

Comparación de métodos de remuestreo en el contexto de regresión

Diego F. Fonseca. V. & Rodolfo A. Quintero *

Resumen. En el presente documento compararemos dos métodos de remuestreo en el contexto de un modelo de regresión múltiple, dichos métodos son el remuestreo por residuos y el remuestreo por pares y sus respectivas versiones no-parámétrica y suavizada.

Palabras clave: Elipse de mínimo volumen, estimación, distancia de Mahalanobis, remuestreo.

1 Introducción

Los métodos de remuestreo intuitivamente presentan un fenómeno que desde una visión no matemática puede llegar a ser desconcertante, la primera disyuntiva que viene a la mente es cómo a partir de una muestra de una población podemos inferir conclusiones acerca de esta última, tal inquietud en el campo de la estadística queda plenamente esclarecida dado que la muestra hace parte de la población, entonces algo debe poder decir acerca de ella, esto es justificado en una serie de teoremas que establecen la convergencia de distribuciones de probabilidad generadas por la muestra hacia la distribución que rige a la población. Una forma de inferir conclusiones acerca de la población a partir de una muestra es el remuestreo, que teóricamente está justificado por dichos teoremas de convergencia, en este documento compararemos dos métodos de remuestreo en el contexto de un modelo de regresión múltiple.

Para contextualizar lo que significa un modelo de regresión múltiple nos situamos en un ámbito general: dado un conjunto de variables aleatorias Y_1, \dots, Y_n donde cada Y_i depende de un conjunto de variables aleatorias X_{i1}, \dots, X_{in} . Una manera de mostrar esta dependencia es:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, \dots, n.$$

Donde β_j son constantes que en algunos casos nos indican el peso de la influencia de X_{ik} en Y_i , y ε_i introduce una medida de error que se puede cometer al estimar los β_k . Una mejor forma de escribir este modelo, habitualmente conocido como modelo de regresión múltiple, es de forma matricial:

$$Y = X\beta + \varepsilon \quad (1)$$

donde

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{pmatrix} \quad \text{y} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

El objetivo en este tipo de situaciones es estimar los β_k , dicho proceso es a lo que llamamos *regresión*. Para estimar los β_k existen dos métodos (no necesariamente los únicos) a considerar acá. Uno se concentra en los errores ε y asume características que este debe satisfacer, este método es conocido como el basado en el *remuestreo de residuos*. El otro método no asume condiciones sobre ε y se concentra en los Y_i y las respectivas filas de X , este método es conocido como el basado en el *remuestreo de pares*. A su vez, dentro de estas dos clases de remuestreo podemos considerar otros dos tipos de remuestreo generales (es decir, que se usan en otros contextos no solo en modelos de regresión múltiple) que son el *no-paramétrico clásico* y el *suavizado*. En lo subsiguiente, hablaremos un poco sobre cada tipo de remuestreo a considerar en este texto.

* Universidad de los Andes, Bogotá, Colombia.

Remuestreo no-paramétrico clásico y remuestreo suavizado

Dada una muestra X_1, \dots, X_n , el *remuestreo no-paramétrico clásico* consiste en elegir subconjuntos de n elementos de la muestra, es decir, X_{i_1}, \dots, X_{i_n} donde este último conjunto puede tener elementos repetidos.

Por otro lado, el *remuestreo suavizado* consiste en generar una muestra de tamaño n que tenga distribución asociada cuya función de densidad aproximada se obtiene de la fórmula

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \mathcal{E} \left(\frac{x - X_i}{h} \right)$$

donde \mathcal{E} es una función de densidad normal estándar. Los teoremas de convergencia garantizan que \hat{f} para n suficientemente grande es aproximadamente la distribución verdadera de la muestra en un sentido especial.

Ahora volviendo con nuestro modelo de regresión múltiple, tenemos las siguientes dos formas de estimar β , en realidad, lo que estimamos es un conjunto en donde β pertenece con una probabilidad alta.

Método 1: Remuestreo por residuos

1. Considere $\hat{\beta} = (X^t X)^{-1} Y$. Estas en realidad son las constantes que permiten expresar la proyección de Y en el espacio generado por las columnas de X como combinación lineal de dichas columnas.
2. Calculamos $\hat{Y} = X \hat{\beta}$, de acuerdo a 1 esta será la proyección de Y en el espacio generado por las columnas de X .
3. Calculamos $\mathbf{e} = Y - \hat{Y}$. Esta es una especie de aproximación de ε . Si todos los ε_i tienen varianza σ^2 constante, entonces decimos que estamos en el caso *Homocedástico*, en caso contrario decimos que estamos en el caso *Heterocedástico*. En este método se asume que se está en el caso homocedástico. Note que como $\mathbf{e} = Y - \hat{Y} = (I - X(X^t X)^{-1} X^t) Y$, si llamamos $H := X(X^t X)^{-1} X^t$, entonces, al hacer el cálculo de la varianza se tendrá $\text{var}(e_j) = \sigma^2(1 - h_j)$ donde h_j es el elemento j,j en la matriz H y e_j es la componente j -ésima de \mathbf{e} , lo ideal sería que dicha varianza fuera σ^2 para que se asemejara a ε , lo que obliga a ajustar e_j , dicho ajuste es a lo que llamamos *residuos ajustados*.
4. Los residuos ajustados son $r_i = \frac{e_i}{\sqrt{1-h_j}}$, llamamos $r = \{r_1, \dots, r_n\}$ el conjunto de dichos residuos, también lo podemos ver como un vector.
5. Elegimos uno de los métodos de remuestreo generales y aplicamos dicho remuestreo a r , es decir, obtenemos un nuevo conjunto r^* , que también podemos ver como un vector.
6. Construimos Y^* dada por

$$Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik} + r_i^*.$$

7. Calculamos $\beta^* = (I - X(X^t X)^{-1} X^t) Y^*$.
8. Elegimos un número natural B suficientemente grande y repetimos B veces los pasos 5, 6 y 7, eso significa que cada vez estamos obteniendo un vector β^* , entonces, al final se obtiene un conjunto de vectores $\beta^{*,1}, \dots, \beta^{*,B}$.
9. Generamos la elipse de menor volumen que contiene por lo menos el 95% de los $\beta^{*,1} - \hat{\beta}, \dots, \beta^{*,B} - \hat{\beta}$, llamamos dicha elipse $\mathcal{E}l$.

En esta última elipse el parámetro $\hat{\beta} - \beta$ está con una alta probabilidad (en realidad 0.95), es decir, β pertenece al conjunto $\hat{\beta} - \mathcal{E}l$ con alta probabilidad, entre más pequeña sea la elipse mejor es la aproximación.

Método 2: Remuestreo por pares

1. Considere $X_i = (1, X_{i1}, \dots, X_{ik})$, es decir, la fila i -ésima de X .
2. Considere los pares $(Y_1, X_1), \dots, (Y_n, X_n)$.
3. Elegimos uno de los métodos de remuestreo generales y aplicamos dicho remuestreo a $(Y_1, X_1), \dots, (Y_n, X_n)$, es decir, obtenemos un nuevo conjunto $(Y_1^*, X_1^*), \dots, (Y_n^*, X_n^*)$.
4. Construimos Y_* el vector cuyas componente i -ésima es Y_i^* y construimos X^* la matriz cuya fila i -ésima es X_i^* .
5. Calculamos $\beta^* = (I - X^*((X^*)^t X^*)^{-1}(X^*)^t)Y_*$.
6. Elegimos un numero natura B suficientemente grande y repetimos B veces los pasos 3, 4 y 5, eso significa que cada vez estamos obteniendo un vector β^* , entonces, al final se obtiene un conjunto de vectores $\beta^{*,1}, \dots, \beta^{*,B}$.
7. Generamos la elipse de menor volumen que contiene por lo menos el 95% de los $\beta^{*,1} - \hat{\beta}, \dots, \beta^{*,B} - \hat{\beta}$, llamamos dicha elipse $\mathcal{E}l$.

En esta ultima elipse queremos verificar que el parámetro $\hat{\beta} - \beta$ esta con una alta probabilidad (en realidad 0.95) dentro de la elipse, o en otras palabras, β pertenece al conjunto $\hat{\beta} - \mathcal{E}l$ con alta probabilidad. Entre mas pequeña sea la elipse mejor es la aproximación (si se satisface que lo contiene con alta probabilidad).

El objetivo de este documento es comparar los dos métodos anteriores en el contexto del modelo (1) pero en este caso haremos dicha comparación para un modelo de regresión específico con $n = 90$ y $k = 4$, es decir, nuestro modelo (propuesto en el archivo `bootstrap regresion.R`) es dado por una matriz X de tamaño 90×4 y $\beta = (\sqrt{2}, 1, -2, \pi)$. El propósito es estimar β_2 y β_3 sabiendo que sus valores reales son -2 y π respectivamente.

Dentro de cada uno de estos métodos, consideramos el remuestreo no-paramétrico clásico y remuestreo suavizado. Otros parámetros que vamos a variar son los siguientes: el vector de errores ε con la desviación estándar especificada en el archivo sugerido por el profesor “`bootstrap regresion.R`” y el vector de errores ε pero con cuatro veces la desviación estándar sugerida.

Ahora, para el caso del error ε consideraremos dos tipos: errores homocedásticos normales y errores heterodásticos no-normales. En el “`bootstrap regresion.R`” se especifica como expresar Y en términos de estos errores. Para cada tipo de error generaremos un conjunto de 500 datos, es decir, 500 copias de Y , luego generamos el conjunto de estimación para cada método de remuestreo (residuos y pares) que como se mencionó son elipses y determinará si el verdadero par (β_2, β_3) cae en el conjunto. Nuestro procedimiento estima la probabilidad que el verdadero par pertenezca a la elipse. Seguido determinaremos también el volumen de la elipse estimada, que nos da información acerca de la exactitud de la aproximación. El mejor escenario sería obtener elipses “pequeñas” con alta probabilidad de contener el parámetro verdadero.

