

Detection Power when outliers are present at or near the end of time series

이종선*, 안미혜*, 이재준*

요약

시계열 모형을 따르는 자료의 예측(Forecasting)이나 공정조정(Process Adjustment)의 경우, 자료의 마지막 부분에 발생한 이상치(Outlier)에 의해 크게 영향 받을 수 있다. 그러나 지금까지 제안된 이상치 탐지 방법은 주로 자료의 중간 부분에 발생한 이상치를 검출하는데 효율적이라고 알려져 왔다. 본 연구에서는 자료의 마지막 부분에 발생한 이상치에 대한 기존 탐지 방법의 검출력을 모의 실험을 통해 분석하였다. 또한, 이를 개선할 수 있는 방안을 제시하고, 모의 실험을 통해 기존의 검출력과 비교하였다.

주요 용어: 이상치, 이상원인, 공정관리, 공정조정

1. 서론

시계열 자료에서의 이상치의 발생을 판단하는 방법으로는 Box & Tiao(1975), Chang, Tiao & Chen(1988)이 그 방법들을 제시하고 있다. 특히, Chen & Liu(1993)는 여러 이상치가 발생한 것에 대한 검출 및 그 영향을 제거한 모형의 추정에 대해 제시하였다. Chen & Liu(1993)에 의한 시뮬레이션 결과를 보면 그림 1. a, b, c과 같이 시계열 자료의 끝에서 이상치가 발생한 경우에, 자료의 중간 부분에 이상치가 발생한 경우에 비해 이상치의 검출력이 떨어짐을 확인할 수 있다.

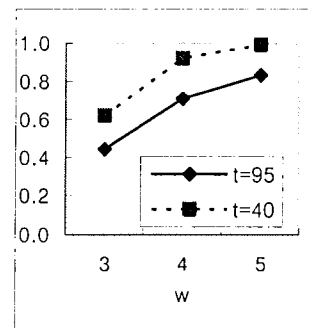
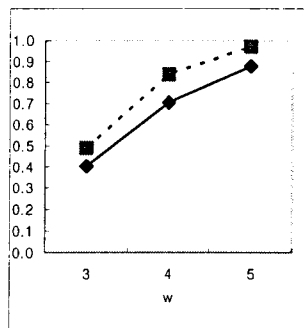
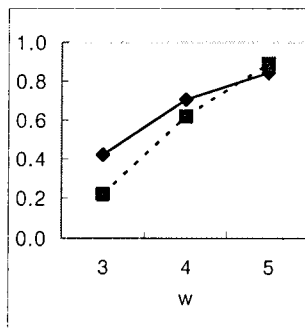


그림 1.a: MA(1)모형, AO type,

그림 1.b: AR(1)모형, IO type

그림 1.c: IMA(1,1)모형, LS type

* (402-751) 인천광역시 남구 용현3동, 인하대학교 이과대학 통계학과

2. 최근 자료만을 이용한 이상치의 검출

이상치 검출의 범위를 시계열 자료 전체에 대해 적용하는 방법은 실제 공정관리에 있어서는 효율적이지 못하다. 최근에 관측된 자료에 이상원인에 의한 이상치가 포함되었는지를 판단하여 공정 조정(Process Adjustment)의 필요성을 판단해야 하기 때문이다. 따라서 Chen & Liu(1994)가 제안한 통계량을 이용하여 최근 자료에서 발생한 이상치의 검출 비율을 비교하기 위하여 크기가 105인 AR(1), MA(1), IMA(1,1) 각 세 가지 모형을 따르는 자료를 임의로 생성하여, 이상치를 검출하는 시뮬레이션을 각각 1000회 실시하였다. 이 시뮬레이션에서는 100번째에 임의로 이상치를 주었다. 최근 95개의 자료를 이용한 이상치 판단 결과(nblk=95)와 최근 5개 자료만을 사용하여 이상치를 판단한 결과(nblk=5)가 다음 그림 2.a, b, c, d, e, f에 나타나있다.

<그림 2>의 결과에 의하면, 이상치의 크기(w)를 3으로 주었을 경우에는 최근 자료만을 이용한 이상치 판정이 전체 자료를 활용하여 이상치를 판단하는 것보다 검출력 뛰어남을 알 수 있다. 또한 검출 시점에 상관없이 최근 5개의 자료만을 이용하는 것이 이상치의 유형을 판단하는 검출력이 우수하게 나타났다.

