CEMA Instituto Universitario
Av. Córdoba 374
1054 Capital Federal

Tel.: 314-2269
Fax: 314-1654

REGRESION ROBUSTA

Victor J. Yohai

Nª 9

REGRESION ROBUSTA

por

Víctor J. Yohai*

C.E.M.A.

SINTESIS

En este trabajo se muestra que el estimador de mínimos cuadrados para los coeficientes de una regresión múltiple, es muy sensible a la normalidad de los errores o a perturbaciones en el modelo, pudiendo unas pocas observaciones atípicas aumentar enormemente su error cuadrático medio. Se introduce el concepto de estimador robusto que corresponde a un estimador que se comporta en forma estable frente a pequeñas perturbaciones en los errores o el modelo. Dos clases de estimadores robustos son revisadas: estimadores del tipo de máxima verosimilitud (M-estimadores) y estimadores del tipo de máxima verosimilitud generalizados (GM-estimadores). Se muestra que los primeros son robustos frente a perturbaciones en los errores, y los segundos frente a perturbaciones en los errores o en el modelo. Se realizan comparaciones numéricas entre los diferentes estimadores robustos y el estimador de mínimos cuadrados que muestran claramente la conveniencia de utilizar los primeros, ya que en presencia de unas pocas observaciones con perturbaciones, pueden dar lugar a enormes ganancias de eficiencia, pero sólo a despreciables pérdidas en ausencia de las mismas. Se describen algoritmos que permiten el cómputo de los estimadores robustos.

*Se agradece la colaboración de Ester Lagomarsino en el cómputo de las varianzas asintóticas de la Sección 8, como también los comentarios recibidos de parte de miembros del C.E.M.A.

0. Introducción

Sea el modelo de regresión múltiple:

$$Y = \theta_1 Z_1 + \theta_2 Z_2 + \dots + \theta_K Z_K + U$$

donde Y es la variable independiente, Z_1, Z_2, ..., Z_K son K variables dependientes y U es el error. Supongamos además que se han observado las variables durante T períodos.

Usualmente θ_1, ..., θ_K son estimados por el método de mínimos cuadrados. Es decir, se eligen de manera de minimizar la suma de los cuadrados de los residuos. Es bien conocido que este método es óptimo si los errores de los distintos períodos son independientes con distribución normal de media 0 y una misma varianza en todos los períodos. Sin embargo la hipótesis de normalidad de los residuos es poco justificable en series económicas. Especialmente en las series económicas argentinas, son frecuentes errores groseros en las observaciones, por lo que es de esperar que las colas de las distribuciones de los errores sean más pesadas que las que corresponden al caso normal.

Otra hipótesis en la que se basa la optimalidad del método de mínimos cuadrados es que el modelo de regresión múltiple se cumpla en todos los períodos. Esto tampoco puede garantizarse en los modelos económicos, donde es frecuente encontrar períodos donde los mismos son perturbados por acontecimientos institucionales o políticos que no son contemplados explícitamente.

Esto hace que resulte importante estudiar el comportamiento del estimador de mínimos cuadrados frente a 2 tipos posibles de perturbaciones: a) cuando los errores no son normales, por ejemplo si existe un cierto porcentaje de observaciones con errores groseros, b) el modelo puede dejar de cumplirse en algunos períodos anormales.

En este trabajo se muestra que el estimador de mínimos cuadrados es sumamente sensible a los 2 tipos de perturbaciones mencionadas, y basta que por ejemplo haya 10% de errores groseros o un 10% de períodos anormales para que su error cuadrático medio aumente enormemente, por ejemplo más del 900%.

Frente a este situación se hace imperativo buscar otros métodos de estimación que sean más estables frente a las perturbaciones mencionadas, y que al mismo tiempo sean eficientes cuando no las haya. Estos métodos de estimación se denominan robustos.

En este trabajo se hace una revisión de los 2 tipos de estimadores robustos que consideramos más eficaces y que al mismo tiempo son simples de implementar, los estimadores del tipo de máxima verosimilitud (M-estimadores), y estimadores del tipo de máxima verosimilitud generalizados (GM-estimadores). Los primeros son robustos frente a perturbaciones del tipo a), y los segundos frente a perturbaciones del tipo a) y b). Lo interesante de estos estimadores es que el aumento de su eficiencia respecto al estimador de mínimos cuadrados (medido por ejemplo en la disminución de su error cuadrático medio) en presencia por ejemplo de un 10% de observaciones con perturbaciones, puede ser enorme e inclusive infinita, y al mismo tiempo, en el caso que no haya perturbaciones (errores normales y el modelo siempre se cumple), la pérdida de eficiencia es sólo del 5%. Por lo tanto la utilización de estimadores robustos puede producir enormes ganancias pero a lo sumo muy pequeñas pérdidas.

El siguiente modelo muestra la sensibilidad del estimador de mínimos cuadrados frente a la presencia de unas pocas observaciones anormales. Consideremos el siguiente modelo de regresión lineal múltiple que puede considerarse una versión simple del modelo de Harberger:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \theta_3$$

donde:

Y es la tasa de variación del índice de precios al por mayor, nivel general (fuente INDEC);

X_1 es la tasa de variación de la oferta monetaria (M2), (fuente International Financial Statistics, Fondo Monetario Internacional);

X_2 es la tasa de variación del producto bruto interno (fuente Banco Central de la República Argentina).

Consideramos datos trimestrales correspondientes a los períodos que van desde el segundo trimestre del año 1971 hasta el cuarto trimestre de 1978.

La estimación de mínimos cuadrados de θ_1, θ_2, θ_3, da los siguientes valores. Los números entre paréntesis son las desviaciones standard estimadas.

$$Y = \underset{(0,33)}{0,56X_1} - \underset{(0,73)}{0,98X_2} + \underset{(0,09)}{0,12}$$

El error standard de la regresión es 0,26.

De acuerdo a estos resultados, ningún coeficiente es significativamente distinto de 0 a un nivel de confianza de 0,05.

Los residuos correspondientes están dados en el siguiente cuadro:

Tri \ Año	1971	1972	1973	1974	1975	1976	1977	1978
1		-0,04	-0,04	-0,24	0,00	1,03	-0,11	-0,14
2	0,06	0,06	-0,11	0,00	0,65	-0,12	-0,03	0,07
3	-0,04	-0,05	-0,29	-0,10	0,20	-0,15	0,02	-0,00
4	0,00	-0,15	-0,19	-0,03	0,00	-0,11	-0,10	0,02

Lo que se observa es que los errores correspondientes al segundo trimestre de 1975 y al primero de 1976 son exageradamente grandes, más de 2 errores standard de regresión en el primero y aproximadamente 4 en el segundo, y por lo tanto se puede sospechar que se trata de períodos atípicos. Esta sospecha se confirma si se recuerda que en dichos períodos ocurrieron hechos no tenidos en cuenta en el modelo, especialmente relacionados con las expectativas de inflación.

Apliquemos ahora un método robusto de estimación, que será explicado en la Sección 4. Los resultados son los siguientes:

$$\underset{}{Y} = \underset{(0,10)}{0,40X_1} - \underset{(0,22)}{0,54X_2} + \underset{(0,03)}{0,11}$$

Se puede observar que los coeficientes estimados cambiaron notablemente y hubo una mejora muy importante en la precisión, si se mide ésta por la disminución de los errores standard, ya que el cociente entre ambas es 3,03. Es decir el método robusto sería 3 veces más eficiente que el de mínimos cuadrados. Esto se debe a que el método robusto determina cuáles observaciones son sospechosas y les da menos peso en el proceso de estimación. Volveremos a este ejemplo al final de la Sección 5.

1. Modelo de Regresión Múltiple

Consideremos el modelo general de regresión múltiple:

$$(1.1)\quad y_t = \sum_{k=1}^{K} \theta_k z_{kt} + u_t \qquad t = 1, 2, \ldots, T$$

donde z_{kt} es el valor numérico de una variable independiente Z_k, $1 \leq k \leq K$, (K variables independientes), u_t es el error del período t, y $\theta_1, \theta_2, \ldots, \theta_K$ son los coeficientes de regresión.

En forma vectorial escribiremos:

$$(1.2)\quad y_t = \underset{\sim}{z}_t' \underset{\sim}{\theta} + u_t \qquad t = 1, 2, \ldots, T$$

donde $\underset{\sim}{z}_t = (z_{1t}, \ldots, z_{Kt})'$, $\underset{\sim}{\theta} = (\theta_1 \ldots, \theta_K)'$.

Notación: En general, los vectores serán vectores columna, el símbolo ' indicará transpuesto de un vector o una matriz. Por lo tanto, un vector seguido de ' será un vector fila. El símbolo ~ debajo de una letra indicará que se trata de un vector o una matriz.

En forma matricial (1.1) también se puede escribir como:

$$(1.3)\quad \underset{\sim}{y} = \underset{\sim}{Z}' \underset{\sim}{\theta} + \underset{\sim}{u}$$

donde $\underset{\sim}{Z}$ es la matriz cuyo elemento k, t es z_{kt}, $\underset{\sim}{y} = (y_1, \ldots, y_T)'$, $\underset{\sim}{u} = (u_1, \ldots u_T)'$.

Las hipótesis usuales que se formulan sobre los errores u_t son las siguientes.

H1: $E(u_t) = 0$, $t=1, \ldots T$, donde el símbolo E indica esperanza o valor medio.

H2: Los errores u_t, $t=1, \ldots, T$, tienen todos una misma distribución que se simbolizará con F.

H3: Los errores u_t, $t=1, \ldots, T$ son independientes.

H4: Los errores u_t, $t=1, \ldots, T$ tienen distribución normal.

El estimador utilizado generalmente para $\underset{\sim}{\theta}$ es el estimador de mínimos cuadrados (EMC) que indicaremos por $\hat{\underset{\sim}{\theta}}_{MC}$. El EMC se define como el valor de $\underset{\sim}{\theta}$ que minimiza:

$$(1.4) \quad \sum_{t=1}^{T} (y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta})^2$$

El EMC viene dado por $\hat{\underset{\sim}{\theta}}_{MC} = (\hat{\theta}_{MC1}, \ldots, \hat{\theta}_{MCK})'$ con:

$$(1.5) \quad \hat{\underset{\sim}{\theta}}_{MC} = (\underset{\sim}{Z}\underset{\sim}{Z}')^{-1}\underset{\sim}{Z}\underset{\sim}{y}$$

La matriz de covarianza del EMC viene dada por:

$$(1.6) \quad \operatorname{cov}(\hat{\underset{\sim}{\theta}}_{MC}) = \operatorname{var}(F)(\underset{\sim}{Z}\underset{\sim}{Z}')^{-1}$$

donde var(F) es la varianza de la distribución F. Para una justificación de (1.5) y (1.6) ver por ejemplo Theil (1971).