2 Implementación de las simulaciones realizadas

En esta sección abordaremos los detalles de la implementación de las simulaciones realizadas, tales simulaciones fueron realizada en el software de análisis estadístico R, y describiremos parte del código desarrollado para cada uno de los problemas mencionados anteriormente y que ocupan el tema central de este artículo.

Es importante aclarar que para el correcto funcionamiento de las rutinas es necesario instalar en R y cargar los paquetes `rrcov` (para generar elipses de menor volumen), `ellipse` (para extraer puntos de las elipses) y `prodlm` (amplia el manejo de las matrices). Además, el comando `getDet` calcula el determinante de la matriz de covarianza de los β^* , esta determina la elipse de menor volumen y dicho volumen será ese determinante por una constante, de modo que en adelante cuando hacemos referencia al volumen en realidad estamos refiriéndonos a dicho determinante.

2.1 Generación de conjuntos de datos

Un dato en el sentido de este documento es un vector Y de 90 componentes de la forma definida anteriormente, el objetivo es generar 500 de estos vectores de manera optima, es decir, sin usar bucles como `for`, para ello usamos las primeras lineas del archivo `bootstrap regresion.R` y los comandos matriciales de `R` que nos permiten generar los 500 datos y donde cada dato es una columna de una matriz, dicha matriz será de tamaño 90×500 .

```
# primero generamos la matriz de diseño X, incluyendo los unos
n=90
unos=rep(1,n)
x1=seq(0,10,length=n)
x2=rep(c(-1,-.5,0,.5,1),length=n)
tmp=seq(0,5,length=10)
x3=rep(tmp,length=n)
X=cbind(unos,x1,x2,x3)
beta0=sqrt(2);beta1=1;beta2=-2;beta3=pi
beta=c(beta0,beta1,beta2,beta3)
# Matriz para estimacion de parametros, de respuesta y de residuos
A=solve(t(X)%*%X) # la matriz A se usará
k=1 # ó =2, en ese caso se usara para multiplicar por cuatro la varianza del
    error, lo que significa multiplicar por 2 el error
```

Estas ultimas lineas de código siempre estarán presentes en adelante.

Datos con errores homocedásticos normales La matriz obtenida de tamaño 90×500 la notamos por `YMhm` (en adelante `M` se refiere a que es una matriz y `hm` a que es el caso homocedástico), la rutina para generar dicha matriz en el caso de errores homocedásticos normales es el siguiente

```
YMhm<-matrix(rep(X%*%beta,each=500),ncol=500,byrow=TRUE)+k*0.5*
    matrix(rnorm(n*500),ncol=500,byrow=TRUE)
```

Datos con errores heterocedásticos no-normales Al igual que el caso anterior, usamos los comandos matriciales de `R`, los cuales permiten generar los 500 conjuntos de datos donde cada conjunto es una columna de una matriz, dicha matriz es de tamaño 90×500 y la notamos por `YMht` (en adelante `M` se refiere a que es una matriz y `ht` a que es el caso heterocedásticos), la rutina para generar dicha matriz en el caso de errores heterocedásticos no-normales es el siguiente

```
# Primero se define la distribución que regirá a los errores
rdexp=function(n){
y=rexp(n)
u=runif(n)
y[u>0.5]=-y[u>0.5]
return(y)
} # fin de rdexp

# Ahora si se define la matriz con el conjunto de datos con los errores descritos
YMht<-matrix(rep(X%*%beta,each=500),ncol=500,byrow=TRUE)
```

```
YMht<-matrix(rep(X%%beta,each=500),ncol=500,byrow=TRUE)+k*0.025*
  YMht*matrix(rdexp(n*500),ncol=500,byrow=TRUE)
```

2.2 Remuestreo por residuos caso no-paramétrico

Debemos generar vectores $\hat{\beta}$, \hat{Y} , \mathbf{e} (que llamaremos residuos) y luego vectores \mathbf{r} (que serán los residuos ajustados) expresados en el método 1 (ver introducción) por cada conjunto de datos, es decir, generamos 500 de estos vectores. La rutina para generar esos vectores caso por caso es la siguiente:

Caso datos con errores homocedásticos normales En este caso notamos dichas matrices como `betahatMhm`, `YhatMhm`, `resiMhm` (este es análogo de \mathbf{e}) y `resiadjMhm` (este es análogo de \mathbf{r}) y se generan como sigue:

```
betahatMhm<- (A%%t(X))%%YMhm
YhatMhm<-X%%betahatMhm
resiMhm<-YMhm-YhatMhm
# Generamos la matriz H para calcular los residuos ajustados (esta sirve tambien
  para el caso Heterocedástico)
H<- diag(nrow=n)-(X%%A)%%t(X)
# Extraemos la diagonal de H (esta misma la usaremos en el caso heterocedástico)
h=diag(H)
# Ahora si producimos la matriz de residuos ajustados, cada columna es el vector
  de residuos ajustados correspondiente a la columna en YMhm
resiadjMhm<- resiMhm/sqrt(h)
```

En este caso, el remuestreo no-paramétrico clásico es generado con la siguiente rutina

```
B=20000
refer=matrix(0,B,4)
AreasElip=rep(0,500) #guardará las areas de las elipses para cada conjunto
  de datos
CountVerd=rep(0,500) #Será 1 en la i-esima compnente si el la elipse formada
  en el remuestreo del i-esimo Y de nuestro conjunto de datos contiene a los
  parametros verdaderos.

# Realizamos un remuestreo clásico no-paramétrico por cada columna de YMhm
for(j in 1:500){
  for(b in 1:B){
    Ystar=X%%betahatMhm[,j]+sample(resiadjMhm[,j],n,replace=T)
    betastar=(A%%t(X))%%Ystar
    refer[b,]=betastar-betahatMhm[,j]
  }# fin del for interno indexado por b.
# Seguio generamos la elipse de menor volumen que contienen los puntos obtenidos
  por el remuestreo
c1=CovMve(refer[,3:4],alpha=.95)
# Extraemos el centro y la matriz de covarianza que determinan la elipse
centro=getCenter(c1)
```

```

MatCov=getCov(c1)
# Calculamos el area de esa elipse.
AreasElip[j]=getDet(c1)
#Generamos 100 puntos de la elipse
listaElipse=ellipse(MatCov,centre=centro)
#Se calcula la matriz inversa de MatCov
invMatCov=solve(MatCov)
#Se calcula la distancia de Mahalanobis de un punto de la elipse (el primero en la
  lista), este se usa para determinar el radio del circulo determinado por la
  elipse en el cambio de cordenadas, este valor permite inferir si los
  parámetros verdaderos estan en la elipse o no.
RadioMahala= t(listaElipse[1,] -centro)%*%invMatCov%*(listaElipse
  [1,] -centro)
#Calculamos la distancia de Mahalanobis del parametro verdadero
distMahalaVerPar=t(beta[3:4]-betahatMhm[3:4,j] -centro)%*%invMatCov%
  %*(beta[3:4]-betahatMhm[3:4,j]-centro)
#Determinamos si el parametro verdadero esta en la elipse (se asigna el valor 1) o
  no (se asigna valor cero)
if( distMahalaVerPar <= RadioMahala){
  CountVerd[j]=1;
} else {
  CountVerd[j]=0;
}
}#fin del for

```

Los datos que nos interesan son los consignados en los vectores CountVerd y AreasElip.

Caso datos con errores heterocedásticos no-normales En este caso, de manera similar al caso anterior, notamos dichas matrices como `betahatMht`, `YhatMht`, `resiMht` (este es análogo de `e`) y `resiadjMht` (este es análogo de `r`) y se generan como sigue:

```

betahatMht<- (A%*%t(X))%*%YMht
YhatMht<-X%*%betahatMht
resiMht<-YMht-YhatMht
#Generamos la mnatriz H para calcular los residuos ajustados (esta sirbe para el
  caso Homocedástico)
H<- diag(nrow=n)-(X%*%A)%*%t(X)
#Extraemos la diagonal de H (la misma del caso homocedástico)
h=diag(H)
# Ahora si producimos la matriz de residuos ajustados, cada columnda es el vector
  de residuos ajustados correspondiente a la columna en YMht
resiadjMht<- resiMht/sqrt(h)

```

En este caso, el remuestreo no-parámétrico clásico es idéntico al caso homocedástico, es decir, se replican las líneas de dicho remuestreo cambiando `betahatMhm`, `YhatMhm`, `resiMhm` y `resiadjMhm` por `betahatMht`, `YhatMht`, `resiMht` y `resiadjMht` respectivamente, de modo que vemos poco productivo reescribir dichas líneas. De nuevo, los datos que nos interesan son los consignados en los vectores CountVerd y AreasElip.

2.3 Remuestreo por residuos caso suavizado

Como en los casos anteriores enfocaremos nuestro análisis caso por caso.

