Journal of Forecasting, Vol. 13, 37–49 (1994)

On Robust Estimation of Threshold Autoregressions

WAI-SUM CHAN
National University of Singapore, Republic of Singapore

SIU-HUNG CHEUNG
Chinese University of Hong Kong, Hong Kong

ABSTRACT

We investigate the effects of additive outliers on the least squares (LS) estimation of threshold autoregressive models. The class of generalized-M (GM) estimates for linear time series is modified and applied to non-linear threshold processes. A Monte Carlo experiment is carried out to study the robust properties of these estimates. Their relative forecasting performances are also examined. The results indicate that the GM method is preferable to the LS estimation when the observations are contaminated by additive outliers. A real example is also given to illustrate the proposed method.

KEY WORDS Non-linear time series Additive outliers Robust estimation Threshold autoregression

INTRODUCTION

Non-linear time series analysis has attracted considerable research interest in recent years. Tong (1990) and De Gooijer and Kumar (1992) have described a very large growth in the literature in this area. One of the useful classes of non-linear time series models is threshold autoregression, which was proposed by Tong (1978). Threshold autoregressive models have been extensively applied to diverse fields, ranging from water pollution (Tong, 1990, p. 278) to stock market returns (Chan, 1990). For a comprehensive account of this class of models, see Tong (1983, 1990).

Although Tsay (1988) and Hau and Tong (1989) have encountered aberrant observations in several non-linear data sets, the outlier problem in threshold autoregressive modelling has not been fully explored. Hau (1984) first attempted to use the M-estimation and robust filtering techniques in estimating threshold models. Later, Hau and Tong (1989) developed an outlier detection statistic based on hat matrix. In this paper, we consider the problem of outliers in non-linear threshold time series models. Simple additive type of outlier models are first defined. The generalized-M estimation techniques of linear time-series modelling are extended to piecewise linear threshold models. We conduct a simulation experiment to study the robust properties of the generalized-M estimates for threshold models. The results show that the GM

0277–6693/94/010037–13\$11.50

Received February 1992
Revised June 1993

estimation method is effective and promising. It can reduce the influence of extraordinary observations in the series and provides better forecasting results.

We are concerned with the estimation of self exciting threshold autoregressive models of the form:

$$X_t = \sum_{l=0}^{p_j} \phi_l^{(j)} X_{t-l} + a_t \qquad \text{if } X_{t-d} \in (r_{j-1}, r_j] \tag{1}$$

where $j = 1, 2, \ldots, k$, $-\infty = r_0 < r_1 < \cdots < r_k = \infty$ are the threshold values, d, k, and $(p_1, p_2, \ldots, p_k)$ are positive integers, and a_t is a sequence of i.i.d. random variables with zero mean and constant variance $\sigma_a^2 < \infty$. We denote model (1) by SETAR(d; $p_1, p_2, \ldots, p_k$) as suggested by Tong (1983).

A straightforward generalization of the definition of an additive time series outlier by Chang *et al.* (1988) is given as follows:

$$Z_t = X_t + \omega I_t^{(T)} \tag{2}$$

where Z_t is the observed time series, X_t follows model (1),

$$I_t^{(T)} = \begin{cases} 1 & \text{if } t = T \\ 0 & \text{otherwise} \end{cases}$$

indicates the location of the outlier, and ω represents the magnitude of the outlier. On the other hand, we can also extend the definition of additive outlier in linear time series by Denby and Martin (1979) in which both the number and the location of outliers are determined by a random mechanism:

$$Z_t = X_t + \eta_t \tag{3}$$

where Z_t's are the observations, η_t's are i.i.d. with density $(1-\alpha)\delta_0(\cdot) + \alpha\pi$, π follows $N(0, \Delta^2\sigma_a^2)$, $\delta_0(\cdot)$ represents a degenerate density at 0, and α is the percentage of contamination $(0 \leqslant \alpha \leqslant 1)$.

The plan of the rest of the paper is as follows. In the next section we describe the least squares estimation method of SETAR models. The class of generalized-M estimates of linear time series is extended to non-linear threshold models in the third section. Simulation studies in the fourth section compare the estimation and forecasting performances of these two methods. The fifth section demonstrates the proposed method through a real example. Some conclusions are given in the final section.

LEAST SQUARES ESTIMATES

In this section we shall describe the procedure of least squares estimation for a two-regime SETAR model. However, the method can be easily generalized to the k-regime case $(k > 2)$ without difficulty when k is known. Let us therefore consider the following SETAR(d; p_1, p_2) model:

$$X_t = \begin{cases} \phi_0^{(1)} + \phi_1^{(1)} X_{t-1} + \cdots + \phi_{p_1}^{(1)} X_{t-p_1} + a_t \text{ if } X_{t-d} \leqslant r \\ \phi_0^{(2)} + \phi_1^{(2)} X_{t-1} + \cdots + \phi_{p_2}^{(2)} X_{t-p_2} + a_t \text{ if } X_{t-d} > r \end{cases} \tag{4}$$

Suppose that we have observations $\{X_1, X_2, \ldots, X_n\}$ and (d, p_1, p_2) is given. We want to estimate the parameter $\Theta = (\Phi^{(1)'}, \Phi^{(2)'}, r)$ where $\Phi^{(j)} = (\phi_0^{(j)}, \ldots, \phi_{p_j}^{(j)})'$, $j = 1, 2$, by the conditional least squares method.