La justificación para el empleo del EMC se basa en las siguientes propiedades.

P1. Si se cumplen H1 a H4, entonces $\hat{\underset{\sim}{\theta}}_{MC}$ es el estimador insesgado de $\underset{\sim}{\theta}$ con mínima matriz de covarianza. Esto implica que la varianza de cualquier combinación lineal de $\hat{\underset{\sim}{\theta}}_{MC}$ tiene menor varianza que la misma combinación lineal de cuqlqüier otro estimador insesg -do de $\underset{\sim}{\theta}$. En particular $\hat{\theta}_{\sim MCk}$ es el estimador insesgado de θ_k, $1 \leq k \leq K$, con menor varianza. Ver por ejemplo teorema 8.3 de Theil (1971).

P2. Si se cumplen H1 a H4, entonces $\hat{\theta}_{\sim MC}$ es el estimador de máxima verosimilitud (EMV) de $\underset{\sim}{\theta}$, y por lo tanto, bajo condiciones muy generales sobre la matriz $\underset{\sim}{Z}$ es asintóticamente normal y eficiente. Ver por ejemplo teorema 8.2 de Theil (1971)

P3. (Gauss-Markov) Si se cumple H1 a H3, entonces $\hat{\theta}_{\sim MC}$ es el estimador de mínima matriz de covarianza entre todos los estimadores insesgados de $\underset{\sim}{\theta}$ que son combinaciones lineales de $\underset{\sim}{y}$. Ver por ejemplo teorema 3.4 de Theil (1971).

Las propiedades P1 y especialmente P2 justifican la utilización de $\hat{\theta}_{\sim MC}$ cuando se cumplen H1 a H4. Puede quedar en la propiedad P1 alguna duda sobre el significado de imponer la restricción que los estimadores sean insesgados, y efectivamente para muestras chicas pueden encontrarse estimadores sesgados mejores que el EMC (con menor error cuadrático medio), especialmente cuando hay multicolinealidad en las variables independientes. No nos ocuparemos en este trabajo de la posibilidad de mejorar la estimación de $\underset{\sim}{\theta}$ usando estimadores sesgados. Puede consultarse para esto por ejemplo Hocking (1976).

Por otro lado la propiedad P3 se ha utilizado para justificar la elección del EMC cuando no hay normalidad. Sin embargo no resulta esta propiedad un argumento satisfactorio, ya que no hay ninguna razón para restringirse a estimadores que sean combinaciones lineales de y. Aunque esta restricción se justificase en el pasado en virtud de las dificultades de cálculo que implicaba utilizar estimadores no lineales, actualmente el argumento ha perdido vigencia debido a las posibilidades de acceso a los sistemas de cómputo e incluso a las calculadoras electrónicas manuales.

Por lo tanto, si desechamos P3 como justificación para el uso del EMC solamente nos quedan las propiedades P1 y P2, las cuales requieren que se satisfaga H1, H2, H3 y H4. Como H1 no es una restricción en el caso usual que la última variable Z_k sea constante 1, y como las hipótesis H2 y H3 han sido abundantemente discutidas en la literatura, en este trabajo discutiremos solamente los problemas que surgen con el EMC cuando H4 no es satisfecha. Previamente estudiaremos la verosimilitud de esta hipótesis.

2. Razonabilidad de la Hipótesis de Normalidad de los Errores

El fundamento teórico para suponer que los errores son normales es el Teorema Central del Límite. Este teorema afirma, por ejemplo en el enunciado de Lindeberg, que si una variable aleatoria es suma de muchas variables independientes, ninguna de las cuales predomina, entonces puede aproximarse por una variable normal. Para un enunciado más preciso, ver por ejemplo Breiman (1968).

Es fácil imaginar, especialmente en series económicas, motivos por los cuales los supuestos del Teorema Central del Límite pueden ser violados. Muchas veces en el término u_t de la ecuación (1.1) hay una causa que predomina, por ejemplo una medida de política económica o un factor externo que influyó sobre el valor de y_t. Puede ocurrir que estos factores que distorsionan el modelo se conozcan y puedan ser considerados explícitamente como variables mudas, pero también puede suceder que se desconozcan, y por lo tanto convenga considerarlos incluidos en el término de error u_t. En este caso es muy probable que la distribución de los u_t no sea normal, y que la probabilidad de errores groseros sea mayor que la que correspondería a errores normales.

3. Robustez de un Estimador y Robustez del EMC

Esta incertidumbre sobre la distribución de los u_t, conduce a preguntarse sobre la eficiencia del EMC cuando la distribución no es normal o al menos en una vecindad de la distribución normal. La teoría de la robustez trata con esta última posibilidad, es decir estudia la eficiencia de un estimador en la vecindad de una distribución dada, en particular en una vecindad de la distribución normal. Un estimador se dirá robusto para una distribución dada F, cuando sea eficiente no sólo para esa distribución, sino también para distribuciones próximas a la misma, es decir en una vecindad de F.

Como el concepto de robustez introducido depende de la noción de vecindad de una distribución F. vamos a precisar este concepto.

Una forma de introducir una vecindad de una distribución F_0, que fue introducida por Tukey (1950), es la siguiente. Sea $\varepsilon > 0$, se definirá vecindad de la distribución F_0 con contaminación de tamaño ε, que indicaremos con $V_\varepsilon(F_0)$ como:

$V_\varepsilon(F_0)$ = {todas las distribuciones F de la forma $F = (1-\ )F_0 + \ G$, donde G es una distribución arbitraria}

En caso de que F_0 sea simétrica, puede interesar definir una vecindad de distribuciones simétricas. En este caso se define la vecindad simétrica de F_0 con contaminación de tamaño ε, que indicaremos con $V_\varepsilon^S(F_0)$ como el subconjunto de $V_\varepsilon(F_0)$ correspondiente a G simétrica.

Sea F una distribución de $V_\varepsilon(F_0)$ dada por $F = (1-\varepsilon)F_0 + \varepsilon G$, luego un error con distribución F puede interpretarse como una variable que con frecuencia $(1-\varepsilon)$ tiene distribución F_0 (podrían corresponder a los períodos normales con por ejemplo, F_0 normal), y que con frecuencia ε tiene una distribución arbitraria G (por ejemplo en los períodos anormales). Sea por ejemplo $\varepsilon = 0,1$, $F = N(0,1)$ ($N(\mu,\sigma^2)$, indica una distribución normal con media 0 y varianza σ^2), y $G = N(0,16)$. Luego si $F = (1-\varepsilon)F_0 + \varepsilon G$, es la distribución de los errores, se tendrá que el 90% de los períodos son normales y un 10% son anormales, con errores con una varianza 16 veces superior. La distribución final F resultante tendrá colas más pesadas que las de una distribución normal, es decir las colas decrecerán más rápidamente. Esto hará que los errores groseros sean más probables que los que corresponden a una distribución normal. Por ejemplo se tendrá que $\text{var}(F) = 0,9\text{x}1 + 0,1\text{x}16 = 2,5$, el error standard $\sigma = 1,58$ y $P(|u| \geq 2,57\sigma) \simeq 0,02$, contrariamente al valor 0,01 para distribuciones normales.

Estudiaremos ahora la robustez del EMC. De acuerdo a la teoría asintótica (ver por ejemplo teorema 8.2 de Theil (1971)), bajo condiciones muy generales sobre la matriz $\underset{\sim}{Z}$, la distribución del EMC es aproximadamente normal, con matriz de covarianza dada por (1.6), y por lo tanto la eficiencia de este estimador quedará determinada por esta matriz, especialmente por su traza, es decir por la suma de las varianzas de sus diferentes componentes. Como la matriz (1.6) es proporcional a var(F), ya que la matriz $\underset{\sim}{Z}\underset{\sim}{Z}'$ es constante en este modelo, de acuerdo al criterio de robustez enunciado anteriormente, para estudiar cuán robusto es el EMC en F_0, se deberá estudiar el comportamiento de var(F) en una vecindad $V_\varepsilon(F_0)$. Pero si $F = (1-\varepsilon)F_0 + \varepsilon G$, entonces $\text{var}(F) = (1-\varepsilon)\text{var}(F_0) + \varepsilon\text{var}(G)$. Como G es arbitraria, var(G) puede ser arbitrariamente grande e inclusive infinita. Por lo tanto en cualquier vecindad de F_0, $V_\varepsilon(F)$, con ε tan pequeño como se quiera hay distribuciones F con varianzas tan grandes como se quiera e inclusive infinita. Esto muestra cuán sensible es el EMC a la hipótesis de normalidad de los residuos. Una pequeña contaminación en la distribución de los residuos, aumenta su varianza enormemente y por lo tanto las varianzas de las componentes del EMC. En el ejemplo dado anteriormente de $F_0 = N(0,1)$, $\varepsilon = 0,1$ y $G = N(0,16)$, como hemos visto var(F) = 2,5. Es decir una contaminación del 10% produce un aumento de la varianza del 150%. Si G fuera una distribución de Cauchy, y por lo tanto $\text{var}(G) = \infty$ se tendría que $\text{var}(F) = \infty$. Esta falta de estabilidad del EMC en la vecindad de una distribución dada muestra que no es robusto: unas pocas observaciones con errores groseros hacen que el estimador aumente su varianza enormemente.

Un enfoque tradicional para remediar esta situación consiste en identificar las observaciones con errores groseros (outliers), utilizando los residuos, y luego tratar de entender la causa del desajuste o surpimirlas. La desventaja de esta metodología consiste en que la identificación de las observaciones con errores groseros se basa en el estudio de los residuos estimados:

$$\hat{u}_t = y_t - \underset{\sim}{z}_t'\hat{\underset{\sim}{\theta}}$$

donde $\hat{\underset{\sim}{\theta}}$ es el estimador disponible, y por lo tanto para que estos residuos estén bien estimados, se deberá tener que $\hat{\underset{\sim}{\theta}}$ debe ser un buen estimador de $\underset{\sim}{\theta}$. En caso que $\hat{\underset{\sim}{\theta}}$ sea el EMC, esta última condición no se cumple, y por lo tanto la identificación de los errores groseros puede ser problemática. Luego inclusive para identificar las observaciones con errores groseros en forma eficiente es preciso contar con un estimador robusto.