Caso datos con errores homocedásticos normales En este caso, la única diferencia con las líneas de código del caso de remuestreo por residuos caso no-parámétrico clásico con *datos con errores homocedásticos normales* esta en las líneas que determinan el remuestreo, luego, basta con exhibir dichas líneas:

```
B=20000
refer=matrix(0,B,4)
AreasElip=rep(0,500) #guardará las areas de las elipses para cada conjunto
de datos
CountVerd=rep(0,500) #Será 1 en la i-esima compnente si el la elipse formada
en el remuestreo del i-esimo Y de nuestro conjunto de datos contiene a los
parametros verdaderos.

# Realizamos un remuestreo suavizado por cada columna de YMhm
for(j in 1:500){
  for(b in 1:B){
    hopt=(4/(3*n))^(0.2)
    h=hopt*sd(resiadjMhm[,j])
    ind=1:n
    indstar=sample(ind,n,replace=TRUE)
    rstar=rnorm(n,mean=resiadjMhm[indstar,j],sd=0.75*h)
    Ystar=X%*%betahatMht[,j]+rstar
    betastar=(A%*%t(X))%*%Ystar
    refer[b,]=betastar-betahatMhm[,j]
  }# fin del for interno indexado por b.
# Sequido generamos la elipse de menor volumen que contienen los puntos obtenidos
por el remuestreo
c1=CovMve(refer[,3:4],alpha=.95)
# Extraemos el centro y la matriz de covarianza que determinan la elipse
centro=getCenter(c1)
MatCov=getCov(c1)
# Calculamos el area de esa elipse.
AreasElip[j]=getDet(c1)
#Generamos 100 puntos de la elipse
listaElipse=ellipse(MatCov,centre=centro)
#Se calcula la matriz inversa de MatCov
invMatCov=solve(MatCov)
#Se calcula la distancia de Mahalanobis de un punto de la elipse (el primero en la
lista), este se usa para determinar el radio del circulo determinado por la
elipse en el cambio de cordenadas, este valor permite inferir si los
parámetros verdaderos estan en la elipse o no.
RadioMahala= t(listaElipse[1,] -centro)%*%invMatCov%*(listaElipse
[1,] -centro)
#Calculamos la distancia de Mahalanobis del parametro verdadero
distMahalaVerPar=t(beta[3:4]-betahatMhm[3:4,j] -centro)%*%invMatCov%
*(beta[3:4]-betahatMhm[3:4,j]-centro)
```

```
#Determinamos si el parametro verdadero esta en la elipse (se asigna el valor 1) o
  no (se asigna valor cero)
if( distMahalaVerPar <= RadioMahala){
  CountVerd[j]=1;
} else {
  CountVerd[j]=0;
}
}#fin del for
```

El h optimo tomado fue $h_{op} = \left(\frac{4}{3n}\right)^{0.2}$, esta elección fue propuesta en uno de los archivos enviados por el profesor para el caso uni-variado. Como siempre, los datos que nos interesan son los consignados en los vectores `CountVerd` y `AreasElip`.

Caso datos con errores heterocedásticos no-normales En este caso, la única diferencia con las líneas de código del caso de remuestreo por residuos caso no-parámétrico clásico con *datos con errores heterocedásticos no-normales* está en las líneas que determinan el remuestreo, en ese sentido, basta con exhibir dichas líneas, pero, aquí el remuestreo suavizado es idéntico al caso homocedástico (el caso anterior), es decir, se replican las líneas de dicho remuestreo cambiando `betahatMhm`, `YhatMhm`, `resiMhm` y `resiadjMhm` por `betahatMht`, `YhatMht`, `resiMht` y `resiadjMht` respectivamente. Omitimos de nuevo esas líneas y recordamos que los datos que nos interesan son los consignados en los vectores `CountVerd` y `AreasElip`.

2.4 Remuestreo por pares caso no-paramétrico

La idea es replicar todos los pasos expresados en el método 2 (ver introducción) por cada conjunto de datos, es decir, cada conjunto de los 500 conjuntos de datos replicaremos la rutina del Método 2, la idea es generar $\beta^{*,1}, \dots, \beta^{*,B}$ por cada columna de conjunto de datos, este proceso no fue posible vectorizarlo ya que implicaba el almacenamiento de matrices de tamaño considerable, en ese sentido, se gana mas usando comandos como `for` que almacenado tanta información, la rutina para abordar esta situación la desarrollamos caso por caso de la siguiente manera:

Caso datos con errores homocedásticos normales Usaremos la matriz `betahatMhm` que fue definida en este mismo caso pero para remuestreo por residuos, teniendo en cuenta esto se tiene lo siguiente:

```
indi=1:n
B=20000
referpa=matrix(0,B,4)
AreasElipPa=rep(0,500) #guardará las areas de las elipses para cada conjunto
de datos
CountVerdPa=rep(0,500) #Será 1 en la i-esima compnente si el la elipse
formada en el remuestreo del i-esimo Y de nuestro conjunto de datos contiene a
los parametros verdaderos

#El remuestreo por pares es lo que sigue:
for(j in 1:500){
  for(b in 1:B){
    indstar=sample(indi,n,replace=T)
    Xstar=X[indstar,]
```



```

    Ystar=YMhm[indstar,j]
    Astar=solve(t(Xstar)%*%Xstar)
    betastar=Astar%*%t(Xstar)%*%Ystar
    referpa[b,]=betastar-betahatMhm[,j]
} # fin del remuestreo indexado por b.

# Sequido generamos la elipse de menor volumen que contienen los puntos obtenidos
# por el remuestreo
c1Pa=CovMve(referpa[,3:4],alpha=.95)
# Extraemos el centro y la matriz de covarianza que determinan la elipse
centropa=getCenter(c1Pa)
MatCovpa=getCov(c1Pa)
# Calculamos el area de esa elipse.
AreasElipPa[j]=getDet(c1Pa)
#Generamos 100 puntos de la elipse
listaElipsePa=ellipse(MatCovpa,centre=centropa)
#Se calcula la matriz inversa de MatCovpa
invMatCovpa=solve(MatCovpa)
#Se calcula la distancia de Mahalanobis de un punto de la elipse (el primero en la
# lista), este se usa para determinar el radio del circulo determinado por la
# elipse en el cambio de cordenadas, este valor permite inferir si los
# parámetros verdaderos estan en la elipse o no.
RadioMahalaPa= t(listaElipsePa[1,] -centropa)%*%invMatCovpa%*%(
    listaElipsePa[1,] -centropa)
#Calculamos la distancia de Mahalanobis del parametro verdadero
distMahalaVerParPa=t(beta[3:4]-betahatMhm[3:4,j] -centropa)%*%
    invMatCovpa%*%(beta[3:4]-betahatMhm[3:4,j]-centropa)
#Determinamos si el parametro verdadero esta en la elipse (se asigna el valor 1) o
# no (se asigna valor cero)
if( distMahalaVerParPa <= RadioMahalaPa){
    CountVerdPa[j]=1;
} else {
    CountVerdPa[j]=0;
}
} # Fin del for

```

De nuevo, los datos que centran nuestra atención son los consignados en los vectores `CountVerdPa` y `AreasElipPa`.

Caso datos con errores heterocedásticos no-normales En este caso, usaremos la matriz `betahatMht` que fue definida en este mismo caso pero para remuestreo por residuos, teniendo en cuenta esto se tiene que la rutina para estimar el porcentaje de las 500 elipses que contienen el parámetro verdadero y el promedio de las areas de las 500 elipses es idéntica a la rutina del caso anterior, es suficiente con reemplazar esas lineas `betahatMhm` por `betahatMht`. De nuevo, los datos que nos interesan son los consignados en los vectores `CountVerdPa` y `AreasElipPa`.

2.5 Remuestreo por pares caso suavizado

Al igual que en la subsección anterior la idea es replicar todos los pasos expresados en el Método 2 (ver introducción) por cada conjunto de datos, este proceso no fue posible vectorizarlo, se intentó

pero las funciones de R que lo permitían hacían la rutina demasiado lento. Otro inconveniente es que dicho intento implicaba el almacenamiento de matrices de tamaño considerable, en ese sentido, se gana más usando comandos como `for` que almacenando tanta información en matrices. El proceso para abordar esta situación es similar al de la subsección anterior, sólo difiere en las líneas del remuestreo, y las desarrollamos caso por caso de la siguiente manera:

Caso datos con errores homocedásticos normales Usaremos la matriz `betahatMhm` que fue definida en este mismo caso pero para remuestreo por residuos, teniendo en cuenta esto se tiene lo siguiente:

```
indi=1:n
B=20000
referpa=matrix(0,B,4)
AreasElipPa=rep(0,500) #guardará las areas de las elipses para cada conjunto
de datos
CountVerdPa=rep(0,500) #Será 1 en la i-esima compnente si el la elipse
formada en el remuestreo del i-esimo Y de nuestro conjunto de datos contiene a
los parametros verdaderos

#El remuestreo por pares es lo que sigue:
for(j in 1:500){
  for(b in 1:B){
    #Creamos una matriz cuyas columnas son YMhm y las columnas que no son 1 de
    X
    W=cbind(YMhm[,j],X[,2:4])
    #Este es el optimo de Foster pero modificado ya que sin modificar no
    funciono
    hopt=0.4*(4/(n*(6)))^(1/8)
    # generamos un vector de desviaciones estandar
    std=apply(W,2,sd)
    #Generamos el vetor de los h para generear la densidad por kernel
    h=hopt*std
    indstar=sample(indi,n,replace=T)
    #aqui se genera el remuestreo
    Wstar=matrix(0,n,4)
    for(k in 1:4){
      Wstar[,k]=rnorm(n,mean=W[indstar,k],sd=h[k])
    } #Fin del for indexado por k
    Xstar=cbind(rep(1,n),Wstar[,2:4])
    Ystar=Wstar[,1]
    Astar=solve(t(Xstar)%*%Xstar)
    betastar=Astar%*%t(Xstar)%*%Ystar
    referpa[b,]=betastar-betahatMhm[,j]
  } # fin del remuestreo indexado por b.

# Sequido generamos la elipse de menor volumen que contienen los puntos obtenidos
por el remuestreo
c1Pa=CovMve(referpa[,3:4],alpha=.95)
# Extraemos el centro y la matriz de covarianza que determinan la elipse
centropa=getCenter(c1Pa)
MatCovpa=getCov(c1Pa)
```

```

# Calculamos el area de esa elipse.
AreasElipPa[j]=getDet(c1Pa)
#Generamos 100 puntos de la elipse
listaElipsePa=ellipse(MatCovpa,centre=centropa)
#Se calcula la matriz inversa de MatCovpa
invMatCovpa=solve(MatCovpa)
#Se calcula la distancia de Mahalanobis de un punto de la elipse (el primero en la
  lista), este se usa para determinar el radio del circulo determinado por la
  elipse en el cambio de cordenadas, este valor permite inferir si los
  parámetros verdaderos estan en la elipse o no.
RadioMahalaPa= t(listaElipsePa[1,] -centropa)%*%invMatCovpa%*%(
  listaElipsePa[1,] -centropa)
#Calculamos la distancia de Mahalanobis del parametro verdadero
distMahalaVerParPa=t(beta[3:4]-betahatMhm[3:4,j] -centropa)%*%
  invMatCovpa%*%(beta[3:4]-betahatMhm[3:4,j]-centropa)
#Determinamos si el parametro verdadero esta en la elipse (se asigna el valor 1) o
  no (se asigna valor cero)
if( distMahalaVerParPa <= RadioMahalaPa){
  CountVerdPa[j]=1;
} else {
  CountVerdPa[j]=0;
}
} # Fin del for

