



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학석사학위논문

서포트 벡터 머신을 활용한
변화점 탐지

2019년 2월

서울대학교 대학원

통계학과

문민음

초록

본 논문에서는 시계열 자료가 정상적인 선형 자기회귀이동평균모형을 따르지 않으리라고 기대되는 경우 비선형, 비모수적 방법인 서포트 벡터 회귀를 이용하여 변화점 검정을 하는 방법을 소개한다. 더불어 캐나다 달러/미국 달러 환율 자료와 영국 아카데미 백본망 통신량 자료를 통하여 실제 자료를 분석하는 방법을 제시한다. 시계열 모형을 적합하기 위해 두 자료 모두 변환 과정을 거쳤으며, 선형 모형과 서포트 벡터 회귀를 이용한 모형을 비교한다.

주요어 : 변화점 탐지, 서포트 벡터 회귀, ARMA 모형, ORCUSUM 검정, 환율 예측, 통신량 예측

학번 : 2017-27707

목차

| | |
|------------------------------------|----|
| 1. 서론 | 1 |
| 2. 이론적 배경 | 3 |
| 2.1 서포트 벡터 회귀 | 3 |
| 2.2 선형 ARMA 모형에서 ORCUSUM 검정 | 4 |
| 2.3 SVR-ARMA 모형에서 ORCUSUM 검정 | 5 |
| 3. 실제 자료 분석 사례 | 6 |
| 3.1 캐나다 달러/미국 달러 환율 자료 | 6 |
| 3.2 영국 아카데미 백본망 통신량 자료 | 14 |
| 4. 결론 | 21 |

그림목차

| | |
|--|----|
| 3.1.1: 환율 원 자료 | 7 |
| 3.1.2: 환율 자료의 표본자기상관계수 | 7 |
| 3.1.3: 로그 차분한 환율 자료 | 8 |
| 3.1.4: 로그 차분한 환율 자료의 표본자기상관계수 | 8 |
| 3.1.5: 환율 자료에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 10 |
| 3.1.6: 환율 자료에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 11 |
| 3.1.7: SVR-ARMA 모형을 통해 추정된 변화점을 표시한 환율 자료 | 12 |
| 3.1.8: SVR-ARMA 모형을 통해 추정된 변화점을 표시한 로그 차분된 환율 자료 | 12 |
| 3.1.9: 환율 자료에서 2002년 12월 23일 이전 모형의 잔차에 대한 Ljung-Box 검정 결과 | 13 |
| 3.1.10: 환율 자료에서 2002년 12월 23일 이후 모형의 잔차에 대한 Ljung-Box 검정 결과 | 13 |

| | |
|--|----|
| 3.2.1: 통신량 자료의 시계열그림 | 16 |
| 3.2.2: 통신량 자료의 표본자기상관계수 그림 | 16 |
| 3.2.3: 통신량 자료의 차분 직후 그림 | 17 |
| 3.2.4: 통신량 자료의 차분 후 앞, 뒤 일부를 제외한 시계열그림 | 17 |
| 3.2.5: 통신량 자료의 차분 후 앞, 뒤 일부를 제외한 자료의 표본자기상관계수 그림 | 18 |
| 3.2.6: 통신량 자료의 최종 시계열 자료; ARIMA(0, 0, 0)(1, 0, 0) 모형을 적합시킨 뒤 모형의 잔차 | 18 |
| 3.2.7: 통신량 자료 앞부분에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 19 |
| 3.2.8: 통신량 자료 뒷부분에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 19 |
| 3.2.9: 통신량 자료 앞부분에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 20 |
| 3.2.10: 통신량 자료 뒷부분에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과 | 20 |

1. 서론

시계열 자료는 자료가 생성된 확률모형이 시간이 흐르더라도 유지된다는 가정에 따라 분석을 하며, 이에 기초하여 확률모형에 대한 이해 및 해석을 진행하며 미래에 대한 예측값을 얻는다(조신섭 등, 2016). 시계열 분석 방법은 다양하나, 공통으로 필요한 점은 가정한 확률모형이 실제 자료를 잘 설명해야 한다는 것이다. 따라서 시계열 자료가 어떤 사건에 의해 중간에 영향을 받아 자료를 생성한 시계열 모형이 바뀐다면 이를 분석에 반영해야 한다. 실제로 많은 시계열 자료가 구조의 변화를 갖는 것으로 알려져 있고, 이를 무시하고 잘못된 모형에 적합하는 경우 이후 분석에 문제가 생기므로 변화점 탐지(change point detection)는 시계열 자료 분석에서 중요한 부분이다.

시계열 자료에 영향을 미칠 만한 사건의 시점이 알려진 경우 한 가지 방법은 외부 효과를 인위적으로 분석에 반영시키는 Box and Tiao (1975)가 제안한 개입모형(intervention model)이 있다. 한편 모형이 변화하는 시점이 알려지지 않은 경우에도 개입모형을 이용하거나(Tsay, 1988)(Chang 등, 1988) 강건한(Robust) 추정 방법을 이용하는 방법(Denby and Martin, 1979) 등 많은 방법이 개발되어 왔다. 이 중 Page (1954)가 통계적 공정 관리(statistical quality control)를 위해 제안한 CUSUM 검정(cumulative sum test)은 대표적인 변화점 탐지 방법이며, IID(independent and identically distributed) 가정하에서 평균(mean) 및 분산(variance)을 추정(estimate)하는 방식으로 주로 쓰여왔다.

이후 기존의 CUSUM 방법을 개선하기 위해 residual-based CUSUM과 이를 개선한 ORCUSUM(observation and residual-based CUSUM) 등이 등장하였다(Lee, 2018). ORCUSUM은 일단 잔차(residual)를 구할 수 있는 방법이라면 항상 검정을 위한 통계량을 계산할 수 있는 특징이 있기 때문에, 일반적으로 쓰이는 선형 모형(linear model)뿐만 아니라 비선형 모형(nonlinear model), 그중에서도 비모수적 모형(nonparametric model),

준모수적 모형(semiparametric model) 등으로 확장 가능하다는 장점 또한 가지고 있다.

한편 비모수적 방법을 통해 시계열 자료를 예측하려는 시도는 예전부터 있었다. Chen 등 (2008)은 서포트 벡터 머신(Support vector machine)을 이용하여 비선형 ARMA 모형을 추정했고, Khashei and Bijari (2010)는 ARIMA 모형(auto-regressive integrated moving average)의 잔차에 인공신경망(artificial neural network)을 적합시켜 예측의 정확도를 향상하려 하였다. 또한, 변화점 탐지에도 서포트 벡터 머신(Camci, 2010), 인공신경망을 이용한 사례들이 존재한다(Oh 등, 2005)(Amiri 등, 2015).

인공신경망은 전통적인 시계열 모형과 비교할 때 모형을 수립하는 이론적인 배경이 부족하므로, 모형을 수립할 때 시행착오를 거쳐야 하며 결과를 해석할 때 어느 정도 한계가 존재한다(Moreno 등, 2011). 한편 서포트 벡터 머신과 ARMA 모형(auto-regressive moving average)을 결합하는 경우 차수를 결정하는 과정 등 모형을 설정하는 과정에 대한 논의가 부족한 경우가 많다. 그리고 현재 서포트 벡터 머신, 인공신경망 등 머신러닝을 이용해 변화점을 검정하는 방법은 일반적으로 평균, 분산 등 기초적인 통계량에 대한 검정이라는 한계가 있다. 본 논문에서는 Lee 등 (2018)과 같이 ARMA 모형의 차수를 사전에 설정하지 않고 자료를 통해 최적의 차수를 결정하며, 모수의 변화에 대한 검정과 같은 보다 일반적인 변화점 검정 방법을 사용하고자 한다.

