P. M. ROBINSON
London School of Economics

Martin and Yohai's paper is a fine technical achievement, developing an interesting tool, the influence functional, for describing an aspect of time series behaviour, and continuing the authors' work on the difficult and important problem of time series analysis in the presence of outliers. I have two points, one being a suggestion prompted by their discussion of hypothesis testing, and motivated by the need for test statistics with both good robustness and good power properties against given alternatives. My first and main point concerns Martin and Yohai's approach towards dealing with the outlier behaviour described by their general replacement model, and to some extent this impacts on the use of their influence functional.

Martin and Yohai's general replacement model (2.2) is indeed "general," and even in the pure replacement (PR) and additive outliers (AO) special cases it presents an identifiability problem to which GM and RA rules need not necessarily provide a useful solution. The non-Gaussian character of y and the nonlinear character of the GM and RA rules severely hinders a proper analysis of the identifiability problem. While Martin and Yohai's results embrace w and v with no moments, even bad contamination can be modelled by w and v with finite variance, in which case, if their core x process is indeed "usually Gaussian," a second moment analysis may gain some insight into the identifiability problem in the LS case, and conceivably also into the possible impact of GM and RA estimators on the problem. Denoting means by m_x, etc., and lag-j autocovariances by $c_x(j)$, etc., for the PR model

$$m_y = m_x + (m_w - m_x)m_z, \tag{1}$$

$$\begin{aligned} c_y(j) = {} & (1 - m_z)^2 c_x(j) + m_z^2 c_w(j) \\ & + \{(m_w - m_x)^2 + c_x(j) + c_w(j)\} c_z(j). \end{aligned} \tag{2}$$

For the AO model with v independent of x (as assumed by Martin and Yohai in Section 5)

$$m_y = m_x + m_v m_z, \tag{3}$$

$$c_y(j) = c_x(j) + m_z^2 c_v(j) + \{m_v^2 + c_v(j)\} c_z(j). \tag{4}$$

Note that x's ARMA coefficients are functionals of the $c_x(j)$.

It is easily seen that the $c_x(j)$ can be quite unrecognisable from the $c_y(j)$, leading in general not only to inconsistent estimation of x's ARMA coefficients but also to incorrect order determination via criteria such as AIC. Can robustification alleviate these problems? Note that c_y is determined not only by c_x and c_w or c_v, and the frequency of contamination m_z, but also by c_z. We cannot choose $m_x = m_w$ or $m_v = 0$ (thereby eliminating $(m_w - m_x)m_z$, $(m_w - m_x)^2 c_z(j)$, $m_v m_z$, and $m_v^2 c_z(j)$ from (1)–(4), respectively) without loss of generality because without further information only y can be mean-corrected; substantially different m_x and m_w, or nonzero m_v (by no means unlikely, it seems) could conceivably result in y reflecting z's autocovariance structure more than x's, or w's or v's for that matter. By down-weighting extreme y's, GM estimation (e.g., the Mallows variant, rather than M estimation) seems to take a step in the right direction, albeit in an ad hoc fashion, and there is evidence (e.g., Martin and Jong (1977)) that it does help; its not clear to me whether the RA or GRA rules are effective in reducing the inconsistency, notwithstanding Section 7 and their evident relevance to innovations outlier (IO) problems.

Because ARMA autocovariance structure is closed under addition and multiplication, generally (2) and (4) imply that x, w, v, and z with ARMA-like autocovariances imply y with ARMA-like autocovariances, with ARMA orders at least as high as x's. In Martin and Yohai's "independent outliers" case, (2) and (4) become