```

Nota. Es importante resaltar que para este caso usamos el $h_{opt} = 0.4 \left(\frac{4}{6n}\right)^{1/8}$, y para el caso de 4 veces la varianza usamos $h_{opt} = 0.1 \left(\frac{4}{6n}\right)^{1/8}$ con el fin de obtener las elipses deseadas. Como antes, los datos que centran nuestra atención son los consignados en los vectores `CountVerdPa` y `AreasElipPa`.

Caso datos con errores heterocedásticos no-normales En este caso, usaremos la matriz `betahatMht` que fue definida en este mismo caso pero para remuestreo por residuos, y se tiene que el código para estimar el porcentaje de las 500 elipses que contienen el parámetro verdadero y el promedio de las áreas de las 500 elipses es idéntica al caso anterior. Por tanto, reemplazamos en esas líneas `betahatMhm` por `betahatMht`, de modo que vemos poco productivo reescribir dichas líneas.

Nota. El h_{opt} empleado en este caso es $h_{opt} = 0.1 \left(\frac{4}{6n}\right)^{1/8}$ y en el caso de 4 veces la varianza fue $h_{opt} = 0.01 \left(\frac{4}{6n}\right)^{1/8}$.

3 Discusión de los resultados

Los resultados presentados en esta sección son referentes las siguientes situaciones: datos con errores homocedásticos normales y datos con errores heterodásticos no-normales, en ambos casos aplicamos dos métodos de remuestreo (por residuos y por pares), y dentro de cada caso de remuestreo se aplica un remuestreo interno (para generar los β^*) que en nuestro caso es el no-parámétrico clásico o el suavizado. Además, también analizaremos el modelo originalmente planteado pero con cuatro veces su varianza.

Los siguientes resultados fueron obtenidos generando un conjunto de 500 datos (es decir, 500 del vector Y en el modelo de regresión (1) con las especificaciones dadas al final de la introducción)

se calculo para cada cada uno de estos datos la elipse generada (elipse de menor volumen) por los puntos en $\mathcal{D} - (\hat{\beta}_2, \hat{\beta}_3)$ donde

$$\mathcal{D} = \left\{ (\beta_2^{*,b}, \beta_3^{*,b}) \mid b = 1, \dots, B \right\}.$$

Los resultados que se expondrán son el porcentaje de veces que las elipses generadas contienen el parámetro verdadero; otro aspecto es el tamaño de estas 500 elipses, este dato lo condensamos en el area promedio de las 500 elipses y por último se presenta una gráfica de dispersión de $\mathcal{D} - (\hat{\beta}_2, \hat{\beta}_3)$, el objetivo de esta última gráfica es sólo ilustrar la nube de dichos puntos y poder comparar con los generados en el caso heterocedástico. Todos estos resultados son presentados caso por caso.

3.1 Caso homocedástico con remuestreo por residuos

Dado que el remuestreo interno puede ser no-paramétrico clásico o remuestreo suavizado, entonces tenemos dos casos, los resultados son consignados en la Tabla 1 y las figuras 3.1 y 3.2. La linea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

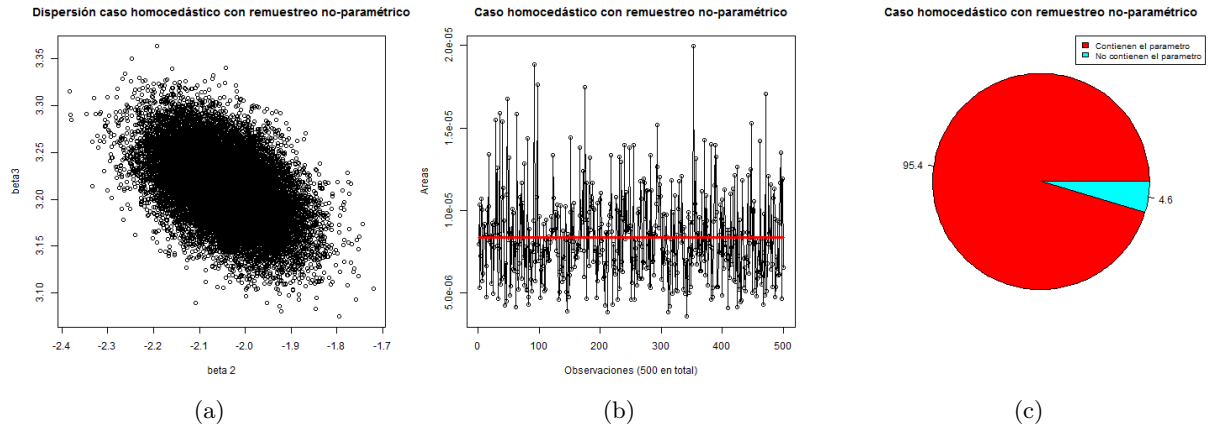


Fig. 3.1. Caso homocedástico con remuestreo por residuos y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 1: Caso homocedástico con remuestreo por residuos

	No-paramétrico clásico	Suavizado
Área promedio de elipses	8.2922×10^{-6}	1.012508×10^{-5}
Porcentaje de contención del parámetro verdadero	95.4%	97.4%

Caso homocedástico con remuestreo por residuos pero 4 veces la varianza Bajo las mismas convenciones anteriores se tiene para este caso en el que se modificó la varianza anterior multiplicándola por 4 (o lo que es lo mismo, multiplicar el error por 2) se obtienen los resultados que son consignados en la Tabla 2 y las figuras 3.3 y 3.4. La linea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

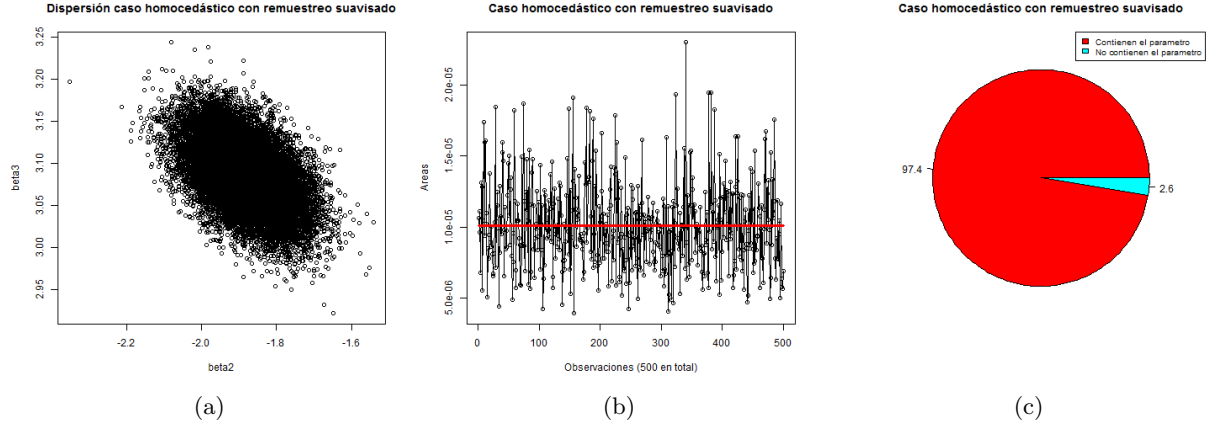


Fig. 3.2. Para el Caso homocedástico con remuestreo por residuos y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

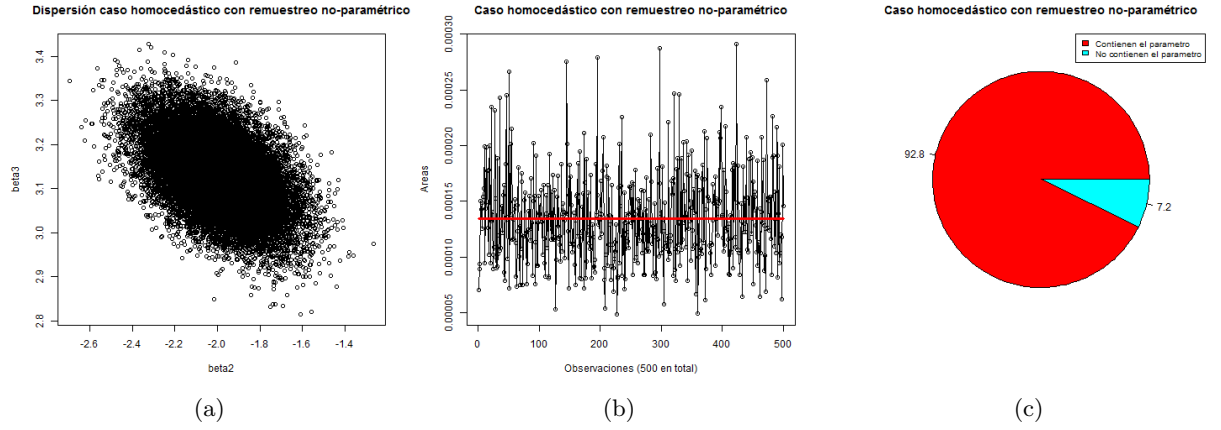


Fig. 3.3. Para el Caso homocedástico con 4 veces la varianza con remuestreo por residuos y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 2: Caso homocedástico con remuestreo por residuos con 4 veces la varianza

