Robustez

Basado en el libro Robust Statistics: Theory and Methods (with R), (2019) 2nd Edition, Ricardo A. Maronna, R. Douglas Martin, Victor J. Yohai, Matías Salibián-Barrera

Al final de cada capítulo, el libro trae una sección *Recommendations and software* que indica procedimientos recomendados por los autores y las funciones de R que los implementan.

Estimadores como funcionales: caso general

Un funcional es una función $T : \mathcal{C} \longrightarrow \mathbb{R}$ o más generalmente, $T : \mathcal{C} \longrightarrow \mathbb{R}^k$, donde $\mathcal{C}$ es el espacio de las funciones de distribución.

Por ejemplo, $T(F) = \mathbb{E}_F(x)$, la media o esperanza de la distribución es un funcional.

Sea $\{x_1, \ldots, x_n\}$ una muestra de tamaño n de variables iid con distribución F. La función de distribución empírica se calcula por

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,t]}(x_i)$$

Para toda función g continua, tenemos $\mathbb{E}_{F_n} g(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i)$.

Por la Ley de los Grandes Números, sabemos que para todo t, $F_n(t)$ converge en probabilidad a $F(t)$. Es más, esta convergencia vale en sentidos más fuertes que éste.

Usando este hecho, un estimador consistente de $T(F)$ será $T(F_n)$. En el ejemplo anterior, $T(F_n) = \mathrm{E}_{F_n}(x) = \overline{x}_n$ (la media muestral)

Estimadores como funcionales: LS, estimador de mínimos cuadrados para regresión

Al estudiar teoría asintótica vimos que, bajo el modelo de regresión (al que llamamos Modelo de Proyección) dado por $(\boldsymbol{x}, y) \sim F$ con $\boldsymbol{x} \in \mathbb{R}^p, y \in \mathbb{R}$

$$y = \boldsymbol{x}'\boldsymbol{\beta} + u \quad \text{con}$$
$$\mathbb{E}_F(\boldsymbol{x}u) = \mathbf{0}$$

el funcional que da origen al estimador de mínimos cuadrados resulta ser

$$T(F) = \left(\mathbb{E}_F\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}_F(\boldsymbol{x}y).$$

El estimador, basado en una muestra $(\mathbf{x}_i, y_i), i = 1, \ldots, n$, es

$$T(F_n) = \left(\mathbb{E}_{F_n}\left(\boldsymbol{x}\boldsymbol{x}'\right)\right)^{-1}\mathbb{E}_{F_n}(\boldsymbol{x}y)$$
$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i y_i\right) = \widehat{Q}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\widehat{Q}_{\boldsymbol{x}y}$$
$$= \left(X^T X\right)^{-1}X^T\mathbf{y} \quad \text{con } X \in \mathbb{R}^{n\times p}, \mathbf{y} \in \mathbb{R}^{n\times 1}$$

Medidas de robustez de un estimador, y del funcional asociado

Como asumimos que F es solo aproximadamente conocida, nos interesa estudiar el comportamiento del estimador cuando F se mueve en una vecindad de una distribución prefijada. Hay muchas formas de caracterizar las vecindades. La más sencilla es trabajar con un *entorno de contaminación*.

$$\mathcal{F}(F, \varepsilon) = \{(1-\varepsilon)F + \varepsilon G : G \in \mathcal{G}\}$$

donde $\mathcal{F}$ es un conjunto apropiado de distribuciones. Muchas veces es la masa puntual en x_0, es decir, la función de distribución de la variable aleatoria constante, x tal que $P(x = x_0) = 1$. La notamos δ_{x_0}.

Función de influencia de un funcional

La función de influencia (*influence function, IF*) de un funcional (Hampel, 1974) es una función que describe una característica de su comportamiento asintótico. Nos da una idea de la sensibilidad (asintótica) del estimador. Es una aproximación del comportamiento del estimador cuando la muestra contiene una pequeña fracción ε de outliers idénticos. Se define por

$$\mathrm{IF}_T(x_0, F) = \lim_{\varepsilon \to 0^+} \frac{T\left((1-\varepsilon)F + \varepsilon\delta_{x_0}\right) - T(F)}{\varepsilon} \tag{1}$$

donde δ_{x_0} es la distribución de la constante igual a x_0 (masa puntual en x_0) y $\varepsilon \to 0^+$ indica que el límite se toma a derecha.

Sensibilidad a contaminaciones: Función de Influencia de un funcional

La cantidad $T((1-\varepsilon)F + \varepsilon\delta_{x_0})$ es el valor del funcional cuando la distribución subyacente de los datos es F y una fracción ε de la muestra está compuesta por outliers iguales a x_0. Si ε es suficientemente chico, por (1) este valor puede ser aproximado por

$$T((1-\varepsilon)F + \varepsilon\delta_{x_0}) \approx T(F) + \varepsilon\,\mathrm{IF}_T(x_0, F)$$

y la diferencia entre ellos es $T((1-\varepsilon)F + \varepsilon\delta_{x_0}) - T(F)$, que mide el cambio en el funcional por la inclusión de una proporción ε de outliers iguales a x_0 en la muestra se puede aproximar por $\varepsilon\,\mathrm{IF}_T(x_0, F)$

Función de influencia del funcional LS (*least squares*)

La versión empírica de la función de influencia es la función de influencia empírica, EIF que se obtiene sustituyendo a F por F_n.

$$EIF_T(x_0) = \mathrm{IF}_T(x_0, F_n)$$

Bajo condiciones bastante generales, para LS tenemos

$$\mathrm{IF}_{LS}((\mathbf{x}_0, y_0), F) = (y_0 - \mathbf{x}_0'\beta) Q_{\mathbf{xx}}^{-1}\mathbf{x}_0$$

con $Q_{\mathbf{xx}} = \mathrm{E}\mathbf{x}\mathbf{x}'$.