When r is known

Let $p = \max(p_1, p_2, d)$. For fixed r, the effective observations $\{X_{p+1}, \ldots, X_n\}$ may be divided into two regimes by the rule:

$$\begin{cases} X_i \in \text{the first regime iff } X_{i-d} \leqslant r \\ X_i \in \text{the second regime iff } X_{i-d} > r \end{cases} \tag{5}$$

Let $\{X_{1i_1}, X_{1i_2}, \ldots, X_{1i_{n_1}}\}$ and $\{X_{2i_1}, X_{2i_2}, \ldots, X_{2i_{n_2}}\}$ denote the data in the first and second regimes, respectively, after the division. Note that $n_1 + n_2 = n - p$. With each regime of data, we have a linear model of the form:

$$\mathbf{X_j} = \mathbf{A_j}\mathbf{\Phi}^{(\mathbf{j})} + \mathbf{a_j} \tag{6}$$

where

$$\mathbf{X_j} = (X_{ji_1}, \ldots, \mathbf{X}_{\mathbf{ji}_{\mathbf{n}_j}})' \tag{7}$$

$$\mathbf{\Phi}^{(\mathbf{j})} = (\phi_0^{(j)}, \ldots, \phi_{p_j}^{(j)})' \tag{8}$$

$$\mathbf{a_j} = (a_{ji_1}, \ldots, a_{ji_{n_j}})', \text{ and}$$

$$\mathbf{A_j} = \begin{pmatrix} 1 & X_{ji_1-1} & X_{ji_1-2} & \cdots & X_{ji_1-p_j} \\ 1 & X_{ji_2-1} & X_{ji_2-2} & \cdots & X_{ji_2-p_j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{ji_{n_j}-1} & X_{ji_{n_j}-2} & \cdots & X_{ji_{n_j}-p_j} \end{pmatrix} \tag{10}$$

for $j = 1, 2$. Let $\hat{\mathbf{\Phi}}^{(\mathbf{j})}$ denote the least squares estimate of $\mathbf{\Phi}^{(\mathbf{j})}$. Then

$$\hat{\mathbf{\Phi}}^{(\mathbf{j})} = (\mathbf{A_j'}\mathbf{A_j})^{-1}(\mathbf{A_j'}\mathbf{X_j}) \tag{11}$$

for $j = 1, 2$.

When r is unknown

Let $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ be the sorted observations (in ascending order). For any $X_{(1)} < r < X_{(n)}$, the data can be divided into two regimes according to equation (5) and the least squares estimates of $\mathbf{\Phi}^{(1)}$ and $\mathbf{\Phi}^{(2)}$ can be computed from equation (11). The associated residual sum of squares (SSE) is given by

$$\text{SSE}(r) = \sum_{j=1}^{2} \| \mathbf{X_j} - \mathbf{A_j}\hat{\mathbf{\Phi}}^{(\mathbf{j})} \|^2 \tag{12}$$

Following Tong and Lim (1980), we consider the empirical percentiles as candidates for r. We also assume that r is not too close to the zeroth or hundredth percentile. For these cases there are not enough observations to obtain efficient estimates. Moeanaddin and Tong (1988) suggested that r is allowed to vary from $X_{(Q1)}$ to $X_{(Q3)}$, where $X_{(Q1)}$ and $X_{(Q3)}$ are the first and the third quartile of the data, respectively. Therefore, we can search $\hat{r}$, the least squares estimate of the threshold value, by the least squares principle, i.e.

$$\text{SSE}(\hat{r}) = \min_{r \in \{X_{(Q1)}, X_{(Q1+1)}, \ldots, X_{(Q3)}\}} \text{SSE}(r) \tag{13}$$

After obtaining the $\hat{r}$, it is easy to get the corresponding $\hat{\mathbf{\Phi}}^{(1)}$ and $\hat{\mathbf{\Phi}}^{(2)}$ through equations (5) to (11).

GENERALIZED-M ESTIMATES

Denby and Martin (1979), Miller (1980), and Chang *et al.* (1988) showed that the least squares estimates for linear autoregressive parameters not only lack robustness in terms of variability but also suffer from a severe bias problem when the observations are contaminated by outliers. Therefore it is expected that the existence of additive outliers may also cause similar problems in estimating threshold autoregressive models which are piecewise linear. In this section we shall modify and apply the generalized-M estimation techniques to obtain robust estimates for threshold autoregressions.

When r is known

We consider the class of generalized-M estimates (GM estimates) of the SETAR model in equation (4). For given r, we first separate the data into two regimes and obtain the matrices of $\mathbf{X_j}$ and $\mathbf{A_j}$ as in equations (7) and (10), respectively. Then the GM estimate, $\tilde{\mathbf{\Phi}}^{(j)} = (\tilde{\phi}_0^{(j)}, \ldots, \phi_{p_j}^{(j)})'$, of $\mathbf{\Phi}^{(j)}$ is defined as the solution of the following equation:

$$\sum_{k=1}^{n_j} \Lambda\left(\frac{X_{ji_k} - M^{(j)}}{C_X S_X^{(j)}}, \frac{X_{ji_k} - (\tilde{\phi}_0^{(j)} + \Sigma_{l=1}^{p_j} \tilde{\phi}_l^{(j)} X_{ji_k - l})}{C_a S_a^{(j)}}\right) = 0 \tag{14}$$

for $j = 1$ and 2, where $\Lambda(\cdot, \cdot)$ is a bounded robustifying function, $M^{(j)}$ is a robust estimate of the location parameter μ_X in the jth regime, and $S_X^{(j)}$ and $S_a^{(j)}$ are robust estimates of the scale parameters σ_X and σ_a in the jth regime, respectively. C_X and C_a are known as tuning constants because they can be chosen to fine-tune the GM estimators so that they can obtain high efficiencies at Gaussian and non-Gaussian distributions simultaneously (cf. Andrews *et al.* 1972).