	No-paramétrico clásico	Suavizado
Área promedio de elipses	0.0001346132	0.0001607094
Porcentaje de contención del parámetro verdadero	92.8%	95.2%

3.2 Caso homocedástico con remuestreo por pares

Al igual que antes, dado que el remuestreo interno puede ser no-paramétrico clásico o remuestreo suavizado, entonces tenemos dos casos, los resultados son consignados en la Tabla 3 y las figuras 3.5 y 3.6. La línea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

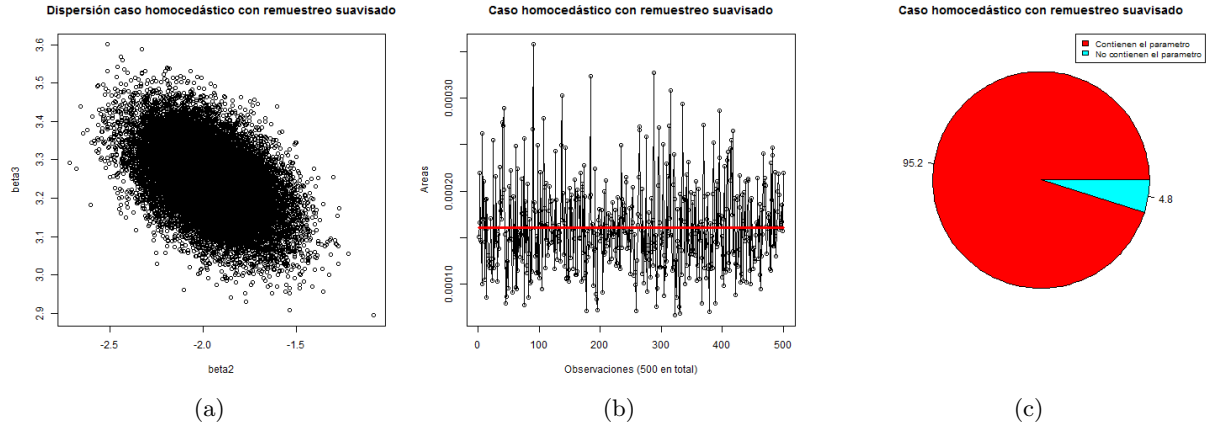


Fig. 3.4. Para el Caso homocedástico con 4 veces la varianza con remuestreo por residuos y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

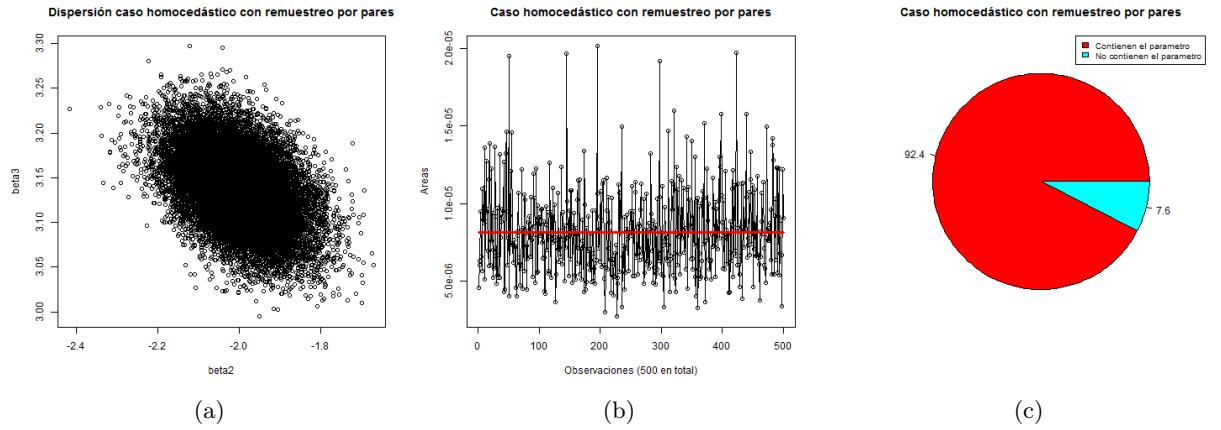


Fig. 3.5. Para el Caso homocedástico con remuestreo por pares y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 3: Caso homocedástico con remuestreo por pares

	No-paramétrico clásico	Suavizado
Área promedio de elipses	8.151423×10^{-6}	0.00131834
Porcentaje de contención del parámetro verdadero	92.4%	99.4%

Caso homocedástico con remuestreo por pares pero 4 veces la varianza Bajo las mismas convenciones anteriores se tiene para este caso en el que se modificó la varianza anterior multiplicándola por 4 (o lo que es lo mismo, multiplicar el error por 2) se obtienen los resultados mostrados en la Tabla 4 y las figuras 3.7 y 3.8. La línea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

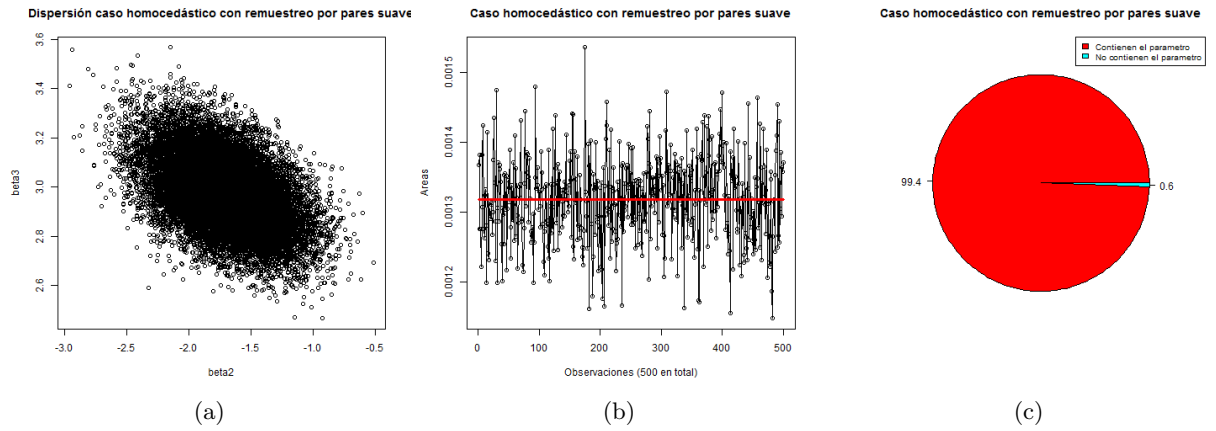


Fig. 3.6. Para el Caso homocedástico con remuestreo por pares y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

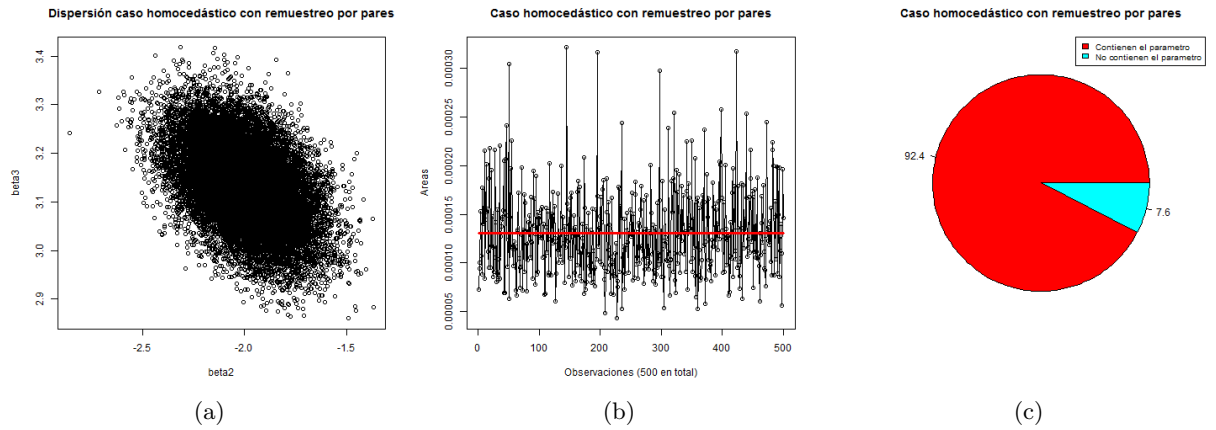


Fig. 3.7. Para el Caso homocedástico con 4 veces la varianza con remuestreo por pares y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 4: Caso homocedástico con remuestreo por pares con 4 veces la varianza

	No-paramétrico clásico	Suavizado
Área promedio de elipses	0.000130347	0.0001836368
Porcentaje de contención del parámetro verdadero	92.4%	96%

3.3 Caso heterocedástico con remuestreo por residuos

De nuevo, el remuestreo interno puede ser no-paramétrico clásico o remuestreo suavizado, tenemos dos casos a considerar y los resultados se exhiben en la Tabla 5 y las figuras 3.9 y 3.10. La línea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