4. Estimadores Robustos de Regresión: M-Estimadores

Supongamos que la distribución de los errores F fuese conocida, luego en este caso se podría utilizar para estimar $\underset{\sim}{\theta}$ el estimador de máxima verosimilitud (EMV), del cual se conoce que en general, salvo casos patológicos muy raros, es asintóticamente óptimo. Supongamos además que F tiene derivada f. Luego la función de verosimilitud de los u_t, $1 \leq t \leq T$ estará dada por:

$$L(u_1, \ldots, u_T, \underset{\sim}{\theta}) = \Pi_{t=1}^{T} f(u_t)$$

y como $u_t = y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta}$, la función de verosimilitud de las observaciones y_t está dada por:

$$L(y_1, \ldots, y_T, \underset{\sim}{\theta}) = \Pi_{t=1}^{T} f(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})$$

Tomando logaritmos naturales se tendrá:

$$\ln L(y_1, \ldots, y_T, \underset{\sim}{\theta}) = \sum_{t=1}^{T} \ln f(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})$$

Luego si llamamos:

$$\rho_f(u) = -\ln f(u)$$

el EMV será el valor de $\underset{\sim}{\theta}$ que minimiza:

$$(4.1) \quad \sum_{t=1}^{T} \rho_f(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})$$

En el caso de que f sea la densidad correspondiente a una distribución $N(0,\sigma^2)$, se tendrá $\rho_f(u) = u^2/2\sigma^2$ y el EMV será el EMC.

Derivando (4.1) respecto a cada componente θ_k de $\underset{\sim}{\theta}$, se tendrá que el EMV deberá satisfacer:

$$(4.2) \quad \sum_{t=1}^{T} \psi_f(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta}) z_{kt} = 0 \qquad k = 1, \ldots, K$$

o en forma vectorial:

$$(4.3) \quad \sum_{t=1}^{T} \psi_f(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta}) \underset{\sim}{z}_t = \underset{\sim}{0}$$

donde $\psi_f = \rho_f'$. En el caso particular de que f sea la densidad correspondiente a una $N(0,\sigma^2)$, $\psi_f = u/\sigma^2$, y se obtienen las ecuaciones normales del EMC:

$$\sum_{t=1}^{T} (y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta}) z_{kt} = 0 \qquad k = 1, \ldots, K$$

Sin embargo, el EMV, que se obtiene minimizando (4.1) o resolviendo (4.2), es de poca aplicabilidad, ya que la distribución verdadera F y por lo tanto también f son generalmente desconocidas. De todos modos (4.1) y (4.3) sugieren una forma de construir estimadores robustos: elegir la función ρ o su derivada ψ independiente de F, pero de manera que el estimador resultante sea robusto en un entorno de una distribución F dada. De acuerdo a esta idea se definen estimadores $\hat{\underset{\sim}{\theta}}$ tales que minimicen:

(4.4) $\sum_{t=1}^{T} \rho(y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta})$

o como solución del sistema:

(4.5) $\sum_{t=1}^{T} \psi(y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta}) \underset{\sim}{z}_t = \underset{\sim}{0}$

Estos estimadores que se denominan estimadores del tipo de máxima verosimilitud (M-estimadores), fueron propuestos por Huber (1964) para estimar un parámetro de locación, y luego extendidos para el caso de regresión por Relles (1968), Huber (1972), Yohai (1974), Yohai y Maronna (1979), Klein y Yohai (1979).

Se puede mostrar, ver por ejemplo Yohai y Maronna (1979), que si se cumple:

(4.6) $E_F(\psi(u_t)) = 0$

entonces bajo condiciones muy generales, la solución $\hat{\theta}$ de (4.5) es asintóticamente normal multivariada con media $\underset{\sim}{\theta}$ y matriz de covarianza:

(4.7) $V(\psi,F)(\underset{\sim}{Z}\underset{\sim}{Z}')^{-1}$

donde:

(4.8) $V(\psi,F) = E_F(\psi^2(u_t))/(E_F(\psi'(u_t)))^2$.

Luego la matriz de covarianza asintótica de un M-estimador será proporcional a $V(\psi,F)$. Por lo tanto, si se quiere que un M-estimador sea robusto en un entorno de una distribución F_0 simétrica, ψ deberá elegirse de manera que $V(\psi,F)$ sea estable en $V_\varepsilon^S(F)$. Una forma de medir la estabilidad de $V(\psi,F)$ en ese entorno está dada por:

$$\overline{V}(\psi,F_0) = \max_{F\varepsilon V_\varepsilon^s(F_0)} V(\psi,F)$$

donde max indica valor máximo (en este caso en el entorno $V_\varepsilon^S(F_0)$).

Una forma de elegir ε será por lo tanto minimizando $\overline{V}_\varepsilon(\psi,F_0)$. Huber (1964) demuestra que existe ψ^* dependiendo de F_0 y ε, tal que $\overline{V}_\varepsilon(\psi^*, F_0)$ es mínima. Para el caso en que F_0 sea la distribución $N(0,\sigma^2)$, se puede mostrar que la función ψ^* óptima es la de la forma:

(4.9) $\psi^*(u) = \psi_m(u/\sigma)$

donde:

(4.10) $$\psi_m(u) = \begin{cases} u & \text{si } |u| \leq m \\ -k & \text{si } u < -m \\ k & \text{si } u > m \end{cases}$$

y donde el valor m depende del grado de contaminación ε a través de una expresión complicada. La función ψ^* dada por (4.9) corresponde a una función ρ^* dada por:

(4.11) $\rho^*(u) = \sigma^2\rho_m(u/\sigma)$

con:

(4.12) $$\rho_m(u) = \begin{cases} u^2/2 & \text{si } |u| \leq m \\ (3/2)k^2 - ku & \text{si } u < -m \\ -k^2/2 + ku & \text{si } u > m \end{cases}$$

Esta función, cuya forma puede verse en figura 4.1, es cuadrá-

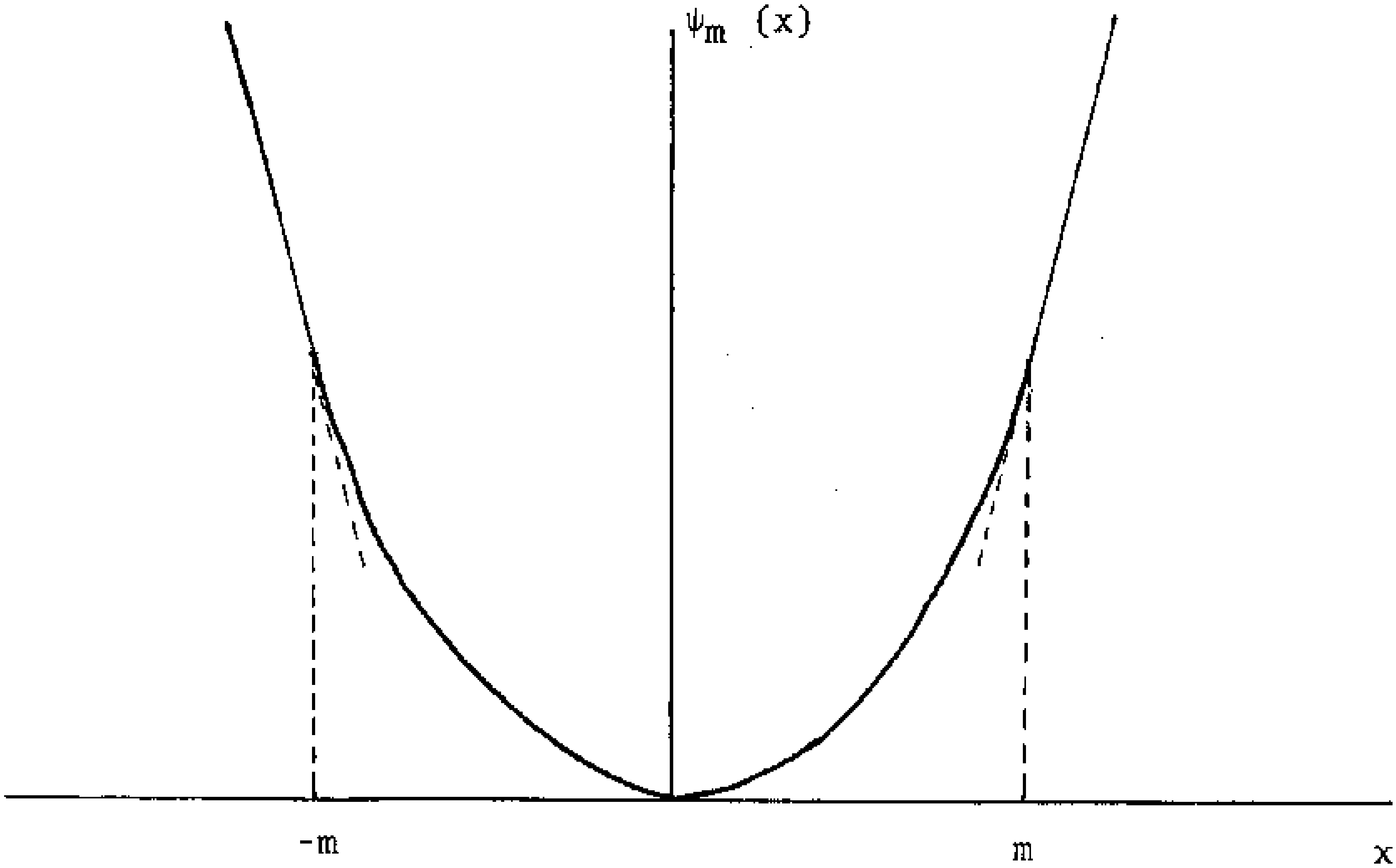

Fig. 4.1.

tica para $|u| \leq m$ y lineal para $|u| > m$, y por lo tanto se puede decir que el estimador asociado a ψ_m da menos peso a los errores groseros que el EMC. Si $\varepsilon \to 1$, entonces $m = 0$ y el estimador correspondiente es el estimador de mínimos valores absolutos (EMVA). Este último estimador está definido como el valor de θ que minimiza:

$$(4.13) \quad \sum_{t=1}^{T} |y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta}|$$

Cuando ψ está dada por (4.9) y (4.10), el sistema (4.5) se transforma en:

$$(4.14) \quad \sum_{t=1}^{T} \psi_m((y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta})/\sigma) \underset{\sim}{z}_t = \underset{\sim}{0}$$

El hecho de que σ aparezca en el denominador dentro de ψ_m hace que el estimador resultante sea invariante respecto del sistema de unidades en que fue medido. Es decir, si todas las observaciones son multiplicadas por un mismo factor λ el estimador $\hat{\underset{\sim}{\theta}}$ también será multiplicado por λ.

Para el caso general de M-estimadores dados por (4.5), si se quiere que el estimador sea invariante, ese sistema deberá ser reemplazado por:

$$(4.15) \qquad \sum_{t=1}^{T} \psi((y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}_t = \underset{\sim}{0}$$

donde σ es un parámetro de dispersión. En general σ no es conocido y debe ser estimado. Para estimar σ hay 2 alternativas: (a) estimar σ a partir de un estimador inicial de $\underset{\sim}{\theta}$ invariante; (b) estimar $\underset{\sim}{\theta}$ y σ simultáneamente.

