

시계열 모형을 이용한 주가지수 방향성 예측[†]

박인찬¹ · 권오진² · 김태윤³

¹부자아빠증권연구소 · ²³계명대학교 성서캠퍼스 통계학과

접수 2009년 7월 30일, 수정 2009년 10월 27일, 게재확정 2009년 11월 10일

요 약

본 논문은 주가지수선물거래 등에서 유용한 역할을 하는 시계열 데이터의 방향성 예측 문제를 다룬다. 여기서 시계열의 방향성 예측이란 시계열 값의 상승 혹은 하락을 예측하는 문제를 뜻한다. 방향성 예측을 위해 본 연구에서는 시계열 요소분해모형과 자기회귀 누적 이동평균 과정 모형을 고려한다. 특히 방향성 예측의 주된 통계량으로서 모형 외 편차와 모형 내 편차를 고려하며 모형 내 편차가 좀 더 유용함을 보인다.

주요용어: 모형 내 편차, 모형 외 편차, 방향성 예측, 시계열 모형.

1. 서론

금융 시장이 발전하고 복잡해짐에 따라 금리, 환율, 주가 등 각종 경제지표들의 변화에 따른 위험을 효율적으로 관리하기 위한 금융선물거래의 필요성이 증대하고 있다. 우리나라에서는 1996년 5월 최초로 증권거래소에서 KOSPI200을 대상으로 한 주가지수선물거래가 시작되었으며 그 이듬해인 1997년 7월에는 주가지수선물과 동일한 기초자산을 대상으로 한 KOSPI200옵션 상품이 증권거래소에 상장되었다. 주가지수선물거래의 가장 핵심적인 기능은 주식시장의 체계적 위험을 관리하는 것이다.

주가지수선물거래의 경우 레버리지 (leverage, 지렛대 효과)가 크고 (여기서 레버리지란 적은 투자금액으로 큰 손실과 수익이 나타날 수 있는 현상을 의미한다.) 익일의 지수상승 혹은 하락을 예측하기 어렵기 때문에 당일 청산이 많이 이루어지고 있다 (백관호, 2002). 즉 익일 주가지수에 대한 방향성 예측이 틀릴 경우 발생할 수 있는 큰 손실을 피하기 위해 당일 청산이 많이 이루어지고 있다. 본 논문에서는 포지션을 다음 날까지 유지하는 거래 (당일 청산을 하지 않는 거래)를 위한 방향성 예측기법을 연구 제안한다. 이를 위해 두 가지 모형, 시계열 요소분해모형과 ARIMA모형을 고려하며 t 시점에서 $t+1$ 시점의 방향성 예측을 위해 두 가지 통계량 δ_{1t} 과 δ_{2t} 를 고려하여 어느 것이 더 높은 정확도를 보이는지 알아본다. 여기서 δ_{1t} 는 예측값 (\hat{Z}_{t+1})과 실제값 (Z_t)차이를 이용하는 기법이며 δ_{2t} 는 추정값 (\hat{Z}_t)과 예측값 (\hat{Z}_{t+1})차이를 이용하는 기법으로서 다음과 같이 정의된다.

$$\delta_{1t} = \begin{cases} 1(\text{상승예측}), & \hat{Z}_{t+1} - Z_t > 0 \text{인 경우} \\ 0(\text{하락예측}), & \hat{Z}_{t+1} - Z_t \leq 0 \text{인 경우} \end{cases} \quad (1.1)$$

[†] 이 논문은 2009년도 교육과학기술부의 재원으로 한국연구재단의 지원을 받아수행된 연구임 (No. 2009-0065645).

¹ (150-874) 서울특별시 영등포구 여의도동 17, 부자아빠증권연구소, 연구원.

² (704-701) 대구광역시 달서구 달구벌대로 2800, 계명대학교 성서캠퍼스 통계학과, 석사과정.

³ 교신저자: (704-701) 대구광역시 달서구 달구벌대로 2800, 계명대학교 통계학과, 교수.