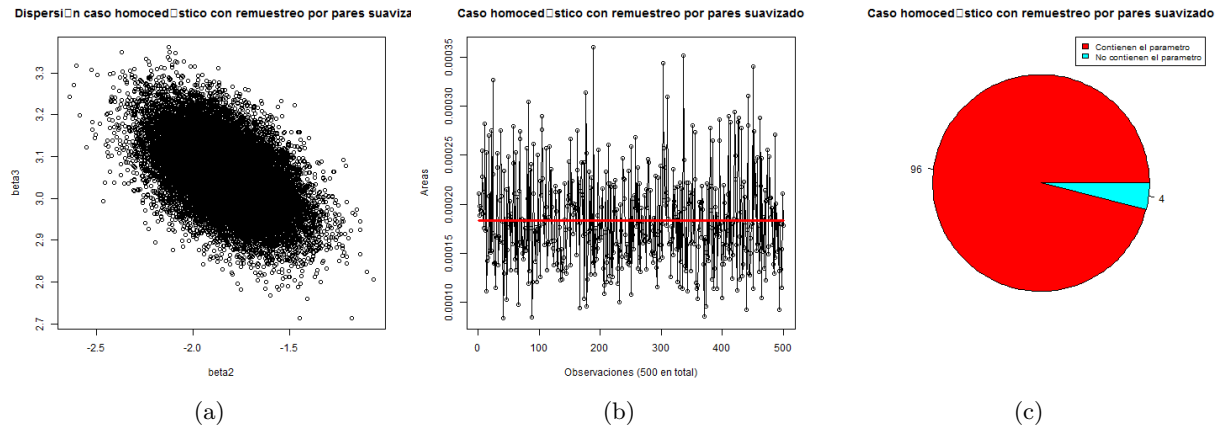


Fig. 3.8. Para el Caso homocedástico con 4 veces la varianza con remuestreo por pares y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

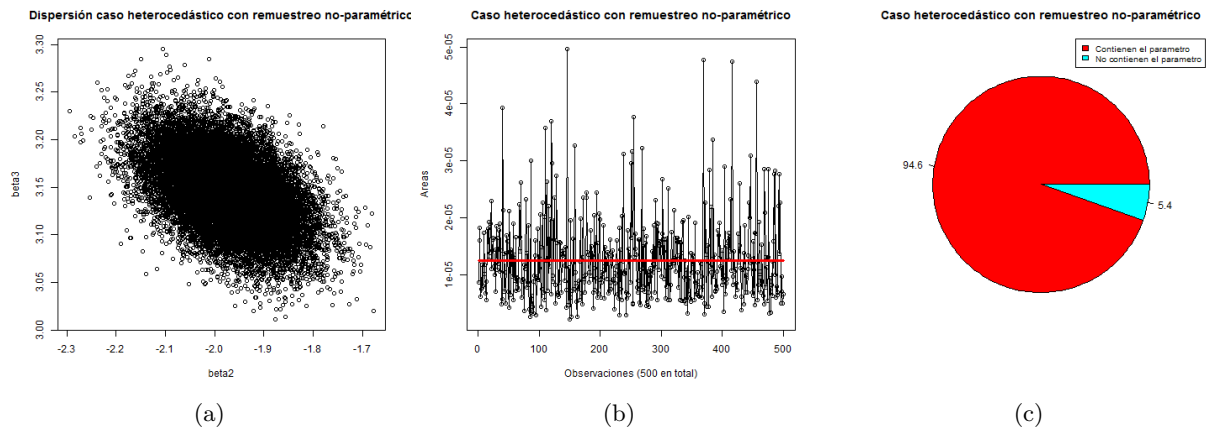


Fig. 3.9. Para el Caso heterocedástico con remuestreo por residuos y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 5: Caso heterocedástico con remuestreo por residuos

	No-paramétrico clásico	Suavizado
Área promedio de elipses	1.241702×10^{-5}	1.439052×10^{-5}
Porcentaje de contenencia del parámetro verdadero	94.6%	96.2%

Caso heterocedástico con remuestreo por residuos pero 4 veces la varianza Bajo las mismas convenciones anteriores se tiene para este caso en el que se modificó la varianza anterior multiplicándola por 4 (o lo que es lo mismo, multiplicar el error por 2) se obtienen los resultados que son consignados en la Tabla 6 y las figuras 3.11 y 3.12. La línea roja en la Tabla indica el área promedio de las elipses.

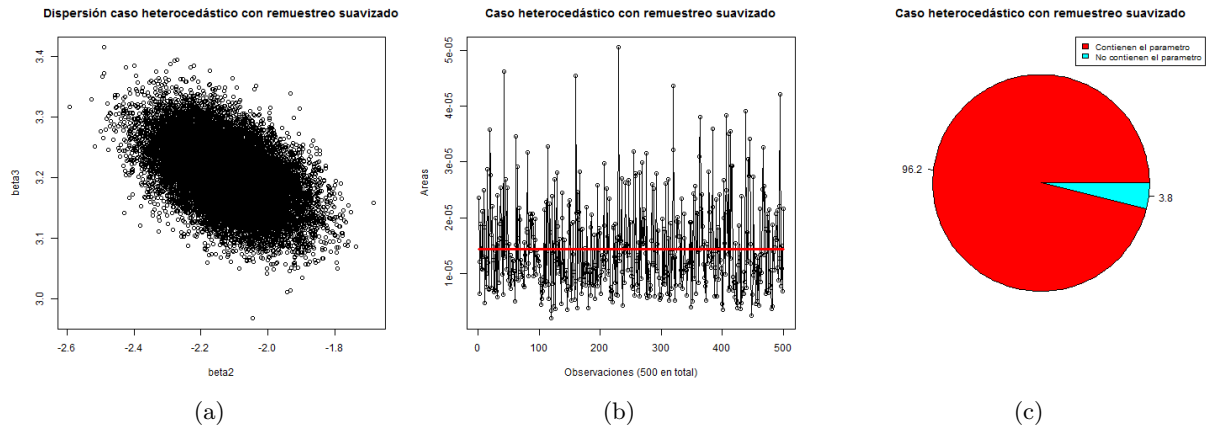


Fig. 3.10. Para el Caso heterocedástico con remuestreo por residuos y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

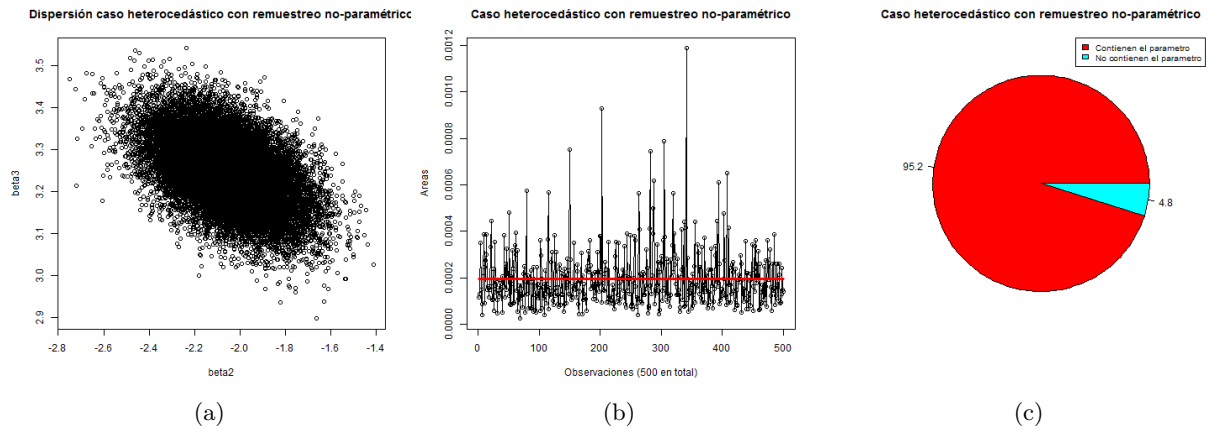


Fig. 3.11. Para el Caso heterocedástico con remuestreo por residuos y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 6: Caso heterocedástico con remuestreo por residuos con 4 veces la varianza

	No-paramétrico clásico	Suavizado
Área promedio de elipses	0.0001952339	0.0002342539
Porcentaje de contención del parámetro verdadero	95.2%	94.8%

3.4 Caso heterocedástico con remuestreo por pares

Como en los demás casos, los resultados obtenidos son consignados en la Tabla 7 y las figuras 3.13 y 3.14. La línea roja en la gráfica (b) de ambas figuras indica el área promedio de las elipses.