Equations (14) may be conveniently solved using the iterative weighted least squares (IWLS) technique suggested by Beaton and Tukey (1974). To obtain the GM estimate of $\Phi^{(j)}$, we use the following IWLS iteration:

$$\tilde{\mathbf{\Phi}}_{g+1}^{(j)} = \tilde{\mathbf{\Phi}}_g^{(j)} + (\mathbf{A}_j'\mathbf{W}_g\mathbf{A}_j)^{-1}\mathbf{A}_j'\mathbf{W}_g(\mathbf{X}_j - \mathbf{A}_j\tilde{\mathbf{\Phi}}_g^{(j)}) \tag{15}$$

where $\tilde{\mathbf{\Phi}}_{g+1}^{(j)}$ is the GM estimate after the gth iteration from an initial estimate $\tilde{\mathbf{\Phi}}_0^{(j)}$ and $\mathbf{W}_g$ is a diagonal matrix with elements

$$\mathbf{W}_g(k, k) = \Lambda\left(\frac{X_{ji_k} - M^{(j)}}{C_X S_X^{(j)}}, \frac{X_{ji_k} - (\tilde{\phi}_0^{(j)} + \Sigma_{l=1}^{p_j} \tilde{\phi}_l^{(j)} X_{ji_k - l})}{C_a S_a^{(j)}}\right) \tag{16}$$

The function $\Lambda(\cdot, \cdot)$ is a 'down-weighting' function to discount the effects of extreme observations and extreme residuals. One commonly used choice of a Λ function is

$$\Lambda(u_1, u_2) = w_0(u_1)w_0(u_2) \tag{17}$$

where w_0 is the redescending Tukey bisquare weight function,

$$w_0(u) = \begin{cases} (1 - u^2)^2 & \text{if } |u| \leqslant 1 \\ 0 & \text{if } |u| > 1 \end{cases} \tag{18}$$

Several other choices of $w_0(\cdot)$ function are available in Hoaglin *et al.* (1983, p. 366).

When r is unknown

For fixed r, the algorithm in equation (15) will converge to the GM estimate of $\Phi^{(j)}$. In fact, the GM estimate is a solution of minimizing the following objective function:

$$\rho^{(j)}(r) = \sum_{k=1}^{n_j} w_0\left(\frac{X_{ji_k} - M^{(j)}}{C_X S_X^{(j)}}\right) L_0\left(\frac{X_{ji_k} - (\phi_0^{(j)} + \Sigma_{l=1}^{p_j} \phi_l^{(j)} X_{ji_k - l})}{C_a S_a^{(j)}}\right) \tag{19}$$

where L_0 is a loss function with respect to w_0. The $L_0(\cdot)$ for Tukey bisquare weights is given by Hoaglin *et al.* (1983, p. 366):

$$L_0(u) = \begin{cases} [1 - (1 - u^2)^3]/6 & \text{if } |u| \leqslant 1 \\ 1/6 & \text{if } |u| > 1 \end{cases} \tag{20}$$

For given r, we have an objective function $\rho^{(j)}(r)$ for each of the two regimes. Thus, we may define an overall objective function by

$$\rho(r) = \sum_{j=1}^{2} \rho^{(j)}(r) \tag{21}$$

A robust estimate $(\tilde{r})$ of r, therefore, can be obtained by searching over the set $\{X_{(Q1)}, X_{(Q1+1)}, \ldots, X_{(Q3)}\}$ such that $\rho(r)$ is minimized, i.e.

$$\rho(\tilde{r}) = \min_{r \in \{X_{(Q1)}, \ldots, X_{(Q3)}\}} \rho(r) \tag{22}$$

After obtaining $\tilde{r}$, $\tilde{\mathbf{\Phi}}^{(1)}$ and $\tilde{\mathbf{\Phi}}^{(2)}$ can easily be computed from algorithm (15).

MONTE CARLO RESULTS

Time-series data of a simple SETAR process,

$$X_t = \begin{cases} \phi^{(1)} X_{t-1} + a_t \text{ if } X_{t-d} \leqslant r \\ \phi^{(2)} X_{t-1} + a_t \text{ if } X_{t-d} > r \end{cases} \tag{23}$$

are generated for a fixed sample size of $n = 100$ with 1000 replications and $\sigma_a^2 = 1$. We consider 18 parameter combinations for $(\phi^{(1)}, \phi^{(2)}, r, d)$. Some of them are taken from Petruccelli (1990). Other combinations are chosen such that there are adequate observations in both regimes for efficient parameter estimation. The start-up value, X_0, is set to zero. It is important to discard a sufficient large number of observations to remove transient effects in generating SETAR time series (Moeanaddin and Tong, 1988). Therefore the first 1500 observations are omitted in each replication. The observations $Z_1, \ldots, Z_n$ are then contaminated according to models (2) or (3). The results for these two models are very similar and hence we only report the case of model (2) in this section. Two separated outlier situations are considered: the single and multiple-outlier case. For the single-outlier case, an additive outlier is located at $T = n/2$ with magnitude ω times the standard deviation of the process. When $|\omega| > 5$, the eyeball method can easily detect most of the additive outliers. To study the practical usefulness of the proposed GM method, we only consider $\omega = 0$, 3, 4, and 5. For the multiple-outlier case, three outliers are fixed at $T = n/4$, $n/2$, and $3n/4$ with magnitude $-\omega, \omega, -\omega$, respectively.

The GM estimates are obtained via the iterative weighted least squares (IWLS) algorithm in equation (15). We employ the sample median as robust estimator $(M^{(j)})$ for location. The scale parameter σ_a in each regime is estimated by the median of the absolute values of the residuals divided by 0.6745, and the scale parameter σ_X in each regime is estimated by the median of the absolute deviations of observations from their sample median divided by 0.6745. The tuning constants are set as $C_X = 6.0$ and $C_a = 3.9$. Following Denby and Martin (1979), four iterations are calculated with the LS estimates as initial points for the GM estimates with Huber weights. The latter values are then used as starting points for the final GM estimates with Tukey bisquare weights. The iterations are stopped when the absolute difference between two consecutive estimates of the parameter is less than 0.0001. Since we assume that both (p_1, p_2) and d are known, the LS and GM estimates can be computed for the case of (1) when r is known and (2) when r is unknown.