En el caso (a), a partir de un estimador inicial invariante $\underset{\sim}{\theta}^{(0)}$ de $\underset{\sim}{\theta}$, se estiman los residuos $\hat{u}_t$ por:

$$\hat{u}_t = y_t - \underset{\sim}{z}_t'\hat{\underset{\sim}{\theta}}^{(0)} \qquad\qquad t = 1, \ldots, T$$

y utilizando estos residuos se estima el parámetro de dispersión σ. Algunos estimadores propuestos son:

$\hat{\sigma}_1 = (\sum_{t=1}^{T} \hat{u}_t^2 /(T-K))^{1/2}$ (desviación standard muestral)

$\hat{\sigma}_2 = \text{mediana}\ (|u_t|)/\phi^{-1}(3/4)$, con $\phi(x)$ la función de distribución de una N(0,1). (Mediana muestral de los valores absolutos normalizados)

La constante de normalización se elige de manera que si la distribución de los u_t es N(0,1), entonces $\hat{\sigma}_2 \to 1$.

$\hat{\sigma}_3 = (\hat{u}([0{,}75T]) - \hat{u}([0{,}25T])/(2\phi^{-1}(3/4))$, (intervalo intercuartil muestral normalizado),

donde $\hat{u}(1) < \hat{u}(2) < \ldots < \hat{u}(T)$ son los valores $\hat{u}_t$ ordenados y [ ] indica parte entera. La constante de normalización fue elegida de manera que si los u_t son N(0,1), entonces $\hat{\sigma}_3 \to 1$.

De estos 3 estimadores, el primero, $\hat{\sigma}_1$, no es robusto, ya que es muy sensible a errores groseros. Por lo tanto se aconseja utilizar $\hat{\sigma}_2$ ó $\hat{\sigma}_3$.

En cuanto al estimador inicial $\hat{\theta}^{(0)}$, deberá ser como se dijo un estimador invariante de $\underset{\sim}{\theta}$. Dentro de los M-estimadores, los únicos invariantes que no dependen de σ están dadas por la función ρ siguiente:

$$(4.16) \quad \rho(u) = |u|^p$$

De estos estimadores, los más utilizados son los correspondientes a p = 2 y a p = 1, es decir el EMC y el EMVA respectivamente.

El EMC es más fácil de calcular, pero como hemos visto es muy sensible a errores groseros. El EMVA por otra parte es mucho menos sensible a la presencia de errores groseros, y por lo tanto se recomienda como estimador inicial. Un algoritmo para su cómputo puede verse por ejemplo en Amstrong y Frome (1976).

A pesar de que el EMVA es poco sensible a errores groseros, no se lo aconseja utilizar como estimador final de $\underset{\sim}{\theta}$ ya que es poco eficiente para el caso normal. Si la distribución de los u_t es N(0,σ^2), el valor de V(ψ,F) para el EMC es σ^2 y para el EMVA es $\frac{\pi}{2}\sigma^2 \simeq 1{,}57\,\sigma^2$, es decir resulta 57% menos eficiente que el EMC.

En el caso (b) de estimación simultánea de $\underset{\sim}{\theta}$ y σ se debe agregar al sistema (4.15) una ecuación más. Huber (1972) propuso que la ecuación suplementaria sea de la forma:

$$(4.17) \quad \sum_{t=1}^{T} X^2((y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta})/\sigma)/(T-K) = \beta$$

con:

$$\beta = E(X^2(u))$$

donde u es una variable con distribución N(0,1), y X una función monótona no decreciente e impar, por ejemplo en la familia dada por (4.10)

5. Interpretación de los M-Estimadores como Estimadores de Mínimos Cuadrados Ponderados Iterados. Procedimiento de cálculo

Consideremos la ecuación (4.15) que define los M-estimadores invariantes y supongamos que σ es estimado por $\hat{\sigma}$. Es fácil ver que esta ecuación también puede escribirse como:

$$(5.1) \quad \sum_{t=1}^{T} (y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta}) \underset{\sim}{z}_t w_t = \underset{\sim}{0}$$

con

$$(5.2) \quad w_t = w((y_t - \underset{\sim}{z}_t' \underset{\sim}{\theta})/\hat{\sigma}) \qquad y \quad w(u) = \psi(u)/u$$

Luego resolviendo (5.1) el M-estimador $\hat{\underset{\sim}{\theta}}$ se podrá escribir como:

$$(5.3) \quad \hat{\underset{\sim}{\theta}} = \left(\sum_{t=1}^{T} w_t \underset{\sim}{z}_t \underset{\sim}{z}_t' \right)^{-1} \sum_{t=1}^{T} w_t y_t \underset{\sim}{z}_t$$

Este estimador corresponde a la estimación por mínimos cuadrados ponderados con pesos w_t, es decir al valor que minimiza:

$$\sum_{t=1}^{T} (y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})^2 w_t$$

La ecuación (5.3) no permite el cálculo directo de $\hat{\underset{\sim}{\theta}}$, ya que los valores w_t a su vez dependen de $\hat{\underset{\sim}{\theta}}$. Sin embargo, el sistema (5.3) sugiere el siguiente método iterativo para la resolución de (4.15). Se parte de un estimador inicial $\hat{\underset{\sim}{\theta}}^{(0)}$ que puede ser el EMC o el EMVA, y se calculan los pesos $w_t^{(0)}$ por $w((y_t - \underset{\sim}{z}_t'\hat{\underset{\sim}{\theta}}^{(0)})/\hat{\sigma})$, luego se calcula un nuevo estimador:

$$(5.4)\quad \hat{\underset{\sim}{\theta}}^{(1)} = \left(\sum_{t=1}^{T} w_t^{(0)} \underset{\sim}{z}_t \underset{\sim}{z}_t'\right)^{-1} \sum_{t=1}^{T} w_t^{(0)} y_t \underset{\sim}{z}_t$$

Con este valor $\hat{\underset{\sim}{\theta}}^{(1)}$ se calculan nuevos pesos $w^{(1)} = w((y_t - z_t'\hat{\underset{\sim}{\theta}}^{(1)})/\hat{\sigma})$ y se computa:

$$\hat{\theta}^{(2)} = \left(\sum_{t=1}^{T} w_t^{(1)} \underset{\sim}{z}_t \underset{\sim}{z}_t'\right)^{-1} \sum_{t=1}^{T} w_t^{(1)} y_t \underset{\sim}{z}_t$$

Iterando este procedimiento se construye una sucesión de estimadores $\hat{\underset{\sim}{\theta}}^{(j)}$ que en general va a converger a un vector solución de (4.15). Condiciones para la convergencia de este procedimiento pueden encontrarse en Klein y Yohai (1979). El valor $\hat{\sigma}$ puede reemplazarse por $\hat{\sigma}_2$ ó $\hat{\sigma}_3$ descriptos en la Sección 4, basados en el estimador inicial $\hat{\underset{\sim}{\theta}}^{(0)}$. Si se quiere estimar simultáneamente $\underset{\sim}{\theta}$ y σ, se calcula un estimador inicial de σ que llamaremos $\hat{\sigma}^{(0)}$ a partir del estimador inicial $\hat{\underset{\sim}{\theta}}^{(0)}$. Luego con este valor se calcula $w_t^{(0)} = w((y_t - \underset{\sim}{z}'\hat{\underset{\sim}{\theta}}^{(0)})/\hat{\sigma}^{(0)})$. Luego se calcula $\hat{\underset{\sim}{\theta}}^{(1)}$ por (5.4). El proceso iterativo continúa ahora calculando un nuevo estimador de σ que llamaremos $\sigma^{(1)}$. Para esto se utiliza la (4.17) que también puede escribirse:

$$\hat{\sigma}^2 = \hat{\sigma}^2 \sum_{t=1}^{T} X^2((y_t - \underset{\sim}{z}'\hat{\underset{\sim}{\theta}})/\hat{\sigma})/((T-K)\beta)$$

Reemplazando en el segundo miembro $\hat{\underset{\sim}{\theta}}$ por $\hat{\underset{\sim}{\theta}}^{(1)}$ y $\hat{\sigma}$ por $\hat{\sigma}^{(0)}$ obtenemos $\hat{\sigma}^{(1)}$

$$\hat{\sigma}^{(1)} = \hat{\sigma}^{(0)} [\sum_{t=1}^{T} X^2((y_t - \underset{\sim}{z}_t' \hat{\underset{\sim}{\theta}}^{(1)})/\hat{\sigma}^{(0)})/((T-K)\beta)]^{1/2}$$

Usando $\sigma^{(1)}$ y $\hat{\underset{\sim}{\theta}}^{(1)}$ se calculan los $w_t^{(1)}$ y así se itera hasta que en 2 iteraciones sucesivas se obtengan diferencias despreciables.

Si se usa ψ como una función en la familia ψ_m dada por (4.10) la función de pesos dada por (5.2) estará dada por:

$$w_t = \begin{cases} 1 & \text{si } |y_t - \underset{\sim}{z}_t'\hat{\underset{\sim}{\theta}}| \leq m\hat{\sigma} \\ m\hat{\sigma}/|y_t - \underset{\sim}{z}'\hat{\underset{\sim}{\theta}}| & \text{si } |y_t - \underset{\sim}{z}'\hat{\underset{\sim}{\theta}}| > m\hat{\sigma} \end{cases}$$

Es decir los pesos son constantemente iguales a 1 si los errores estimados son menores o iguales que $m\hat{\sigma}$, luego comienzan a decrecer, tendiendo a 0 cuando estos errores tienden a infinito. Ver figura 5.1.