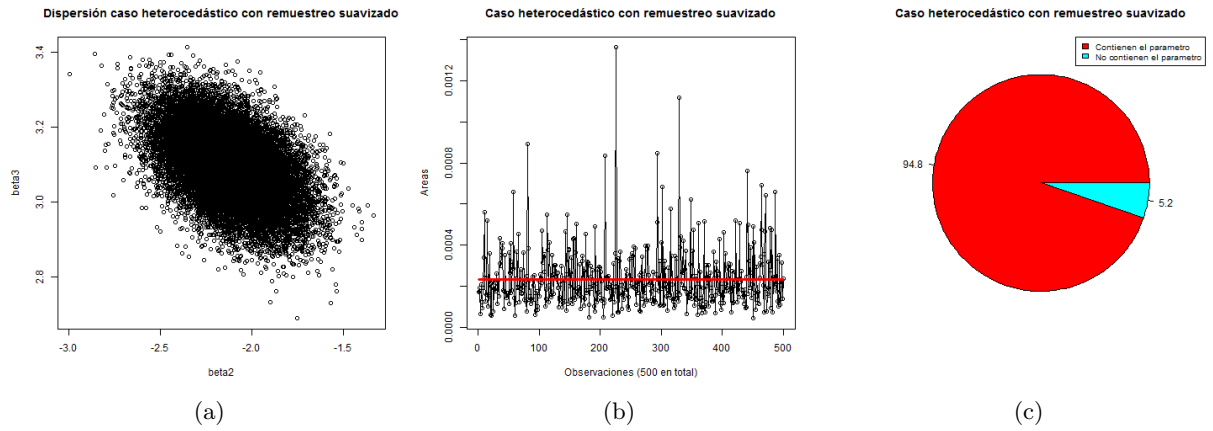


Fig. 3.12. Para el Caso heterocedástico con 4 veces la varianza con remuestreo por residuos y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

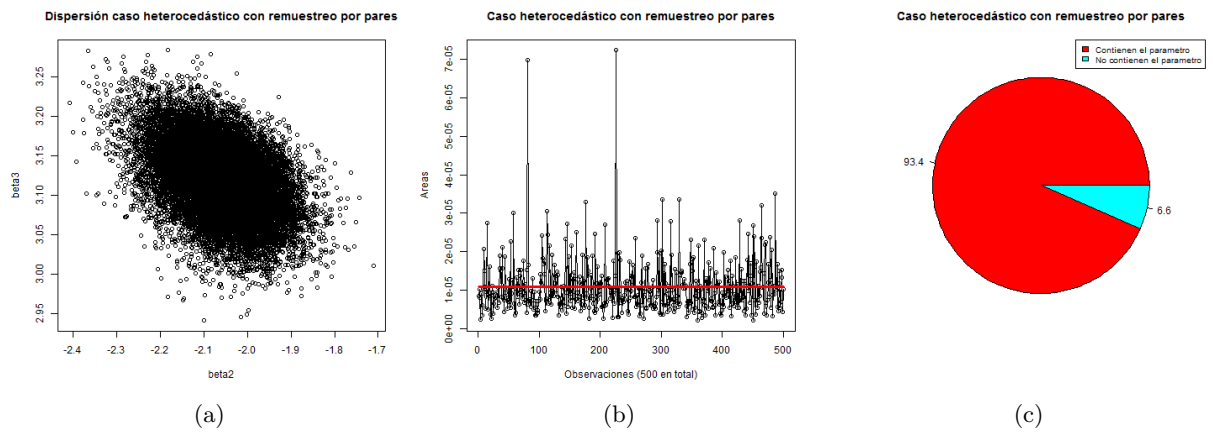


Fig. 3.13. Para el Caso heterocedástico con remuestreo por pares y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 7: Caso heterocedástico con remuestreo por pares

	No-paramétrico clásico	Suavizado
Área promedio de elipses	1.07725×10^{-5}	3.156157×10^{-5}
Porcentaje de contención del parámetro verdadero	93.4%	99.6%

Caso heterocedástico con remuestreo por pares pero 4 veces la varianza Bajo las mismas convenciones anteriores se tiene para este caso en el que se modificó la varianza anterior multiplicándola por 4 (o lo que es lo mismo, multiplicar el error por 2) se obtienen los resultados que son consignados en la Tabla 7 y las figuras 3.15 y 3.16. La línea roja en la Tabla indica el área promedio de las elipses.

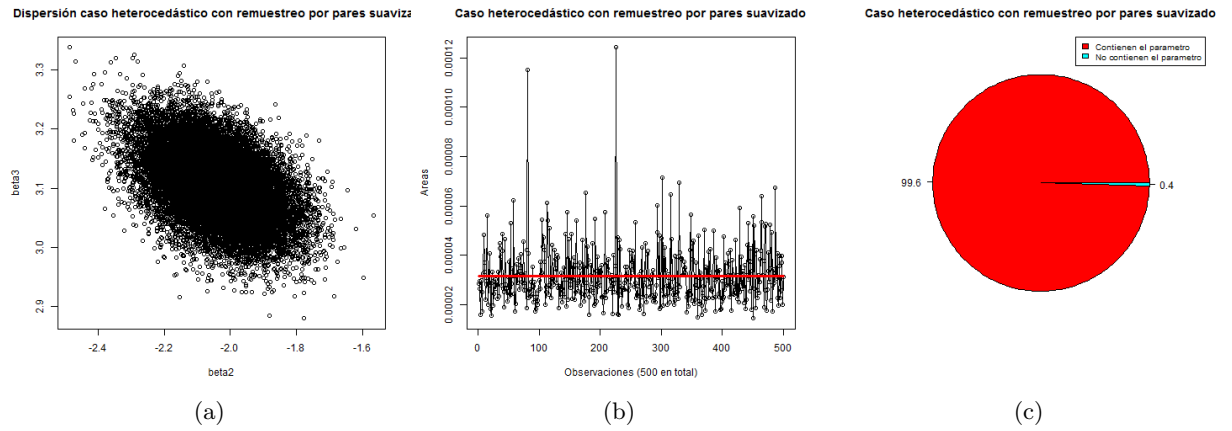


Fig. 3.14. Para el Caso heterocedástico con remuestreo por pares y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

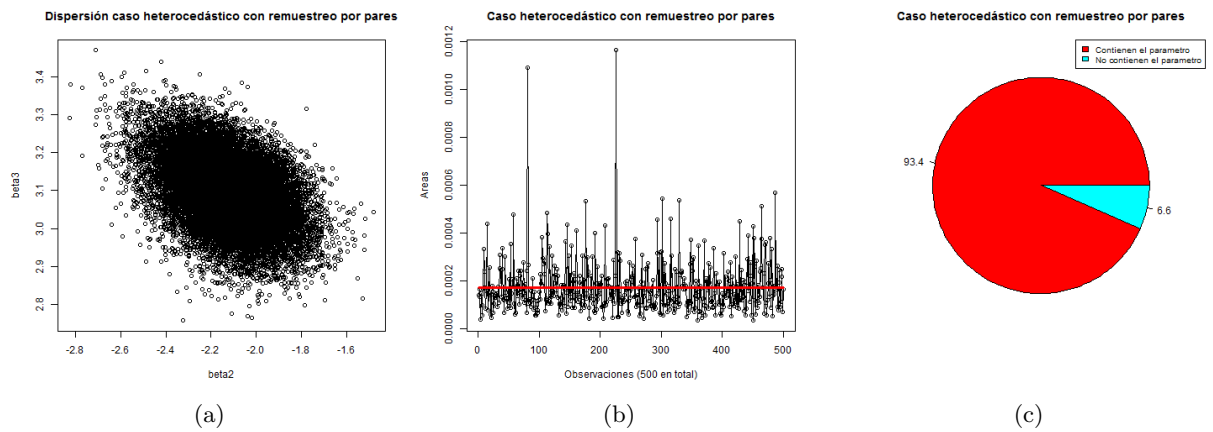


Fig. 3.15. Para el Caso heterocedástico con remuestreo por pares y remuestreo interno no-paramétrico (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

Tabla 8: Caso heterocedástico con remuestreo por pares con 4 veces la varianza

	No-paramétrico clásico	Suavizado
Área promedio de elipses	0.0001724547	0.0002386653
Porcentaje de contenencia del parámetro verdadero	93.4%	96.4%

3.5 Resumen de los resultados

Los resultados expuestos en las gráficas anteriores se resumen en las siguientes diagramas (Figuras 3.17 y 3.18) que comparan el porcentaje de elipses que contienen el parámetro verdadero en el proceso de remuestreo y el área promedio de dichas elipses, específicamente, en **azul** se representa el área promedio y en rojo el porcentaje de contenencia o cobertura, la barra derecha corresponde

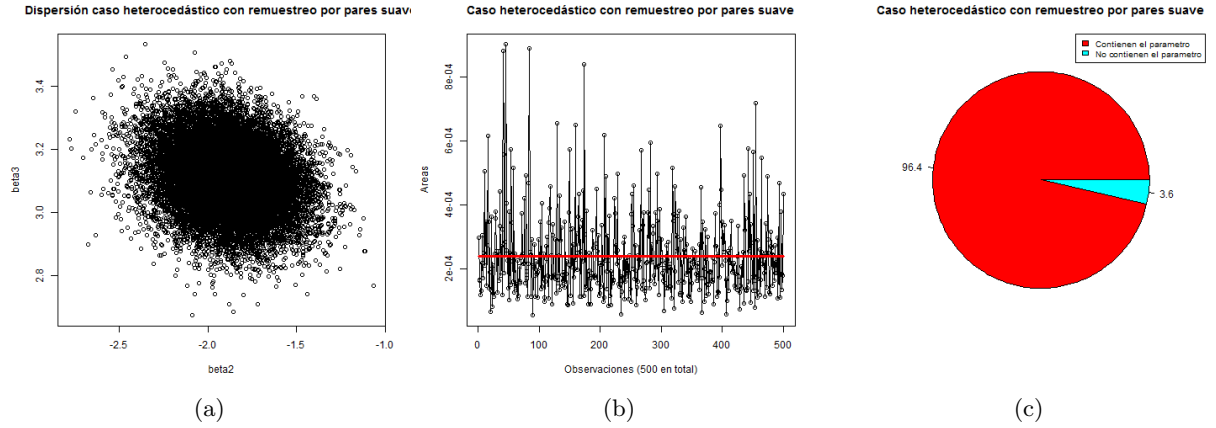


Fig. 3.16. Para el Caso heterocedástico con remuestreo por pares y remuestreo interno suavizado (a) Gráfico de dispersión de los β^* . (b) Áreas de las elipses en cada uno de los 500 datos. (c) Porcentaje en que el parámetro verdadero estuvo en la elipse generada en cada uno de los 500 datos

a los datos en azul y la izquierda a los datos en rojo. Además, *residuos NP*, *residuos S*, *pares NP* y *pares S* se refiere a los tipos de remuestreo de residuos con no-parámétrico, residuos con suavizado, pares con no-parámétrico y pares con suavizado.

