-covB[3]
covx1x2 <- covaabA - covabbB - vabB
varx3 <- vNabA+vNabB
covx1x3 <- covaNabA - covbNabB - covabBNabB
covx2x3 <- covabANabA + covabBNabB
varx2 <- vabA + vabB
p1 <- ((covx1x2+varx3) - (covx1x3+covx2x3))/((covx2x3^2)-(varx3+varx2))
p2 <- ((-covx1x2+covx2x3) + (varx2+covx1x3))/((covx2x3^2)-(varx3+varx2))
if(is.na(p1) || !is.numeric(p1)){p1<-0.5}
if(is.na(p2) || !is.numeric(p2)){p2<-0}
pFB <- cbind(p1,p2)
names(pFB)<- c("p1FB", "p2FB")
return(pFB)
}
#Función que calcula el total para marcos duales usando el estimador
#de Fuller-Burmeister
#Requiere la función DomainEstimate aplicada al marco A y al marco B
#el campo NabFB indica si se quiere estimar el tamaños del traslape
#requiere de los valores p1 y p2
tyFB_est <- function(tA,tB,NabFB=FALSE,p1,p2)
{
tya <- tA$estimation[1,2]
tyabA<- tA$estimation[2,2]
tyb <- tB$estimation[1,2]
tyabB<- tB$estimation[2,2]
NabA<-tA$estimation[2,3]
NabB<-tB$estimation[2,3]
pFB1<p1
if(tyabA==0){pFB1<-0}
if(tyabB==0){pFB1<-1}
if((tyabA==0 && tyabB==0) || tA$estimation[1,3]==0 || tB$estimation[1,3]==0){return(tyH<- -9999999)}
if (NabFB){
NabA<-NabFB
NabB<-NabFB}
tyFB <- tya + pFB1*tyabA + (1-pFB1)*tyabB + tyb + p2 *(NabA - NabB)
names(tyFB)<- c("tyFB")
return(tyFB)
}
#Estima la varianza del estimador de Fuller-Burmeister
#Requiere la función DomainEstimate aplicada al marco A y al marco B
#Requiere la función CovarDomain aplicada al marco A y al marco B
#Requiere los valores de p1 y p2
VarFB <- function(tA,tB,covA,covB,p1,p2)
{
va <- (tA$estimation[1,4])^2
vb <- (tB$estimation[1,4])^2
vabA <- (tA$estimation[2,4])^2
vabB <- (tB$estimation[2,4])^2
covaabA <-covA[1]
covabbB <-covB[2]
vNabA <- (tA$estimation[2,5])^2
vNabB <- (tB$estimation[2,5])^2
covaNabA <-covA[2]

```

```

covbNabB <-covB[2]
covabANabA <-covA[3]
covabBNabB <-covB[3]
covx1x2 <- covaab - covabBb - vabB
varx3 <- vNabA+vNabB
covx1x3 <- covaNabA - covbNabB - covabBNabB
covx2x3 <- covabANabA + covabBNabB
varx2 <- vabA + vabB
varx1 <- va+vb+2*covabBb
VFB <- varx1+2*p1*covx1x2+2*p2*covx1x3+(p1^2)*varx2+2*p1*p2*covx2x3+(p2^2)*varx3
names(VFB)<- c("VFB")
return(VFB)
}
#Función que estima el tamaño del traslape para el estimador PML
#Requiere la función DomainEstimate aplicada al marco A y al marco B
NabPML_est <- function(tA,tB)
{
  sA<-tA$sample
  sB<-tB$sample
  N_A<-tA$estimation[1,3]
  NB<- tB$estimation[1,3]
  NabA<-tA$estimation[2,3]
  NabB<-tB$estimation[2,3]
  nA<-nrow(sA)
  nB<-nrow(sB)
  z1<-z2<-z3<-0
  z1 <-nA*NabA*NB+nB*NabB*N_A
  z2 <-nA*NB+nB*N_A+nA*NabA+nB*NabB
  z3 <-nA*nB
  if(missing(z1) || is.null(z1) || is.na(z1) || !is.numeric(z1)){z1<-0}
  if(missing(z2) || is.null(z2) || is.na(z2) || !is.numeric(z2)){z2<-0}
  if(missing(z3) || is.null(z3) || is.na(z3) || !is.numeric(z3)){z3<-0}
