

An Improved Iterative Procedure for Outlier Detection in Time Series

Anh Tuan Bui · Chi-Hyuck Jun[†]

Department of Industrial and Management Engineering, Pohang University of Science and Technology,
Pohang, 790-784, Korea

시계열 이상치 탐지를 위한 개선된 반복적 절차

Tuan Anh Bui · 전치혁

포항공과대학교 산업경영공학과

We address some potential problems with the existing procedures of outlier detection in time series. Also we propose modifications in estimating model parameters and outlier effects in order to reduce the number of tests and to increase the detection accuracy. Experiments with some artificial data sets show that the proposed procedure significantly reduces the number of tests and enhances the accuracy of estimated parameters as well as the detection power.

Keywords: Additive Outlier, Innovational Outlier, ARMA

1. Introduction

Outlier detection in time series was first introduced by Fox (1972) with two types of outliers, additive and innovational. When the timing of outliers is known, they can be detected using test of hypothesis (Wei, 2006). Chang (1982) proposed an iterative procedure to detect multiple outliers when the timing of outliers is unknown. Tsay (1988) generalized this procedure to detect four types of outliers. Chen and Liu (1993) proposed a procedure (T-C) using three stages in order to jointly estimate model parameters and outlier effects. Stage I initially detects potential outliers. Stage II jointly estimates outliers' effect and model parameters through iterations. Stage III identifies outliers using the final model parameter estimate obtained in Stage II. This procedure is now implemented in the SCA

statistical system (SCA, 1994). Many authors also adapt this procedure to their outlier detection procedure for multivariate time series (Galeano *et al.*, 2006) and GARCH processes (Charles and Darné, 2005; Franses and Ghijssels, 1999).

However, there are still potential problems incurred by T-C procedure. Sánchez and Peña (2003) made some refinements for three of these problems: swamping effect caused by bias initial parameters estimation, confusion between innovational outliers and level shifts in the presence of a level shift, and finally masking effect. In this paper, we also address the first problem but in a different aspect concerning the joint estimation iterative procedure. In addition, we raise another potential problem encountered by not only T-C, but also other iterative procedures for outlier detection for their redundant test statistics used throughout their procedures. We then propose two solutions for the former problem,

This work was supported by Basic Science Research Program through National Research Foundation of Korea funded by the Korea government (MEST)(Project No.2011-0012879).

[†] Corresponding author : Professor Chi-Hyuck Jun, Department of Industrial and Management Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang, 790-784, Korea, Fax : +82-54-279-2870, E-mail : chjun@postech.ac.kr

Received July 14, 2011; Accepted November 9, 2011.

and introduce a new lower bound to resolve the latter one while maintaining accuracy.

The organization of this paper is as follows : in Section 2, we address potential problems with existing procedures. Then, in Section 3, we suggest solutions and propose our new procedure. Experiments with our proposed procedure are presented in Section 4. Discussion and conclusion is shown in Section 5.

2. Existing Procedures and Potential Problems

In general, a time series may contain multiple outliers of different types. Outliers are usually divided into four types : innovational outlier (IO), additive outlier (AO), level shift (LS) and temporary change (TC). Here, we will consider IO and AO types for simplicity. LS and TC types can be considered with a little modification. Let $\{X_t\}$ be an outlier free ARMA process as follows : $\phi(B)X_t = \theta(B)a_t$ where $\phi(B)$ is the AR polynomial, $\theta(B)$ is the MA polynomial, and a_t is a white noise. Roots of $\phi(B)$ and $\theta(B)$ lie outside unit circle to satisfy stationary and invertible conditions. Suppose that there are k outliers of either IO type or AO type. Let t_i be the timing of the i^{th} outlier ($i = 1, \dots, k$). Then, the observation at time t , Z_t , can be represented by

$$Z_t = \sum_{i=1}^k w_i v_i(B) I_t^{(t_i)} + X_t \quad (1)$$

where w_i denotes the initial magnitude (effect) of outlier i , $v_i(B) = \theta(B)/\phi(B)$ for an IO type, $v_i(B) = 1$ for an AO type, and

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

Stage I of T-C is essentially the procedure proposed by Tsay (1988). It initially detects potential outliers using the following tests of hypotheses for all observations when the null hypothesis (H_0) is that there are no outliers at Z_t : (i) Existence of an IO : H_0 vs. H_1 ; (ii) Existence of an AO : H_0 vs. H_A .