E-mail: tykim@kmu.ac.kr

$$\delta_{2t} = \begin{cases} 1(\text{상승예측}), & \hat{Z}_{t+1} - \hat{Z}_t > 0 \text{인 경우} \\ 0(\text{하락예측}), & \hat{Z}_{t+1} - \hat{Z}_t \leq 0 \text{인 경우.} \end{cases} \quad (1.2)$$

앞으로 본 논문에서 δ_{1t} 를 모형 외 편차 (extra-model deviation), δ_{2t} 를 모형 내 편차 (intra-model deviation)로 부른다. 직관적으로 모형 외 편차 (δ_{1t})와 모형 내 편차 (δ_{2t})의 기본적인 차이점은 δ_{2t} 가 δ_{1t} 보다 좀 더 시점에 관계없는 고정적인 성격 (시간불변성, time-invariant property)의 규칙을 제공한다는 데 있다. 즉 δ_{1t} 는 Z_t 를 사용함으로써 Z_t 가 내포하는 t 시점의 iid 오차의 영향을 직접적으로 받으나 δ_{2t} 는 \hat{Z}_t 를 사용함으로써 t 시점의 iid 오차를 0으로 간주하게 되며, 그에 따라 δ_{2t} 가 δ_{1t} 보다 상대적으로 좀 더 시간 불변성을 갖게 된다. 참고로 방향성 예측과 관련된 기존의 연구결과로서 지혜영과 조완현 (2009) 등이 있다.

2. 시계열 예측모형

본 연구에서는 예측 기법으로 두 가지 시계열모형, 즉 시계열 요소분해모형과 ARIMA모형을 고려한다. 물론 이외에도 잘 알려진 시계열 모형으로서 변동성을 이용하는 예측기법 등 (김삼용과 김진아, 2009) 다양한 예측 기법이 존재한다. 잘 알려진 대로 시계열 요소분해모형은 시계열의 시간에 따른 변화를 (시간)회귀함수 (추세, 순환함수 등)로 모형화하여 이를 외삽 (extrapolation)함으로써 예측을 시도하는 것이며 ARIMA모형은 시간이 흘러도 변하지 않는 정상적인 (stationary) 현재와 과거간의 선형 관계를 찾아내어 예측을 시도하는 것이다. 본 연구의 시계열 요소분해모형은 요소로서 회귀함수 (추세함수와 순환함수)와 상관된 정상 오차 등을 포함한다.

2.1. 시계열 요소분해모형과 ARIMA 모형에 의한 예측

본 연구에서는 KOSPI200지수 움직임을 회귀 모형화 하는데 있어서 삼각함수를 활용한다. 즉 지수의 추세적 변동을 나타내는 이차함수 $f(t)$ 를 삼각함수로 감싸는 식 (2.1) 형태의 모형을 활용한다.

$$Z_t = f(t) + \alpha_1 \sin\left(\frac{2\pi t}{L}\right) + \alpha_2 \cos\left(\frac{2\pi t}{L}\right) + \epsilon_t \quad (2.1)$$

여기서 ϵ_t 는 정상과정인 오차항이다. 식 (2.1)을 사용하여 KOSPI200지수를 예측을 할 때 중요한 절차는 삼각함수의 계수 (α_1 과 α_2)추정, 주기 L 의 결정과 $f(t)$ 를 추정하는 일이다. 여기서 주목할 점은 식 (2.1)을 주기적으로 (1주 혹은 2주단위로) 최근의 데이터들을 사용하여 최신화 (update)한다면 결과적으로 지역 적합 (local fitting)을 통해 회귀함수를 비모수적으로 추정하는 효과를 갖게 된다는 점이다.

식 (2.1)의 추정을 위해 $f(t)$ 는 $\beta_0 + \beta_1 t + \beta_2 t^2$ 로 정하며 시도표 (time plot) 등에 의거 적절한 주기 L 을 찾은 후 잘 알려진 회귀분석 기법을 사용하여 다음과 같은 회귀함수 (중심선) 추정값을 얻는다.

$$\hat{Z}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \hat{\alpha}_1 \sin\left(\frac{2\pi t}{L}\right) + \hat{\alpha}_2 \cos\left(\frac{2\pi t}{L}\right) \quad (2.2)$$

여기서 모형추정 후 잔차 분석을 통해 잔차가 독립이 아닐 경우 잔차를 추가로 분석하여 최종 추정값 혹은 최종모형을 얻는다.

ARIMA 모형은 자기회귀이동평균모형에 시계열 데이터의 정상성 확보를 위한 차분절차가 추가된 모형으로서 차분후 적절한 AR과 MA의 차수를 정하는 절차를 거쳐 추정된다. ARIMA 모형에 의한 구체적 예측 절차는 박유성과 김기환 (2004)등에 상세히 기술되어 있다. 본 연구에서 방향성 예측은 모형 외 편차 δ_{1t} 와 모형 내 편차 δ_{2t} 에 의해 각각 결정된다.

3. 실제 데이터 적용

본 연구에 사용된 자료는 2007년 5월1일부터 2008년 2월29일까지 일별 KOSPI200지수이다. 훈련용 데이터는 2007년 5월1일부터 10월31일까지 KOSPI200으로서 123개이며 평가용 데이터는 2007년 11월1일부터 2008년 2월29일까지의 KOSPI200으로서 74개이다. 이들에 대한 시도표는 그림 3.1과 그림 3.2에 주어져 있다.

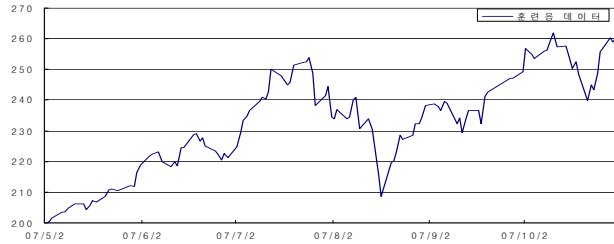


그림 3.1 훈련용 데이터의 시도표



그림 3.2 평가용 데이터의 시도표

3.1. 요소분해모형

그림 3.1의 훈련용 데이터에 식 (2.1)의 모형의 추세와 순환 부분을 적합한 결과는 다음과 같으며 이에 근거한 시도표는 그림 3.3에 주어져 있다.

$$\hat{Z}_t = 206.108 + 0.65t - 0.003t^2 - 7.985\sin\left(\frac{2\pi t}{60}\right) + 4.967\cos\left(\frac{2\pi t}{60}\right) \quad (3.1)$$

식 (3.1)에 따른 잔차의 시도표 그림 3.4를 보면 잔차가 독립이 아니며 뚜렷한 패턴을 나타내고 있음을 알 수 있다. 또한 이는 잔차의 포트만토 검정 통계값을 통해서도 확인할 수 있었다.

잔차의 독립성을 달성하기 위해 ARIMA 모형을 적합하였으며 그 결과 잔차를 AR(4)모형으로 적합하였다. 추가 적합 후 최종 잔차에 대한 독립성 여부를 검정하였으며 그 결과는 표 3.1에 주어져 있는데 최종 잔차들이 독립임을 보여주고 있다. 따라서 최종요소분해모형은

$$\hat{Z}_t^{(1)} = \hat{Z}_t - 0.12165 + 0.98578e_{t-1} - 0.24878e_{t-2} + 0.37716e_{t-3} - 0.33411e_{t-4} \quad (3.2)$$

이며 이에 대한 시도표는 그림 3.5에 주어져 있다. 식 (3.2)에서 e_t 는 t 시점의 잔차를 나타낸다.

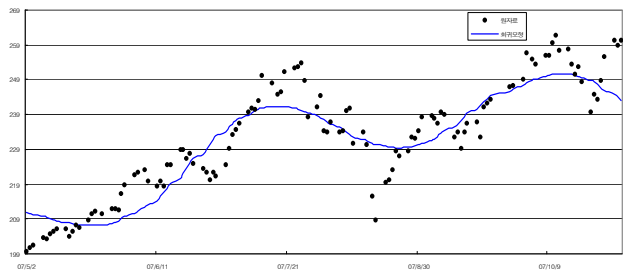


그림 3.3 식 (3.1) \hat{Z}_t



그림 3.4 식 (3.1) 회귀분석 후 잔차의 시도표

표 3.1 최종잔차의 포트만토 검정통계값

시차	카이제곱 통계량	자유도	유의확률
6	5.07	6	0.5344
12	9.04	12	0.6998
18	15.52	18	0.6262
24	26.38	24	0.3341

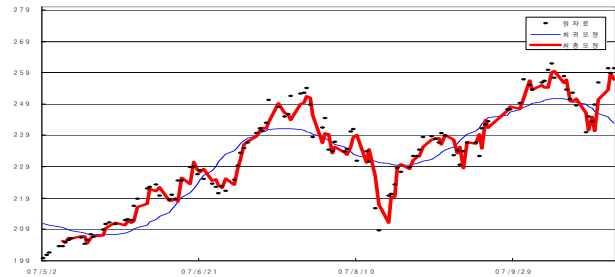


그림 3.5 식 (3.2) $Z_t^{(1)}$ 의 시도표

3.2. ARIMA 모형

훈련용 데이터에 ARIMA 모형추정의 표준화된 절차들 (차분, 모형식별, 모형추정, 모형진단)을 적용

한 결과 1차 차분 후 AR(5)모형이 선택되었으며 추정된 모형은 식 (3.3)과 같다.

$$\begin{aligned}\hat{Z}_t = & 0.47322 + Z_{t-1} + 0.09145(Z_{t-1} - Z_{t-2}) - 0.11698(Z_{t-2} - Z_{t-3}) \\ & + 0.23119(Z_{t-3} - Z_{t-4}) - 0.03672(Z_{t-4} - Z_{t-5}) - 0.23574(Z_{t-5} - Z_{t-6})\end{aligned}\quad (3.3)$$

표준화된 절차들에 대한 자세한 내용은 Box와 Jenkins (1976) 등을 참조할 수 있다. 아래 그림 3.6은 모형식별 절차에서 주요한 역할을 하는 차분된 데이터의 편자기상관함수 (Partial Auto Correlation Function: PACF) 그림이다.

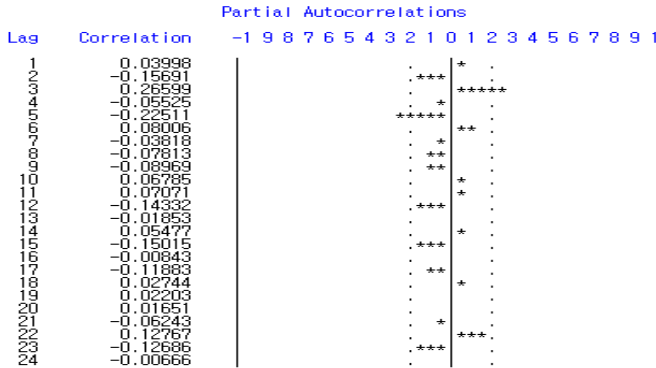


그림 3.6 1차 차분 된 혼련용 데이터의 PACF

3.3. 방향성 예측

요소분해모형 식 (3.2)과 ARIMA모형 식 (3.3)을 토대로 KOSPI200지수의 방향성을 예측해보자. 본 절에서는 두 가지 방향성 예측기법, 모형 외 편차 δ_{1t} (식 (1.1))과 모형 내 편차 δ_{2t} (식 (1.2))를 사용하여 방향성을 예측한 후 두 기법의 유용성에 대해 살펴본다. 혼련용 데이터로 추정된 요소분해모형 식 (3.2)과 ARIMA 모형 식 (3.3)을 사용하여 2008년 2월 22일까지 방향성 예측을 실시하였다. 먼저 요소분해모형과 ARIMA 모형을 이용한 예측 값들은 그림 3.7과 그림 3.8에 각각 나타나 있다. 그림 3.7과 그림 3.8에는 실제 값은 점들로, 예측 값은 실선으로 각각 나타나 있으며 그림 중간에 보이는 기둥 점선은 그 이전과 이후가 각각 혼련 및 평가기간임을 나타내기 위해 도입되었다. 먼저 기대할 수 있는 것은 추정된 요소분해모형이나 ARIMA모형 모두 일정기간이 지나면 변화된 미지의 실제 모형 (the underlying model)을 제대로 예측 (추정)하지 못할 것이라는 점인데 실제로 요소분해모형 (3.2)의 경우 2월 22일 이후 예측값이 실제값과 현저히 달라졌으며 ARIMA 모형 (3.3)의 경우 1월 8일 이후 예측값이 실제값과 현저히 달라졌다. 이는 요소분해모형이 장기예측에 좀 더 유용하다는 잘 알려진 사실과 부합된다. 이러한 이유로 그림 3.7의 요소분해모형 모형의 경우 예측값을 2월 22일, 그림 3.8의 ARIMA 모형의 경우 예측값을 1월 8일까지만 제시하였다.

그림 3.9에는 평가용 데이터에 대한 요소분해모형과 ARIMA모형에 의한 방향성 예측결과가 주어져 있다. 즉 예측률의 시간에 따른 변화가 주어져 있다. 여기서 예측률이란 평가기간 중 성공예측수를 총 예측수로 나눈 값이다. 그림 3.9를 살펴보면 요소분해모형에 δ_{2t} 를 사용한 경우 (굵은 실선) 2008년 2월 22일까지 60%의 예측률을 유지하였으며 δ_{1t} 를 사용한 경우 (가는 실선)보다 대부분의 구간에 걸쳐 우월한 예측률을 보이고 있다. 또한 ARIMA 모형에 δ_{2t} 를 사용한 경우 (굵은 점선)와 δ_{1t} 를 사용

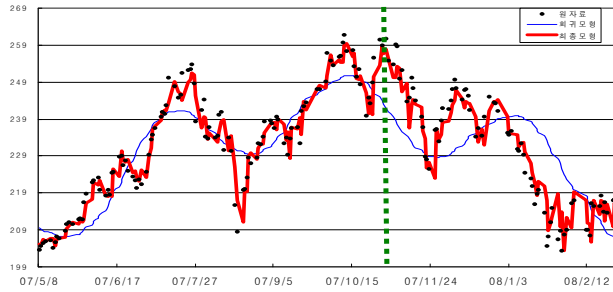


그림 3.7 요소분해모형의 예측 결과

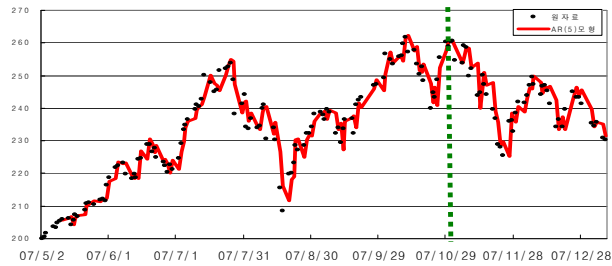


그림 3.8 AR(5)모형 예측 결과

한 경우 (굵은 점선)가 그림 3.9에 함께 나타나 있는데 이 결과 역시 δ_{2t} 의 우월성을 보여주고 있다. 이와 같이 그림 3.9의 방향성 예측률 시도표를 통해 알 수 있는 것은 δ_{2t} 의 유용성이며 이는 방향성 예측의 경우 모형 내 편차를 이용하는 것이 모형 외 편차를 이용하는 것보다 바람직하다는 것을 보여주고 있다. 참고로 모형의 예측값이 정확성을 상실하는 시점 (요소분해모형의 경우 2월22일, ARIMA 모형의 경우 1월8일)까지의 방향성 예측결과를 살펴보면 요소분해모형의 경우 δ_{1t} 를 사용했을 경우 56%의 예측률을, δ_{2t} 를 사용했을 때 60%의 예측률을 각각 유지하였으며, ARIMA 모형의 경우 2008년 1월 8일까지 δ_{1t} 를 사용했을 경우 55%의 예측률을, δ_{2t} 를 사용했을 때 60%의 예측률을 유지하였다.

