SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.

TUKEY, J. W. (1984). *The Collected Notes of John Tukey* **2** (D. R. Brillinger, ed.). Wadsworth, Monterey, Calif.

UNIVERSITY OF FRANKFURT
DEPARTMENT OF MATHEMATICS
JOHANN WOLFGANG GOETHE UNIVERSITY
P.O. BOX 111319, 6000 FRANKFURT
WEST GERMANY

AUSTRALIAN NATIONAL UNIVERSITY
DEPARTMENT OF STATISTICS
MATHEMATICS BUILDING, I.A.S.
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4, CANBERRA, 2601
AUSTRALIA

HANS R. KÜNSCH

ETH, Zurich

Martin and Yohai provide an interesting study on the effect of atypical observations on the behavior of estimators in time series. The influence functional given by Definition 4.2 is the infinitesimal asymptotic bias in a one-parameter family of contaminations of a given model. The bias was also the starting point of my own paper (1984, cf. Section 1.2), but I treated only a smaller class of estimators and I focused on different aspects. So let me explain the differences between the two approaches and discuss their advantages and disadvantages.

Heuristically the connection between ICH and IF is as follows. ICH is the derivative in all directions, i.e., the gradient of $\mathbf{T}$. Hence by the chain rule of differential calculus one gets, formally,

$$\mathrm{IF} = \frac{d}{d\gamma}\mathbf{T}(\mu_y^\gamma) = < \operatorname{grad}\mathbf{T}, \frac{d}{d\gamma}\mu_y^\gamma > = \int \mathrm{ICH}(\mathbf{y}_1)\frac{d}{d\gamma}\mu_y(d\mathbf{y}_1).$$

If $\mathbf{T}$ depends only on the m-dimensional marginal, we can find $(d/d\gamma)\mu_y^\gamma$ in the model (2.4) by the following argument. Ignoring terms of order $o(\gamma)$, there is at most one block of outliers intersecting with $(1, 0, \dots, 2-m)$, and the initial point of this block is distributed uniformly over $(1, 0, \dots, 3-m-k)$. To me, the most important theoretical contribution of Martin and Yohai is Theorem 4.2 where they show that the same result also holds for $m = \infty$, at least if $\tilde{\psi}$ depends only weakly on values far away. Since the uniform distribution on all integers is not finite, a bounded $\tilde{\psi}$ is not sufficient for the boundedness of $(d/d\gamma)\mathbf{T}(\mu_y^\gamma)$.

Some of the arguments in the proof of Theorem 4.2 involve the specific contamination model while others are valid more generally. Since the latter may be useful in other situations, I propose to split it in the following way.

THEOREM 4.1′. *Let* $\mathbf{T}$ *be a* $\tilde{\psi}$ *estimate with* $\mathbf{t}_0 = \mathbf{T}(\mu_x)$ *and put* $\mathbf{m}(\gamma, \mathbf{t}) = E[\tilde{\psi}(\mathbf{y}_1^\gamma, \mathbf{t})]$. *If*

(a′) $\mathbf{T}(\mu_y^\gamma) - \mathbf{t}_0 = O(\gamma)$,

(b′) $\mathbf{m}(0, \mathbf{t})$ *is differentiable at* $\mathbf{t} = \mathbf{t}_0$ *and the derivative* C *is nonsingular*,

(c′) $\mathbf{b}(\mathbf{t}) = \lim(\mathbf{m}(\gamma, \mathbf{t}) - \mathbf{m}(0, \mathbf{t}))/\gamma$ *exists and the convergence is uniform for* $|\mathbf{t} - \mathbf{t}_0| \le \varepsilon_0$, *then* $\lim(\mathbf{T}(\mu_y^\gamma) - \mathbf{t}_0)/\gamma$ *exists and is equal to*

$$-C^{-1}\mathbf{b}(\mathbf{t}_0) = \lim_\gamma E\left[\mathrm{ICH}(\mathbf{y}_1^\gamma)\right]/\gamma.$$

PROOF (extracted from Martin and Yohai). First note that because of (b′) and (c′) $\mathbf{b}$ is continuous at $\mathbf{t}_0$. Using $\mathbf{m}(\gamma, \mathbf{T}(\mu_y^\gamma)) = 0$ we thus obtain

$$\begin{aligned}\left(\mathbf{m}\left(0, \mathbf{T}\left(\mu_y^\gamma\right)\right) - \mathbf{m}(0, \mathbf{t}_0)\right)/\gamma &= -\left(\mathbf{m}\left(\gamma, \mathbf{T}\left(\mu_y^\gamma\right)\right) - \mathbf{m}\left(0, \mathbf{T}\left(\mu_y^\gamma\right)\right)\right)/\gamma \\ &= -\mathbf{b}(\mathbf{t}_0) + \mathbf{b}(\mathbf{t}_0) - \mathbf{b}\left(\mathbf{T}\left(\mu_y^\gamma\right)\right) + o(1) \\ &= -\mathbf{b}(\mathbf{t}_0) + o(1).\end{aligned}$$

On the other hand by (b′) and (a′)

$$\left(\mathbf{m}\left(0, \mathbf{T}\left(\mu_y^\gamma\right)\right) - \mathbf{m}(0, \mathbf{t}_0)\right)/\gamma = C\left[\mathbf{T}\left(\mu_y^\gamma\right) - \mathbf{t}_0\right]/\gamma + o(1). \qquad \square$$

In the situation of Theorem 4.2, the condition (c′) above follows from (9.10), (9.13), and (9.4)–(9.6).