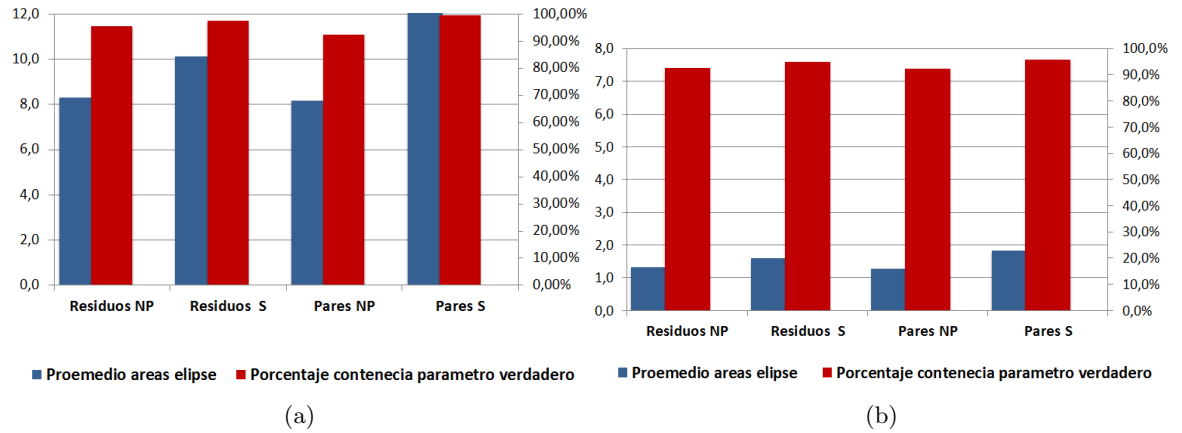


Fig. 3.17. Comparación del porcentaje de contención del parámetro verdadero en las elipses versus el área promedio de estas elipses generadas en cada uno de los métodos. **a)** Datos con errores homocedásticos (el eje izquierdo esta re-escalado por 10^6), **b)** Datos con errores homocedásticos con 4 veces la varianza anterior (el eje izquierdo esta re-escalado por 10^4).

Es importante recalcar que las áreas promedios en las figuras **a)** de 3.17 y 3.18 son del orden de 10^{-6} , mientras que los datos de las figuras **b)** de 3.17 y 3.18 son del orden 10^{-4} . Tratamos los esquemas homocedástico, homocedástico con 4 veces la varianza, heterocedástico y heterocedástico con 4 veces la varianza de manera independiente y tenemos lo siguiente:

1. Para el caso **homocedástico simple** se observa en la figura 3.17 (a) que el remuestreo por *pares no-parámétrico* es en el que menor área promedio tienen las elipses que genera, y el de

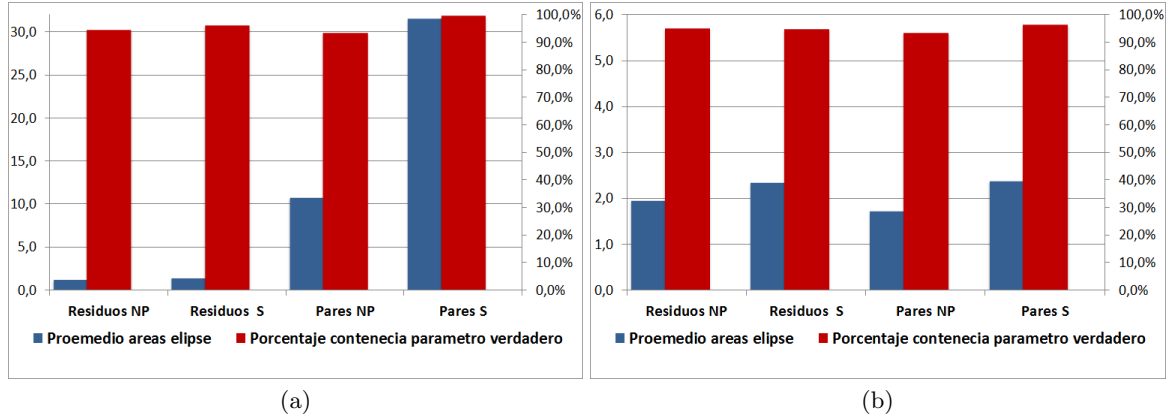


Fig. 3.18. Comparación del porcentaje de conteneencia del parámetro verdadero en las elipses versus el área promedio de estas elipses generadas en cada uno de los métodos. **a)** Datos con errores heterocedásticos (el eje izquierdo esta re-escalado por 10^6), **b)** Datos con errores heterocedásticos con 4 veces la varianza anterior (el eje izquierdo esta re-escalado por 10^4).

mayor porcentaje de conteneencia del parámetro verdadero es el método de remuestreo por *pares suavizado*. Cabe aclarar que el área promedio en "pares suavizado" se sale de la gráfica pues su valor es demasiado grande respecto a las áreas promedio de los demás tipos de remuestreo, lo cual explica por qué su porcentaje de conteneencia es el más grande.

- Para el caso **homocedástico con 4 veces la varianza anterior** se observa en la figura 3.17 (b) que el remuestreo por *pares suavizado* es en el que menor área promedio tienen las elipses que genera (en realidad observando las tablas se reafirma la observación), y *residuos suavizado* es el de mayor porcentaje de conteneencia del parámetro verdadero.
- Para el caso **heterocedástico simple** se observa en la figura 3.18 (a) que el remuestreo por *residuos no-parámétrico* es en el que menor área promedio tienen las elipses que genera (en la respectiva tabla se reafirma esta observación), y el de mayor porcentaje de conteneencia del parámetro verdadero es el método de remuestreo por *pares suavizado*, aclarando que su área promedio en ese caso es la más grande respecto a los otros métodos.
- Para el caso **heterocedástico con 4 veces la varianza anterior** se observa en la figura 3.17 (b) que el remuestreo por *pares no-parámétrico* es en el que menor área promedio tienen las elipses que genera, y *pares suavizado* es el de mayor porcentaje de conteneencia del parámetro verdadero.

En todos los casos se observa que los porcentajes de conteneencia están por encima de 90%, para los casos 1 y 3 en la numeración anterior existe una especie de inestabilidad en el promedio de las áreas de las elipses siendo en ambos caso el remuestreo *por pares suavizado* el menos eficiente dado que sus elipses son demasiado grandes.

No deja de ser llamativo que al multiplicar por 4 la varianza (ítems 2 y 4) el comportamiento del promedio de las áreas de las elipses se estabiliza, es decir, los valores son cercanos, no obstante, esas áreas son del orden de 10^{-4} , es decir, son mas grandes que la menor área promedio en los casos 1 y 3, de modo que lo que se gana en estabilidad se pierde en precisión.

4 Conclusiones

Un método ideal en nuestro concepto es aquel en el que el área de las elipses son pequeñas y el porcentaje de conteneencia del parámetro verdadero es alto. En ese sentido, para el caso **homocedástico simple** en nuestro criterio los métodos de remuestreo por *residuos no-parámétrico*

y *pares no-parámtrico* son la mejor elección, las diferencias entre los dos son numéricas y no son significativas para decantarse por uno u otro, lo que sí se puede asegurar es que el método menos recomendable es el de *pares suavizado* dado su inconveniente para elegir un h óptimo en la estimación por núcleos. Por ejemplo, intentamos usar el h óptimo sugerido por [2] pero el porcentaje de contenencia era cero, ante tal inconveniente se multiplicó dicho valor por un número entre cero y uno hasta obtener uno en el que el porcentaje de contenencia era alto, otro aspecto en contra es su costo computacional.

En el caso **homocedástico con 4 veces la varianza** es prácticamente inoportuno decantarse por un método de remuestreo, quizá por ligera ventaja numérica en el área promedio de las elipses la mejor elección sea el remuestreo por *pares no-parámtrico*, pero dicha diferencia pareciera ser poco significativa, lo mismo ocurre cuando queremos determinar cual es el método menos recomendable. Ahora, de acuerdo a las pequeñas diferencias numéricas el menos recomendable es el remuestreo por *residuos suavizado*, ya que su porcentaje de contenencia no es tan alto y su área promedio es ligeramente de las mas grandes.

Para el caso **heterocedástico simple** curiosamente la mejor elección es el remuestreo por *residuos no-parámtrico*, decimos curiosamente ya que se pensó que la hipótesis de homocedasticidad era fuerte, no obstante ocurrió todo lo contrario: aunque el porcentaje de contenencia del parámetro verdadero es casi 90%, en el caso homocedástico dichos porcentajes están casi por encima del 95%, es decir, en el caso heterocedástico hay casi un 5% de probabilidad menos de estimar el parámetro verdadero que en el caso homocedástico. Por otro lado, el método menos eficiente es el de *pares suavizado*, el inconveniente con este remuestreo es igual al mencionado anteriormente.

Por ultimo, para el caso **heterocedástico con 4 veces la varianza** se presenta un comportamiento similar a su par homocedástico, las áreas promedio (**eficiencia**) de las elipses son similares, y los porcentajes de contenencia (**cobertura**) del parámetro verdadero son altos en todos los casos, de modo que cualquier elección que se haga estará basada en los pequeños detalles, en este caso, serían las pequeñas diferencias numéricas. En ese orden de ideas la mejor elección es el remuestreo por *pares no-parámtrico* ya que su área promedio de las elipses es pequeña y su porcentaje de contenencia es alto, y el método menos recomendable es el remuestreo por *pares suavizado* debido a su costo computacional e imposibilidad de tener un h concreto único para la estimación por núcleos.

Referencias

1. R. Randles & D. Wolfe, *Introduction to Theory of Nonparametric Statistic*, Krieger Publishing Company, 3-114 (1979)
2. Bowman, AW and Foster, PJ, *Adaptive smoothing and density-based tests of multivariate normality*, Journal of the American Statistical Association, 529–537 (1993)