$$c_y(j) = (1 - m_z)^2 c_x(j) + m_z^2 c_w(j) + a_1\delta_{j0}, \tag{2'}$$

$$c_y(j) = c_x(j) + m_z^2 c_v(j) + a_2\delta_{j0}, \tag{4'}$$

where $a_1 > 0$, $a_2 > 0$, and δ is the Kronecker delta. For example an AR(p) x and white noise w or v implies an ARMA(p, q) y, $q \le p$, so y's AR order (and coefficients) matches x's, but y generally has an MA component so AR(p) fitting to y, by GM, RA, and other rules, leads to inconsistency; an ARMA w or v with positive AR order implies y's AR order generally exceeds x's, so the inconsistency problem is if anything more serious. In Section 2.2 Martin and Yohai suggest a z process generating "patchy outliers"; it may be shown that this implies z has an MA($k - 1$) representation, with $c_z(j) = (1 - p)^k \{(1 - p)^j - (1 - p)^k\}$, $0 \le j \le k$. The effect may be to further increase y's MA order (though not its AR order) relative to the "independent outliers" case. Martin and Yohai's "patchy outliers" model is only one such, and to the extent that we can identify outlier occurrence in real data sets it might be worth investigating whether this particular model warrants emphasis. For example, one can model binary time series to have a more general MA structure than theirs, or to have AR and ARMA structure; in the latter cases y's AR order will generally be increased, causing inconsistency in the GM and RA estimators as well as leading to different forms of the influence functional.

I hasten to add that Martin and Yohai are well aware that AO and other outlier models (though not IO ones) cause inconsistency, and Martin has to some extent addressed the problem in earlier work. I feel that the identification problem should be faced up to more squarely by making a conscious effort to take outlier models such as Martin and Yohai's sufficiently seriously to allow them to determine the form of model to be robustly estimated, via arguments such as mine. In case (4′) with x AR(1) and v white noise for example, we could fit (albeit inefficiently) an ARMA(1, 1); or we could estimate only the (correctly identified) AR coefficients (again inefficiently) using the following modification of Martin and Yohai's (3.6):

$$\tilde{\psi}_i(\mathbf{y}_i^1; \phi) = \eta\left(y_i - \phi y_{i-1}, y_{i-2}(1 - \phi^2)^{1/2}\right), \qquad i \ge 2,$$

where y_{i-2}, unlike y_{i-1}, is uncorrelated with the MA(1) disturbance in y_i

(though not necessarily independent of it, so non-LS estimators may still suffer from some inconsistency). More ambitious approaches would be to exploit Gaussianity of x and non-Gaussianity of outliers in a manner analogous to some solutions to the classical errors-in-variable problem, or to approach the non-Gaussian modelling problem head-on. I do not underestimate the complications and pitfalls in these alternatives, but I do not think that Martin and Yohai's use of the same estimators in both independent and patchy outliers cases should be taken to imply that approximate knowledge of z's properties (or of w's or v's for that matter) should not influence estimator choice. Their approach to estimation could be said to take for granted that we have almost no information about the character of w, v, and z. While this may more or less often be realistic, the authors also seem able to characterise some outlier patterns occurring in practice, and calculation of their influence functional with a real data set in mind itself requires considerable knowledge of serial dependence and other distributional structure of w, v, and z, as well as of x (though normality of x does not seem crucial to most of their theoretical results). Note that if we base the estimation rule on the "true" model for y, derived from stochastic assumptions on w, v, and z possibly as described above, we could still study the corresponding influence functional, more complicated though it may be; there is interest in the influence of outliers on consistent rules, as well as on inconsistent ones.

Let me finally turn briefly to the question of hypothesis testing. In Section 8 Martin and Yohai apply their influence functional to the Box–Pierce portmanteau statistic

$$V_n^L = \sum_{i=1}^{L} r_i^2.$$

It is known that V_n^L is asymptotically equivalent to the score test statistic against AR(L) or MA(L) alternatives, based on a Gaussian likelihood, and is thus asymptotically locally most powerful against such alternatives. Not surprisingly, therefore, Martin and Yohai find that V_n^L is not robust. A robust alternative, that maintains good power properties against specified time series alternatives, could be obtained by applying the score principle (or Wald or likelihood ratio principles) to an appropriate robustified loss function. While this should work well in IO cases I must echo my earlier reservations in the PR and AO cases; white noise x may be far from synonymous with white noise y.

REFERENCE

MARTIN, R. D. and JONG, J. (1977). Asymptotic properties of generalized M-estimates for the first-order autoregressive parameter. Bell Labs. Technical Memo., Murray Hill, N.J.

DEPARTMENT OF ECONOMICS
LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
ENGLAND