앞으로의 내용 순서는 다음과 같다. 2.1에서는 서포트 벡터 회귀(support vector regression)의 개념과 관련된 용어를 정의한다. 2.2에서는 선형 ARMA(linear auto-regressive moving average) 모형에서 ORCUSUM 통계량 식을 정의한다. 2.3에서는 Lee 등 (2018)을 참고하여 SVR이 어떻게 ARMA 모형과 ORCUSUM 검정에 적용될 수 있는지 간략히 살펴본다. 이후 시계열 자료로 3.1 환율 자료와 3.2 인터넷 통신량 자료 두 가지 경우에 대해 기존의 선형 ARMA와 SVR-ARMA를 이용한 비모수적 방법을 각각 적용해보며 두 방법론을 비교한다. 마지막으로 4에서는 결론을 제시한다.

2. 이론적 배경

2.1 서포트 벡터 회귀

서포트 벡터 회귀(Support vector regressor)는 서포트 벡터 머신(Support vector machine)의 확장으로, 다양한 함수의 근사와 시계열 예측에서 우수한 일반화 성능을 보여 왔다(Mukherjee 등, 1997)(Müller 등, 1997).

서포트 벡터 회귀를 이용해 함수를 근사할 때는 주어진 훈련 데이터 세트(training data set) $(x_i, y_i) (i = 1, \dots, M)$ (x_i 는 i 번째 m 차원 입력 벡터, y_i 는 i 번째 스칼라)에 대해 입력 공간(input space)을 고차원의 특징 공간(feature space)로 보내어 최적의 초평면 $f(x) = w^T \phi(x) + b$ 를 찾게 된다. 여기서 w 는 l 차원 가중치 벡터이며, $\phi(x)$ 는 x 를 l 차원 특징 공간으로 사상시키는(mapping) 함수이며, b 는 편향(bias) 항이다. 서포트 벡터 회귀에서 잔차(residual) r 에 대한 손실 함수(loss function)는 양수 ϵ 에 대하여 $E(r) = I(|r| - \epsilon)$ 으로 주어지며, 최적화 문제는 이차 계획법(quadratic programming)을 통하여 풀 수 있다(Abe, 2010).

2.2 선형 ARMA 모형에서 ORCUSUM 검정

ARMA (p,q) 모형 $y_t = \sum_{i=1}^p \phi_i y_{t-1} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$ 가 정상성(stationarity)

을 만족할 때, 적당한 조건 하에서 ORCUSUM 검정을 위한 통계량 \widehat{T}_n 은 아래와 같이 주어진다(Lee, 2018).

$$\widehat{T}_n = \max_{1 \leq k \leq n} \left(\frac{1}{n\widehat{\tau}_{1,n}^2} \left| \sum_{t=1}^k (y_t - \widehat{\epsilon}_t) \widehat{\epsilon}_t - \frac{k}{n} \sum_{t=1}^n (y_t - \widehat{\epsilon}_t) \widehat{\epsilon}_t \right|^2 \right. \\ \left. + \frac{1}{n\widehat{\tau}_{2,n}^2} \left| \sum_{t=1}^k \widehat{\epsilon}_t^2 - \frac{k}{n} \sum_{t=1}^n (y_t - \widehat{\epsilon}_t) \widehat{\epsilon}_t \right| \right)$$

$$\widehat{\tau}_{1,n}^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \widehat{\epsilon}_t)^2 \widehat{\epsilon}_t^2$$

$$\widehat{\tau}_{2,n}^2 = \frac{1}{n} \sum_{t=1}^n |\widehat{\epsilon}_t|^4 - \left(\frac{1}{n} \sum_{t=1}^n |\widehat{\epsilon}_t|^2 \right)^2$$

모수(parameter)의 변화가 없다는 귀무가설 하에서 \widehat{T}_n 은 점근적으로 2차원 표준 브라운 운동(Brownian motion)의 상한(Supremum)임을 유도할 수 있고(Lee, 2018), \widehat{T}_n 이 사전에 정해진 유의수준(significant level)에 해당하는 상방 분위수(upper quantile)보다 값이 커지면 귀무가설을 기각하게 된다.

2.3 SVR-ARMA 모형에서 ORCUSUM 검정

선형 ARMA 모형과 비선형 모형까지 포함하는 일반적인 ARMA(p,q) 모형은 $y_t = f(y_{t-1}, \dots, y_{t-p}, \epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-q})$ 와 같이 나타낼 수 있다. Lee 등 (2018)에서는 그 중에서 오차항에 대해 가법(additive) 모형인 $y_t = f(y_{t-1}, \dots, y_{t-p}, \epsilon_{t-1}, \dots, \epsilon_{t-q}) + \epsilon_t$ 의 모형을 가정하고 $\hat{y}_t = \hat{f}(y_{t-1}, \dots, y_{t-p}, \epsilon_{t-1}, \dots, \epsilon_{t-q}) = \hat{f}(x) = w^T \phi(x) + b$ 를 서포트 벡터 회귀를 통해 추정한다. 본 논문에서는 $\phi(x)$ 를 위한 커널(kernel)로 일반적으로 많이 쓰이는 방사 기저 함수(radial basis function) $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ 을 사용하였으며, 최적의 초모수(hyperparameter)는 ARMA 모형의 차수 (p,q)는 Lee 등 (2018)의 방식을, 서포트 벡터 회귀에 필요한 ϵ, γ, C (비용;cost) 등은 그리드 검색(grid search)을 이용하였다.

일단 서포트 벡터 회귀를 통해 모형을 추정하면, 관측치와 잔차를 결합하여 2.2의 ORCUSUM 통계량 \hat{T}_n 을 선형 ARMA 모형과 마찬가지로 계산할 수 있다. 3.1과 3.2에서는 \hat{T}_n 값을 통해 변화점 유무를 살펴보고, 통계량 계산에 사용된 모형이 실제로 잘 적합 되었는지 Ljung-Box 검정을 통해 확인한다.

3. 실제 자료 분석 사례

3.1 캐나다 달러/미국 달러 환율

첫 번째 사례는 캐나다 달러/미국 달러 환율(Canada/U.S. Foreign Exchange Rate)이다. 자료는 FRED(<https://fred.stlouisfed.org/>)에서 확인할 수 있으며, 1 미국 달러에 해당하는 캐나다 달러를 기록하였다. 1971년 1월 1일부터 2018년 11월 30일까지 12,500개의 관측치가 기록되어 있다. 본 논문에서는 변동성(volatility)을 제어하기 위해 자료를 시차 14(2주) 간격으로 재추출하였다.

시계열 그림(그림 3.1.1)을 보면 특별히 증가하거나 감소하는 추세는 보이지 않으나 강한 자기상관(autocorrelation)이 의심되며, 그림 3.1.2를 볼 때 표본자기상관계수(sample autocorrelation function)가 서서히 감소하고 있으므로 단위근이 존재하는 것으로 보인다. Trapletti and Hornik (2018)을 참고하여 $[(n-1)^{1/3}]$ (단, n 은 시계열 관측치의 개수)을 시차 차수(lag order)로 설정한 뒤 실시한 Augmented Dickey-Fuller(ADF) 단위근 검정에서 유의확률이 0.6367로 단위근이 존재한다는 귀무가설을 유의수준 0.05에서 기각할 수 없었다. 금융 및 경제 시계열 자료에서 많은 경우 원 자료를 로그 변환 후 차분한 자료 $r_t = \log(x_{t+1}) - \log(x_t)$ 를 이용하기 때문에(이상열, 2013) 본 분석에서도 로그변환 후 차분을 한 자료를 이용한다. 로그 차분 결과 시계열자료(그림 3.1.3) 및 표본자기상관계수(그림 3.1.4)가 상당히 안정된 것으로 보이며, ADF 단위근검정 결과 유의확률이 0.01 미만으로 단위근이 존재한다는 귀무가설을 유의수준 0.05에서 기각할 수 있었다.