Stage II jointly estimates outliers' effect and model parameters through iterations. First, using information of outliers and model parameters obtained from Stage I, outliers' effects are estimated simultaneously by the following multiple regression model based on residuals \hat{e}_t 's.

$$\hat{e}_t = \sum_{i=1}^k w_i \pi(B) v_i(B) I_t^{(t_i)} + a_t \quad (2)$$

where $\pi(B) = \phi(B)/\theta(B)$. After the effects of significant outliers are removed based on equation (1), model parameters are re-estimated. This process is repeated until stop condition is met. The stop condition is a user-defined tolerance constant, which is the relative change of the residual standard error from the previous estimate.

Stage III identifies outliers using the final model parameter estimates obtained in Stage II.

However, there are potential problems with this procedure. First, like all other iterative procedures in literature, this procedure uses hypothesis testing for every observation in all iterations. Since there are many normal observations in a series, this strategy leads to unnecessary heavy computation when there are many observations, multivariate series, outliers, or complicated tests applied.

Second, the joint estimation of outliers' effect and model parameters is now a very popular term and commonly used technique in recent research. Nevertheless, there is a potential problem with this iterative procedure that may lead to wrong estimation of model parameters as well as outliers detection. When outliers are not correctly identified (e.g. swamping effect) and there are interactions among fake outliers and true outliers, the joint estimation will produce biased estimates. As a result, the iterative procedure will lead to wrong estimation of model parameters. Consequently, outliers are wrongly detected in the final stage. To illustrate this problem, we generate an artificial series from AR(2) model with $\Phi(B) = 1 - 1.1B + 0.5B^2$, where white noise follows a Gaussian distribution with mean 0 and variance 1. The length of this series is 100 observations <Figure 1>. Four outliers of AO and IO types are added randomly to this series. The resulting artificial series is called as Series 1. <Table 1> shows these true outliers in Series 1 as well as outliers detected in Stage I and III of T-C. <Table 2> shows the maximum likeli-

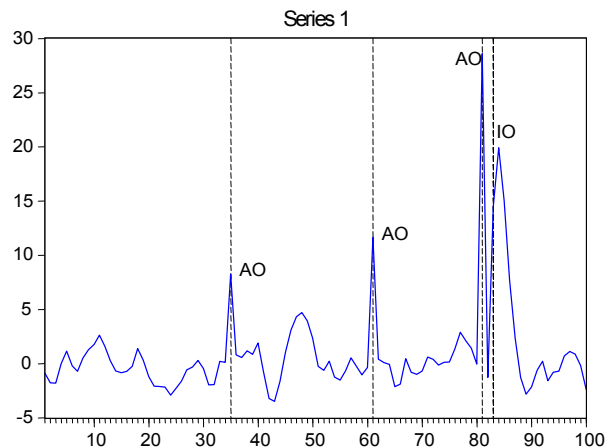


Figure 1. Plot of Series 1

Table 1. Outliers detected from T-C for Series 1

True outliers			T-C(Stage I)			T-C(Stage III)		
Time	Type	Magnitude	Time	Type	Magnitude	Time	Type	Magnitude
35	AO	8.28	35	AO	7.91	35	AO	7.77
61	AO	11.20	61	AO	11.89	61	AO	11.43
81	AO	27.95	81	IO	28.18	81	AO	29.57
83	IO	16.94	82	AO	-9.06	83	IO	15.48
			83	AO	4.74	84	IO	-12.10
			84	IO	16.41	86	IO	9.54
			85	AO	6.23	87	IO	11.28
						88	IO	6.64

hood estimates (MLE) of model parameters of : original series without outliers, series contaminated by those four outliers, and outcomes of Stage I and II of T-C. Notice that we will use the maximum likelihood method for all estimations in this paper.

Table 2. Parameters estimates according to methods for Series 1

Model parameter	Φ_1	Φ_2
MLE with outliers deleted	1.0835	-0.4812
MLE with outliers	0.2676	0.3037
T-C(Stage I)	1.1363	-0.4458
T-C(Final)	2.0656	-1.6529

As can be seen in <Table 2> the estimates of model parameters in the presence of outliers are distorted. This leads to a wrong detection of outliers in Stage I and causes the problem of joint estimation. The outcome of this stage is a wrong model parameter for Stage III. Consequently, the final result of outlier detection may not be correct.