Entonces vemos que la función de influencia del funcional LS depende del error $y_0 - \mathbf{x}_0'\beta$ de la observación $(\mathbf{x}_0, y_0)$ por un lado y de $Q_{\mathbf{xx}}^{-1}\mathbf{x}_0$ por otro, y que por lo tanto es no acotada. Se hace arbitrariamente grande tanto si $\mathbf{x}_0$ es grande como si y_0 lo es.

$$EIF_{LS}(\mathbf{x}_0, y_0) = \left(y_0 - \mathbf{x}_0'\widehat{\beta}\right) n(X'X)^{-1}\mathbf{x}_0$$

M-estimadores robustos de regresión

El enfoque para controlar la influencia de observaciones atípicas que pueden estar tanto en la variable respuesta como en las covariables consiste en definir al M-estimador $\widehat{\boldsymbol{\beta}}$ como aquel valor que

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}}\right) = \sum_{i=1}^{n} \rho\left(\frac{r_i(\widehat{\boldsymbol{\beta}})}{\widehat{\sigma}}\right) \tag{2}$$

donde $r_i(\widehat{\boldsymbol{\beta}}) = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$, ρ es una función que se comporta como el cuadrado en una vecindad del cero, pero que es acotada, y $\widehat{\sigma}$ es una escala previamente calculada. La escala deberá cumplir ciertos requisitos.

Si ρ tiene derivada ψ, (2) es equivalente a hallar $\widehat{\boldsymbol{\beta}}$ que resuelva

$$\sum_{i=1}^{n} \psi\left(\frac{r_i(\widehat{\boldsymbol{\beta}})}{\hat{\sigma}}\right) \mathbf{x}_i = \mathbf{0} \tag{3}$$

