

시차구조의 설정에 따른 시장변동의 조정과정 분석

김태호¹⁾ 이청림²⁾

요약

서로 연관관계에 있는 실제의 통계자료들은 동태적, 확률적 동시발생적으로 유발되며, 이로 인해 한 자료의 변동이 다른 자료에 미치는 영향은 같은 기간 뿐 아니라 시차를 두고 여러 기간에 걸쳐 지속되며 조정되어 간다. 그러나 일반적인 선형,비선형 통계모형을 사용하여 현실동향을 분석하는 경우 자료의 이러한 특성에서 오는 시차관계를 통상 무시함으로써 변수 사이의 관계는 같은 기간 내에 결정되어야 하는 제약이 가해지게 된다. 그 결과 시간이 흐름에 따라 이들의 관계가 변화하는 과정이나 한 변수의 변동이 다른 변수에 미치는 장기적 영향도 추정할 수 없을 뿐 아니라 현실여건의 변동이나 전개과정을 설명하는 데도 큰 결함을 갖게 된다. 시차관계가 존재하는 변수에 실제 여건에 합당한 시차구조가 설정되면 현실이 정확히 반영되고, 모형에 내재된 변수들의 장단기 변동상황과 동태적 적응과정이 파악됨과 동시에 다양한 분석이 가능해지므로 모형의 활용도는 높아지게 된다.

주요용어: 시차구조, 다항시차분포, 인공변수, 등락예측, Theil계수.

1. 서론

서로 연관관계에 있는 변수들 가운데 일부에 변동이 발생함에 따라 다른 변수가 반응을 보이기까지는 현실적으로 시차가 존재한다. 즉 기간 t 에 자본을 투자함에 따른 수익은 $t+1, t+2, \dots$ 에 걸쳐 유발되며, 이는 예를 들어 어느 시점에 투자된 자본은 같은 기간뿐만 아니라 미래에까지 영향을 미친다는 뜻이다. 우리의 일상생활과 연관시키면 소비행위는 습관을 바꾸기 어렵거나 시차적 지연으로 인해 현재의 소비 Y_t 는 현재와 과거의 소득 $(X_t, X_{t-1}, X_{t-2}, \dots)$ 의 함수로 설정해야 현실적일 것이다. 이는 한 설명변수가 변화하면 종속변수에 끼친 영향은 그 기간 내에 끝나는 것이 아니라 시간이 흐르는 가운데 계속된다는 것으로 설명변수들의 시차시리즈를 모형에 표기하여 시간에 따른 적응과정이 명시되어야 모형이 현실을 더욱 정확히 반영하게 된다.

그러나 현실을 설명하기 위해 일반적인 통계모형을 사용하는 경우 각 변수 사이의 관계는 같은 기간 내에 결정된다는 제약이 가해진다. 그 결과 서로 다른 시점 변수들간의 관계를 알 수 없게 되고, 시간이 흐름에 따라 변수들간의 관계가 변화하는 과정에 대해서도 설명할 수 없을 뿐만 아니라 한 변수의 변동이 다른 변수에 미치는 단기적, 장기적 영향 또한 추정이 불가능하게 된다.

1) (361-763) 충북 청주시 개신동 산48, 충북대학교 통계학과, 교수

E-mail : thkim@trut.chungbuk.ac.kr

2) (361-763) 충북 청주시 개신동 산48, 충북대학교 통계학과, 전 대학원생

E-mail : cllee@netian.com

시차구조가 포함되지 않은 통상적인 모형은 특정 설명변수들이 종속변수에 미치는 영향을 예측하기 위한 우선적인 접근방법이긴 하지만 시간관계를 무시함으로써 현실상황을 연구하는데 많은 결함을 갖는다. 즉 시간 개념이 체계 내에 명확히 도입되지 않아 현실여건의 변동이나 전개과정에 대한 연구에 사용되기에는 한계가 있으므로 시간이 흐름에 따라 종속변수가 적응해 가는 과정을 파악하기 어려울 뿐만 아니라 현실여건의 변화로 인한 영향이 얼마나 계속되는가 역시 분석할 수 없게 된다. 대부분의 실제 자료는 동태적, 확률적, 동시 발생적 관계를 가지며 생성된다. 통계모형을 사용하여 현실상황을 분석하려 할 때 자료의 이러한 특성이 충분히 반영된다면 모형의 현실적 활용도를 높일 수 있을 것이다.

2. 모형의 시차관계

증권예탁원이 2000년 말 기준으로 상장사와 코스닥, 제3시장 등록기업 보유지분을 조사한 통계에 따르면 주식투자자는 429만 5,754명이며 이중 기관을 제외한 순수 개인투자인구는 428만 2,459명으로 전체 국민의 9.3%에 이르고, 투자자들은 1인당 평균 3.1종목, 3,631주의 주식을 보유하고 있는 것으로 집계됐다. 따라서 주식시장의 동향은 점차 국민적인 관심사가 되고 있으며, 또한 국가 경제에도 큰 영향을 미치고 있다. 국내 증시는 미국 증시의 영향을 많이 받으며, 국내의 거래소시장과 코스닥시장간에도 동조화현상이 존재하는 것으로 알려져 있다. 그러나 이러한 상호 연관관계는 한 시장에 변동이 발생하면 다른 시장에 미치는 영향이 같은 기간 내에 마무리되는 것이 아니라 시차를 두고 여러 기간에 걸쳐 조정되어 가지만 주식시장의 동적 변동관계를 분석한 연구는 국내외를 막론하고 거의 이루어지지 않고 있다. 본 연구에서는 과연 종합주가지수(KOSPI)와 코스닥지수는 동조화현상을 보이는가, 또 국내 증시는 미국 증시의 변동에 민감하게 반응하는가를 검정해 보고자 한다.