3. Proposed Procedure

To deal with the first problem mentioned in Section 2, we propose the use of a lower bound (CL, say) to identify normal observations in the series. The idea is simple : after the first iteration, all observations whose test statistic values smaller than CL will be considered as normal ones (those observations will be called as *normal list*) and will not be tested in the next iteration. In general, CL could take any value in the range (0, C) without affecting performance of our proposed procedure (C is the critical value used for the statistical tests). We suggest choosing CL of around 2 because in our experience, most of the normal observations have

test statistics value < 2 . To avoid the masking effect, all observations will be tested after no more outliers are found in the previous iteration.

To improve the iterative procedure for the joint estimation, we propose adding the initial outlier detection step that re-detects outliers using the original series and the most recent model parameters obtained in Stage I of T-C. The purpose of Stage I is then divided into two separated tasks : initial parameter estimation (the same with Stage I of T-C) and then, initial outlier detection. The initial outlier detection step starts with a less outlier-contaminated model parameter estimate from (as we see on <Table 2>, estimate of model parameter obtained after Stage I of T-C is quite close to the outlier free series'). Therefore, it is expected that the outcome of this modification will be more accurate.

However, using an additional step does not ultimately guarantee to resolve the problem of joint estimation. In case the initial estimated model is seriously distorted by outliers, the outcome of Stage I may be still biased. We propose another simple but effective method to deal with this problem by taking advantage from the following observation : residual standard error of the model parameter estimation will decrease as outliers are correctly detected and their effects are removed, but when the problem of joint estimation stage happens, this standard error will increase. <Figure 2> (a) shows residual standard error of model parameter estimations throughout the procedure for the above series. Notice that, this value first decreases to nearly 1 as outliers are detected and their effects are removed, but then increases when the problem of joint estimation occurs. <Figure 2>(b) shows a similar plot for another artificial series for which the problem of joint estimation doesn't happen. The residual standard error decreases over time and eventually converges to 1. In both cases, the residual standard errors of the series without outliers are approximately 1 because these series were generated with white noises following standard normal distribution.

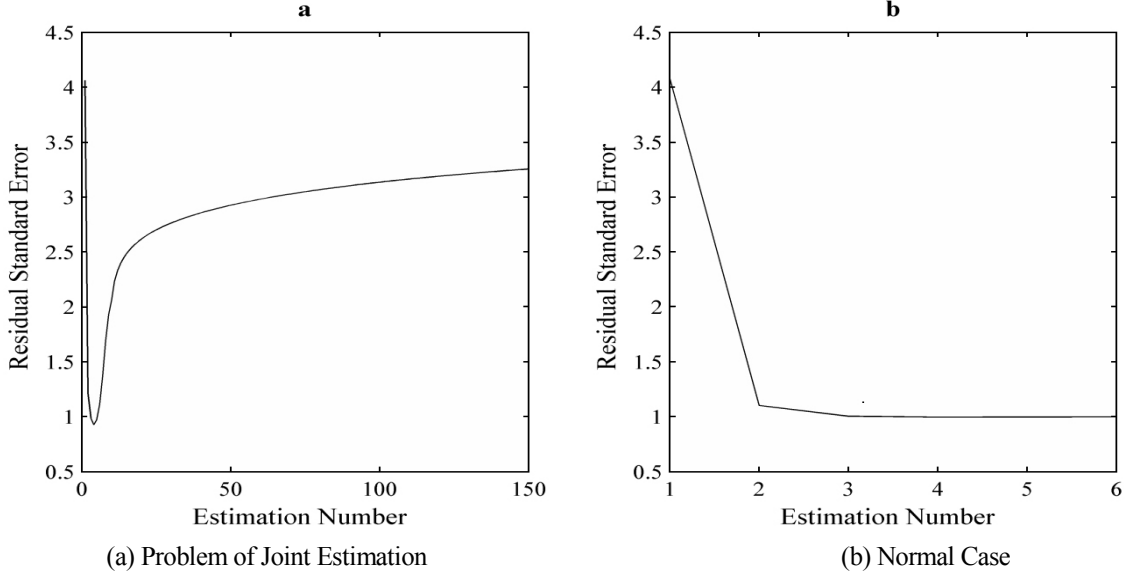


Figure 2. Residual Standard Errors Versus Estimation Numbers in T-C

Therefore, our proposal is to keep track of this residual standard error. If this error in Stage II increases to the value larger than that of Stage I, stops Stage II and use estimated model parameters corresponding to minimum residual standard error as the final model parameter.