First, we study the relative estimation performances of the proposed GM method. When r

Table I. Ratios of the RMSE of the GM estimate to the LS estimate (when r is known)

Parameters				$\hat{\phi}^{(1)}$				$\hat{\phi}^{(2)}$			
$\phi^{(1)}$	$\phi^{(2)}$	r	d	$\omega=0$	3	4	5	$\omega=0$	3	4	5
(a) Single-outlier case											
0.9	−0.1	0.0	1	1.042	0.915	0.851	0.796	0.992	0.930	0.913	0.878
0.9	−0.77	0.0	1	1.051	0.921	0.852	0.795	1.076	0.951	0.879	0.826
−0.5	−1.0	0.0	1	1.072	0.992	0.933	0.873	1.056	0.637	0.429	0.315
−1.0	−0.5	0.0	1	1.084	0.993	0.929	0.866	1.063	0.888	0.757	0.642
0.3	0.8	0.0	1	1.052	0.972	0.914	0.857	1.038	0.968	0.810	0.636
0.5	0.8	0.0	1	1.048	0.956	0.898	0.842	1.036	0.940	0.777	0.614
−0.3	0.8	0.0	1	1.084	1.012	0.955	0.895	1.042	0.999	0.846	0.668
−0.5	0.8	0.0	1	1.065	1.024	0.987	0.945	1.066	1.052	1.032	0.997
0.8	0.3	0.0	1	1.046	0.949	0.891	0.840	1.046	1.016	0.992	0.970
0.8	0.5	0.0	1	1.043	0.940	0.884	0.832	1.044	0.978	0.924	0.843
0.8	−0.3	0.0	1	1.056	0.962	0.901	0.846	1.065	1.019	1.003	0.970
0.8	−0.5	0.0	1	1.058	0.964	0.902	0.847	1.074	1.032	0.992	0.946
0.3	0.8	0.1	1	1.052	0.971	0.913	0.856	1.042	0.968	0.812	0.639
0.3	−0.8	−0.1	1	1.070	0.985	0.961	0.909	1.064	0.901	0.892	0.789
0.3	0.8	0.0	2	1.074	1.067	1.037	0.995	1.012	0.978	0.880	0.752
0.3	−0.8	0.0	2	0.957	0.948	0.940	0.936	1.033	1.004	0.947	0.954
0.3	0.8	0.1	2	1.052	1.033	1.012	0.975	1.016	0.985	0.888	0.757
0.3	−0.8	−0.1	2	0.962	0.950	0.941	0.936	1.037	1.025	1.002	0.988
(b) Multiple-outlier case											
0.9	−0.1	0.0	1	1.042	0.724	0.509	0.361	0.992	0.866	0.822	0.801
0.9	−0.77	0.0	1	1.051	0.720	0.496	0.343	1.076	0.943	0.878	0.851
−0.5	−1.0	0.0	1	1.072	0.804	0.700	0.632	1.056	0.747	0.505	0.360
−1.0	−0.5	0.0	1	1.084	0.574	0.388	0.273	1.063	0.998	0.937	0.808
0.3	0.8	0.0	1	1.052	0.985	0.971	0.967	1.038	0.830	0.666	0.523
0.5	0.8	0.0	1	1.048	0.976	0.921	0.865	1.036	0.816	0.650	0.516
−0.3	0.8	0.0	1	1.084	1.002	0.921	0.863	1.042	0.847	0.671	0.520
−0.5	0.8	0.0	1	1.065	1.006	0.987	0.965	1.066	0.847	0.670	0.519
0.8	0.3	0.0	1	1.046	0.880	0.714	0.562	1.046	0.905	0.848	0.825
0.8	0.5	0.0	1	1.043	0.859	0.697	0.558	1.044	0.875	0.801	0.738
0.8	−0.3	0.0	1	1.056	0.902	0.718	0.554	1.065	0.969	0.944	0.945
0.8	−0.5	0.0	1	1.058	0.902	0.714	0.546	1.074	1.014	0.990	0.928
0.3	0.8	0.1	1	1.052	0.966	0.903	0.833	1.042	0.833	0.668	0.525
0.3	−0.8	−0.1	1	1.070	0.924	0.914	0.906	1.064	0.902	0.888	0.828
0.3	0.8	0.0	2	1.074	1.043	1.007	0.987	1.012	0.808	0.658	0.526
0.3	−0.8	0.0	2	0.957	0.939	0.935	0.934	1.033	1.003	0.986	0.935
0.3	0.8	0.1	2	1.052	1.031	1.003	0.961	1.016	0.806	0.649	0.525
0.3	−0.8	−0.1	2	0.962	0.949	0.935	0.923	1.037	1.015	0.996	0.947

Table II. Ratios of the RMSE of the GM estimate to the LS estimate (when r is unknown)