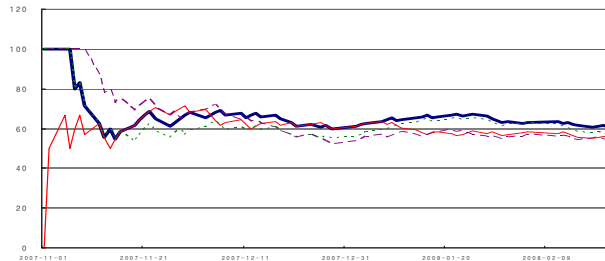


그림 3.9 모형들의 방향성 예측률

본 절의 δ_{1t} 와 δ_{2t} 를 적용한 연구결과는 다음과 같이 요약될 수 있다. “방향성 예측을 위해 요소분해

모형이 ARIMA에 비해 우월하며, 모형에 관계없이 모형 내 편차인 δ_{2t} 가 모형 외 편차인 δ_{1t} 보다 유용하다.” 모형 외 편차가 현재 시점의 데이터를 활용하는 반면 모형 내 편차는 현재 시점의 데이터를 활용하지 않는다는 사실을 고려할 때 이러한 결과는 흥미로운 결과라고 할 수 있다. 이러한 결과가 관찰되는 주된 이유는 시간불변성의 관점에서 δ_{2t} 의 우월성에 있다고 판단된다. 즉 방향성 예측의 경우 정확한 예측값 사용이 필수적이지만 이외에도 어느 정도 시간 불변적인 성격을 갖는 기법을 사용하는 것이 바람직한 것으로 보인다.

4. 결론

주가지수 선물거래 등 방향성 예측은 다양한 금융문제에서 중요한 문제로 등장하고 있으나 이에 대한 연구 결과는 그리 많지 않다. 본 논문에서는 KOSPI200지수의 방향성 예측문제를 기존의 시계열분석 관점을 통해 분석하였다. 본 논문의 주된 연구결과는 모형 내 편차를 통한 방향성 예측이 정확한 결과를 낼 확률이 높다는 사실을 밝힌데 있으며 이 결과는 추후 좀 더 상세히 연구해 볼 만한 가치가 있다.

참고문헌

- 김삼용, 김진아 (2009). 일반 자기회귀 이분산 모형을 시계열 자료분석. <한국데이터정보과학회지>, **20**, 475-484.
- 박유성, 김기환 (2004). <SAS/ETS를 이용한 시계열 자료 분석 I>, 자유아카데미, 서울.
- 백관호 (2002). 시물레이션에 의한 주가지수선물의 데이터트레이닝. <선문대학교 사회과학논집>, **5**, 53-67.
- 지혜영, 조완현 (2009). 인터넷 쇼핑물에서 구매품목을 이용한 고객의 예측모델 설계. <한국데이터정보과학회지>, **20**, 475-484.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (2nd ed.), Holden-Day, San Francisco.

KOSPI directivity forecasting by time series model

In Chan Park¹ · O Jin Kwon² · Tae Yoon Kim³

¹Boojaappa securities of research laboratory

²³Department of Statistics, Keimyung University

Received 30 July 2009, revised 27 October 2009, accepted 10 November 2009

Abstract

This paper deals with directivity forecasting of time series which is useful for futures trading in stock market. Directivity forecasting of time series is to forecast whether a given time series will rise or fall at next observation time point. For directional forecasting, we consider time regression model and ARIMA model. In particular, we study two statistics, intra-model and extra-model deviation and then show usefulness of intra-model deviation.

Keywords: Directivity forecasting, extra-model, intra-model, time series.

¹ Researcher, Boojaappa securities of research laboratory, Seoul 150-874, Korea.

² Master Student, Department of Statistics, Keimyung University, Daegu 704-701, Korea.

³ Corresponding author: Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea. E-mail: tykim@kmu.ac.kr