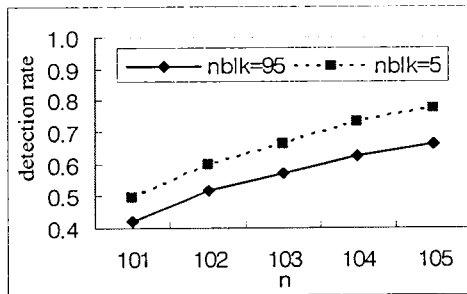


그림 2.a: AR(1) - w=3, LS type

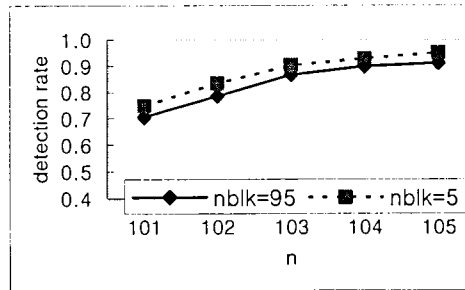


그림 2.b: AR(1) - w=4, LS type

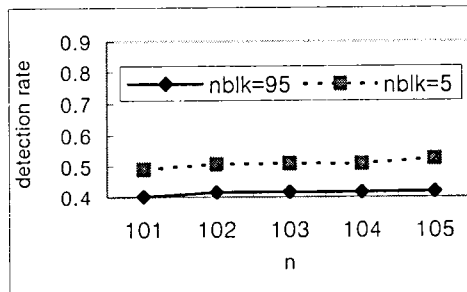


그림 2.c: MA(1) - w=3, IO type

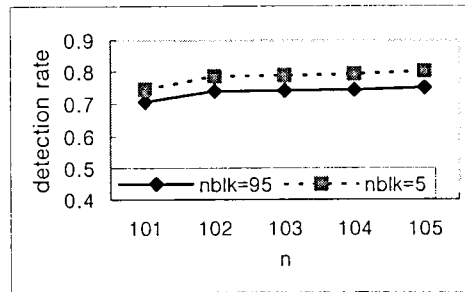


그림 2.d: MA(1) - w=4, IO type

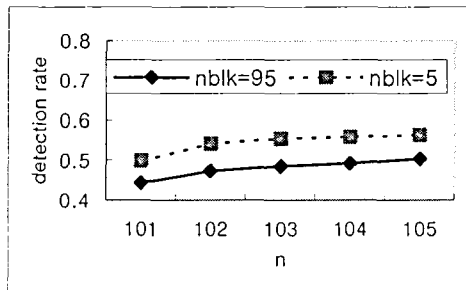


그림 2.e: IMA(1) - w=3, AO type

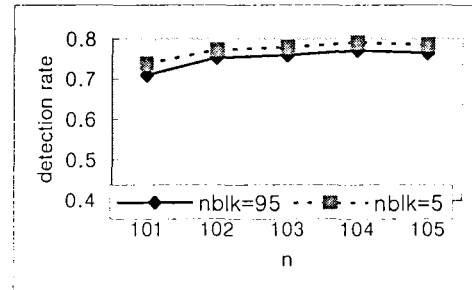


그림 2.f: IMA(1) - w=4, AO type

3. 결론 및 향후 연구 방향

시계열 모형을 따르는 특성치에 대한 공정조정에 있어서, 이상원인의 판단 여부는 아주 중요하다. 본 연구를 통해 제시된 방법은 최근 자료에 대한 이상치의 검출 및 유형 판단이 보다 효율적으로 가능함을 보였다. 따라서 향후 공정조정 혹은 예측에 있어, 최근 자료의 이상치 판정에 유용하게 활용될 수 있으리라 본다. 향후에 다양한 시계열 모형에 대한

참고 문헌

- [1] Box, G.E.P. & Tiao, G. C. (1975). Intervention Analysis with applications to economic and environmental problems, *J. Amer. Statist. Assoc.*, **70**, 70-79.
- [2] Chang, I., Tiao, G. C., and Chen C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, **30**. 193-204
- [3] Chung Chen and Lon-Mu Liu (1993). Joint estimation of model parameters and outlier effects in time series model. *J. Amer. Statist. Assoc.* **88**. 284-297.