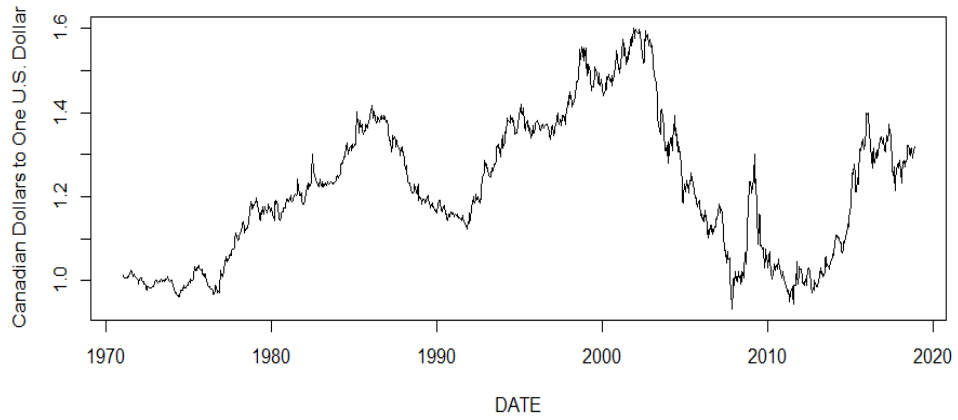


그림 3.1.1: 환율 원 자료

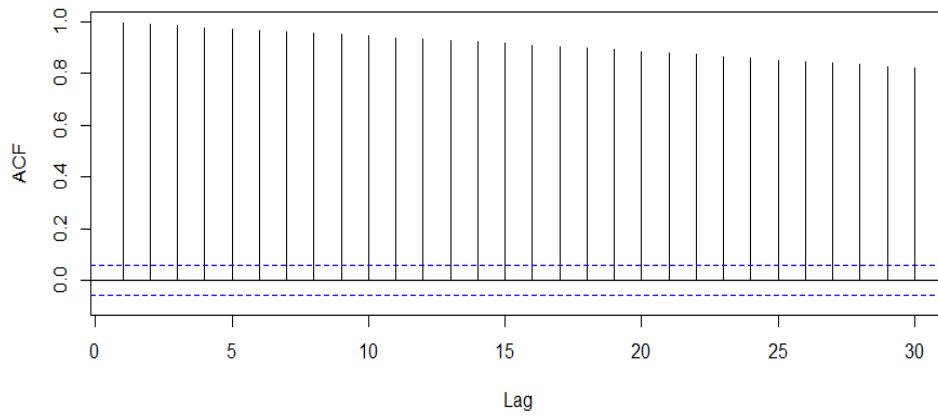


그림 3.1.2: 환율 자료의 표본자기상관계수

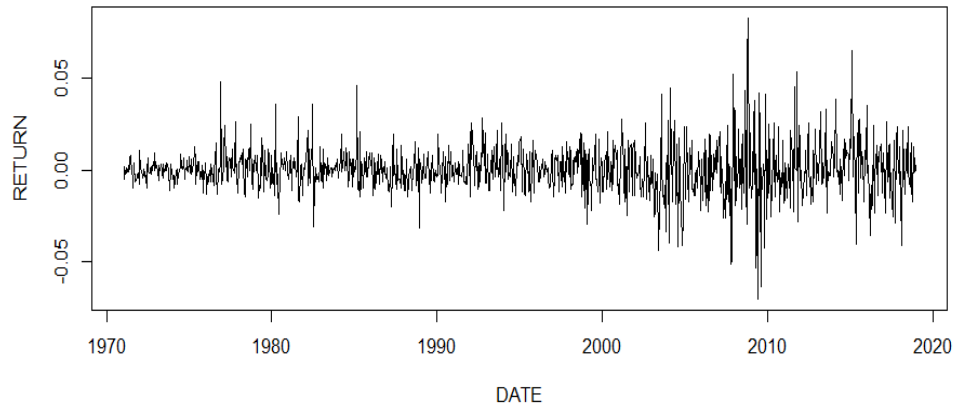


그림 3.1.3: 로그 차분한 환율 자료

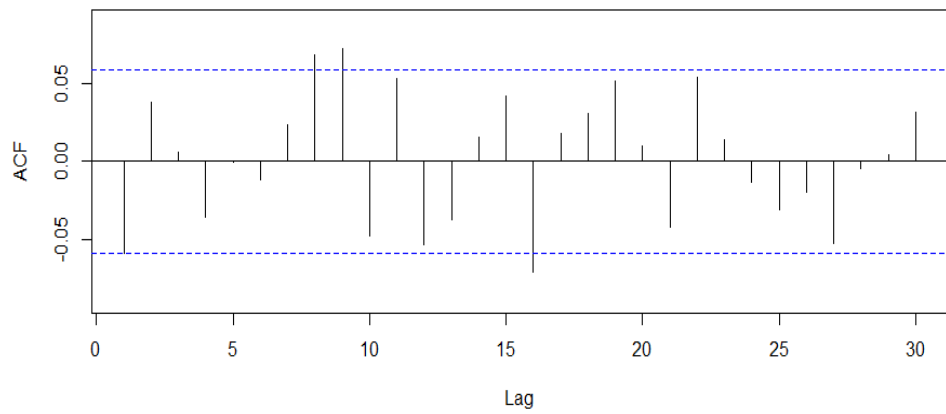


그림 3.1.4: 로그 차분한 환율 자료의 표본자기상관계수

귀무가설에 사용될 모형을 만들기 위해 모형의 변화가 없는 자료 구간이 필요한데, 환율 자료의 경우는 시계열 자료만 가지고 판단하기는 힘들다고 생각되어 다음과 같은 과정을 거쳤다.

1. 임의로 구간을 선택한다.
2. 절반씩 구간을 나누어 (앞의 절반은 모형의 변화가 없다고 가정하고) 모형을 적합시킨다.
3. 뒤의 절반에 대해 2의 모형을 활용하여 예측값을 구한다.
4. ORCUSUM 검정을 한 뒤 변화점이 없으면 구간 전체에서 모형의 변화가 없다고 판단하고 이를 이후 변화점 탐지에 활용한다.

위 과정을 거치면 변화가 없다고 가정하는 구간이 절반으로 줄어든다. 우선 일반적인 선형 ARMA 모형을 이용해 변화점을 탐지해 보았다. 먼저 1971년 1월 ~ 1980년 12월까지 10년에 해당하는 앞 235개의 자료를 선택한뒤 AIC(Akaike information criterion)를 기준으로 하여 ARMA 모형을 적합시킨 결과 평균 0인 ARMA(0, 1) 모형이 선택되었다. 적합된 모형을 가지고 1981년 1월 ~ 1990년 12월까지 10년에 대해 변화점이 있었는지 ORCUSUM 검정을 시행해 보면 변화점이 없다고 나오지만, 20년에 해당하는 구간 전체의 잔차에 대해 Ljung-Box 검정을 시행하면(그림 3.1.5) 시차 1부터 유의확률이 매우 낮으므로 잔차가 백색잡음(white noise)이라고 볼 수 없으므로 선형 ARMA 모형을 활용하긴 힘들다.