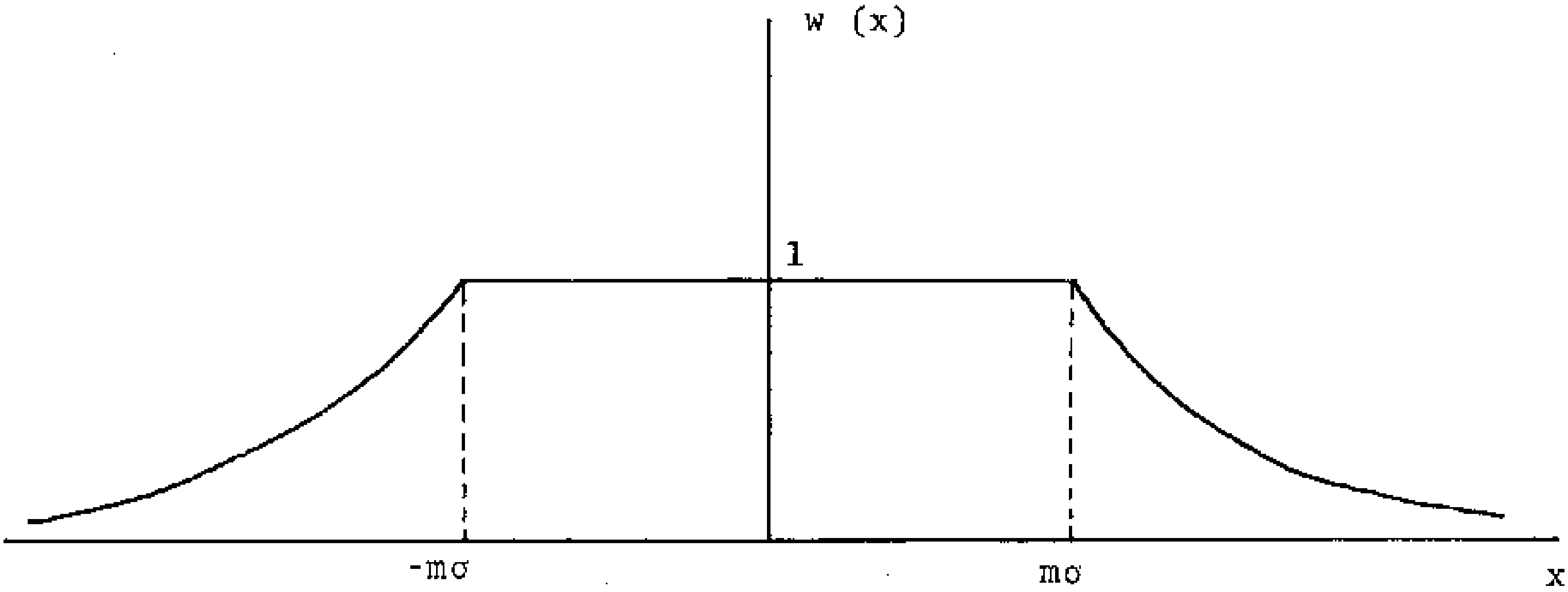

Figura 5.1.

En el ejemplo dado en la introducción se utiliza el estimador basado en la función ψ_m de Huber con m = 1,345 (este valor se justificará en la Sección 8) y la estimación simultánea del parámetro de escala σ. El estimador de σ, resultó $\hat{\sigma}$ = 0,08 que es completamente diferente al valor 0,25 obtenido con EMC. El algoritmo de cómputo utilizado fue el descripto anteriormente y la convergencia se obtuvo en 4 iteraciones. El sistema de pesos finales fue el siguiente:

Año / Tri	1971	1972	1973	1974	1975	1976	1977	1978
1		1,00	1,00	0,65	1,00	0,10	1,00	1,00
2	1,00	1,00	1,00	1,00	0,17	1,00	1,00	1,00
3	1,00	1,00	0,52	1,00	0,39	1,00	1,00	1,00
4	1,00	1,00	0,72	1,00	1,00	1,00	1,00	1,00

Como se ve en este cuadro, las observaciones con menos peso son el primer cuatrimestre de 1976 (0,10) y segundo cuatrimestre de 1975 (0,17), que correspondían a los períodos con mayor residuo luego del ajuste con el EMC. Sin embargo este cuadro muestra que hay otras observaciones con ponderación reducida por ejemplo el tercer cuatrimestre de 1975 (0,39), tercero de 1973 (0,52), primero de 1978 (0,65) y cuarto de 1973 (0,72). Estas observaciones no resultaban sospechosas cuando se utilizan los residuos del EMC, pero en cada iteración que se realiza la mejora en los estimadores de los coeficientes induce una mejora en los estimadores de los residuos permitiendo de esta forma descubrir nuevas observaciones sospechosas.

6. Familias de Funciones ψ no Monótonas

Ya hemos visto en la Sección 4 que de acuerdo al criterio de minimizar la máxima varianza en una vecindad de la distribución normal, la elección de la función debería hacerse en la familia ψ_m dada por (4.10). Sin embargo en muchos casos, ni siquiera podemos estar seguros que la distribución F de los errores u_t están en una vecindad de la normal, es decir en principio la función F puede ser totalmente arbitraria, por ejemplo una distribución de Cauchy. Para estos casos se han propuesto otras clases de funciones ψ, cuyos estimadores se comportan aún más establemente que los correspondientes a los ψ_m cuando la indeterminación de la F es total. La idea básica es que como puede esperarse que los errores tengan una distribución F con colas muy pesadas (alta probabilidad de errores groseros), directamente darle peso 0 a las observaciones muy sospechosas, es decir a aquellas que aparentan errores groseros. Por lo tanto de acuerdo a (5.2), la función $\psi(u)$ debe anularse para u suficientemente grande. Estas funciones ψ no son monótonas (contrariamente a lo que ocurría con las ψ_m), y se denominan redescendientes. Algunas familias de funciones redescendientes propuestas son las siguientes.

(a) Familia de Hampel. Fue introducida por Hampel (1968). Depende de 3 parámetros reales no negativos a, b, c. La función ψ viene dada por:

$$\psi_{a,b,c}(u) = \begin{cases} u & \text{si } |u| \leq a \\ a\,sg(u) & \text{si } a < |u| \leq b \\ a(c-|u|)sg(u)/(c-b) & \text{si } b < |u| \leq c \\ 0 & \text{si } |u| > c \end{cases}$$

La función $sg(u)$ vale 1 si $u > 0$, 0 si $u = 0$ y -1 si $u < 0$

La forma de estas funciones puede verse en la figura 6.1.

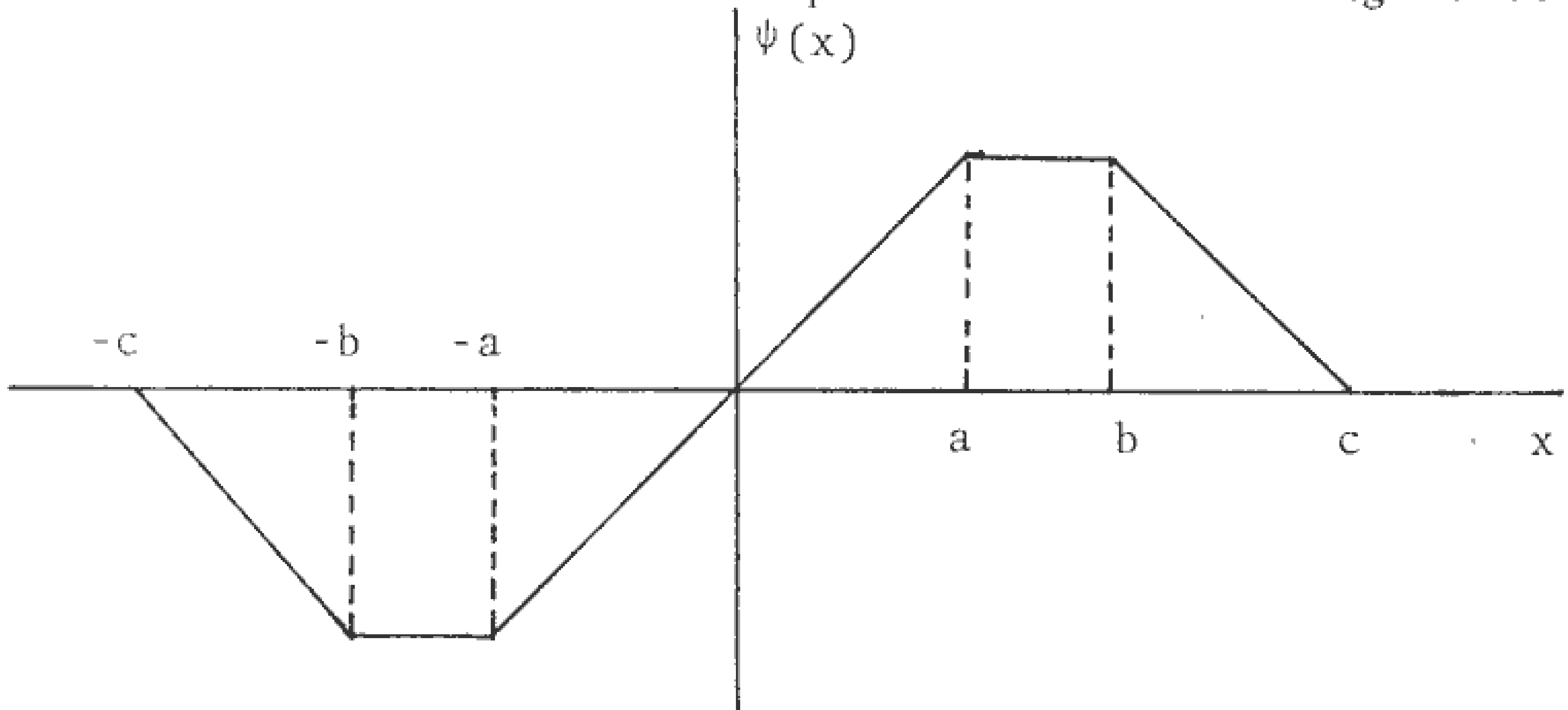

Fig. 6.1.

La forma de la función de peso correspondiente $w(u) = \psi(u)/u$ puede verse en la figura 6.2

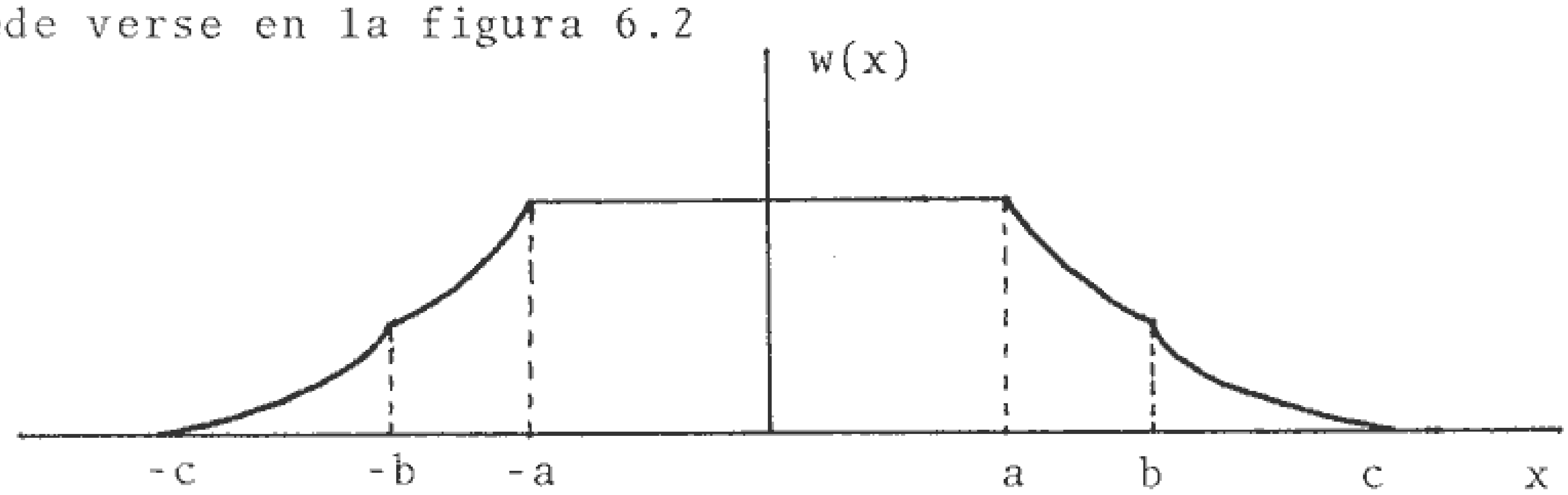

Fig. 6.2.