  roots<-polyroot(c(z1,-z2,z3))
  NabPMV <- min(Re(roots))
  names(NabPMV)<- c("NabPMV")
  return(NabPMV)
}
#Función que calcula el p-óptimo para el estimador PML
#Requiere la función DomainEstimate aplicada al marco A y al marco B
poptPML <- function(tA,tB)
{
  Na <- tA$estimation[1,3]
  Nb <- tB$estimation[1,3]
  vNabA <- (tA$estimation[2,5])^2
  vNabB <- (tB$estimation[2,5])^2
  N_A<-tA$estimation[1,3]+tA$estimation[2,3]
  NB<- tB$estimation[1,3]+tB$estimation[2,3]
  pPML <- (Na*NB*vNabB)/((Na*NB*vNabB)+(Nb*N_A*vNabA))
  if(is.na(pPML) || !is.numeric(pPML)){pPML<-0.5}
  names(pPML)<- c("pPML")
  return(pPML)
}
#Función que calcula el total para marcos duales usando el estimador PML
#Requiere la función DomainEstimate aplicada al marco A y al marco B
#Requiere la función NabPML_est y el valor de p-optimo
tyPML_est <- function(tA,tB,NabPMV_est,p3)
{
  tya <- tA$estimation[1,2]
  tyabA<- tA$estimation[2,2]
  tyb <- tB$estimation[1,2]
  tyabB<- tB$estimation[2,2]
  N_A<-tA$estimation[1,3]+tA$estimation[2,3]
  NB<- tB$estimation[1,3]+tB$estimation[2,3]
  NabA<-tA$estimation[2,3]
  NabB<-tB$estimation[2,3]
  Na <- tA$estimation[1,3]
  Nb <- tB$estimation[1,3]
  if(NabA==0 || is.na(NabA)){
    tyabA<-0}
  if(NabB==0 || is.na(NabB)){
    tyabB<-0}
  pPML<-p3
  if(tyabA==0){pPML<-0}
  if(tyabB==0){pPML<-1}
  if((tyabA==0 && tyabB==0) || (NabB==0 && NabA==0) || Na==0 || Nb==0){return(tyH<- -9999999)}
  tyPML <- ((N_A-NabPMV_est)*tya/Na) + (NabPMV_est *(pPML*tyabA + (1-pPML)*tyabB))/(pPML*NabA + (1-pPML)* NabB) + ((NB-NabPMV_est)*tyb/Nb)
  names(tyPML)<- c("tyPML")
  return(tyPML)
}
#Estima la varianza del estimador PML
#Requiere la función DomainEstimate aplicada al marco A y al marco B
#Requiere la función CovarDomain aplicada al marco A y al marco B
VarPML <- function(tA,tB)
{
  sA<-tA$sample
  sB<-tB$sample
  dA<-tA$domains
  dB<-tB$domains
  dab<-tA$overlap
  da<- tA$estimation[1,1]
  db<- tB$estimation[1,1]
  yA<-tA$data
  yB<-tB$data
  pikA<-1/tA$weights

```

```

pikB<-1/tB$weights
vabA <- (tA$estimation[2,4])^2
vabB <- (tB$estimation[2,4])^2
vNabA <- (tA$estimation[2,5])^2
vNabB <- (tB$estimation[2,5])^2
tya <- tA$estimation[1,2]
tyabA<- tA$estimation[2,2]
tyb <- tB$estimation[1,2]
tyabB<- tB$estimation[2,2]
NabA<-tA$estimation[2,3]
NabB<-tB$estimation[2,3]
Na <- tA$estimation[1,3]
Nb <- tB$estimation[1,3]
N_A<-tA$estimation[1,3]+tA$estimation[2,3]
NB<- tB$estimation[1,3]+tB$estimation[2,3]
nA<-nrow(sA)
nB<-nrow(sB)
mua <- tya/Na
mub <- tyb/Nb
muab <- ((nA/N_A)*tyabA)+(nB/NB)*tyabB)/((nA/N_A)*NabA)+(nB/NB)*NabB)
theta <- (Na*NB*vNabB)/((Na*NB*vNabB)+(Nb*N_A*vNabA))
fi <- (nA*Nb)/((nA*Nb)+(nB*Na))
lambda <- muab-mua-mub
zA <- matrix(1,nrow=nA,ncol=1)
zB <- matrix(1,nrow=nB,ncol=1)
iAa<-which(dA==da)
zA[iAa]=yA[iAa,]-mua
iAab<-which(dA==dab)
zA[iAab]=theta*(yA[iAab,]-muab)+(lambda*fi)
iBb<-which(dB==db)
zB[iBb]=yB[iBb,]-mub
iBab<-which(dB==dab)
zB[iBab]=(1-theta)*(yB[iBab,]-muab)+(lambda*(1-fi))
zA <-data.frame(as.numeric(zA))
names(zA)<-c("zA")
sZA<-data.frame(sA,zA)
zB <-data.frame(as.numeric(zB))
names(zB)<-c("zB")
sZB<-data.frame(sB,zB)
desA <-svydesign(tA$iden,variables=zA,weights=1/pikA, data=sZA)
EstZA <- svytotal(~zA, design=desA)
desB <-svydesign(tB$iden,variables=zB,weights=1/pikB, data=sZB)
EstZB <- svytotal(~zB, design=desB)
VPML <- (SE(EstZA)^2)+(SE(EstZB)^2)
names(VPML)<- c("VPML")
return(VPML)
}
#Función que calcula las probabilidades de inclusión de primer orden (pik)
#para un diseño general con remplazamiento
#Requiere los valores de las probabilidades de selección y el tamaño de
#muestra efectivo m (para muestreo con remplazamiento)
PikWR<- function(pk,m){
1-((1-pk)^m)
}
#Función que calcula las probabilidades de inclusión de segundo (pikl)
#para un diseño general con remplazamiento
#Requiere de las probabilidades de selección de las unidades y
#el tamaño de muestra efectivo m (para muestreo con remplazamiento)
PiklWR <- function(pk,pl,m)
{
PIKL=matrix(nrow=length(pk),ncol=length(pl))
for (k in 1:length(pk)){
for (l in 1:length(pl)){
PIKL[k,l]<- 1-((1-pk[k])^m)-((1-pl[l])^m)+((1-pk[k]-pl[l])^m)
}
}
return(PIKL)
}
#Función que calcula las covarianzas entre indicadoras (deltakl)
#para un diseño general con remplazamiento
#Requiere de las probabilidades de inclusión de primer y segundo orden
deltakl<- function(Pikl,Pik,Pil)
{
DELTAKL=matrix(nrow=length(Pik),ncol=length(Pil))
for (k in 1:length(Pik)){
for (l in 1:length(Pil)){
DELTAKL[k,l] <- Pikl[k,l]- Pik[k]*Pil[l]
}
}
return(DELTAKL)
}
#Calcula el total usando el estimador HT (pi-estimador)
#Requiere los datos de la variable "yk" y las probabilidades de inclusión de
#primer orden
Pi_est <- function(yk,Pik)
{
sum(yk/Pik)
}
#Calcula el sesgo, el sesgo relativo el error cuadrático medio
#el sesgo relativo de la varianza y la probabilidad de cobertura
error=function(tyest,varty,ty,alfa)
{
m<-length(tyest)
RB<-(1/m)*sum((tyest-ty)/ty)
}

```

```

B<-mean(tyest-ty)
MSE<-(1/m)*sum((tyest-ty)**2)
### MSE=VAR+BIAS^2 ; VAR=MSE-BIAS^2 #####
VAR<-MSE-(B^2)
RBV<-(1/m)*sum((varty-VAR)/VAR)
Zr<-(tyest-ty)**2
Zbar<-mean(Zr)
S2MSE<-(1/m)*(1/(m-1))*sum((Zr-Zbar)**2)
per<- qnorm(alfa)
ICI<-tyest-per*sqrt(varty)
ICS<-tyest+per*sqrt(varty)
ind<-matrix(0,m)
for(i in 1:m){
if(ty <= ICS[i] && ty >=ICI[i]){ind[i]=1}
COB=sum(ind)/m
result<-data.frame(B,RB,MSE,VAR,RBV,S2MSE,COB)
return(result)
}
##### CARGA DE DATOS #####
#Marco de hortalizas (marco de lista)