Ljung-Box test for residuals

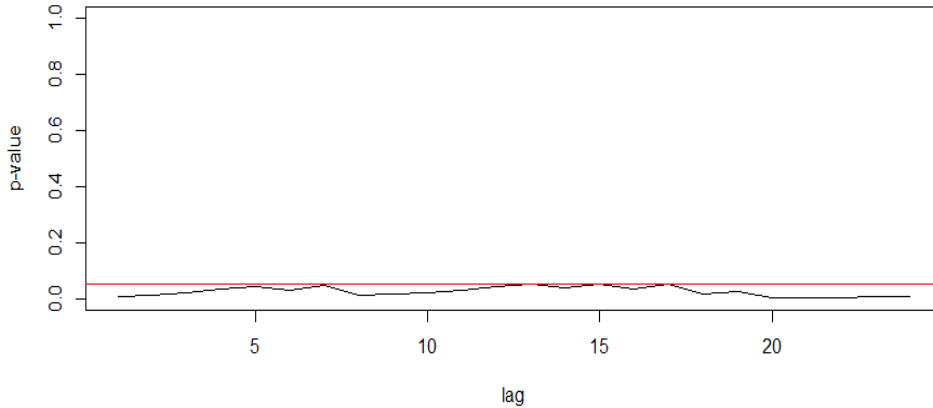


그림 3.1.5: 환율 자료에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

이어서 SVR-ARMA 모형을 앞의 10년에 적합시킨 결과 ARMA(0, 0) 모형이 선택되었다. Meese and Rogoff (1983), Chinn and Meese (1995)를 비롯한 많은 연구에서는 환율을 단기적으로 예측할 때 근본적인 모형(fundamental exchange rate model)이 임의보행모형(random walk model)임을 밝힌 바 있는데, 로그차분된 자료에 ARMA(0, 0)을 적합시키는 것이 아래 식(이상열, 2013)에 의해 ARIMA(0, 1, 0)을 적합시킨 것과 비슷한 효과를 주기 때문에 모형 적합 결과는 알려진 사실과도 부합한다고 볼 수 있다.

$$r_t = \log(x_{t+1}) - \log(x_t) = \log\left(\frac{x_{t+1}}{x_t}\right) \simeq \frac{x_{t+1}}{x_t} - 1$$

(단, $\left| \frac{x_{t+1} - x_t}{x_t} \right| < 1$ 을 만족하는 경우)

즉, ARMA(0, 0)이 이 경우엔 지배적인(dominant) 모형이라고 생각할 수 있다. 뒤의 10년에 대해 ORCUSUM 검정을 시행한 결과 변화점이 없다고 나오고, 20년에 해당하는 구간 전체의 잔차에 대해 Ljung-Box

검정을 시행하면(그림 3.1.6) 몇몇 시차에서 유의확률이 0.05보다 작다고 나오지만 선형 ARMA의 경우보단 개선되었으므로 이를 이용해 1991년 1월 ~ 2010년 12월까지 20년에 대해 변화점이 있는지 확인해본다.

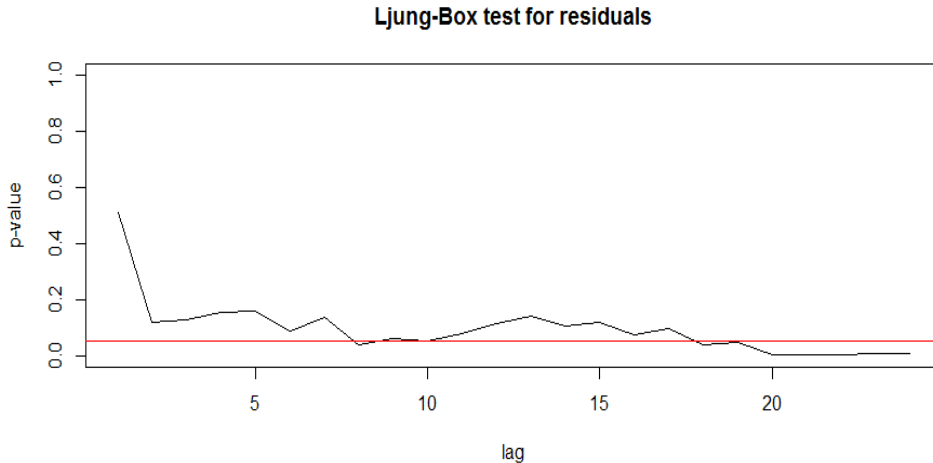


그림 3.1.6: 환율 자료에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

1971년 1월 ~ 1990년 12월까지 20년에 대해 SVR-ARMA 모형을 적합시킨 결과 마찬가지로 ARMA(0, 0) 모형이 적합되었으며, 변화점 검정 결과 2002년 12월 23일이(그림 3.1.7, 3.1.8) 변화점으로 나타났다.

2002년 12월 23일 이전까지 SVR-ARMA 모형을 적합시킨 결과 마찬가지로 ARMA(0, 0) 모형이 적합되었으며, 잔차에 대한 Ljung-Box 검정을 시행한 결과(그림 3.1.9) 모형이 잘 적합된 것을 확인할 수 있었다. 2002년 12월 23일 이후에 SVR-ARMA 모형을 적합시킨 결과 마찬가지로 ARMA(1, 1) 모형이 적합되었으며, 잔차에 대한 Ljung-Box 검정을 시행한 결과(그림 3.1.10) 역시 모형이 잘 적합된 것으로 보인다. 따라서 이 기간에 일시적으로 환율을 설명하는 확률과정에 변화가 있었다고 볼 수 있다.