• **Proposed procedure :**

Stage I: Initial parameter estimation and outlier detection

I.1 Initial parameter estimation

I.1.1 Estimate the model parameters using MLE method based on the original or the adjusted series. The original series is used only for the first iteration of I.1. Compute the residuals :

$$\hat{e}_t = \hat{\pi}(B)Z_t = \hat{\phi}(B)/\hat{\theta}(B)$$

Keep track of the residual standard error every time the estimation is made. Set normal list = ϕ .

Inner Loop :

I.1.2 For each observation that does not belong to normal list, perform the outlier tests :

- Existence of an IO, H_I vs. H_0 : $\lambda_{I,T} = \frac{\hat{\omega}_{I,T}}{\sigma_a}$
- Existence of an AO, H_A vs. H_0 : $\lambda_{A,T} = \tau_{A,T} \frac{\hat{\omega}_{I,T}}{\sigma_a}$

where $\hat{\omega}_{I,T} = e_T$; $\hat{\omega}_{A,T} = \frac{e_T - \sum_{i=1}^{n-T} \pi_j e_{T+j}}{\tau_{A,T}^2}$;
 $\tau_{A,T}^2 = \sum_{j=0}^{n-T} \pi_j^2$; π_j is the coefficient of B^j in polyno-

mial $\pi(B)$. $\hat{\sigma}_a$ is estimated using median absolute deviation method :

$$\hat{\sigma}_a = 1.4826 \times \text{median}|\hat{e}_t - \text{median}(\hat{e}_t)|$$

An observation whose test statistic is at maximum and larger than C is declared as an outlier (either IO or AO). If no outliers are found in the very first iteration of this inner loop, then stop the procedure-the observed (original) series is free from outlier effects.

I.1.3 If there is an outlier found, then remove the effect of this outlier from the residuals and the observations according to its type, add observations whose test statistic is smaller than CL to normal list, set doublecheck = false, then go back to I.1.2 to check if an additional outlier can be found. If there is no additional outlier found, then go to I.1.4.

I.1.4 If outliers are found in the inner loop and doublecheck = false, then reset normal list = ϕ , set doublecheck = true, and go back to I.1.2. If doublecheck = true, then go to I.1.1 to revise the parameter estimates. If no additional outliers are detected after revising parameter estimates, go to I.2.

I.2 Initial outlier detection

Repeat I.1 using the original series after fixing the model parameters to the final estimates from I.1 in order to detect the timing and the type of an outlier if any.

Stage II: Joint Estimation of Outlier Effects and Model Parameters

II.1 Assuming that k outliers are found, estimate outliers' effects simultaneously using multiple linear re-

gressions based on equation (2). Delete the least insignificant outlier s such that $\hat{\tau}_s = \min\{\hat{w}_j / \text{std}(\hat{w}_j)\} \leq C$, $j = 1, 2, \dots, k$. Repeat until all outliers left are found to be significant.

II.2 Remove the outliers' effects that are significant to obtain the adjusted series based on equation (1) and revise the model parameters. Keep track of the residual standard error every time the estimation is made.

II.3 Repeat II.1 and II.2 until either one of the following two conditions is met :

- The relative change in the residual standard error $< \varepsilon$ (defined by users).
- The residual standard error does not increase to the value larger than the smallest one from step I.1. If this is the case, the model parameter corresponding to the smallest residual standard error out of all in the records will be the final one.

Stage III : Final Outlier Detection :

III.1 Perform the same procedure as in I.2 using the final model parameter obtained in Stage II.

III.2 Estimate outliers' effects simultaneously and delete insignificant outliers similarly to Stage II.