(b) Familia seno. Fue estudiada por Andrews (1974). Depende de una constante k positiva. La función ψ viene dada por:

$$\psi_k(u) = \begin{cases} \operatorname{sen}(u/k) & \text{si } |u| \leq k\pi \\ 0 & \text{si } |u| > k\pi \end{cases}$$

La función de peso w(u) correspondiente se puede ver en la figura 6.3.

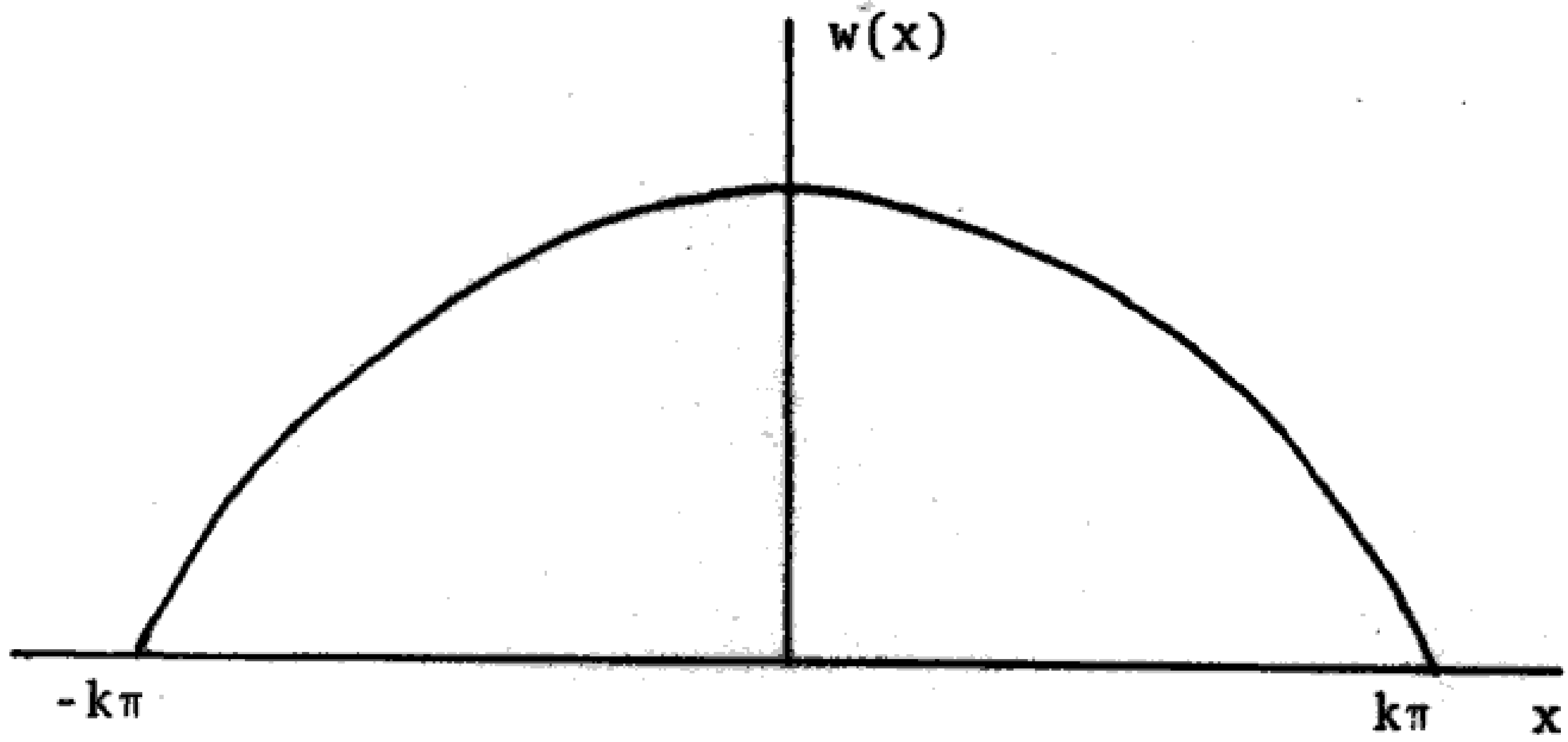

Fig. 6.3.

(c) Función bicuadrada. Fue introducida por Beaton y Tukey (1974). Depende de una constante k positiva. La función ψ está dada por:

$$\psi_k(u) = \begin{cases} u(1-(u/k)^2)^2 & \text{si } |u| \leq k \\ 0 & \text{si } |u| > k \end{cases}$$

La forma de la función de peso correspondiente puede verse en la figura 6.4.

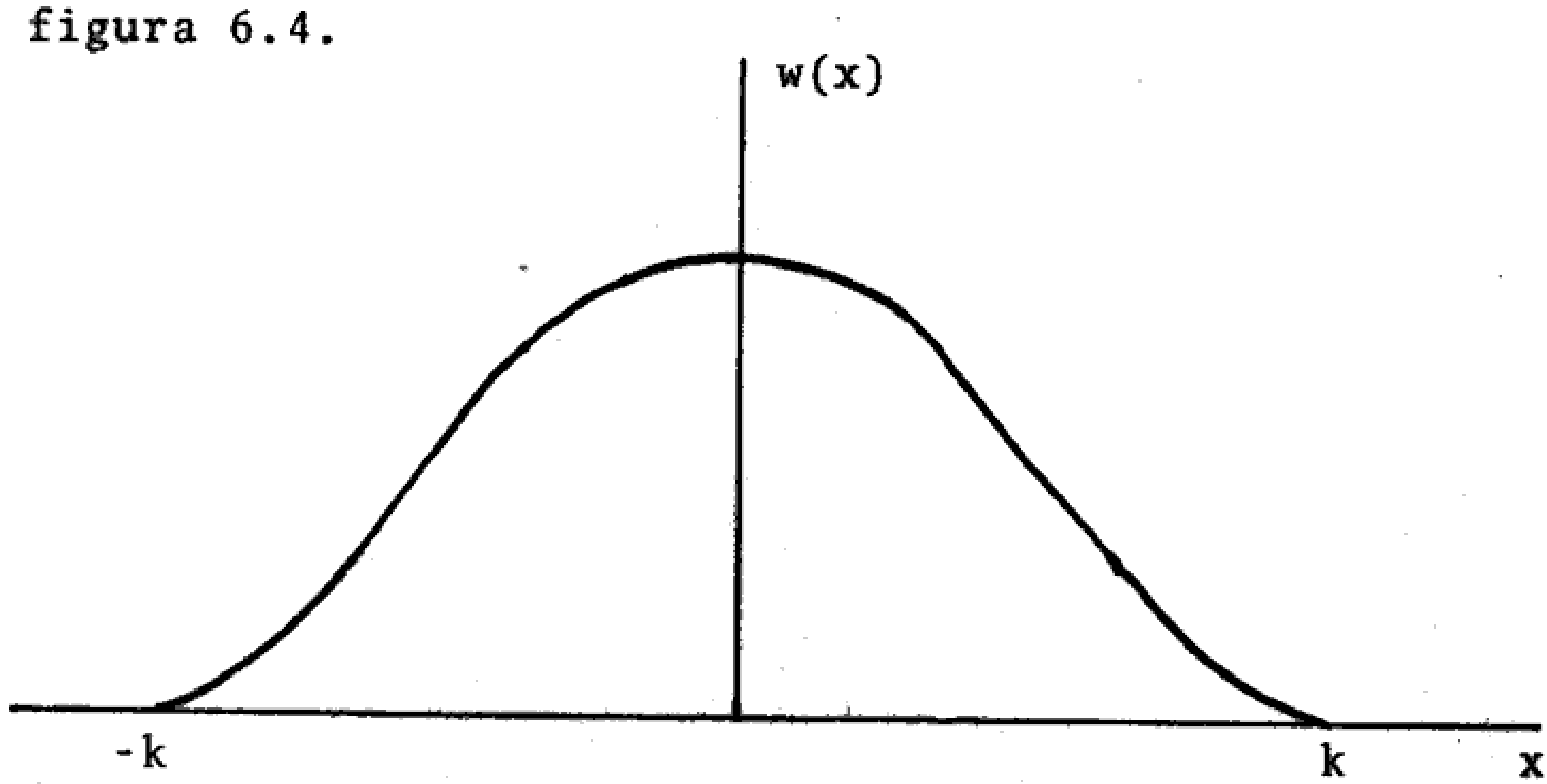

Fig. 6.4.

Otras familias de funciones monótonas y resdescendientes para la elección de la función ψ pueden encontrarse en Holland y Welch

(1977). Para el estudio de las propiedades asintóticas de los M-estimadores basados en funciones ψ redescendientes puede verse Klein y Yohai (1979).

Una forma de elegir los parámetros de estas curvas, es utilizando los valores que dan un valor $V(\psi,\phi) = 1{,}05$, donde ϕ es la función de distribución de una $N(0,1)$. Esto significa que los estimadores correspondientes tienen una varianza 5% mayor que la del EMC cuando los errores son normales. Este pequeño aumento de la varianza de estos estimadores para el caso normal va a ir acompañado por disminuciones mucho más importantes cuando haya contaminación con errores groseros. En la sección 7 se da un cuadro con el comportamiento de diferentes estimadores robustos y el EMC para diferentes distribuciones F.

7. Estimación de la matriz de covarianza de los M-estimadores

La matriz de covarianza asintótica de un M-estimador definido por (4.5) está dada por la fórmula (4.7). Si se usa un estimador invariante dado por (4.15) donde σ está reemplazado por un estimador consistente $\hat{\sigma}$, la fórmula (4.7) seguirá valiendo si se reemplaza la función $\psi(u)$ por $\psi_\sigma(u) = \psi(u/\sigma)$.

Como $\psi_\sigma'(u) = \psi'(u/\sigma)/\sigma$, la fórmula (4.7) se transforma en:

(7.1) $V(\psi_\sigma,F)(\underset{\sim}{Z}\underset{\sim}{Z}')^{-1}$

donde

(7.2) $V(\psi_\sigma,F) = \sigma^2 E_F(\psi^2(u_t/\sigma))/(E_F\psi'(u_t/\sigma))$

y de acuerdo a (1.6), la eficiencia del M-estimador con respecto a EMC está dado por:

(7.3) $e(\psi,F) = \mathrm{var}(F)/V(\psi_\sigma,F)$

Luego para estimar (7.1) y (7.3), basta estimar $V(\psi_\sigma,F)$ y var(F).

En la estimación de $V(\psi_\sigma,F)$, σ^2 se estima por $\hat{\sigma}^2$, que como se explicó en la Sección 4 se puede estimar por separado o simultáneamente con $\underset{\sim}{\theta}$. $E\,(\psi^2(u_t/\sigma))$ se estima por:

$$\sum_{t=1}^{T} \psi^2(\hat{u}_t/\hat{\sigma})/(T-K)$$

donde los errores estimados $\hat{u}_t$ están dados por:

$$\hat{u}_t = y_t - \underset{\sim}{z}'\underset{\sim}{\hat{\theta}}$$

Del mismo modo $E_F(\psi'(u_t/\sigma)$ se estima por:

$$\sum_{t=1}^{T} \psi'(\hat{u}_t/\hat{\sigma})/(T-K)$$

Finalmente var(F) se puede estimar por:

$$\sum_{t=1}^{T} \hat{u}_t^2 \;/(T-K)$$

8. Comparaciones Numéricas de los Diferentes Estimadores

Como hemos visto en la fórmula (7.1), la matriz de covarianza asintótica de los M-estimadores invariantes es proporcional a $V(\psi_\sigma,F)$, y por lo tanto este parámetro nos permitirá compara la eficiencia de los diferentes estimadores. En la tabla siguiente se comparan 5 estimadores: el EMC, el EMVA, el M-estimador de Huber (EHU), el M-estimador seno de Andrews (EAN), y el M-estimador bicuadrático de Beaton-Tukey (EBT). Los parámetros de los 3 estimadores últimos fueron elegidos de manera que la pérdida de eficiencia fuese del 5%

en el caso de distribuciones normales. Los valores fueron calculados suponiendo que el estimador de escala utilizado es $\hat{\sigma}_1$ ó $\hat{\sigma}_2$ descriptos en la Sección 4. Las distribuciones consideradas son N(0,1); 0,9N(0,1) + 0,1N(0,9); 0,9N(0,1) + 0,1N(0,25); 0,9N(0,1) + 0,1N(0,100) y de Cauchy. Esta última tiene densidad:

$$f(x) = 1/(\pi(1+x^2))$$

y tiene varianza infinita. En la tabla también se registra el valor del parámetro que corresponde a la función elegida dentro de la familia respectiva.

Valores de $V(\psi_\sigma, F)$

	N(0,1)	0,9N(0,1) 0,1N(0,9)	0,9N(0,1)+ 0,1N(0,25)	0,9N(0,1)+ 0,1N(0,100)	CAUCHY	Parámetro
EMC	1,00	1,80	3,40	10,9	∞	
EMVA	1,57	1,80	1,85	1,90	2,47	
EHU	1,05	1,30	1,41	1,45	3,52	1,345
EAN	1,05	1,27	1,27	1,22	2,79	1,339
EBT	1,05	1,27	1,27	1,22	2,79	4,685

De esta tabla se desprende que al costo de aumentar sólo un 5% la varianza en el caso normal, se obtienen disminuciones enormes cuando hay presentes un 10% de observaciones groseras si se utilizan los EHU, EAN ó EBT en vez del EMC. Los resultados muestran que las performances del EAN y el EBT son similares con la precisión de 2 decimales. El EMVA se comporta muy bien frente a distribuciones como la de Cauchy, pero es ineficiente en el caso normal donde la varianza aumenta un 57%. Un resultado sorprendente es que el EAN y el EBT se comportan mejor cuando se contamina con una N(0,100) que cuando se contamina con una N(0,9). Sin embargo esto puede explicarse por el carácter no monótono de las funciones ψ correspondientes.

Para realizar comparaciones entre los diferentes estimadores para muestras pequeñas se debe utilizar el método de Monte Carlo. Los estudios realizados, ver por ejemplo Andrews y colaboradores (1972) muestran que se obtienen resultados que son en líneas generales similares a los asintóticos.

9. Función de Influencia

Los M-estimadores descriptos en las secciones precedentes se desarrollaron suponiendo las siguientes hipótesis.

(a) Las variables independientes dadas en los vectores $\underset{\sim}{z}_t$ eran fijas y no aleatorias.

(b) Las variables independientes no estaban sujetas a errores groseros. Los únicos errores groseros posibles afectan a las u_t.

(c) El modelo (1.1) siempre se cumplía, es decir no se consideraba la posibilidad de períodos atípicos donde el modelo lineal dejase de cumplirse.

En las series económicas las hipótesis (a), (b) y (c) son violadas frecuentemente. Por un lado las variables independientes pueden ser aleatorias y observadas con errores groseros, y además puede haber períodos atípicos donde el modelo lineal dejase de cumplirse. Vamos a ver que en estos casos, aún los M-estimadores definidos en

Sección 4 pueden ser muy inestables. Esto querrá decir que unas pocas observaciones atípicas pueden afectar en forma apreciable la estimación de los coeficientes de regresión. Por lo tanto los M-estimadores dejarán de ser robustos y se deberán buscar nuevos estimadores.

Para estudiar la estabilidad de los M-estimadores en presencia de observaciones atípicas determinaremos cómo la presencia de una pequeña proporción de éstas puede modificar el estimador correspondiente.

Supongamos que se tengan (T-1) observaciones $(y_t', \underset{\sim}{z}_t)\ 1 \le t \le T-1$, y supongamos que se agregue una nueva observación $(y^*, \underset{\sim}{z}^*)$. Supongamos que se utilice el M-estimador dado por (4.5), llamemos $\hat{\underset{\sim}{\theta}}$ al estimador obtenido con las primeras (T-1) observaciones y $\hat{\underset{\sim}{\theta}}^*$ el obtenido cuando se agrega la nueva observación. Luego se puede mostrar, ver Krasker y Welsch (1978) que vale la siguiente fórmula aproximada para T grande:

$$(9.1)\quad T(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}^*) = (-1/(T-1)) \sum_{t=1}^{T} \psi'((y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}_t\underset{\sim}{z}_t')^{-1}\sigma\psi((y^* - \underset{\sim}{z}^*\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}^*$$

Si consideramos la observación $(y^*, \underset{\sim}{z}^*)$ como la única observación espuria que no satisface el modelo, la contaminación de observaciones espurias será en la proporción de 1 en T. Luego el primer miembro de (9.1) será la variación del estimador $\hat{\underset{\sim}{\theta}}$ por unidad de contaminación, ya que es igual a $(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}^*)/(1/T)$.

Supongamos ahora que:

$$(9.2)\quad \lim_{T\to\infty} - \sum_{t=1}^{T} \psi'(y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}_t\underset{\sim}{z}_t'/(T-1) \to \underset{\sim}{\Lambda}$$

donde $\underset{\sim}{\Lambda}$ es una matriz definida positiva. En el caso de que las $\underset{\sim}{z}_t$ son aleatorias se tendrá:

$$(9.3) \quad \underset{\sim}{\Lambda} = E(\psi'((y_t - \underset{\sim}{z}_t'\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}_t\underset{\sim}{z}_t')$$

Luego de (1.1), (1.2) y (1.3) resulta que para T grande la variación por unidad de contaminación en el punto $(y^*,\underset{\sim}{z}^*)$ está dado por:

$$(9.4) \quad \lim_{T\to\infty}(\hat{\underset{\sim}{\theta}} - \hat{\underset{\sim}{\theta}}^*)/(1/T) = \underset{\sim}{\Lambda}^{-1}\sigma\psi((y^* - \underset{\sim}{z}^{*\prime}\underset{\sim}{\theta})/\sigma)\underset{\sim}{z}^*$$

Esta expresión que indicamos por INF$(y^*,\underset{\sim}{z}^*)$ y que depende del valor espurio con el que se contamina, se denomina función de influencia del M-estimador y fue introducida por Hampel (1968) como indicador de la robustez de un estimador. Como se ve en la expresión (9.4) la función de influencia puede ser arbitrariamente grande, aún en el caso de que ψ sea acotada, si permitimos que z pueda ser arbitrariamente grande. Esto es bien conocido en el caso del EMC, donde una observación correspondiente a valores altos de las variables independientes puede tener una influencia desmesurada en la estimación de los coeficientes de regresión. De acuerdo a la fórmula (9.4) esto continúa ocurriendo con los M-estimadores. Para solucionar este problema en la sección siguiente se introducirá una nueva clase de estimadores: los M-estimadores generalizados (GM-estimadores).

10. M-Estimadores Generalizados

En la Sección 5 hemos visto que el M-estimador dado por (4.15) puede escribirse como un estimador de mínimos cuadrados ponderados dado por (5.1) En este caso de acuerdo a (5.2) los pesos w_t pueden escribirse como:

(10.1) $w_t = w(\hat{u}_t/\hat{\sigma})$

con:

(10.2) $\hat{u}_t = y_t - \underset{\sim}{z}_t' \underset{\sim}{\hat{\theta}}$

De acuerdo a lo visto en Sección 9, estos estimadores pueden dar un peso excesivo a las observaciones con z_t grande. Por lo tanto parece razonable generalizar (10.1) haciendo depender w_t de $\hat{u}_t/\hat{\sigma}$ y de $\underset{\sim}{z}_t$, penalizando las observaciones con z_t grande. Luego podemos considerar estimadores dados por (5.1) pero con w_t de la forma:

(10.3) $w_t = w(\hat{u}_t/\hat{\sigma}, ||\underset{\sim}{z}_t||)$

donde $||\underset{\sim}{z}_t||$ indica la norma del vector $\underset{\sim}{z}_t$ dada por:

$$||\underset{\sim}{z}_t|| = \left(\sum_{k=1}^{K} z_{kt}^2\right)^{1/2}$$

y donde w es una función monótona no creciente en los 2 argumentos. Si se quiere que estos estimadores sean invariantes por una transformación lineal de las variables independientes, se deberá normalizar los $\underset{\sim}{z}_t$ en (10.3) por una matriz $\underset{\sim}{M}$ de KxK. Los vectores normalizados serán los $\underset{\sim}{\hat{M}} z_t$ y los pesos en (10.3) deberán reemplazarse por:

(10.4) $w_t = w(\hat{u}_t/\hat{\sigma}, ||\underset{\sim}{\hat{M}}\underset{\sim}{z}_t||)$

La matriz $\underset{\sim}{\hat{M}}$ deberá ser tal que si se hace una transformación lineal de los $\underset{\sim}{z}_t$, $\underset{\sim}{z}_t = \underset{\sim}{A}\underset{\sim}{z}_t$, la matriz $\underset{\sim}{\hat{M}}^*$ correspondiente a los $\underset{\sim}{z}_t^*$ deberá ser $\hat{M}_1 A_1^1$ de manera que $\hat{M}z_t = \hat{M}^* z_t$. Una forma de elegir la matriz $\hat{M}$ es usar la raíz cuadrada de la inversa de un estimador de la covarianza del vector z_t. En este caso si $\hat{\Sigma}$ es un estimador de la matriz de covarianza Σ de Z_t, $\hat{M}$ se elegirá tal que:

$$(10.5)\quad \hat{M}\hat{M}' = \hat{\Sigma}^{-1}$$

Una forma de elegir $\hat{\Sigma}$ está dada por la matriz de covarianza muestral:

$$(10.6)\quad \hat{\Sigma} = \sum_{t=1}^{T} z_t z_t'/T$$

si la ecuación (1.1) no tiene constante. Si la ecuación (1.1) tiene constante se puede utilizar:

$$(10.7)\quad \hat{\Sigma} = \sum_{t=1}^{T} (z_t - \bar{z})(z_t - \bar{z})'/T$$

con

$$\bar{z} = \sum_{t=1}^{T} z_t/T$$

Como los estimadores dados por (10.6) y (10.7) no son robustos, convendrá reemplazarlos por estimadores robustos de Σ. Para el estudio de estimadores robustos de una matriz de covarianza puede verse Maronna (1976).

Si se utiliza $\hat{M}$ dado por (10.5), entonces se tendrá:

$$||\hat{M}z_t|| = (z_t' \hat{\Sigma}^{-1} z_t)^{1/2}$$

Con los w_t definidos por (10.4), los MG-estimadores se definirán por:

$$(10.8)\quad \sum_{t=1}^{T} (y_t - z_t'\theta) z_t w_t = 0$$

El comportamiento asintótico de estos estimadores fue estudiado por Maronna y Yohai (1979). Se puede mostrar que asintóticamente estos estimadores son normales multivariados con media 0 y matriz de covarianza:

$$(10.9)\quad \underset{\sim}{\Lambda}^{-1}\ \underset{\sim}{\Omega}\ \underset{\sim}{\Lambda}^{-1}$$

donde

$$(10.10)\quad \underset{\sim}{\Lambda} = E_F(\partial\psi(u_t/\sigma, ||\hat{\underset{\sim}{M}}\underset{\sim}{z}_t||)/\partial u_t\ \underset{\sim}{z}_t\ \underset{\sim}{z}_t')$$

$$(10.11)\quad \Omega = E_F(\psi^2(u_t/\sigma, ||\hat{\underset{\sim}{M}}\underset{\sim}{z}_t||)\underset{\sim}{z}_t\underset{\sim}{z}_t')$$

y donde:

$$(10.12)\quad \psi(u, ||\underset{\sim}{z}||) = w(u, ||\underset{\sim}{z}||)u$$

Por otro lado la función de influencia de estos estimadores es igual a:

$$(10.13)\quad \mathrm{INF}(y^*, \underset{\sim}{z}^*) = \underset{\sim}{\Lambda}^{-1}\psi((y - \underset{\sim}{z}^{*\prime}\underset{\sim}{\theta})/\sigma, ||\hat{\underset{\sim}{M}}\underset{\sim}{z}^*||)\underset{\sim}{z}^*$$

De aquí resulta que la condición para que la función de influencia sea acotada es que la función $\nu(u, ||\underset{\sim}{z}||)$ dada por:

$$(10.14)\quad \nu(u, ||\underset{\sim}{z}||) = w(u, ||\underset{\sim}{z}||)u\underset{\sim}{z}$$

sea acotada.

Una forma de medir cuantitativamente la robustez de un estimador está dado por el máximo de la norma de la función de influencia. Esta medida fue introducida por Hampel (1968) y (1974) y se denomina sensibilidad a los errores groseros (GES). Luego, (10.15) $\quad GES = \max_{y^*,\underset{\sim}{z}^*} INF(y^*,\underset{\sim}{z}^*)$

Algunas propuestas para la función $w(u,\underset{\sim}{z})$ acotada son las siguientes:

(a) Mallows considera funciones de la forma:

$$w(u,\underset{\sim}{z})=w_1(u)w_2(||\underset{\sim}{z}||)$$

donde

$$w_1(u) = \psi_1(u)/u, \quad w_2(z) = \psi_2(z)/z$$

En este caso si ψ_1 y ψ_2 son acotados, entonces la función ν de (10.13) también resulta acotada y por lo tanto el GES será finito. Las funciones ψ_1 y ψ_2 pueden elegirse por ejemplo en la familia de Huber dada por (4.10) o en cualquiera de las familias de funciones redescendientes vistas en Sección 6.

(b) Hampel considera funciones w de la forma:

$$w(u,||\underset{\sim}{z}||) = w(u||\underset{\sim}{z}||)$$

donde

$$w(x) = \psi(x)/x$$

En este caso ψ acotada implica ν acotada. Por lo tanto ψ puede elegirse en la familia dada por (4.10) o en las familias de funciones redescendientes introducidas en Sección 6.

Los GM-estimadores tienen en general varianza un poco mayor que los correspondientes M-estimadores, pero esta mayor varianza se compensa por una influencia acotada de las observaciones atípicas que no satisfacen el modelo. Para algunas comparaciones numéricas puede verse Maronna y Bustos (1979).

Varios autores, Maronna y Yohai (1979), Krasker (1978), Krasker y Welsch (1978) estudian la elección óptima de la función w. El criterio de optimalidad utilizado es generalmente del siguiente tipo: minimizar la suma de las varianzas de las componentes de $\hat{\underset{\sim}{\theta}}$ sujeto a la restricción que el GES sea menor que una constante dada.

11. Otros Enfoques y Problemas de Estimación Robusta

Otros enfoques sobre estimadores robustos de regresión pueden encontrarse en Bickel (1973) que estudia L-estimadores, Jaeckel (1972) y Jureckova (1971) que estudian R-estimadores. Tanto los L- como los R-estimadores son más difícil de computar que los M-estimadores. Yohai (1974) estudia la elección óptima de la función en la clase de funciones (4.10) utilizando las observaciones.

Denby y Martin (1979) y Bustos (1978) extienden los M- y los GM-estimadores para procesos autorregresivos.

Hampel(1968) define el concepto de robustez cualitativa para un estimador de un parámetro de locación. Aunque no está hecha explícitamente,esa definición puede extenderse para la estimación de los coeficientes de regresión en un modelo lineal. Otra medida de robustez cuantitativa introducida por Hampel (1968) y (1974) es el punto de ruptura de un estimador, el cual puede interpretarse como la máxima cantidad de contaminación arbitraria que puede aceptar un estimador sin irse a infinito. El estudio del punto de ruptura para M- y para GM-estimadores de regresión puede verse en Maronna y Bustos (1979).

12. Conclusiones

Tal como se adelantara en la introducción, en este trabajo se presentaron estimadores robustos que permiten obtener estimaciones mucho más confiables que las correspondientes al estimador de mínimos cuadrados cuando se sospecha que hay observaciones con errores groseros o perturbaciones en el modelo. Por otra parte, estos estimadores sólo pierden 5% de eficiencia en ausencia de tales perturbaciones.

Por otra parte C.E.M.A. cuenta actualmente con programas en FORTRAN IV que permiten computar los M-estimadores y los GM-estimadores.

REFERENCIAS

Andrews, D., Bieckel, P., Hampel, F., Huber,P., Rogers, W.& Tukey, J. (1972) Robust estimates of location: survey and advances, Princeton University Press, Princeton.

Armstrong, R.D. & Frome, E.L. (1976) A comparison of two algorithms for absolute deviation curve fitting. Journal of the American Statistical Association. 71, 328-33.

Beaton, A.E. & Tukey, J.W. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16, 147-185.

Bickel, P (1973) On some analogous to linear combinations of order statistics in a linear model. Annals of Statistics 1, 597-616.

Breiman, L. (1968) Probability. Addison Wesley, Reading.

Bustos, O. (1978) Estimación robusta en procesos autorregresivos. Tesis doctoral, Universidad Nacional de San Luis, Argentina.

Denby, L. & Martin, R.D. (1979) Robust estimation of the first order autorregresive parameter. Journal of the American Statistical Association 74, 140-146.

Hampel, F. (1968) Contribution to the theory of robust estimation. Ph.D. tesis, University of California, Berkeley.

Hampel, F. (1974) The influence curve and its role in robust estimation. Journal of the American Statistical Association 69, 383-393.

Hocking, R.R. (1976) The analysis and selection of variables in linear regression. Biometrics.

Huber, P. (1964) Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73-101.

Huber, P. (1973) Robust regressión: asymptotics, conjectures and Monte Carlo, Annals of Statistics 1, 799-821.

Huber, P.J. (1977) Robust Statistical Procedures. Society for Industrial and Applied Mathematics, Philadelphia.

Holland, P.W. & Welsch, R.E. (1977) Robust regression using iteratively reweighted least squares. Communications in Statistics, Theory and Methods 9, 813-827.

Jaeckel, L.A. (1972) Estimating regression coefficients by minimizing the dispersion of the residuals. Annals of Mathematical Statistics 43, 449-458.

Jureckova, J. (1971) Nonparametric estimation of the regression coefficients. Annals of Mathematical Statistics 42, 1328-1338.

Krasker, W.S. (1978) Estimation in linear regression models with disparate data points. Manuscripto no publicado.

Krasker, W.S. & Welsch, R.E. (1978) Efficient bounded-influence regression estimation using alternative definitions of sensitivity. Manuscripto no publicado.

Klein, R. & Yohai, V.J. (1979) Asymptotic behavior of iterative M-estimators for the linear model. Enviado para publicación.

Maronna, R. (1976) Robust M-estimators of multivariate location and scatter. Annals of Statistics, 54-67.

Maronna, R. & Yohai, V.J. (1979) Consistency of general M-estimators for regression and scale with random carriers. Enviado para publicación.

Maronna, R. & Bustos, O. (1979) Bias and efficiency robustness of general M-estimators for regression with random carriers. Research Report N° 20 Fachgruppe fuer Statistik E.T.H., Zurich.

Relles, D.A. (1968) Robust regression by modified least squares. Ph.D. Tesis, Yale University.

Theil, H. (1971) Principles of Econometrics. John Wiley & Sons, New York.

Tukey, J.W. (1960 A survey of sampling from contaminated distributions. Contributions to Probability and Statistics. (ed. Olkin)

Yohai, V.J. (1974) Robust estimation in the linear model. Annals of Statistics 2, 562-567.

Yohai, V.J. & Maronna, R. (1979) Asymptotic behavior of M-estimators for the linear model. Annals of Statistics 7.