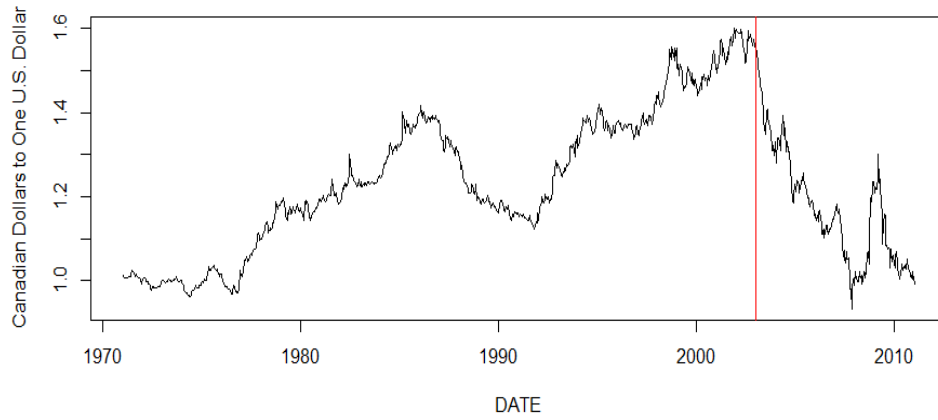


그림 3.1.7: SVR-ARMA 모델을 통해 추정된 변화점을 표시한 환율 자료

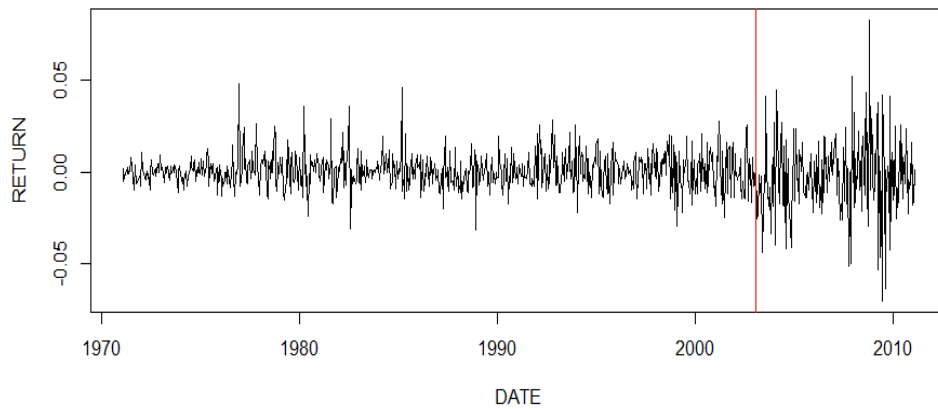


그림 3.1.8: SVR-ARMA 모델을 통해 추정된 변화점을 표시한 로그 차분된 환율 자료

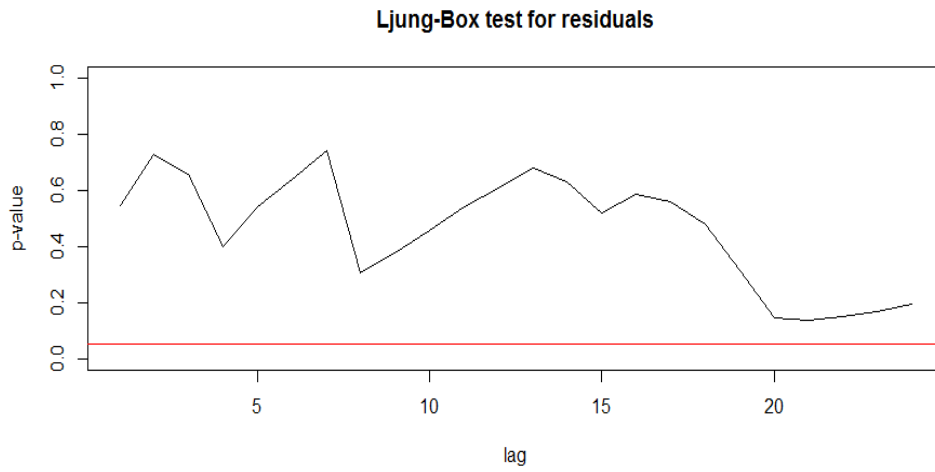


그림 3.1.9: 환율 자료에서 2002년 12월 23일 이전 모형의 잔차에 대한 Ljung-Box 검정 결과

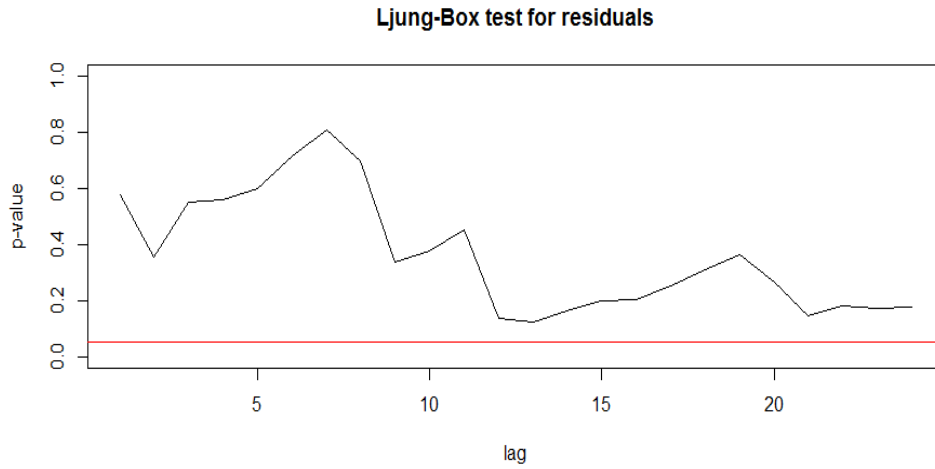


그림 3.1.10: 환율 자료에서 2002년 12월 23일 이후 모형의 잔차에 대한 Ljung-Box 검정 결과

3.2 영국 아카데미 백본망 통신량 자료

두 번째 사례는 UKERNA(United Kingdom Education and Research Networking Association)에서 제공한 영국 아카데미 백본망(United Kingdom academic network backbone)의 통신량 자료이다. 자료는 DataMarket(<https://datamarket.com>)의 Time Series Data Library에서 확인할 수 있다. 자료의 단위는 비트(bit)이며, 2004년 11월 19일 09시 30분부터 2005년 1월 27일 11시 11분까지 5분 간격으로 수집되어 총 19,888개의 관측치가 존재한다.

첫 번째로 원 자료의 시계열 그림(그림 3.2.1)을 보면 통신량 자료는 비정상 시계열로 보이며, 자료가 5분 간격으로 수집되었다는 점과 자료의 특성을 고려할 때 시차 288(1일)에서 주기성을 나타낼 것이라고 예측할 수 있다. 이는 표본자기상관계수 그림(그림 3.2.2)을 통해서도 확인할 수 있으며, ADF 단위근 검정에서 시차를 288로 둔 결과 또한 유의확률이 0.5885로 단위근이 존재하는 것으로 보인다. 앞의 사실에 추가로 2,016(1주일) 시차를 고려해

$$z_t = (1 - B)(1 - B^{288})(1 - B^{2016})y_t$$

와 같이 차분을 한 결과(그림 3.2.3), 자료 앞과 뒷부분에 이상치(outlier)로 보이는 일부 관측치들이 존재하여 일부 구간을 제외하였다. 마지막으로 변동성(volatility)을 제어하기 위해 자료를 시차 12(1시간) 간격으로 재추출하였고, 정상성을 만족하는 1,359개의 관측치를 얻었다(그림 3.2.4). 자료의 표본자기상관계수(그림 3.2.5)를 보면 시차 24에서 유의한 것으로 나타나서, 자료에 ARIMA(0, 0, 0)(1, 0, 0)₂₄ 모형을 적합시킨 뒤 모형의 잔차를 최종 자료로 사용하였다(그림 3.2.6).

우선 일반적인 선형 ARMA 모형을 이용해 변화점을 탐지해 보았다. 귀무가설에 사용될 모형을 만들기 위해 시계열 그림을 보고 정상성을 만족하는 것으로 보이는 앞 200개의 자료를 선택한 뒤 AIC(Akaike information criterion)를 기준으로 하여 ARMA 모형을 적합시킨 결과 평균 0인 ARMA(0, 0) 모형이 선택되었다. 뒤의 200개 구간에 대해서

ORCUSUM 검정 결과 변화점이 존재하지 않는다고 나왔지만 잔차에 대한 Ljung-Box 검정을 시행한 결과 시차 2, 3, 4에서 유의확률이 0.05 미만이므로(그림 3.2.7) 결과에 문제가 있어 보인다. 다시 말하면, 자료의 평균에 대한 경향성을 선형 ARMA 모형은 잘 포착하지 못하는 모습을 보였다.