A limitation of the results obtained by Martin and Yohai lies in the special class of contamination models considered. The general replacement model (2.1) does not contain innovation outliers, and some of the results from Sections 5 and 6 do not generalized from AO's to other types of outliers. For instance if we calculate the GES for pure replacement outliers instead of AO's we get the following values for the HKW estimators in the AR(1) model at $\phi = 0.5$ (cf. Table 6.1):

	$k = 1$	$k = 20$
GM-H	-2.5	2.3
GM-BS	-1.7	2.6

So it is not always true that the bisquare is to be preferred to the Huber function. Even in the case of location estimators with i.i.d. observations, the AO model has some peculiar features. If we choose the constants such that both estimators have 95% efficiency at the Gaussian, the GES for Huber is 1.63 and for bisquare 1.35.

For real data it is impossible to say if one has additive or pure replacement outliers or if there is some dependence between the (x_i), (w_i), and (z_i). Because of this and because one time series often contains different types of outliers, the class P in Definition 6.1 of the GES has to be chosen quite large. The authors do not discuss this point, but I think that one should take at least all joint distributions of (x_i, w_i) with marginal μ_x and possibly also all block lengths k in the model (2.4) for the (z_i). With such a class Hampel's optimality problem becomes intractable, but one possibly will have to live with this situation. Optimality in such small classes as considered in Section 7 is not very helpful in my opinion.

A second limitation comes from the fact that—in contrast to the i.i.d. case—the infinitesimal asymptotic bias does not coincide with the standardized influence of one outlier in a long time series. In order to make this clearer, let us consider a GM estimator for the AR(1) model as an example. If $(x_1, x_2, \ldots, x_n)$ is a sample from the clean process and $1 < i < n$ we obtain (cf. (1.27) of Künsch (1984) and (5.10)):

$$\begin{aligned} &T_n(x_1, \ldots, x_i + v, \ldots, x_n) - T_n(x_1, \ldots, x_n) \\ &\quad \approx n^{-1}B^{-1}(1-\phi^2)^{1/2}\Big(\eta\big(x_i - \phi x_{i-1} + v,\, x_{i-1}(1-\phi^2)^{1/2}\big) \\ &\qquad + \eta\big(x_{i+1} - \phi x_i - \phi v,\, (x_i + v)(1-\phi^2)^{1/2}\big) \\ &\qquad\quad - \eta\big(x_i - \phi x_{i-1},\, x_{i-1}(1-\phi^2)^{1/2}\big) - \eta\big(x_{i+1} - \phi x_i,\, x_i(1-\phi^2)^{1/2}\big)\Big). \end{aligned}$$

If we take averages, only the second term remains in accordance with (5.6). But that does not mean that the first term is small or even zero. In the HKW case it takes any value between $\inf \eta$ and $\sup \eta$. If we only look at the infinitesimal asymptotic bias, we thus underestimate the possible effect of one outlier. These considerations led me to take sup|ICH| as the sensitivity measure. This gives an upper bound for the above expression although it may be too pessimistic and gets worse as the range of $\tilde{\psi}$ increases.

Finally let me point out an important open problem: How can we assess the effects of deviations from the assumption of stationarity? The difficulty is that with most models the effect of the nonstationarity becomes either negligible or dominant as n increases. One possibility would be to rescale the nonstationarity with each n in order to get reasonable asymptotics. In the case of a trend this would mean to consider $y_1 = x_1 + f(1/n)$, $y_2 = x_2 + f(2/n), \ldots, y_n = x_n + f(1)$. At least if $\tilde{\psi}$ has finite range, I can show that the asymptotic value is the solution of

$$\int\int \tilde{\psi}(x_1 + f(u), x_0 + f(u), \ldots, x_{2-m} + f(u), \mathbf{T})\, d\mu(\mathbf{x}_1)\, du = 0$$

and the present techniques could be applied. However, in this way we have returned to the land of stationarity, and I wonder if there is a better approach.

In summary, this is an important contribution, particularly because it covers a very general class of estimators in time series. However, I have some reservations on the use of the AO model, and I do not think that the infinitesimal asymptotic bias alone captures the main effects of outliers in large, but finite samples, since it involves too much averaging.

REFERENCE

KÜNSCH, H. (1984). Infinitesimal robustness for autoregressive processes. *Ann. Statist.* **12** 843–863.

SEMINAR FÜR STATISTIK
ETH–ZENTRUM
CH-8092 ZURICH
SWITZERLAND