Ejemplo de funciones ρ y ψ: familia bicuadrada

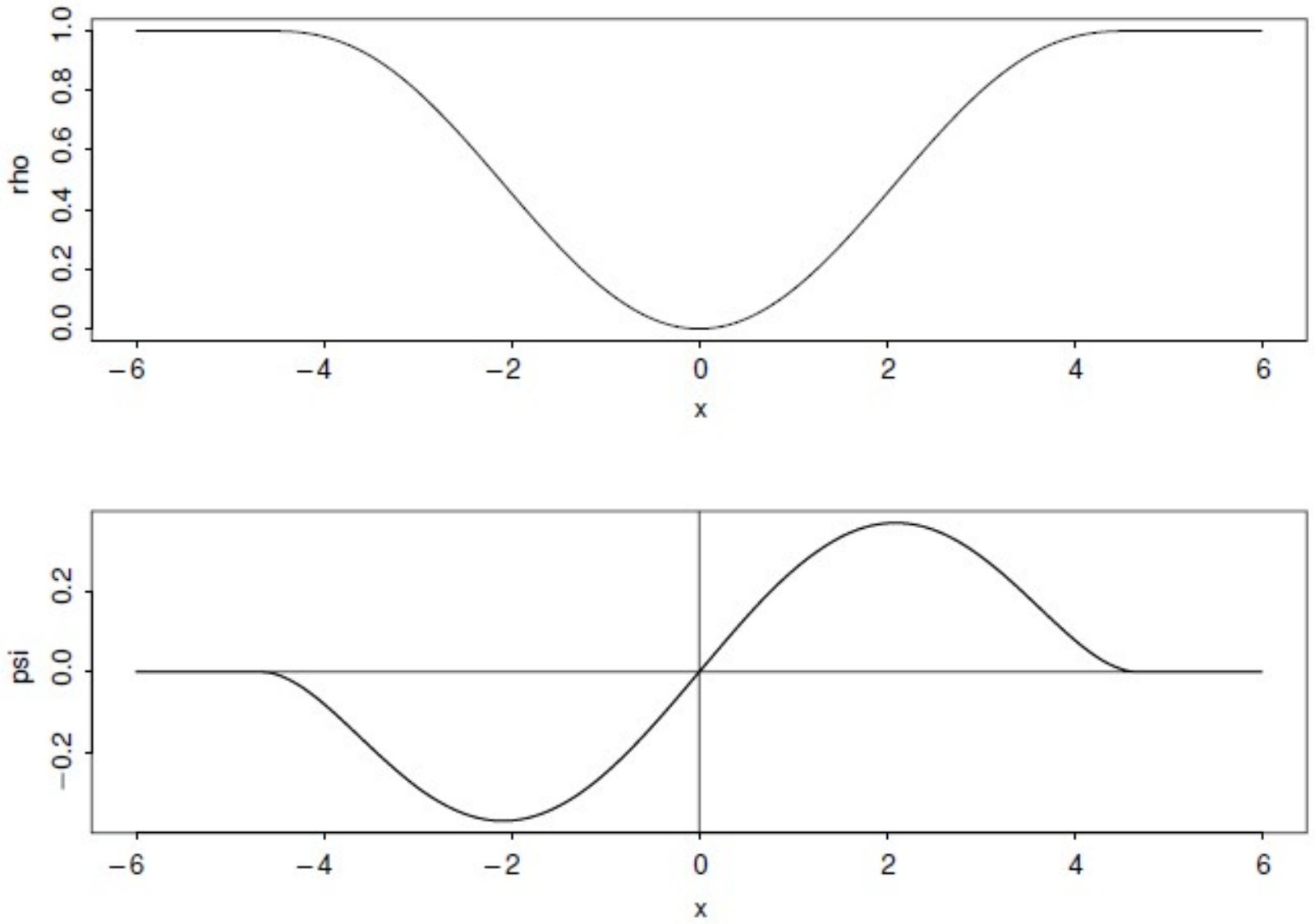


Figure 2.5 ρ- and ψ-functions for bisquare estimate

MM-estimadores de regresión para el modelo lineal con **x** aleatoria

Un MM-estimador de regresión es un método de cálculo particular para aproximar $\widehat{\boldsymbol{\beta}}$ definido por (2). Fueron propuestos en 1987 por Yohai. Los pasos del procedimiento propuesto por el MM-estimador son

1. Calcular un estimador inicial consistente $\widehat{\boldsymbol{\beta}_0}$ que resista bien la presencia de outliers en la muestra (i.e. con alto punto de ruptura) pero que posiblemente tenga baja eficiencia (i.e. performance subóptima en términos de varianza) si los errores tienen distribución normal.
2. Calcular una escala robusta $\widehat{\sigma}$ de los residuos $r_i(\widehat{\boldsymbol{\beta}_0}) = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}_0}$.
3. Encontrar una solución $\widehat{\boldsymbol{\beta}}$ de (2) utilizando un procedimiento iterativo (IRWLS) que comience en $\widehat{\boldsymbol{\beta}_0}$ y utilice la escala $\widehat{\sigma}$ calculada en 2.

MM-estimadores de regresión para el modelo lineal

Tenemos observaciones de vectores aleatorios $(\mathbf{x}_i, y_i) \in \mathbb{R}^{(p+1)}$, $1 \leq i \leq n$ con distribución F. Bajo el modelo

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$$

donde los errores u_i son iid e independientes de las $\mathbf{x}_i$, puede probarse que la función de influencia de un M-estimador con σ conocida es

$$\mathrm{IF}\left((\mathbf{x}_0, y_0), F\right) = \frac{\sigma}{b}\psi\left(\frac{y_0 - \mathbf{x}_0'\boldsymbol{\beta}}{\sigma}\right)\mathbf{Q}_{\mathbf{xx}}^{-1}\mathbf{x}_0$$

con $b = \mathbb{E}\psi'\left(\frac{u}{\sigma}\right)$ y $\mathbf{Q}_{\mathbf{xx}} = \mathbb{E}\mathbf{x}\mathbf{x}'$. Entonces vemos que la función de influencia del funcional M depende de la función ψ del error $y_0 - \mathbf{x}_0'\boldsymbol{\beta}$ estandarizado por σ de la observación $(\mathbf{x}_0, y_0)$.

Si el M-estimador se define a través de una función ρ acotada con derivada ψ tal que $\psi(t) = 0$ para $|t| \geq k$, entonces IF tenderá a infinito sólo cuando x_0 tienda a infinito y $|y_0 - \mathbf{x}_0'\boldsymbol{\beta}|/\sigma < k$, lo que significa que los outliers grandes no tendrán ninguna influencia en el funcional.

Estimador recomendado y rutina de R

En el libro de Maronna, Martin, Salibian-Barrera, Yohai recomiendan usar el MM-estimador de regresión (con función ρ que se denomina óptima y estimador inicial Peña-Yohai) que está implementado en la rutina `lmrobdetMM` (en la librería `RobStatTM` del R). Y también (con otros estimadores iniciales u otros algoritmos para hallar la raíz) en las rutinas `lmrob` (de la librería `robustbase`) y `lmRob` (de la librería `robust`).