이어서 SVR-ARMA 모형을 앞 200개 구간에 적합시킨 결과 ARMA(1, 0) 모형이 선택되었으며, 앞과 마찬가지로 잔차에 대한 Ljung-Box 검정을 시행한 결과(그림 3.2.8) 앞 200개 구간은 모형이 잘 적합된 것을 확인할 수 있었다. 뒤 200개 구간의 예측값을 구하고 ORCUSUM 검정을 시행한 결과 변화점이 존재하지 않는다고 나왔고, 귀무가설 하에서 뒤 200개 구간에 대한 모형의 잔차를 이용하여 Ljung-Box 검정을 시행한 결과 유의수준 0.05에서 별다른 문제가 없어 보인다. 따라서 이 경우에는 전체 400개 구간에서 변화점이 존재하지 않으며 SVR-ARMA(1, 0)모형을 따른다고 말할 수 있다.

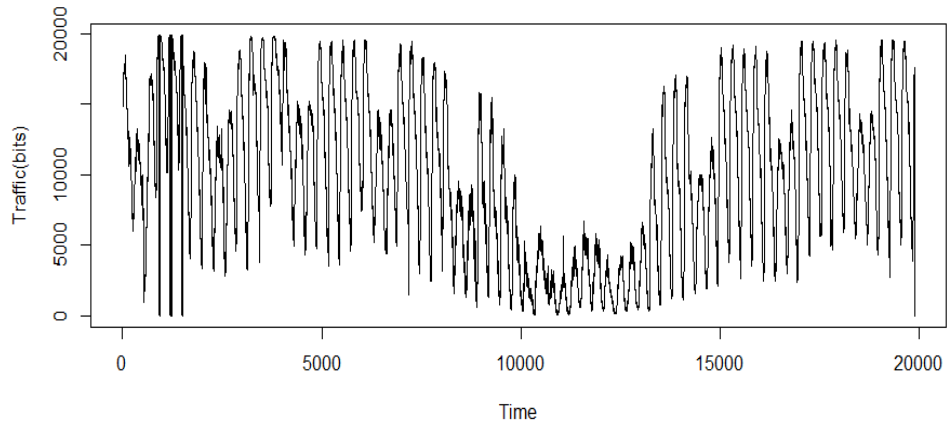


그림 3.2.1: 통신량 자료의 시계열그림

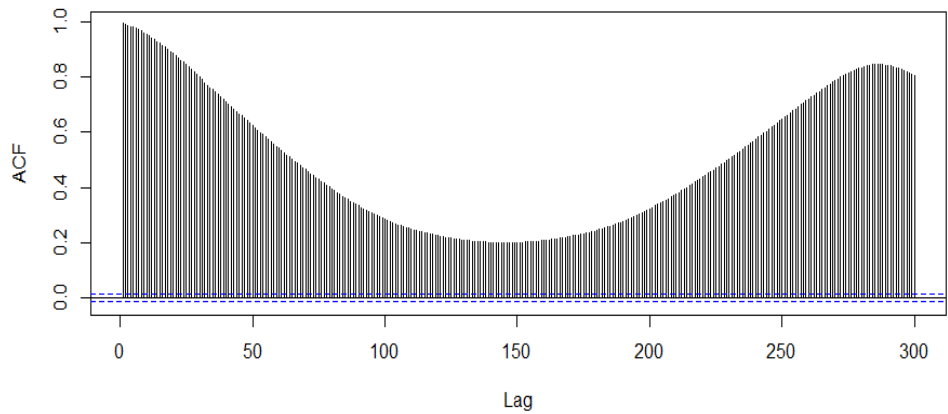


그림 3.2.2: 통신량 자료의 표본자기상관계수 그림

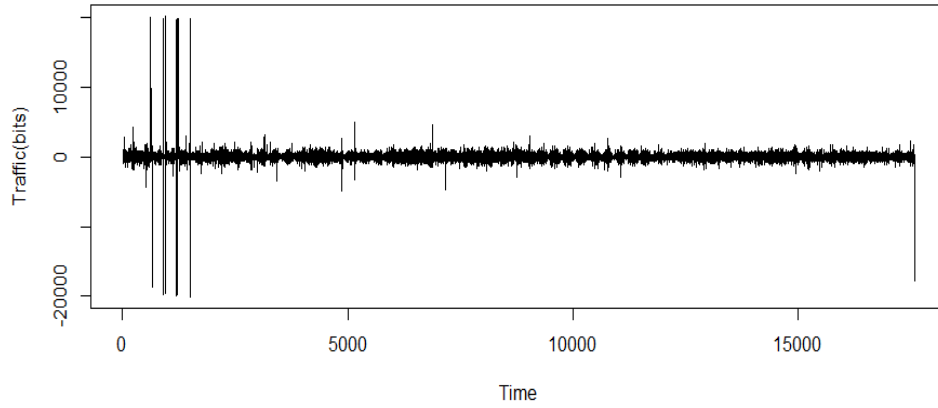


그림 3.2.3: 통신량 자료의 차분 직후 그림

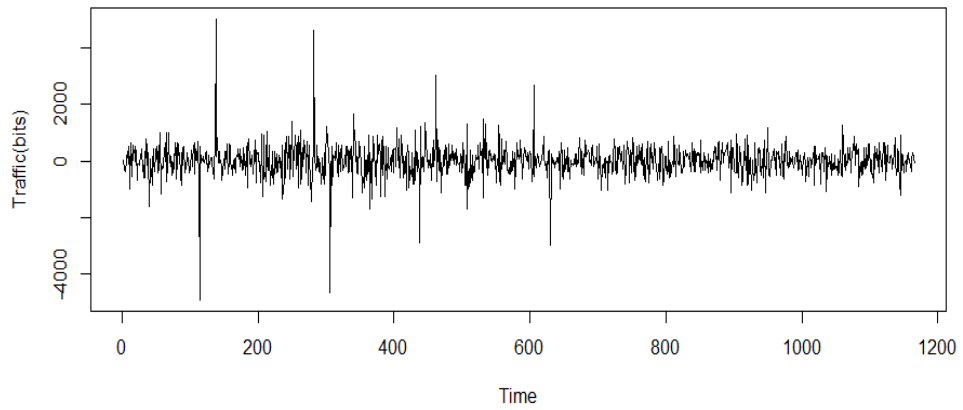


그림 3.2.4: 통신량 자료의 차분 후 앞, 뒤 일부를 제외한 시계열그림

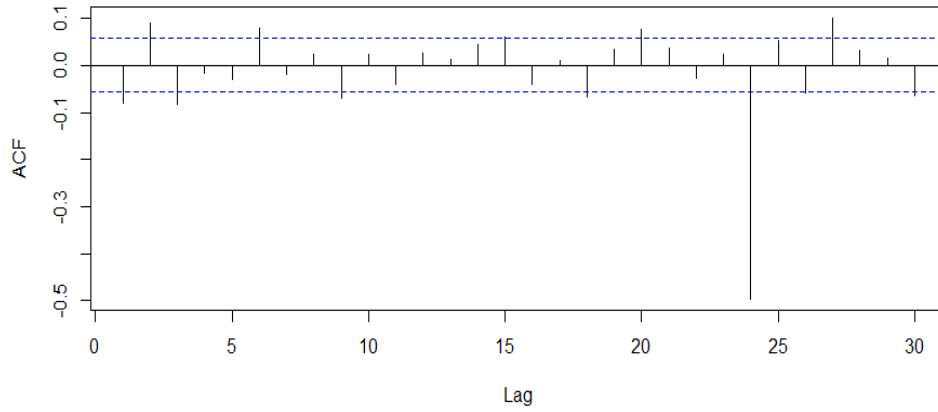


그림 3.2.5: 통신량 자료의 차분 후 앞, 뒤 일부를 제외한 자료의 표본자기상관계수 그림

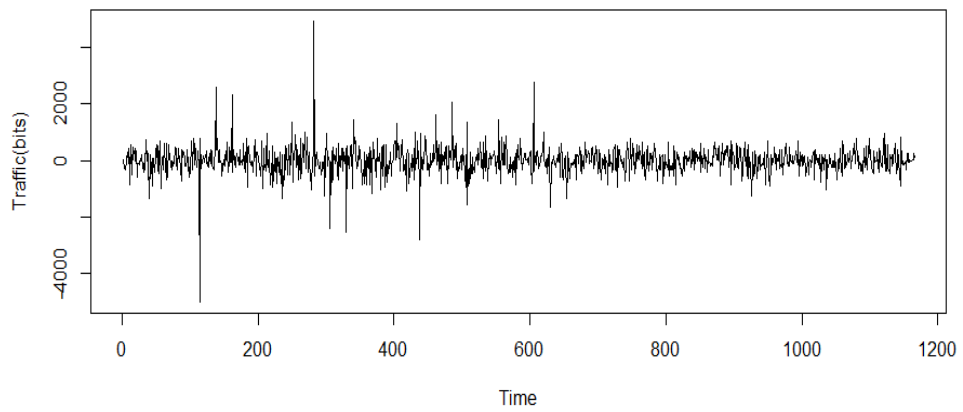