4. Experiments

Our proposed procedure consists of three modifications that can be used independently as well as jointly without any conflict. Nevertheless, for easier understanding and comparison of these modifications' effectiveness, we will show experiments of each modification separately in Section 4.1 through 4.3. The experiment in Section 4.3 is actually conducted using the full proposed procedure. For these experiments, we use Series 1 mentioned in Section 2 and generate two more series (Series 2 and Series 3-<Figure 3>) with white noises following Gaussian distribution with mean 0 and variance 1. They are summarized in <Table 3>. Outliers of IO and AO types are randomly added to these series. The information about these outliers will

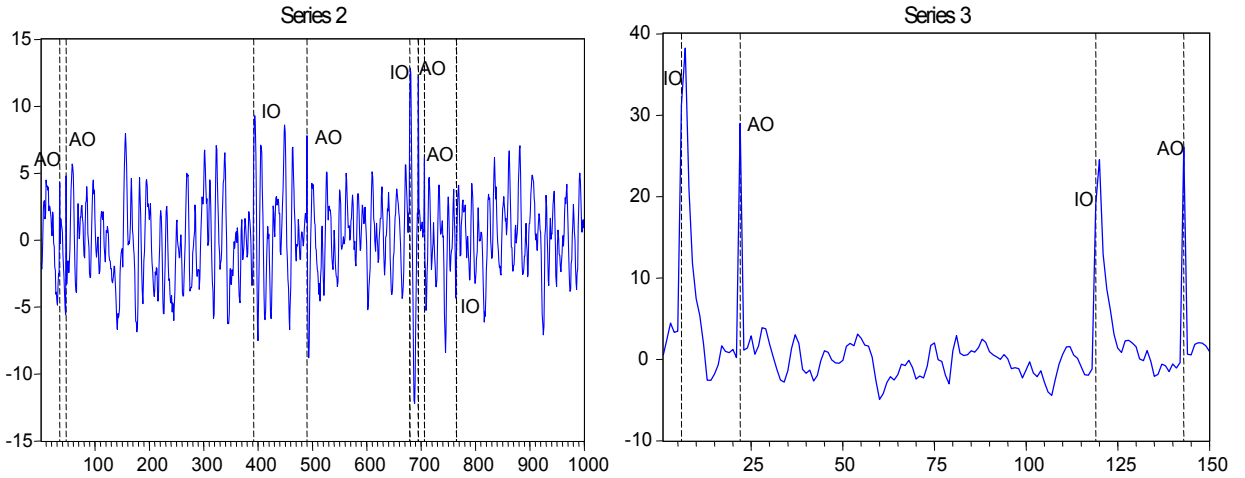


Figure 3. Plot of Series 2 and Series 3

Table 3. Artificial Series Used in Experiments

	Series 1			Series 2			Series 3		
Length	100			1,000			150		
True Model	$(1-1.1B+0.5B^2)Z_t = a_t$			$(1-1.5B+0.7B^2)Z_t = a_t$			$(1-0.6B)Z_t = (1+0.8B)a_t$		
	Time	Type	Magnitude	Time	Type	Magnitude	Time	Type	Magnitude
True Outliers	35	AO	8.28	35	AO	6.29	6	IO	28.28
	61	AO	11.20	47	AO	8.94	22	AO	27.81
	81	AO	27.95	392	IO	5.03	119	IO	21.05
	83	IO	16.94	490	AO	7.88	143	AO	24.81
				679	IO	8.46			
				695	AO	9.70			
				706	AO	5.66			
				765	IO	5.12			

be shown in the results of experiments for easy comparisons.

In Section 4.4, we compare performance of our proposed procedure with T-C on 50 series. These series are generated in the same manner with Series 3. In all of our experiments, we use median absolute deviation method to compute residuals. We set $C = 4$ and $CL = 2$ for experiments in Section 4.1 through 4.3. We use $C = 3.5$ and $CL = 2.5$ for experiment in Section 4.4.

4.1 Effect of Lower Bound

By introducing the lower bound, the proposed procedure will reduce the number of tests required significantly as shown in <Table 4>. The computation time will be reduced accordingly. However, the lower bound does not affect the parameter estimation.