marcohor<- read.table("marco hortalizas.tab", header=TRUE, sep="", na.strings="NA", dec=".",
strip.white=TRUE)
#Datos de área sembrada, área cosechada y producción (variables de estudio)
asemh<- read.table("asemh.tab", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
acosh<- read.table("acosh.tab", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
prodh<- read.table("prodh.tab", header=TRUE, sep="", na.strings="NA", dec=".", strip.white=TRUE)
#Marco de la ENA (marco de áreas)
marcoENA <- read.table("universo1.tab", header=TRUE, sep="", na.strings="NA", dec=".",
strip.white=TRUE)
#Identificador de los segmentos y UPMS para el marco de la ENA
IDESM <-paste(marcoENA$COD_DTO,marcoENA$UPM,marcoENA$sm)
IDEUPM<-paste(marcoENA$COD_DTO,marcoENA$UPM)
marcoENA2 <- cbind(marcoENA,IDEUPM,IDESM)
#Identificador de los segmentos para el marco de hortalizas
IDESM <-paste(marcohor$NDEPARTA,marcohor$NUPM,marcohor$NUM_SM)
IDEUPM<-paste(marcohor$NDEPARTA,marcohor$NUPM)
marcohor2 <- cbind(marcohor,IDEUPM,IDESM)
#Creación de la población artificial para la ENA (áreas)
N_A <- nrow(marcoENA2)
Nab <- 521
Na <- N_A - Nab
domA <- append(matrix(0,Na),matrix(1,Nab))
Aasm <- artipop(nrow(marcoENA2),asemh,"area sembrada")
Aacos <- artipop(nrow(marcoENA2),acosh,"area cosechada")
Aprod <- artipop(nrow(marcoENA2),prodh,"produccion")
marcoA <- data.frame(marcoENA2,Aasm,Aacos,Aprod,domA)
#Creación de la población artificial para hortalizas (lista)
NB <- nrow(marcohor2)
Nb <- NB - Nab
basM <- artipop(Nb,asemh,"area sembrada")
bacos <- artipop(Nb,acosh,"area cosechada")
bprod <- artipop(Nb,prodh,"produccion")
Basm <- append(unlist(basM),unlist(Aasm[(Na+1):N_A,]))
Bacos <- append(unlist(bacos),unlist(Aacos[(Na+1):N_A,]))
Bprod <- append(unlist(bprod),unlist(Aprod[(Na+1):N_A,]))
domB <- append(matrix(0,Nb),matrix(1,Nab))
marcoB <- data.frame(marcohor2,Basm,Bacos,Bprod,domB)
tyasm<-sum(marcoA["area sembrada"])+sum(basm)
##### MODELO SIMULACION #####
mub <-mean(marcoB[1:Nb,"Basm"])
sdb <-sd(marcoB[1:Nb,"Basm"])
muab <-mean(marcoB[(Nb+1):NB,"Basm"])
sdab <-sd(marcoB[(Nb+1):NB,"Basm"])
#muab <-mean(marcoA[(Na+1):N_A,"area sembrada"])
#sdab <-sd(marcoA[(Na+1):N_A,"area sembrada"])
mua <-mean(marcoA[1:Na,"area sembrada"])
sda <-sd(marcoA[1:Na,"area sembrada"])
#MODELO: yij=muk + alpha_i + epsilon_ij
epsb <-as.vector(rnorm(Nb, mean=0, sd=sdb)+mub)
epsab <-as.vector(rnorm(Nab, mean=0, sd=sdab)+muab)
epsa <-as.vector(rnorm(Na, mean=0, sd=sda)+mua)
epsB <-append(epsb,epsab)
epsA <-append(epsa,epsab)
ysB <-epsB+rnorm(NB,mean=0, sd=1)
ysA <-epsA+rnorm(N_A, mean=0, sd=1)
marcoA2<-data.frame(marcoA,ysA)
marcoB2<-data.frame(marcoB,ysB)
tyasmm<-sum(marcoA2["ysA"])+sum(marcoB2[1:Nb,"ysB"])

```