Parameters				$\hat{\phi}^{(1)}$				$\hat{\phi}^{(2)}$				$\hat{r}$			
$\phi^{(1)}$	$\phi^{(2)}$	r	d	$\omega=0$	3	4	5	$\omega=0$	3	4	5	$\omega=0$	3	4	5
(a) Single-outlier case															
0.9	−0.1	0.0	1	1.073	0.977	0.807	0.753	1.014	1.013	1.005	0.979	1.244	1.547	1.561	1.362
0.9	−0.77	0.0	1	1.109	0.946	0.837	0.717	1.056	1.048	1.009	0.944	1.199	1.455	1.454	1.417
−0.5	−1.0	0.0	1	0.961	0.808	0.719	0.661	1.048	0.706	0.526	0.407	0.912	0.846	0.835	0.869
−1.0	−0.5	0.0	1	1.083	0.857	0.760	0.697	1.005	0.790	0.727	0.672	0.935	0.952	0.976	0.987
0.3	0.8	0.0	1	0.932	0.811	0.561	0.515	1.101	1.041	0.904	0.751	1.123	1.032	1.042	1.036
0.5	0.8	0.0	1	0.819	0.508	0.476	0.414	1.021	0.951	0.834	0.711	0.975	0.899	0.957	0.916
−0.3	0.8	0.0	1	1.092	1.051	0.844	0.792	1.147	1.092	0.938	0.777	1.476	1.424	1.368	1.384
−0.5	0.8	0.0	1	1.021	1.015	1.005	0.997	1.148	1.105	0.969	0.780	1.606	1.509	1.581	1.548
0.8	0.3	0.0	1	1.097	0.981	0.826	0.735	0.963	0.870	0.844	0.896	1.232	1.264	1.322	1.267
0.8	0.5	0.0	1	1.012	0.896	0.795	0.706	0.807	0.812	0.792	0.836	1.072	1.184	1.184	1.191
0.8	−0.3	0.0	1	1.132	1.000	0.885	0.776	1.081	1.015	1.007	0.955	1.415	1.581	1.475	1.449
0.8	−0.5	0.0	1	1.106	0.977	0.869	0.791	1.111	1.072	1.047	1.007	1.485	1.428	1.449	1.468
0.3	0.8	0.1	1	0.911	0.822	0.562	0.517	1.089	1.037	0.907	0.751	1.129	1.203	1.047	1.035
0.3	−0.8	−0.1	1	1.056	1.004	0.900	0.799	1.270	1.198	1.067	0.986	1.329	1.180	1.087	0.999
0.3	0.8	0.0	2	1.011	0.905	0.866	0.766	0.962	0.998	0.965	0.888	0.975	0.970	0.963	0.876
0.3	−0.8	0.0	2	0.834	0.851	0.842	0.855	0.944	0.968	0.967	0.984	0.955	0.918	0.907	0.871
0.3	0.8	0.1	2	1.006	0.899	0.860	0.763	0.979	1.000	0.975	0.905	0.968	0.964	0.955	0.891
0.3	−0.8	−0.1	2	0.838	0.852	0.850	0.843	0.949	0.970	0.970	0.984	0.935	0.912	0.907	0.892
(b) Multiple-outlier case															
0.9	−0.1	0.0	1	1.073	0.959	0.881	0.741	1.014	1.005	0.997	0.987	1.244	1.320	1.387	1.421
0.9	−0.77	0.0	1	1.109	0.978	0.829	0.731	1.056	1.009	0.999	0.985	1.199	1.282	1.316	1.387
−0.5	−1.0	0.0	1	0.961	0.783	0.815	0.796	1.048	0.868	0.698	0.551	0.912	0.820	0.827	0.851
−1.0	−0.5	0.0	1	1.083	0.785	0.724	0.621	1.005	0.866	0.849	0.849	0.935	0.902	0.914	0.930
0.3	0.8	0.0	1	0.932	0.800	0.853	0.872	1.101	0.940	0.800	0.717	1.123	1.150	1.161	1.218
0.5	0.8	0.0	1	0.819	0.826	0.853	0.848	1.021	0.877	0.777	0.662	0.975	1.052	1.031	1.094
−0.3	0.8	0.0	1	1.092	1.041	0.957	0.929	1.147	0.985	0.822	0.709	1.476	1.479	1.356	1.306
−0.5	0.8	0.0	1	1.021	1.008	0.986	0.958	1.148	0.981	0.828	0.704	1.606	1.598	1.464	1.427
0.8	0.3	0.0	1	1.097	1.031	0.974	0.844	0.963	0.760	0.724	0.769	1.232	1.507	1.504	1.539
0.8	0.5	0.0	1	1.012	0.971	0.915	0.830	0.807	0.696	0.738	0.774	1.072	1.223	1.212	1.312
0.8	−0.3	0.0	1	1.132	1.139	1.049	0.922	1.081	1.051	1.004	0.962	1.415	1.658	1.663	1.691
0.8	−0.5	0.0	1	1.106	1.139	1.034	0.919	1.111	1.041	1.007	0.963	1.485	1.613	1.700	1.768
0.3	0.8	0.1	1	0.911	0.802	0.856	0.872	0.089	0.940	0.798	0.720	1.129	1.171	1.155	1.206
0.3	−0.8	−0.1	1	1.056	1.003	0.990	0.976	1.270	1.156	1.051	0.957	1.329	1.392	1.283	1.187
0.3	0.8	0.0	2	1.011	1.069	0.905	0.853	0.962	0.990	0.973	0.852	0.975	0.904	0.910	0.901
0.3	−0.8	0.0	2	0.834	0.768	0.712	0.687	0.944	0.921	0.912	0.904	0.955	0.943	0.918	0.913
0.3	0.8	0.1	2	1.006	1.090	0.915	0.852	0.979	0.999	0.972	0.854	0.968	0.914	0.904	0.901
0.3	−0.8	−0.1	2	0.838	0.754	0.711	0.712	0.949	0.917	0.914	0.908	0.935	0.917	0.887	0.866

is known, the ratios of the RMSE of GM estimate to LS estimate are given in Table I. For $\omega = 0$, most of the ratios are greater than one. It indicates that the LS estimates are better than the GM estimates when there are no outliers. On the other hand, when $\omega = 3$, more than 75% of the ratios are less than one. The GM estimate reduces 43% of the RMSE of the LS estimate in the best case. The performances of the GM method improve with increasing value of ω. When $\omega = 5$, all of the ratios are less than one and the GM estimate enjoys an average 23%

Table III. Forecasting performances. (Entries are ratios of average RMSE of GM forecasts to LS forecasts)