Ejemplo: wood data

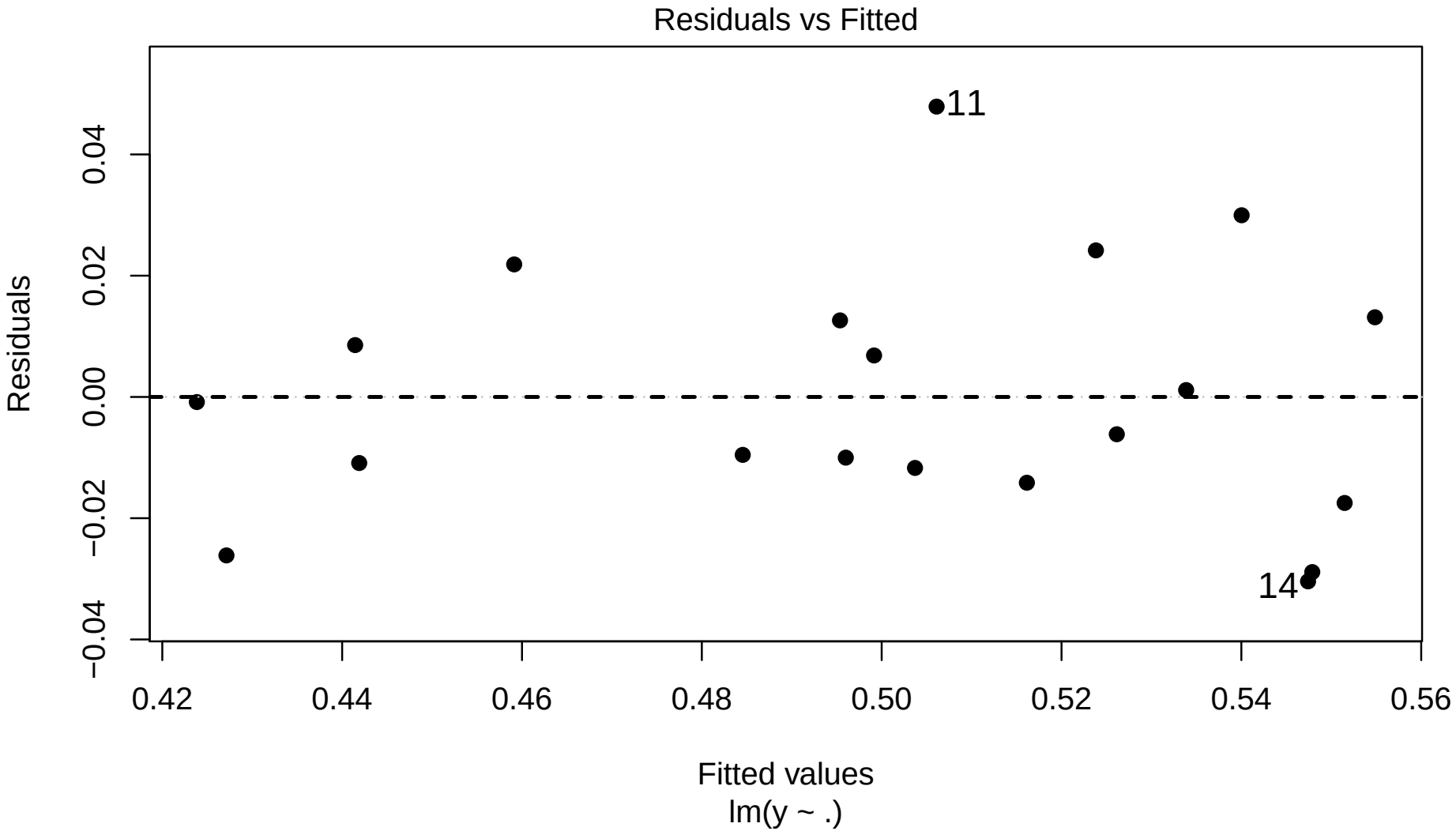
El siguiente ejemplo se basa en los "datos modificados del peso específico de la madera". Los los datos en bruto provinieron del libro de Draper y Smith (1966, p. 227) y se utilizaron para determinar la influencia de factores anatómicos sobre el peso específico de la madera, con 20 casos, cinco variables explicativas y un intercept. Rousseeuw y Leroy (1987) modificaron los datos reemplazando cuatro observaciones (las observaciones 4, 6, 8 y 19) por valores atípicos (dataset `wood`), y son analizados en el libro de Maronna, Martin, Salibián-Barrera y Yohai.

Las instrucciones, datos y demás están en el script `wood.R`, colgado de la página web del libro Maronna et al.,
`www.wiley.com/go/maronna/robust` (Ejemplo 5.2)