주가의 변동은 분기별자료, 연간자료와 같이 여건변화의 단위기간이 긴 자료보다는 일일자료에 의해 설명될 때 더욱 의미가 있을 것이다. 그러나 일간자료가 존재하는 변수는 많지 않으므로 모형설정에 사용될 수 있는 변수의 수에 제약을 받게된다. 이때 사용 가능한 변수들의 시차구조를 모형에 포함시키면 설명변수의 수가 증가하게 되므로 모형의 설명력을 높이게 된다. 따라서 연구의 목적은 국내 주가의 일일변동패턴을 정확히 추적하는 시차분포구조가 갖추어진 통계모형을 설정하여 국내 주식시장의 동태적 구조를 파악하고, 시장요인들의 변화에 따른 주가의 장단기 반응도를 분석하는데 있다.

모형을 사용하여 국내 증시를 분석한 연구로 안태백(1988)은 시장조정모형을 사용하여 신주발행과 주식수익률간의 관계를, 박상용·연강흠(1994)은 VAR모형을 사용하여 외환·증권·금융시장에서의 가격메카니즘간의 상호 관계를, 그리고 고광수(1997)는 회귀진단모형과 GARCH-M모형을 사용하여 기대수익률과 주가변동성간의 관계를 각각 분석하였다. 또한 박재환(1998)은 주식수익률 시계열의 최적 ARIMA모형을 구했으며, 조재범·김호일(1998)은 종합주가지수와 금리의 추이를 분석하는데 마르코프 국면전환모형을 사용하였다.

국내 주식시장을 통계적으로 분석한 연구로 허창수(1988)는 한국과 미국증시의 분산한계검증에 관한 비교연구를 수행하였고, 이정도·김정호(1996)는 확률적 지배에 의한 포

트폴리오전략의 비교우위를 검토한 바 있다. 일별자료를 사용한 연구로는 김석용·이하일(1998), 이일균(1998), 김서경·고광수(2001) 등이 있으며, 시차관계를 다룬 연구로는 박재환(2000), 이일균(2001), 그리고 최정호·나종길(2001)이 있다.

선행연구들은 증시변동에 유의한 영향을 미치는 여러 변수들간의 관계에 대한 분석은 간과해왔다. 일간자료를 사용한 연구들도 일부 시장요인 사이의 관계에 대한 분석에 그쳐 증시 전반의 일일변동에 대한 포괄적인 정보는 제공해 주지 않으며, 시차관계를 분석한 연구들 또한 단순히 한 기간 선후에 불과한 두 변수간의 관계에만 초점을 맞추었을 뿐이다. 그러나 모형 설정시 국내 증시에 영향을 미치는 주요변수들이 모두 고려되어야 시장변수들간의 관계와 증시의 구조에 대한 전반적인 정보가 제공될 수 있을 것이다. 특히 주가 변동과 시차관계가 존재하는 각 변수들에 대해서는 통계적으로 유의한 시차기간이 모두 모형에 포함되어야 시장변동의 현실이 구체적으로 파악될 수 있다.

모형에 시차구조가 표기되면 설명변수의 변동에 따른 종속변수의 각 시차반응들이 포착되므로, 모형에 내재된 변수간의 적응과정이 파악되며, 따라서 일반 정태모형이나 자기회귀모형보다 현실을 더욱 정확히 반영하게 된다. 통계모형에 사용되는 시차구조는 대부분 과거로 거슬러갈수록 현재에 미치는 영향력이 감소해가는 기하적 형태를 가진다. 이러한 시차분포는 현실적으로 합리적이기는 하나 주식시장의 분석에 사용하기는 부적절하다. 주가를 상승시키는 외생적 요인이 변화하면 주가는 상승했다가 그 다음 반발 매수세로 인해 일단 하락하며, 이후 다시 상승하는 등 상승과 하락세가 반복되며 시간이 흐르면서 그 영향력이 소멸되는 것이 증시의 특징으로 알려져 있다. 다항시차분포모형은 이러한 현상을 설명하는데 유용한 구조를 갖추고 있으며, 본 연구에서 국내 주식시장의 일일변동을 설명하는 모형에 적용시켜 보고자 한다.

다항시차분포의 유용성은 Maddala and Rao(1971), Cooper(1972), Schmidt and Waud(1973), Pagano and Hartley(1981), Haynes and Stone(1983), Leamer and Leonard(1983), Warner and Kreinin(1983), