# S-estimation in the nonlinear regression model with long-memory error terms [1]

**Philipp Sibbertsen**
**Fachbereich Statistik, Universität Dortmund**
**D - 44221 Dortmund**
**Germany**
**e-mail: sibberts@amadeus.statistik.uni-dortmund.de**

**Abstract**

In this paper we consider the asymptotic distribution of $S$-estimators in the nonlinear regression model with long-memory error terms. $S$-estimators are robust estimates with a high breakdown point and good asymptotic properties in the iid case. They are constructed for linear regression. In the nonlinear regression model with long-memory errors it turns out, that $S$-estimators are asymptotically normal with a rate of convergence of $n^{1-H}$, $1/2 < H < 1$. But the distribution depends heavily on the unknown parameter vector.

KEY WORDS: Nonlinear regression model; long - range dependence; robustness

## 1 Introduction

In the last years long-memory has been modeled in various financial time series. But even financial time series are quite often a good example for contaminated data. Consider for example financial market price series: the data are typically quite reliable but market crashes could generate large movements well outside the range of typical behaviour. From this it seems natural to consider the behaviour of robust estimation techniques under the assumption of long memory in the data. Also the consideration of nonlinear regression seems natural in this context.

[1]Research supported by "Deutsche Forschungsgemeinschaft" (DFG).

A first model of long-memory time series was given by Mandelbrot/van Ness(1968) by introducing fractional Brownian motion. Another popular model is the extension of the classical ARIMA models of Box/Jenkins(1970) to fractional ARIMA models introduced by Granger/Joyeux(1980) and Hosking(1981).

In general a stationary process $X_t$ is said to have long memory, if

$$Cov(X_t, X_{t+k}) := R_k \sim_{k\to\infty} L(k)|k|^{2H-2}, \quad H \in (1/2, 1), \tag{1}$$

where $L(k)$ is slowly varying for $k \longrightarrow \infty$.

For a more detailed discussion of long-memory time series see for example Beran(1994) or Sibbertsen(1999a).

One possibility to measure the quality of an estimator in the case of contaminated data is the breakdown point introduced by Hampel(1971). The breakdown point is the smallest fraction of contaminated data that can cause an estimator $T$ to take on values arbitrarily far from $T(X)$, where $X$ is the correct sample.

One possible class of estimators with a high breakdown point and also good asymptotic properties are $S$-estimators introduced by Rousseeuw/Yohai(1984). $S$-estimators have an asymptotic breakdown point of 1/2, which is the best possible asymptotic breakdown point. Moreover there are other good robustness properties, such as the exact fit property, which are fulfilled by S-estimators.

$S$-estimators were introduced by Rousseeuw/Yohai(1984) for the linear regression model with iid errors under strong regularity conditions. Davies(1990) generalized the results of Rousseeuw/Yohai(1984) to weak regularity conditions for the regressors including trends. Sibbertsen(1999a,b) proved asymptotic normality of $S$-estimators in the linear regression model with long-memory error terms. Sakata/White(1998) considered $S$-estimators in the nonlinear regression model with dependent and heterogenous observations under quite general conditions. The idea of this paper is to generalize the results of Sibbertsen(1999b) to nonlinear regression with long-memory error terms.

The idea of S-estimators is based on a scale M-estimation, but in contrary to this technique S-estimators first estimate the scale and subsequently the regression parameter. So this estimator is scale invariant in contrary to M-estimators. But because of this definition, there is no loss of efficiency for S-estimators compared to M-estimators.

To define S-estimators, let $\rho$ be a real function satisfying the following assumptions:

1. $\rho$ is symmetric, continuously differentiable and $\rho(0) = 0$;
2. there exists a $c > 0$, such that $\rho$ is monotonously increasing in $[0, c]$ and constant in $[c, \infty)$.

For every set $\{e_1, \ldots, e_n\}$ the scale estimator $s(e_1, \ldots, e_n)$ is then defined as a solution of the equation

$$\frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = K, \tag{2}$$

where the constant $K$ is given by $E_\Phi[\rho] = K$ and $\Phi$ denotes the standard normal distribution. If (2) has more than one solution, $s(e_1, \ldots, e_n)$ is the supremum of all the solutions. If there is no solution, $s(e_1, \ldots, e_n) = 0$.

The S-estimator $\hat{\theta}$ of the regression parameter $\theta$ is defined as

$$\hat{\theta} = \min_{\theta} \{ s[e_1(\theta), \ldots, e_n(\theta)] \} \tag{3}$$

and the scale estimator $\hat{\sigma}$ is

$$\hat{\sigma} = s[e_1(\hat{\theta}), \ldots, e_n(\hat{\theta})]. \tag{4}$$

For a more detailed discussion of robust regression estimates see for example Rousseeuw/Leroy(1987).

In the rest of the paper we consider mainly the function $\psi(x)$, which is the derivative of the function $\rho(x)$ defined above. Because of the definition of $\rho$ $\psi$ is a redescending function. The most popular choice for the function $\psi$ is Tukey's biweight because of its smoothness.

## 2 Asymptotic normality

Let us consider the nonlinear regression model

$$y_t = f(x_t, \theta) + \varepsilon_t, \quad t = 1, \ldots, n, \tag{5}$$

where $x_t$ are the regressors, $\theta$ is the unknown $p$-dimensional parameter vector, $f : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ is the regression function fulfilling regularity conditions mentioned later and $\varepsilon_t$ is a sequence of Gaussian long-memory random variables.

The function $f(x, \theta)$ shall fulfill the following regularity conditions:

**(F1)** $f'(x, \theta) := \frac{\partial}{\partial \theta} f(x, \theta)$ exists for all $\theta$;

**(F2)** $f'(x, \theta) f'(x, \theta)^T$ is finite and regular $\quad \forall x, \theta$.

A necessary tool for the investigation of the asymptotic behaviour of S-estimators is the Hermite rank of a function:

**Definition (Hermite rank)**

*Let $Z$ be a standard normal random variable. A function $G : \mathbb{R} \to \mathbb{R}$ with $E[G(Z)] = 0$ and $E[G^2(Z)] < \infty$, is said to have Hermite rank $m$, if $E[G(Z)P_q(Z)] = 0$ for all Hermite polynomials $P_q, q = 1, \ldots, m-1$ and $E[G(Z)P_m(Z)] := J_G(m) \neq 0$.*

The function $J_G(l)$ is defined by

$$J_G(l) := E[G(Z)P_l(Z)], \qquad l \in \mathsf{IN}. \tag{6}$$

Hermite polynomials, normalized by $q!$ provide an orthonormal basis in the $L_2$ - space with respect to the standard normal distribution. So every such function $G(\cdot)$ can be expanded into a series

$$G(Z) = \sum_{q=m}^{\infty} J_G(q) \frac{P_q(Z)}{q!}.$$

For $G(Z)$ with Hermite rank 1, Taqqu(1975) showed that the normalized sum

$$n^{-H} L_{Var}^{-\frac{1}{2}}(n) \sum_{i=1}^{n} \frac{G(Z_i)}{J_G(1)}$$

is asymptotically standard normal.

We can now establish asymptotic normality for S-estimators in the nonlinear regression model.

**Theorem 1 (asymptotic normality)**

*Let $\hat{\theta}_n$ be a sequence of S-estimators for the regression parameter in the nonlinear regression (5) with long-memory errors. Then if the conditions F1, F2 hold and if*

**(1)** $E\psi' \neq 0 \qquad \sigma > 0$;

**(2)** $J_{\psi(Q)} \neq 0$ ,

*we have*

$$n^{1-H} L_{Var}^{-\frac{1}{2}}(n) J_{\psi(Q)}(1)^{-1} (\hat{\theta}_n - \theta_0) \quad \xrightarrow{d} \quad N(0, \sigma_0^2 \frac{E_\Phi \psi^2}{(E_\Phi \psi')^2} [f'(x,\theta_0) f'(x,\theta_0)^T]^{-1}).$$

**Proof**

From the definition of S-estimators we have

$$\mathbf{0} = n^{-H} \sum_{t=1}^{n} f'(x_t, \hat{\theta}_n) \psi(\frac{y_t - f(x_t, \hat{\theta}_n)}{\hat{\sigma}_n}). \qquad (7)$$

where $\mathbf{0}$ is the $p$ - dimensional vector $(0, \ldots, 0)^T$.

Taylor expansion of the function $f(x_t, \theta_0)$ respectively $\theta_0$ gives

$$f(x_t, \hat{\theta}_n) = f(x_t, \theta_0) + f'(x_t, \theta_0)(\hat{\theta}_n - \theta_0). \qquad (8)$$

From equation (7) together with (8) we have

$$\begin{aligned} \mathbf{0} &= n^{-H} \sum_{t=1}^{n} f'(x_t, \hat{\theta}_n) \psi(\frac{y_t - f(x_t, \hat{\theta}_n)}{\hat{\sigma}_n}) \\ &= n^{-H} \sum_{t=1}^{n} f'(x_t, \hat{\theta}_n) \psi(\frac{y_t - f(x_t, \theta_0) - f'(x_t, \theta_0)(\hat{\theta}_n - \theta_0)}{\hat{\sigma}_n}) \\ &= n^{-H} \sum_{t=1}^{n} f'(x_t, \hat{\theta}_n) \psi(\frac{r_t - f'(x_t, \theta_0)(\hat{\theta}_n - \theta_0)}{\hat{\sigma}_n}), \end{aligned}$$

where $r_t := y_t - f(x_t, \theta_0)$ denotes the residual process.

For the function $\psi$ we have the Taylor expansion

$$\psi(z + h) = \psi(z) + h \int_0^1 \psi'(z + th) dt. \qquad (9)$$

Applying relation (9) twice to $\psi$ we obtain

$$\psi(z + h) = \psi(z) + h \int_0^1 \psi'(z) dt + h^2 \int_0^1 \int_0^t \psi''(z + sh) ds dt. \qquad (10)$$

Applying Fubini's theorem to the right-hand side of equation (10) gives

$$\psi(z+h) = \psi(z) + h[\psi'(z) + h\int_0^1 (1-s)\psi''(z+sh)ds]. \tag{11}$$

Setting $h_t := -f'(x_t,\theta_0)^T(\hat{\theta}_n - \theta_0)/\hat{\sigma}_n$ and $z_t := r_t/\hat{\sigma}_n$ relation (11) gives for (7)

$$\frac{1}{n}\sum_{t=1}^{n} f'(x_t,\hat{\theta}_n)\frac{f'(x_t,\theta_0)^T n^{1-H} L_{Var}^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)}{\hat{\sigma}_n}\times$$
$$\times[\psi'(\frac{r_t}{\hat{\sigma}_n}) + h_t\int_0^1 (1-s)\psi''(z_t+sh_t)ds] \quad = \quad n^{-H} L_{Var}^{-\frac{1}{2}}\sum_{t=1}^{n} f'(x_t,\hat{\theta}_n)\psi(\frac{r_t}{\hat{\sigma}_n}). \tag{12}$$

We now have to prove two assertions:

**1)** $$\frac{1}{n}\sum_{t=1}^{n}[\psi'(\frac{r_t}{\hat{\sigma}_n}) - \frac{f'(x_t,\theta_0)^T(\hat{\theta}_n-\theta_0)}{\hat{\sigma}_n}\int_0^1 (1-s)\psi''(z_t+sh_t)ds]f'(x_t,\hat{\theta}_n)f'(x_t,\theta_0)^T \stackrel{P}{\longrightarrow} E_\Phi\psi'[f'(x_t,\theta_0)f'(x_t,\theta_0)^T] \tag{13}$$

and

**2)** $$n^{-H}L_{Var}^{-\frac{1}{2}}\sum_{t=1}^{n}\psi(\frac{r_t}{\hat{\sigma}_n})f'(x_t,\hat{\theta}_n) \quad \stackrel{d}{\longrightarrow} \quad N(0, E_\Phi[\psi^2][f'(x_t,\theta_0)f'(x_t,\theta_0)^T]).$$

Let us prove assertion 1) first.

To this end, we show first:

$$|\frac{1}{n}\sum_{t=1}^{n}\frac{f'(x_t,\theta_0)^T(\hat{\theta}_n-\theta_0)}{\hat{\sigma}_n}\int_0^1(1-s)\psi''(z_t+sh_t)dsf'(x_t,\hat{\theta}_n)f'(x_t,\theta_0)^T| \xrightarrow{P} 0. \tag{14}$$

This expression follows from (F2) and the consistency of $\hat{\theta}_n$ and $\hat{\sigma}_n$. We also have $\int_0^1(1-s)\psi''(z_t+sh_t) \leq \| \psi''\|_\infty$ and $\|\psi''\|_\infty$ is finite.

Here $\|f\|_\infty$ denotes the $\infty$-norm of the function $f$.

To show (13) the first term on the left-hand side of this equation has to be considered. We have:

$$|\frac{1}{n}\sum_{t=1}^{n}f'(x_t,\hat{\theta}_n)f'(x_t,\theta_0)^T[\psi'(\frac{r_t}{\hat{\sigma}_n})-\psi'(\frac{r_t}{\sigma_0})]| \to 0. \tag{15}$$

This can be seen from the following inequality, which holds because of the consistency of the scale estimation:

$$\begin{aligned} |\frac{1}{n}\sum_{t=1}^{n}f'(x_t,\hat{\theta}_n)f'(x_t,\theta_0)^T[\psi'(\frac{r_t}{\hat{\sigma}_n})-\psi'(\frac{r_t}{\sigma_0})]| &\leq \\ |\frac{1}{\hat{\sigma}_n}-\frac{1}{\sigma_0}|\|\psi''\|_\infty\frac{1}{n}\sum_{t=1}^{n}|f'(x_t,\hat{\theta}_n)f'(x_t,\theta_0)^T||r_t| &\xrightarrow{P} \\ |\frac{1}{\hat{\sigma}_n}-\frac{1}{\sigma_0}|\,\|\psi''\|_\infty\, f'(x_t,\theta_0)f'(x_t,\theta_0)^T\, E|r| &\longrightarrow \quad 0. \end{aligned} \tag{16}$$

The right-hand side also converges to 0 because of the assumed consistency of $\hat{\sigma}_n$ and $\hat{\theta}_n$ and because ofthe finiteness of the other terms. It was also assumed that $\sigma_0 > 0$.

To prove the second assertion let $Q : \mathrm{IR} \to \mathrm{IR}$ be a function defined as follows:

$$Q(\eta) := \frac{\eta}{\hat{\sigma}_n}. \tag{17}$$

This function has Hermite rank 1 (see Taqqu(1975)). Define $J_{\psi(Q)}(1)$ as in (6).

In view of the limit theorem 5.1 from Taqqu(1975) we see that

$$n^{-H} L_{Var}^{-\frac{1}{2}}(n) J_{\psi(Q)}(1)^{-1} \sum_{i=1}^{n} \psi(Q(e_i(n))) f'(x_t, \hat{\theta}_n)$$

is a normal random variable. To compute the mean and the covariance of this variable we get from $\psi$ odd and $r_t$ Gaussian:

$$\begin{aligned} E\psi(\frac{r_t}{\sigma_0}) &= -E\psi(-\frac{r_t}{\sigma_0}) \\ &= -E\psi(\frac{r_t}{\sigma_0}). \end{aligned} \tag{18}$$

Hence

$$E\psi(\frac{r_t}{\sigma_0}) = 0 \tag{19}$$

and consequently

$$E(\psi(\frac{r_t}{\sigma_0}) f'(x_t, \theta_0)) = \mathbf{0}. \tag{20}$$

Therefore we have

$$Cov(\psi(\frac{r_t}{\sigma_0}) f'(x_t, \theta_0)) = E_\Phi(\psi^2(\frac{r_t}{\sigma_0})) f'(x_t, \theta_0) f'(x_t, \theta_0)^T. \tag{21}$$

Altogether we obtain from (13):

$$n^{1-H} L_{Var}^{-\frac{1}{2}}(n) J_{\psi(Q)}(1)^{-1} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma_0^2 \frac{E_\Phi \psi^2}{(E_\Phi \psi')^2} [f'(x_t, \theta_0) f'(x_t, \theta_0)^T]^{-1}), \tag{22}$$

which is the assertion of the theorem. $\diamondsuit$

The theorem above establishes the asymptotic normality of S-estimators in the nonlinear regression model, when the error terms have long-memory. It should be mentioned that these results do not only hold in the case of Gaussian disturbances but also in the case of transformed Gaussian errors where the transformation has Hermite rank 1. By using the theory of Appel polynomials instead of Hermite polynomials the result can even be generalized to all symmetric error distributions. In this paper we confined ourselves to the case of Gaussian disturbances to simplify the proof and to illustrate the idea behind it.

We can also establish a similar result for S-estimators in the short-memory case, that is when $0 < H < 1/2$.

**Corollary 1** Under the same conditions as in the above theorem and $0 < H < 1/2$, that is $\sum_{k=1}^{\infty} R(k) < \infty$, and $\psi$ is nonlinear the following holds

$$n^{-H} L_{Var}^{-1/2}(\hat{\theta}_n - \theta_0) \quad \xrightarrow{d} \quad N(0, \sigma_0^2 \frac{E_\Phi \psi^2}{(E_\Phi \psi')^2} [f'(x_t, \theta_0) f'(x_t, \theta_0)^T]^{-1}).$$

**Proof** This result follows in the same way as in Theorem 1. But instead of using the limit theorem of Taqqu(1975) this result follows from a limit theorem of Breuer/Major(1983). $\diamondsuit$

# REFERENCES


Beran, J. (1994). *Statistics for long-memory processes.* Chapman & Hall, New York.

Box, G.E.P., Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Breuer, P., Major, P. (1983) Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multiv. Anal.* **13** 425 - 441.

Davies, P. L. (1990). The asymptotic of S-estimators in the linear regression model. *Ann. Statist.* **18** 1651 - 1675.

Granger, C., Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis* **1**, 15 - 30.

Hampel, F. R. (1971). General qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887 - 1896.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165 - 176.

Mandelbrot, B., van Ness, J. (1968). Fractional brownian motions, fractional noises and applications. *SIAM Reviews* **10** 422 - 437.

Rousseeuw, P., Leroy, A. (1987). *Robust regression and outlier detection.* Wiley, New York.

Rousseeuw, P., Yohai, V. J. (1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series regression. Springer Lecture Notes in Statist.* **26** 256 - 272, Springer, New York.

Sakata, S., White, H. (1998). S-estimation of nonlinear regression models with dependent and heterogenous observations. *Preprint University of California, San Diego*

Sibbertsen, Ph. (1999a). *Robuste Parameterschaetzung im linearen Regressionsmodell bei Fehlertermen mit langem Gedaechtnis.* Verlag fuer Wissenschaft und Forschung, Berlin (in German).

Sibbertsen, Ph. (1999b). S-estimators in the linear regression model with long-memory error terms. *J. Time Ser. Anal.* to appear.

Taqqu, M. S. (1975). Weak convergence to fractional Brownian Motion and to the Rosenblatt Process. *Probab. Th. Rel. Fields* **31** 287 - 302.