그림 3.2.6: 통신량 자료의 최종 시계열 자료; $ARIMA(0, 0, 0)(1, 0, 0)_{24}$ 모델을 적합시킨 뒤 모형의 잔차

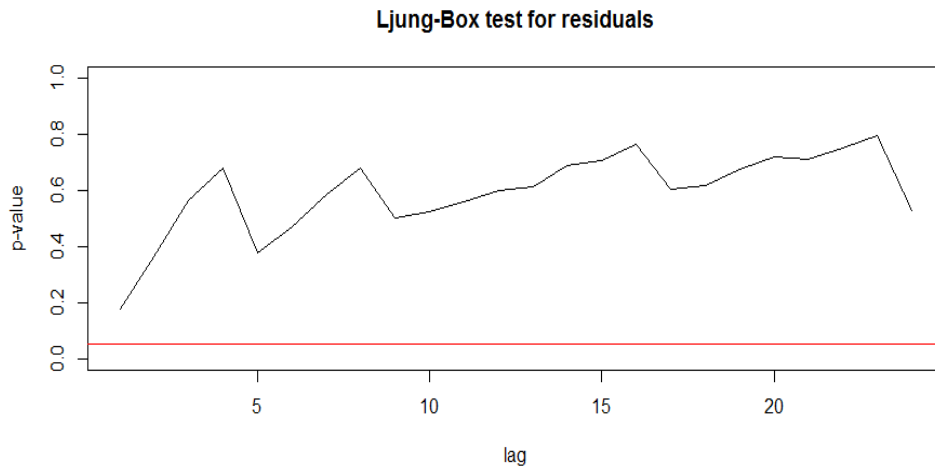


그림 3.2.7: 통신량 자료 앞부분에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

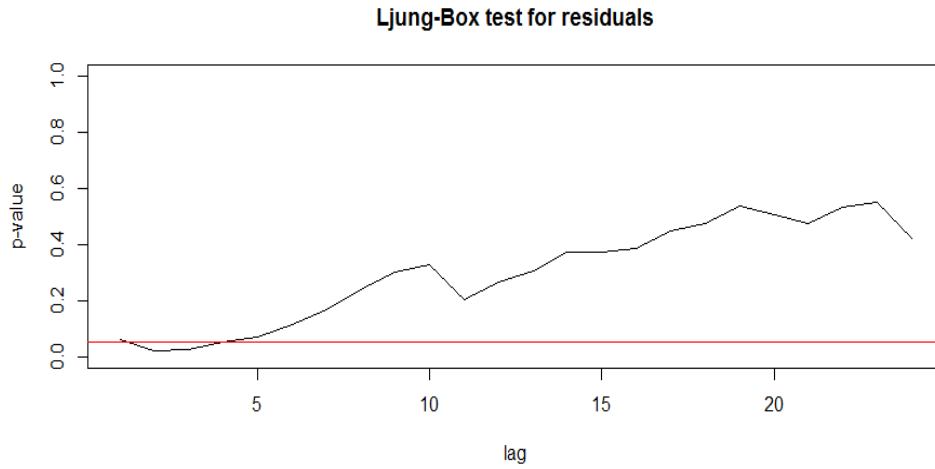


그림 3.2.8: 통신량 자료 뒷부분에서 선형 ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

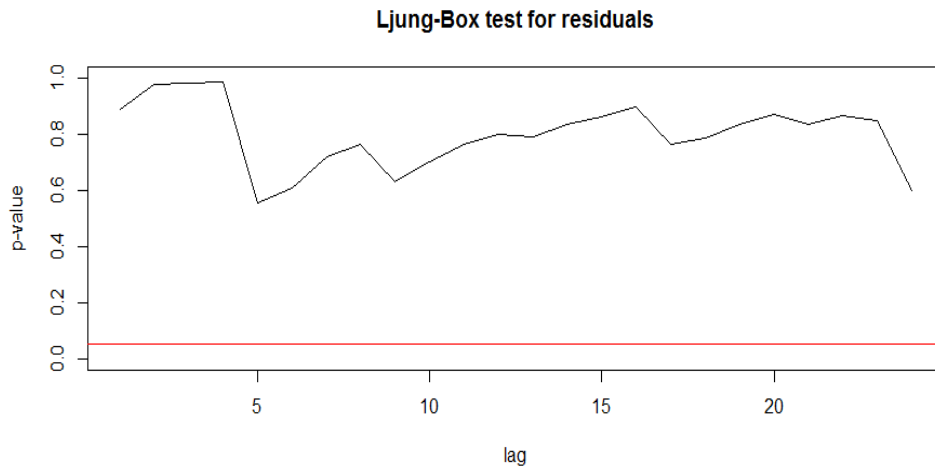


그림 3.2.9: 통신량 자료 앞부분에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

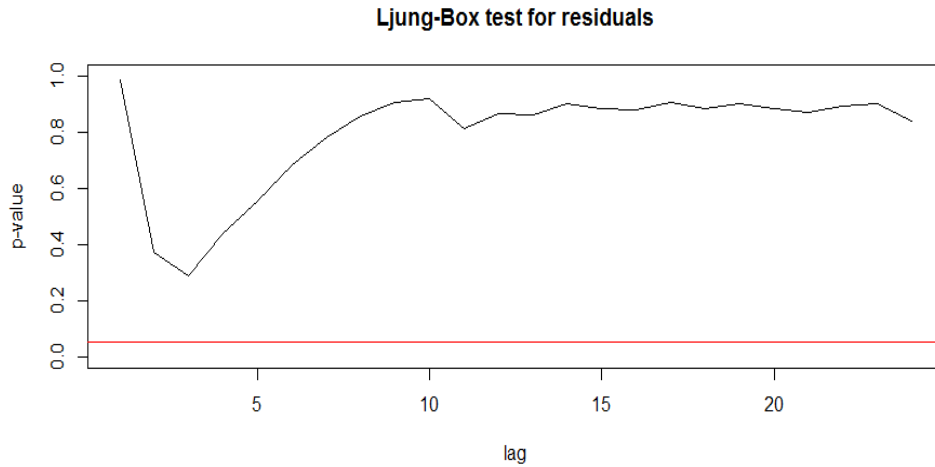


그림 3.2.10: 통신량 자료 뒷부분에서 SVR-ARMA 모형의 잔차에 대한 Ljung-Box 검정 결과

4. 결론

서론에서 변화점 탐지의 중요성과 이를 위한 선행 연구 결과에 대해 언급하였고, 2. 이론적 배경에서 서포트 벡터 회귀, ORCUSUM 검정을 간략하게 살펴보았다. 3. 실제 자료 분석 사례에서는 환율 자료와 통신량 자료 두 개를 분석하였다. 구체적으로 시계열 그림과 표본상관도표, 단위근검정을 통해 적절한 차분을 해주었고, 선형 ARMA 모형이 자료에 적합하지 않음을 확인한 뒤에 SVR-ARMA 모형의 AR-차수 및 MA-차수를 구하였다. 다음으로 적합된 모형에서 잔차에 대한 백색잡음 가정 등을 확인하였고, ORCUSUM 검정을 통해 변화점 유무를 확인하였다.