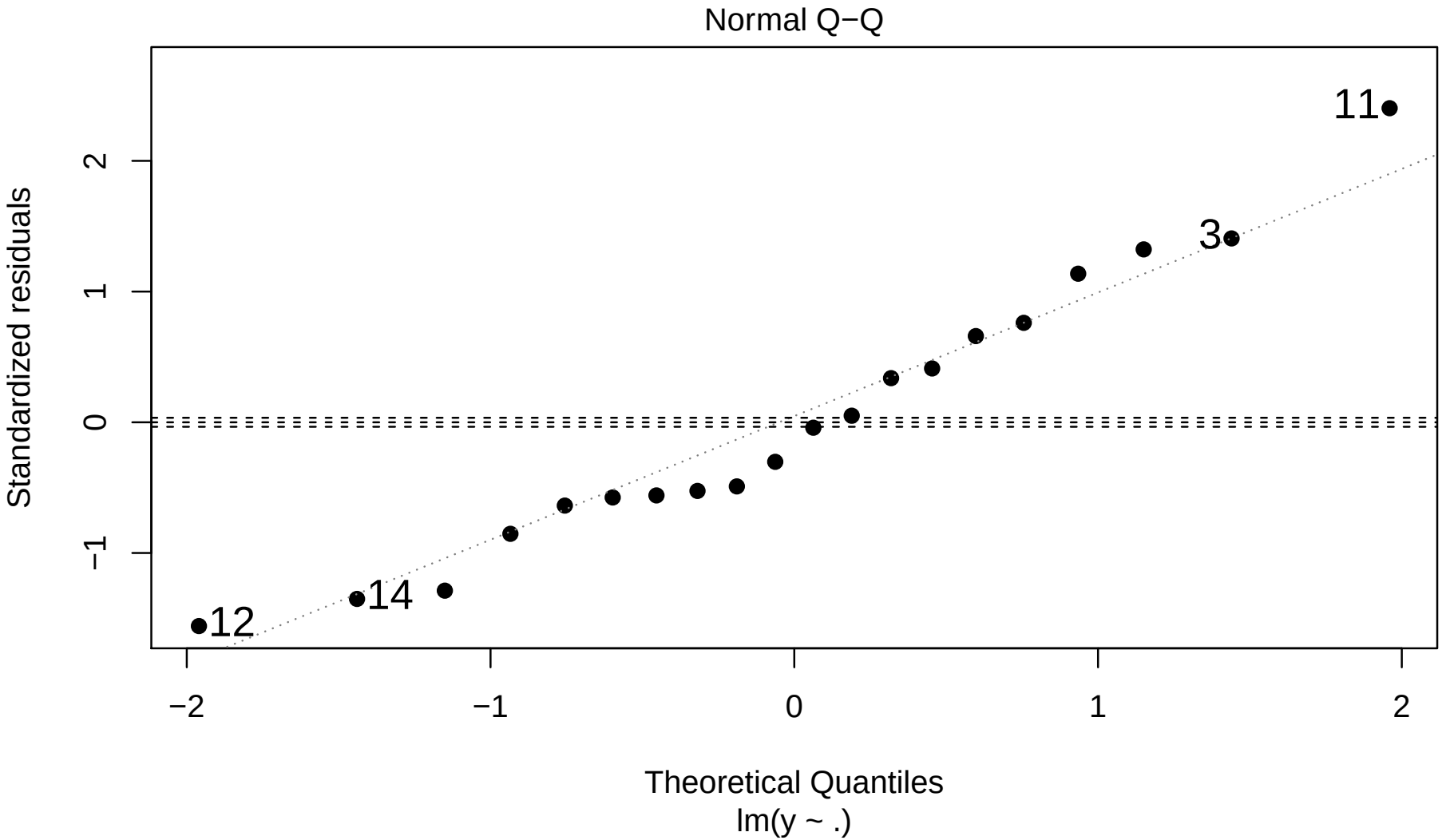
Ejemplo: wood data

```
library(RobStatTM)
data(wood, package='robustbase')
cont <- lmrobdet.control(bb = 0.5, efficiency = 0.85,
family = "bisquare")

#MM fit
woodMM <- lmrobdetMM(y ~ ., data=wood, control=cont)

#LS fit
woodLS <- lm(y ~ ., data=wood)
```