Parameters				When r is known				When r is unknown			
$\phi^{(1)}$	$\phi^{(2)}$	r	d	$\omega = 0$	3	4	5	$\omega = 0$	3	4	5
(a) Single-outlier case											
0.9	−0.1	0.0	1	1.002	0.988	0.986	0.981	1.009	0.985	0.975	0.978
0.9	−0.77	0.0	1	1.005	0.978	0.971	0.952	1.003	0.999	0.993	0.991
−0.5	−1.0	0.0	1	1.003	0.961	0.922	0.883	0.997	0.964	0.926	0.882
−1.0	−0.5	0.0	1	1.003	0.988	0.971	0.954	1.002	0.976	0.955	0.941
0.3	0.8	0.0	1	1.004	0.976	0.951	0.922	1.000	0.982	0.964	0.935
0.5	0.8	0.0	1	1.001	0.988	0.965	0.935	0.995	0.967	0.944	0.920
−0.3	0.8	0.0	1	1.000	0.995	0.979	0.956	1.002	0.989	0.974	0.956
−0.5	0.8	0.0	1	1.007	0.990	0.972	0.948	1.003	0.996	0.992	0.976
0.8	0.3	0.0	1	1.002	0.993	0.987	0.983	1.001	0.989	0.983	0.975
0.8	0.5	0.0	1	1.002	0.991	0.980	0.972	0.996	0.983	0.964	0.960
0.8	−0.3	0.0	1	1.005	0.989	0.980	0.972	1.004	0.999	0.997	0.996
0.8	−0.5	0.0	1	1.002	0.995	0.991	0.987	1.006	0.994	0.988	0.983
0.3	0.8	0.1	1	1.001	0.992	0.971	0.947	0.999	0.982	0.963	0.935
0.3	−0.8	−0.1	1	1.005	0.991	0.982	0.981	1.007	0.997	0.995	0.990
0.3	0.8	0.0	2	1.005	0.999	0.988	0.971	1.003	0.979	0.970	0.968
0.3	−0.8	0.0	2	1.006	1.005	1.001	0.997	0.990	0.990	0.988	0.986
0.3	0.8	0.1	2	1.010	0.988	0.982	0.976	1.002	0.981	0.972	0.971
0.3	−0.8	−0.1	2	1.006	1.004	1.001	0.998	0.988	0.989	0.988	0.984
(b) Multiple-outlier case											
0.9	−0.1	0.0	1	1.002	0.952	0.928	0.841	1.009	0.968	0.946	0.908
0.9	−0.77	0.0	1	1.005	0.968	0.922	0.850	1.003	0.976	0.958	0.930
−0.5	−1.0	0.0	1	1.003	0.968	0.932	0.895	0.997	0.957	0.946	0.910
−1.0	−0.5	0.0	1	1.003	0.956	0.889	0.865	1.002	0.955	0.928	0.885
0.3	0.8	0.0	1	1.004	0.987	0.979	0.950	1.000	0.975	0.953	0.944
0.5	0.8	0.0	1	1.001	0.977	0.943	0.908	0.995	0.964	0.943	0.915
−0.3	0.8	0.0	1	1.000	0.983	0.933	0.824	1.002	0.995	0.972	0.962
−0.5	0.8	0.0	1	1.007	0.996	0.953	0.922	1.003	0.990	0.986	0.966
0.8	0.3	0.0	1	1.002	0.989	0.964	0.934	1.001	0.994	0.976	0.959
0.8	0.5	0.0	1	1.002	0.981	0.926	0.918	0.996	0.980	0.959	0.935
0.8	−0.3	0.0	1	1.005	0.977	0.988	0.965	1.004	0.998	0.995	0.994
0.8	−0.5	0.0	1	1.002	0.995	0.976	0.950	1.006	0.998	0.980	0.978
0.3	0.8	0.1	1	1.001	0.979	0.942	0.920	0.999	0.975	0.954	0.947
0.3	−0.8	−0.1	1	1.005	0.999	0.997	0.991	1.007	0.996	0.994	0.992
0.3	0.8	0.0	2	1.005	0.982	0.952	0.909	1.003	0.986	0.976	0.970
0.3	−0.8	0.0	2	1.006	0.991	0.941	0.896	0.990	0.991	0.982	0.979
0.3	0.8	0.1	2	1.010	0.960	0.944	0.914	1.002	0.986	0.975	0.970
0.3	−0.8	−0.1	2	1.006	1.000	0.999	0.997	0.988	0.990	0.981	0.978

reduction in RMSE for all models considered. In general, the GM method performs better in the multiple-outlier case as compared to the single-outlier case.

When r is unknown, the simulation results are given in Table II. The performances of the GM estimates for the threshold autoregressive parameters ($\phi^{(1)}$ and $\phi^{(2)}$) are very similar to those in Table I. However, the improvement of the GM estimate over the LS estimate for the threshold value r is not so impressive. The reason for this phenomenon is related to the algorithm of searching the least squares $\hat{r}$. Following Moeanaddin and Tong (1988), we search the $\hat{r}$ from the first quartile $X_{(Q1)}$ to the third quartile $X_{(Q3)}$ of the data. This procedure not only guarantees adequate observations in each regime but also 'trims' out the additive outliers as potential candidates for $\hat{r}$. The least squares estimate $\hat{r}$ obtained by this method has a built-in resistance to additive outliers.

Second, we study the relative forecasting performances for these two estimation methods. Ten one-step-ahead forecasts and the associated RMSE are computed for each estimated model. The ratios of the average RMSE of GM forecasts to LS forecasts over the 1000 replications are reported in Table III. The results show that the GM method provides better forecasts when the observations are contaminated by additive outliers.