Table 4. Total Number of Tests Using Lower Bound

	T-C	Using lower bound
Series 1	4,000	1,988
Series 2	42,000	18,552
Series 3	4,800	2,132

4.2 Effect of Adding Initial Outlier Detection stage

<Table 5> summarizes the result of parameter estimates by T-C and the proposed one adding initial outlier detection stage. <Table 6> and <Table 7> show the outlier detection results for 3 series from T-C and proposed procedure, respectively. For Series 2, both procedures produce the same estimates and the same

Table 5. Parameter Estimates According to Methods for 3 Series

Model parameter	Series 1		Series 2		Series 3	
	Φ_1	Φ_2	Φ_1	Φ_2	Φ_1	Θ_1
MLE without outliers added	1.0835	-0.4812	1.4962	-0.6833	0.5476	-0.7756
T-C	2.0656	-1.6529	1.4946	-0.6825	0.6368	0.0092
Proposed(Stage I)	1.1175	-0.4939	1.4946	-0.6825	0.6222	-0.0143

Table 6. Outliers Detected from T-C for 3 Series

Series 1			Series 2			Series 3		
Time	Type	Magnitude	Time	Type	Magnitude	Time	Type	Magnitude
35	AO	7.77	35	AO	6.09	6	IO	29.12
61	AO	11.43	47	AO	9.42	7	IO	18.34
81	AO	29.57	392	IO	5.58	22	AO	28.38
83	IO	15.48	490	AO	8.41	119	IO	19.76
84	IO	-12.10	679	IO	8.22	120	IO	12.40
86	IO	9.54	695	AO	10.26	143	AO	25.95
87	IO	11.28	706	AO	6.42			
88	IO	6.64	764	AO	-2.29			

Table 7. Outliers Detected from Proposed Procedure(Stage I) for 3 Series

Series 1			Series 2			Series 3		
Time	Type	Magnitude	Time	Type	Magnitude	Time	Type	Magnitude
35	AO	7.78	35	AO	6.09	6	IO	29.33
61	AO	11.44	47	AO	9.42	9	IO	16.94
81	AO	29.56	392	IO	5.58	10	AO	9.86
83	IO	15.42	490	AO	8.41	22	IO	28.62
			679	IO	8.22	119	IO	18.92
			695	AO	10.26	122	IO	12.73
			706	AO	6.42	143	AO	25.77
			764	AO	-2.29			

outlier detection results. For Series 1, as seen in the example in Section 2, T-C suffers from the joint estimation problem. But when adding the initial outlier detection stage, this problem is now resolved. Thus, our proposed procedure leads to better estimates and outlier detection results. However for Series 3, adding initial outlier detection stage is not sufficient to resolve this problem; both of procedures produce biased estimates that lead to wrong outlier detection results as shown in the last three columns in <Table 7>.

4.3 Effect of Using Minimum Residual Standard Error

In the previous section, we see that including the initial outlier detection stage may not be sufficient to resolve the problem of joint estimation for Series 3. However, using minimum standard error as a criterion, we are able to detect this problem and obtain quite better parameter estimates of the final model. Consequently, the outlier detection performance of this method is superior. This is shown in <Table 8>. Notice that the results in <Table 8> are conducted using our full procedure. Hence, the total number of tests needed includes tests conducted in the initial outlier detection stage. Nevertheless, it is still less than that of T-C. For Series 1 and Series 2, the problem of joint estimation does not occur or is already resolved; thus our proposed procedure does not provoke the use of minimum standard error. The outlier detection results by our proposed procedure for Series 1 and Series 2 will be the same as the one only using Stage I (<Table 7>).

Table 8. Results of Full Proposed Procedure for Series 3

Total number of tests	Model parameters		Outlier detection		
	Φ_1	Θ_1	Time	Type	Magnitude
3352	0.5646	-0.6504	6	IO	28.12
			22	AO	28.06
			119	IO	19.55
			143	AO	25.21

Table 9. Comparison of Estimation Performance for 50 Series

Case	Method	Φ_1		Θ_1		σ	
		Mean	RMSE	Mean	RMSE	Mean	RMSE
1	C-L	0.58	0.10	0.73	0.21	1.18	0.59
	Proposed	0.58	0.09	0.81	0.08	0.99	0.03
2	C-L	0.61	0.12	0.58	0.36	1.41	0.88
	Proposed	0.59	0.08	0.78	0.07	0.99	0.03
3	C-L	0.55	0.13	0.48	0.48	1.59	0.82
	Proposed	0.59	0.08	0.81	0.07	0.97	0.04

4.4 Comparison of Performance

In this section, we compare performance of our proposed procedure with T-C on 50 series as mentioned above. We delete some series (around two cases in 50 series) for which T-C procedure takes too much iteration in joint estimation stage to converge. There are three cases that are different in the way outliers added. In Case 1, outliers have magnitudes follow uniform distribution $U(3, 20)$ with random types AO or IO. Case 2 is similar to Case 1 but outliers' magnitudes follow $U(20, 30)$. Case 3 has only AO type, and magnitudes follow $U(20, 30)$. <Table 9> shows the estimation performance and <Table 10> shows the outlier detection performance and number of tests. Notice that σ is the estimation of residual standard error. Obviously, the proposed procedure produces better estimation and outlier detection performance than T-C while using fewer tests. In addition, the proposed procedure has similar estimation error in all three cases. Hence, it is more robust to different outliers' types and magnitudes. Interestingly, T-C seems to have worst performance when the series have only AO type.