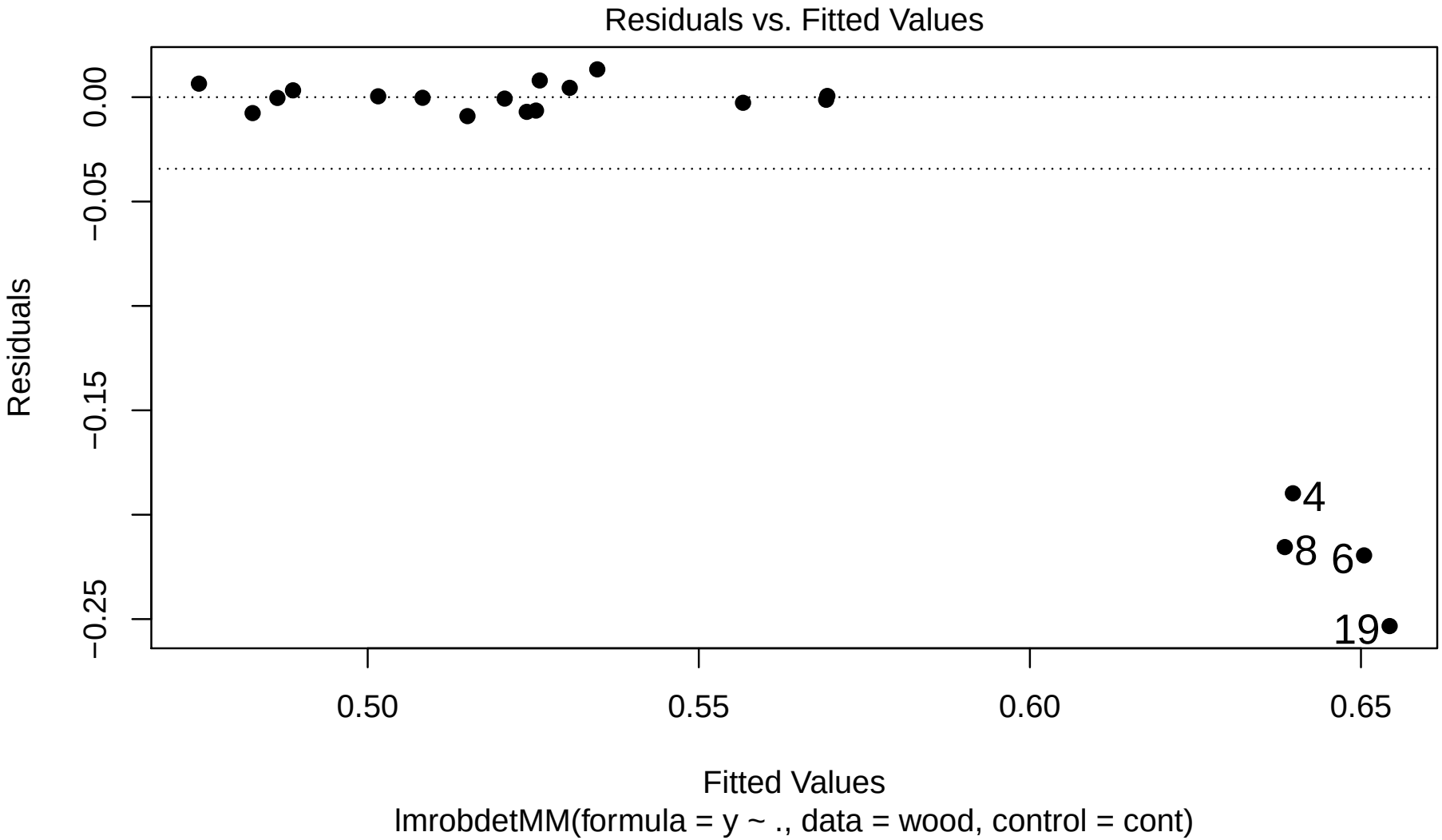
wood: Residuos vs predichos LS



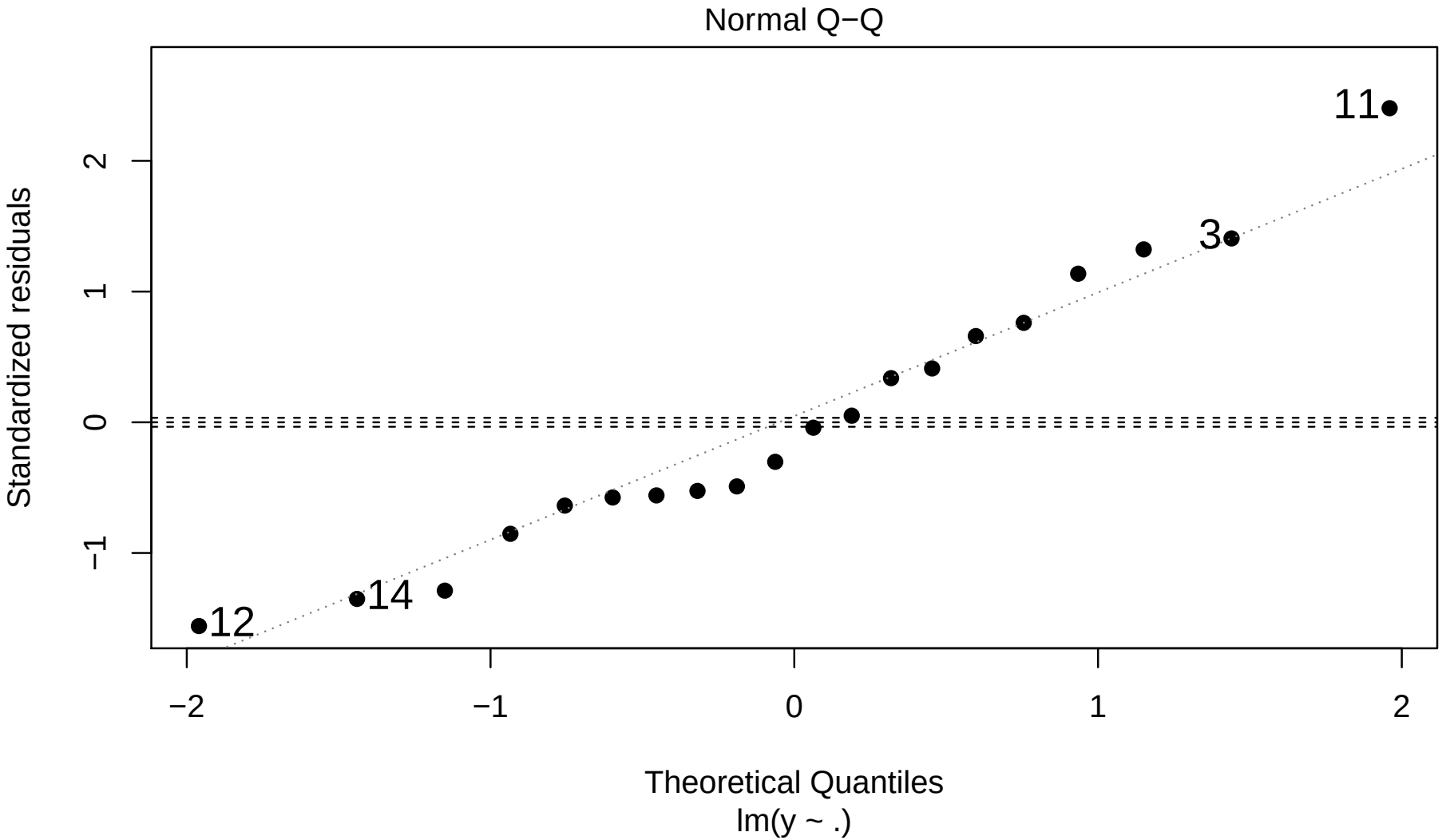
wood: QQplot de los Residuos LS



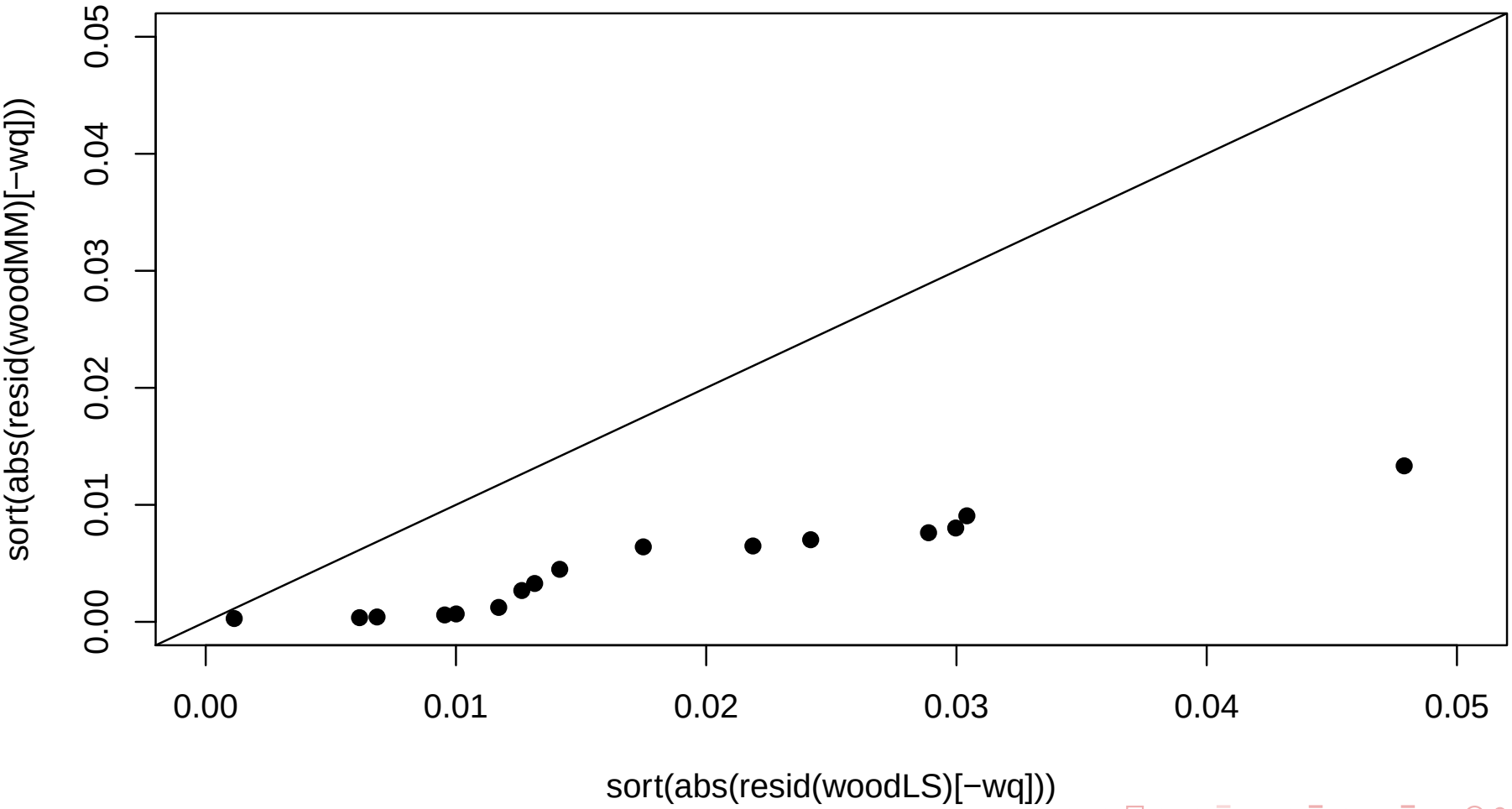
wood: Residuos vs predichos MM-estimador



wood: QQplot de los Residuos MM-estimador



wood: plot de los Residuos MM-estimador vs los del ajuste por LS



wood: análisis del ajuste

Las primeras dos figuras muestran el gráfico de los residuos del ajuste por mínimos cuadrados (LS) a los datos, el primero contra los valores predichos, el segundo el Q-Q plot.

A partir de estos gráficos uno concluye (equivocadamente) que no hay valores atípicos en los datos y confía en el ajuste dado por LS

Las siguientes dos figuras son los gráficos equivalentes para el estimador MM de eficiencia normal del 85%.

A partir de estos gráficos uno descubre acertadamente que hay cuatro observaciones atípicas: que son las observaciones 4, 6, 8 y 19.

Vemos que el ajuste LS se ve completamente confundido por los outliers en este ejemplo, el LS proporciona un ajuste equivocado a los datos. Los outliers enmascaran las conclusiones del LS. Un ajuste robusto permite identificar la presencia de outliers en la muestra.

wood: análisis del ajuste

Si para un conjunto de datos las conclusiones del ajuste LS difieren mucho de las que proporciona un ajuste robusto, esto es una indicación de la presencia de observaciones atípicas.

La forma de identificar a las observaciones atípicas es a través de un ajuste robusto.

En la quinta figura vemos los residuos absolutos ordenados de LS, en el eje horizontal y los del estimador de MM en el eje vertical, y la línea de identidad. Las observaciones con los cuatro residuos absolutos más grandes del estimador MM se omitieron por razones de escala. En el gráfico vemos que los residuos del ajuste del MM son en general más pequeños que los residuos de LS, y por lo tanto, MM da un mejor ajuste a la mayor parte de los datos.

Wood data, ajuste LS

Por supuesto, los coeficientes ajustados y la significatividad de los mismos cambia. Comparemos el resultado del ajuste LS con el del MM.

```
> summary(woodLS)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.42178    0.16912   2.494  0.02576 *
x1           0.44069    0.11688   3.770  0.00207 **
x2          -1.47501    0.48692  -3.029  0.00901 **
x3          -0.26118    0.11199  -2.332  0.03513 *
x4           0.02079    0.16109   0.129  0.89915
x5           0.17082    0.20336   0.840  0.41505

Residual standard error: 0.02412 on 14 degrees of freedom
Multiple R-squared:  0.8084,Adjusted R-squared:   0.74
F-statistic: 11.81 on 5 and 14 DF,  p-value: 0.0001282
```

Wood data, ajuste MM

```
> summary(woodMM)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.37937    0.05583   6.795 8.66e-06 ***
x1           0.21576    0.04343   4.968 0.000207 ***
x2          -0.07674    0.20364  -0.377 0.711941
x3          -0.56358    0.04446 -12.676 4.62e-09 ***
x4          -0.39615    0.06762  -5.858 4.16e-05 ***
x5           0.60202    0.08172   7.367 3.53e-06 ***
Robust residual standard error: 0.01372
Multiple R-squared:  0.8499,Adjusted R-squared:  0.7963
Convergence in 8 IRWLS iterations
```

Wood data, ajuste LS con pesos obtenidos con el MM

Veamos que ajustando por mínimos cuadrados con los pesos obtenidos por el MM-estimador da lo mismo que el ajuste del MM-estimador

```
> #ajuste LS con pesos
> woodLSpesado <- lm(y ~ ., data=wood,
weights = woodMM$rweights)
> summary(woodLSpesado)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.37937    0.05256   7.218 2.86e-05 ***
x1           0.21576    0.04089   5.277 0.000359 ***
x2          -0.07674    0.19170  -0.400 0.697347
x3          -0.56358    0.04186 -13.465 9.82e-08 ***
x4          -0.39615    0.06366  -6.223 9.84e-05 ***
x5           0.60202    0.07693   7.826 1.43e-05 ***

Residual standard error: 0.007133 on 10 degrees of freedom
Multiple R-squared:  0.9608,Adjusted R-squared:  0.9412
F-statistic: 49.03 on 5 and 10 DF,  p-value: 1.032e-06
```

Wood data, pesos obtenidos con el MM

Veamos los pesos que el MM-estimador le da a cada observación

```
> woodMM$rweights
1         2         3         4         5         6
0.9415992 0.9814845 0.9996758 0.0000000 0.8428951 0.0000000
7         8         9        10        11        12
0.9616147 0.0000000 0.9472429 0.9998796 0.9933961 0.9625081
13        14        15        16        17        18
0.9901004 0.9551235 0.9998368 0.9999241 0.9995758 0.9257626
19        20
0.0000000 0.9985913
```

Las observaciones 4, 6, 8 y 19 reciben peso cero y no influyen en la estimación.

Distribución asintótica del MM-estimador de regresión

Asumamos que vale el modelo lineal con covariables aleatorias, que la varianza de $\mathbf{x}$ es finita y que $\widehat{\sigma}$ converge en probabilidad a un valor σ. Entonces se puede probar bajo condiciones bastante generales que el M-estimador de $\boldsymbol{\beta}$ es consistente y asintóticamente insesgado. Ver la Sección 10.10.2 del libro. Más precisamente

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}_p\left(0, vQ_{\mathbf{xx}}^{-1}\right)$$

donde $Q_{\mathbf{xx}} = \mathbb{E}(\mathbf{x}\mathbf{x}')$ y

$$v = \sigma^2 \frac{\mathbb{E}\psi(u/\sigma)^2}{\left[\mathbb{E}\psi'(u/\sigma)\right]^2}$$

Estos resultados permiten construir tests de hipótesis e intervalos de confianza asintóticos para los parámetros del modelo. También pueden obtenerse tests e IC para los parámetros a través del bootstrap, para una discusión del tema ver el libro de Maronna et al. citado antes, sección 5.6.1.

Otro ejemplo de ajuste robusto

El otro ejemplo de ajuste robusto visto en clase (regresión lineal simple con el agregado de puntos A, B, C y D) puede consultarse en la Sección 3.2.3, páginas 95 a 108, del apunte colgado en la dirección que está más abajo. Para esos ejemplos se comparar el ajuste LS con el de robusto de los MM-estimadores y también se calculan los leverage de las observaciones, como herramienta clásica para la detección de puntos atípicos.

http://mate.dm.uba.ar/∼meszre/apunte_regresion_lineal_szretter.pdf