AN APPLICATION

We consider the monthly average wholesale prices of regular leaded gasoline in the United States between January 1973 and December 1987. The data are listed in Liu (1991) and the

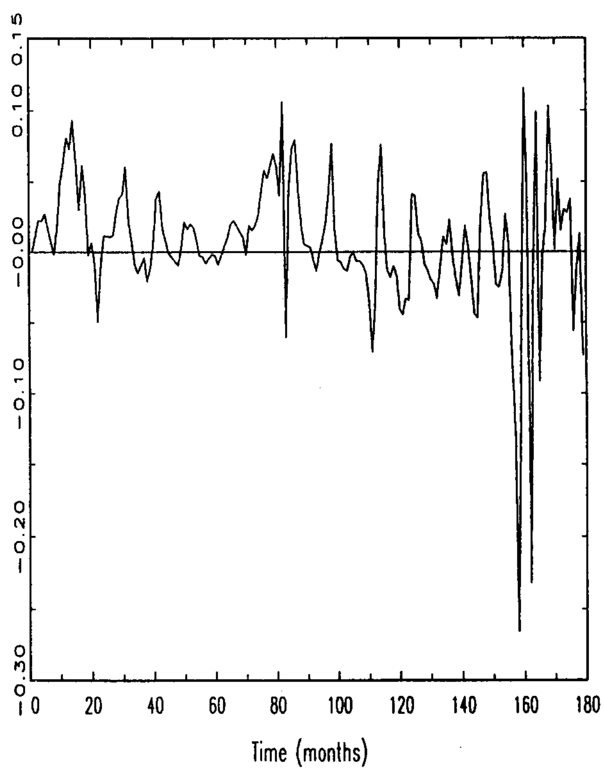

Figure 1. Time plot of the GAS data

series is non-stationary and volatile. Therefore we shall only consider the first differenced series of the logarithmic transformed data in this study. A time-series plot of the transformed series is given in Figure 1. The plot displays some large jumps between $t = 150$ and $t = 168$. It indicates that non-linear threshold and/or outlier models may be useful for forecasting purposes.

As the first step in our analysis, we apply the test for threshold-type non-linearity proposed by Tsay (1989) to the GAS data. The sample autocorrelation function of the series is given in Table IV and it indicates the AR order $p = 2$. Following the suggestions by Tsay (1989), the delay parameter $d = 2$ is selected. The F-statistic of the test is then sequentially computed with $p = 2$, $d = 2$, and n rolling from 120 to 179. The results are plotted in Figure 2. The critical value of the test is around 4.0 at the 1% level.

It is evident that the series is non-linear with some outliers located between $t = 155$ and $t = 168$. These outliers create aberrant F-values in Figure 2. Moeanaddin and Tong (1988) have pointed out that most tests for non-linearity are quite sensitive to outliers. Therefore, we shall employ the first 150 observations ($t = 1$ to 150) for model specification; the first 169 observations ($t = 1$ to 169) for model estimation; and the last 10 observations ($t = 170$ to 179) for calculation of post-sample one-step ahead forecasts.

For comparison purpose, we build a linear time-series model to the GAS data. From the sample autocorrelations in Table IV, an AR(2) process is specified. Both the LS and GM

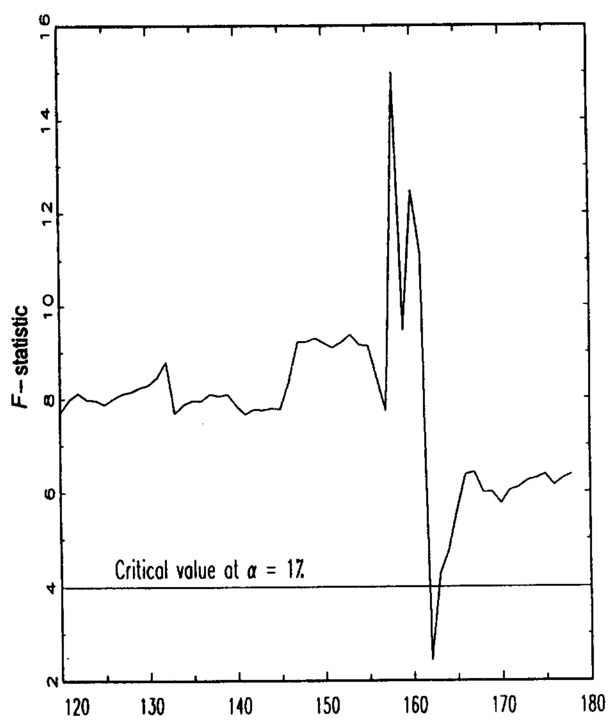

Figure 2. Test for threshold-type non-linearity in the GAS data

Table IV. Sample autocorrelation of the GAS data (first 150 observations)

k	1	2	3	4	5	6	7	8
$\hat{\rho}(k)$	0.62	0.35	0.17	0.10	0.09	0.09	0.03	0.01
Standard error	(0.08)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
k	9	10	11	12	13	14	15	16
$\hat{\rho}(k)$	0.07	0.18	0.23	0.19	0.10	0.03	0.03	0.11
Standard error	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)
k	17	18	19	20	21	22	23	24
$\hat{\rho}(k)$	0.08	0.03	0.02	−0.00	−0.00	0.00	−0.04	−0.12
Standard error	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)

Table V. Estimation results and forecasting performances of various models for the GAS data

	Lag				
Regime	0	1	2	MAE	RMSE
(a) Linear AR(2) using LS estimation					
	0.0050	0.5139	−0.2106	0.0336	0.0437
(b) Linear AR(2) using GM estimation					
	0.0031	0.2333	−0.2422	0.0317	0.0422
(c) Non-linear SETAR(2; 2, 0) using LS estimation (with estimated threshold value $\hat{r} = 0.0249$)					
(1)	−0.0005	0.7185	−0.4331		
(2)	0.0225			0.0311	0.0417
(d) Non-linear SETAR(2; 2, 0) using GM estimation (with estimated threshold value $\tilde{r} = -0.0102$)					
(1)	−0.0017	0.9288	−0.4422		
(2)	0.0186			0.0298	0.0414

Notes:
MAE = Mean Absolute Error
RMSE = Root Mean Squared Error

estimates are computed. The mean absolute error (MAE) and the root mean squared error (RMSE) for ten post-sample one-step-ahead forecasts are obtained.

