RUEY S. TSAY

Carnegie-Mellon University

I am pleased to see an interesting paper on influence functionals for time series and would like to thank Martin and Yohai for giving me the opportunity to read the paper before its publication. It was more than ten years ago that Fox (1972) formally considered the problem of outliers (or contamination) in time series analysis. But only in recent years, did results of rigorous investigations on the effects of outliers or other deviations from normality appear in the literature. To a large extent, this is due to the complicated dynamic structure of the time series process. As clearly pointed out by Martin and Yohai and by others, any investigation of contamination in time series is inappropriate unless it takes into account the time configuration. With this recognition, it is time to investigate rigorously the contamination problem in time series and to consider seriously its practical implication in applications. I hope that the publication of this paper will mark a new beginning for robustness in time series analysis.

Since many discussants are experts in robustness, I shall confine my comment to the time series part. For simplicity, I use the same notation as Martin and Yohai and assume that the mean value of a time series is zero. First, the idea of using contamination measures $\{\mu_y^\gamma: 0 \le \gamma < 1\}$ in defining influence functionals is a good one. However, from the definition (2.2), one must handle the contaminated process y_t with care whenever $\gamma \neq 0$ because in this case the distributions of the "clean" and "contaminated" observations are different. Take the lag-one correlation coefficient ρ for example. Under the stationarity assumption (this is the case when $\gamma = 0$), $\rho = E(y_t y_{t-1})/E(y_t^2)$ which is independent of time t. On the other hand, when $\gamma \neq 0$ the meaning of ρ is time dependent depending on whether y_t or y_{t-1} is contaminated. Consequently, further clarification is needed in using the general replacement model (2.2). It seems to me that the important assumption is the stationarity of the core process x_t, the contaminating process w_t, and the 0–1 process z_t^γ. Notice that this is related to my comment below on forecasting which is concerned with the underlying generating mechanism of a time series.

Second, from the examples shown in the paper, the influence functional is very much model dependent. It depends not only on the form but also on the parameter values of a model. In practice, neither the model nor its parameter values is known. They must be specified from the data. Therefore, from a practical point of view, one should consider the unknown model as part of the problem in studying the influence of contamination in time series analysis. Based on my limited experience, the problem of model specification is often tangled with the fact that contaminated data tend to show certain nonstationary characteristics that in turn might obscure the picture of possible models.

Finally, forecasting sometimes is the main purpose of a time series analysis. In this case parameter estimation becomes an intermediate step from which the forecasts can be obtained. Suppose now that the series under study follows the contaminated structure (2.2). In this situation, should one construct optimal estimates based on the constraint of bounded gross error sensitivity or should