본 논문에서는 선행 연구들보다 통계적 관점에서 좀 더 자료에 일반적으로 적용될 수 있는 분석 기법을 활용하였다. 먼저 여러 가지 차분 기법을 상황에 맞게 적용하여 자료를 정상 시계열 자료로 변환하였으며, ARMA 모형의 차수를 미리 특정하지 않고 자료를 통해 판단하였다. 한편으론 변동성(volatility)을 제어하기 위해 각각 2주, 1시간 간격의 시차를 주거나 통신량 자료에서 계절성 제거를 위해 ARIMA(0, 0, 0)(1, 0, 0) 모형을 적합시킨 뒤 잔차를 이용하는 등 자료의 특성에 맞게 적당히 추가적인 기법을 활용하기도 하였다.

분석 대상이었던 본론의 두 자료를 통하여 다음과 같은 결론을 내릴 수 있다.

- ① 경험적으로 볼 때 변동성(volatility)을 잘 제어한다면 선형 ARMA 모형에서 얻어진 ORCUSUM 통계량을 SVR-ARMA 모형의 잔차에도 활용해 볼 수 있다.
- ② 환율 자료, 통신량 자료와 같이 정상성을 만족하는 선형 ARMA 모형을 따를 것이라고 기대되지 않는 자료의 경우 SVR-ARMA 모형을 이용하면 좀 더 모형이 잘 적합될 수 있다.
- ③ 위의 자료같은 경우에는 결과적으로 ORCUSUM 검정 결과도 더 신뢰할 수 있으므로, 변화점을 탐지하고 결과를 해석할 때에도 SVR-ARMA 모형을 이용하는 경우 결과를 더 신뢰할 수 있다.

한편, 본 논문에서는 많은 자료에 일반적으로 존재하는 변동성을 제어하기 위해 전체 자료 중 일부만 사용하였다. 만약 서포트 벡터 회귀와 같은 비선형 기법을 이용한 GARCH 모형을 ORCUSUM 검정과 사용할 수 있는 조건 또는 상황을 알 수 있다면, 본문의 방식을 응용하여 자료의 더 많은 정보를 고려한 변화점 탐지가 가능해질 것이다.

참고문헌

- [1] 이상열 (2013), 「시계열 분석: 이론 및 SAS 실습」, 자유아카데미, 제 1판.
- [2] 조신섭 · 손영숙 · 성병찬 (2016), 「SAS/ETS를 이용한 시계열분석」, 을곡출판사, 제 4판.
- [3] Abe, S. (2010), 「Support Vector Machines for Pattern Classification: Advances in Computer Vision and Pattern Recognition」, Second Edition, Springer.
- [4] Amiri, A., Niaki, S.T.A. and Moghadam, A.T. (2010), 「A probabilistic artificial neural network-based procedure for variance change point estimation」, *Soft Computing*, 19(3):691-700.
- [5] Box, G.E.P. and Tiao, G.C. (1975), 「Intervention Analysis with Applications to Economic and Environmental Problems」, *Journal of the American Statistical Association*, 70:70-79.
- [6] Camci, F. (2010), 「Change Point Detection in Time Series Data Using Support Vectors」, *International Journal of Pattern Recognition and Artificial Intelligence*, 24(1):73-95.
- [7] Chang, I., Tiao, G.C. and Chen, C. (1988), 「Estimation of Time Series Parameters in the Presence of Outliers」, *Technometrics*, 30(2):193-204.

- [8] Chen, S., Jeong, K. and Härdle, W.K. (2008), 「Recurrent Support Vector Regression for a Nonlinear ARMA Model with Applications to Forecasting Financial Returns」, SFB 649 Discussion Paper 2008-014.
- [9] Chinn, M. and Meese, R. (1995), 「Banking on currency forecasts: How predictable is change in money?」, *Journal of International Economics*, 38(1-2):161-178
- [10] Denby, L. and Martin, R.D. (1979), 「Robust Estimation of the First-Order Autoregressive Parameter」, *Journal of the American Statistical Association*, 74:140-146.
- [11] José Montaña Moreno, J., Palmer, A., and Muñoz Gracia, P. (2011), 「Artificial neural networks applied to forecasting time series」, *Psicothema*, 23(2):322-9.
- [12] Khashei, M. and Bijari, M. (2010), 「An artificial neural network (p, d, q) model for timeseries forecasting」, *Expert Systems with Applications*, 37(1):479-489.
- [13] Lee, S. (2018), Observation and residual-based CUSUM test with application to autoregressive models, Unpublished manuscript.
- [14] Lee, S., Lee, S. and Moon, M. (2018), Hybrid CUSUM test for ARMA models via support vector machine, Unpublished manuscript.

- [15] Lee, S., Tokutsu, Y. and Maekawa, K. (2004), 「The Cusum Test for Parameter Change in Regression Models with ARCH errors」, Journal of the Japan Statistical Society, 34(2):173-188.
- [16] Meese, R.A. and Rogoff, K. (1983), 「Empirical exchange rate models of the seventies: Do they fit out of sample?」, Journal of International Economics, 14(1 - 2):3-24.
- [17] Mukherjee, S., Osuna, E. and Girosi, F. (1997), 「Nonlinear prediction of chaotic time series using support vector machines」, Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop:511-520.
- [18] Müller, K.R., Smola, A.J., Rätsch, G., Schölkopf, B., Kohlmorgen, J. and Vapnik, V. (1997), 「Predicting time series with support vector machines」, International Conference on Artificial Neural Networks, Proceeding ICANN '97 Proceedings of the 7th International Conference on Artificial Neural Networks:999-1004.
- [19] Oh, K.J., Moon, M.S. and K. T.Y. (2005), 「Variance change point detection via artificial neural networks for data separation」, Neurocomputing, 68:239-250.
- [20] Page, E.S. (1955), 「A test for a change in a parameter occurring at an unknown point」, Biometrika, 42(3-4): 523-527.

[21] Trapletti, A. and Hornik, K. (2018), 「tseries: Time Series Analysis and Computational Finance」, R package version 0.10-46.

[22] Tsay, R.S. (1988), 「Outliers, Level Shifts, and Variance Changes in Time Series」, Journal of Forecasting, 7(1):1-20.

Abstract

In this paper, it will be shown that support vector regression, a nonlinear and nonparametric method, can be used to detect the change point of time series data if the data are not from stationary linear ARMA processes. This hybrid method is applied to two data: Canada/U.S. foreign exchange rate data and internet traffic data from an ISP, aggregated in the United Kingdom academic network backbone. Both data are transformed to fit the time series data model. And then we compares the linear ARMA model and our SVR-ARMA model in data.

Keywords : Change point detection, Support vector regression, ARMA model, ORCUSUM test, Exchange rate forecasting, Traffic forecasting

Student Number : 2017-27707