Following the AIC identification procedure by Tong (1983, pp. 228–302), we specify a SETAR(2; 2, 0) model for the series. It is interesting to note that both methods from Tong (1983) and Tsay (1989) are consistent in choosing $p = \max(p_1, p_2) = 2$ and $d = 2$. The LS and the proposed GM estimates are calculated. Post-sample forecasting comparisons are also obtained.

Table V summarizes the results. There is a 4.18% reduction in MAE of the GM threshold model over the LS threshold model and a 11.31% reduction over the LS linear AR model. Results of using RMSE as comparison criterion are very similar. In summary, the overall performance of the proposed GM method is reassuring in this example.

CONCLUSIONS

We have demonstrated that the GM estimates have advantages over the LS estimates in estimating the threshold autoregressive parameters in the presence of outliers. Our study, however, concentrates on the additive type of outliers. The generalization of the innovational outlier model in linear time series to threshold models will be a research topic of some interest. However, Chan (1989, p. 89) recognized the potential difficulties of innovational outliers in threshold models. An innovational outlier in a threshold model not only affects the observations inside the original regime but also may force an erroneous shift of the dynamic structure to another regime. Hence, direct application of the M-estimation techniques as in Hau (1984) may not be appropriate. Furthermore, Moeanaddin and Tong (1988) reported that some tests for non-linearity are quite sensitive to outliers and tend to regard a linear series with outliers as non-linear. Therefore, it is interesting to derive robust tests for non-linearity. Research in some of these topics is in progress.

ACKNOWLEDGEMENTS

We are grateful to Professor Robert H. Shumway for his suggestions that led to the results in the fifth section. We also thank two anonymous referees for their helpful comments on an earlier version of the paper.

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W., *Robust Estimates of Location: Survey and Advances*, Princeton University Press, 1972.

Beaton, A. E. and Tukey, J. W., 'The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data', *Technometrics*, **16** (1974), 147–85.

Chan, W. S., *Some Robust Methods for Time Series Modelling*, unpublished PhD dissertation, Department of Statistics, Temple University, USA, 1989.

Chan, W. S., 'On tests for non-linearity in Hong-Kong stock returns', *Hong Kong Journal of Business Management*, **VIII** (1990), 1–11.

Chang, I., Tiao, G. C. and Chen, C., 'Estimation of time series parameters in the presence of outliers', *Technometrics*, **30** (1988), 193–204.

De Gooijer, J. G. and Kumar, K., 'Some recent developments in non-linear time series modelling, testing and forecasting', *International Journal of Forecasting*, **8** (1992), 135–56.

Denby, L. and Martin, R. D., 'Robust estimation of the first order autoregressive parameter', *Journal of the American Statistical Association*, **74** (1979), 140–46.

Hau, M. C., *Robust-Resistant Estimation in Threshold Autoregressive Time Series Modelling*, Unpublished MPhil Thesis, Chapter 4, Department of Statistics, Chinese University of Hong Kong, 1984.

Hau, M. C. and Tong, H., 'A practical method for outlier detection in autoregressive time series modelling', *Stochastic Hydrology and Hydraulics*, **3** (1989), 241–60.

Hoaglin, D., Mosteller, F. and Tukey, J., *Understanding Robust and Exploratory Data Analysis*, New York: John Wiley, 1983.

Liu, L.-M. 'Dynamic relationship analysis of US gasoline and crude oil prices', *Journal of Forecasting*, **10** (1991), 521–47.

Miller, R. D. 'Comments on "Robust estimation of autoregressive models"', in Brillinger, D. R. and Tiao, G. C. (eds), *Direction in Time Series*, Hayward, CA: IMS, 1980.

Moeanaddin, R. and Tong, H., 'A comparison of likelihood ratio test and CUSUM test for threshold autoregression', *The Statistician*, **37** (1988), 213–25.

Petruccelli, J. D., 'A comparison of tests for SETAR-type non-linearity in time series', *Journal of Forecasting*, **9** (1990), 25–36.

Tong, H., 'On a threshold model', in Chan C. H. (ed.), *Patter Recognition and Signal Processing*, Netherlands: Sythoff and Noordhoff, 1978.

Tong, H., *Threshold Models in Non-linear Time Series Analysis*, New York: Springer-Verlag, 1983.

Tong, H., *Non-linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press, 1990.

Tong, H. and Lim, K. S., 'Threshold autoregression, limit cycles and cyclical data', *Journal of the Royal Statistical Society*, **B42** (1980), 245–292.

Tsay, R. S., 'Non-linear time series analysis of blowfly population', *Journal of Time Series Analysis*, **9** (1988), 247–63.

Tsay, R. S., 'Testing and modeling threshold autoregressive processes', *Journal of the American Statistical Association*, **84** (1989), 231–240.

Authors' biographies:

Wai-sum Chan is Lecturer in the Department of Economics and Statistics of the National University of Singapore. He holds an MPhil in Statistics from the Chinese University of Hong Kong and a PhD in Statistics from Temple University, USA. He has published articles in the *Journal of Forecasting*, *Journal of Applied Statistics*, *Computational Statistics & Data Analysis*, and *Asia Pacific Journal of Management*.

Siu-hung Cheung is Lecturer of Statistics at the Chinese University of Hong Kong. He received a doctorate in Statistics from Temple University, USA. His current research interests include multivariate analysis, simultaneous inference, and time-series analysis. He has published articles in *Biometrics* and *Computational Statistics & Data Analysis*.

Authors' addresses:

Wai-sum Chan, Department of Economics and Statistics, National University of Singapore, Singapore 0511.

Siu-hung Cheung, Department of Statistics, Chinese University of Hong Kong, Hong Kong.