Table 10. Comparison of Outlier Detection Performance for 50 Series

Case	Method	# of tests	Detection rate	Spurious outliers
1	C-L	4932±881	94%	0.72
	Proposed	3678±513	95%	0.48
2	C-L	5466±1275	96%	1.36
	Proposed	3955±1812	96%	0.60
3	C-L	5040±2876	90%	0.26
	Proposed	4232±1950	91%	0.12

5. Discussion and Conclusion

In this paper, we address two potential problems with the existing outlier detection procedure : unnecessary

computation and problem of joint estimation. Then we suggest solutions for these problems. For the first problem, we introduce a lower bound (CL) to reduce the number of tests needed to detect outliers. Our experiment shows the significant effect of this bound. This effect will be even more meaningful when the lower bound notion is applied for procedures with large sample sizes, multivariate series, and/or complex tests. Certainly, the choice of CL value also affects the number of tests. In our experience, most of the normal observations have test statistics value < 2 ; thus we suggest to use CL value of 2.

For the second problem, we first propose the inclusion of the initial outlier detection stage before the joint estimation stage. Our experiment shows the effectiveness of this stage. In fact, our idea is quite similar to that of Sánchez and Peña (2003) in terms of splitting Stage I of T-C procedure into two specialized tasks: initial parameters estimation and initial outlier detection, but different in terms of method used. Then, we suggest using the minimum standard error as a criterion to detect the problem of joint estimation and determine the final model parameter. Our proposal successfully solves the problem of joint estimation as shown in our experiment. The result of estimated model parameter is quite good. In case a higher accuracy is required, we can use this final model parameter to initialize another run of the whole procedure. This requires a small modification for the proposed procedure that is to allow user-defined initialized model parameter. What's more, as far as we know, there is no guarantee for the convergence of joint estimation stage. Hence, our proposal can serve a good stopping criterion for all other procedures that use joint estimation before diverging to bias estimates.

However, there is still a limitation in this research

that is we did not implement LS and TC types in our procedure. It would be interesting to see how LS and TC affect the joint estimation, and how our proposed procedure deals with it. In future, we will work on this issue to develop a full procedure for outlier detection in time series.

References

- Chang, I. (1982), Outliers in Time Series, Unpublished Ph.D. thesis, University of Wisconsin-Madison, Dept. of Statistic.
- Charles, A. and Darné, O. (2005), Outliers and GARCH models in financial data, *Economic Letters*, **86**, 347-352.
- Chen, C. and Liu, L.-M. (1993), Joint estimation of model parameters and outlier effects in time series, *Journal of the American Statistical Association*, **88**, 284-297.
- Fox, A. J. (1972), Outliers in time series, *Journal of the Royal Statistical Society, Ser. B*, **34**, 350-363.
- Franses, P. H. and Ghijsels, H. (1999), Additive Outliers, GARCH and Forecasting Volatility, *International Journal of Forecasting*, **15**, 1-9.
- Galeano, P., Peña, D., and Tsay, R. (2006), Outlier detection in multivariate time series by projection pursuit, *Journal of the American Statistical Association*, **101**, 654-669.
- Sánchez, M. J. and Peña, D. (2003), The identification of multiple outliers in ARIMA models, *Communications in Statistics: Theory and Methods*, **32**, 1265-1287.
- SCA (Scientific Computing Associates) (1994), The SCA statistical system reference manual, version 4.3. Oak Brook, SCA.
- Tsay, R. S. (1988), Outliers, level shifts, and variance changes in time series, *Journal of Forecasting*, **7**, 1-20.
- Wei, W. W. S. (2006), Time Series Analysis: Univariate and Multivariate Methods, second ed. New York, Addison-Wesley.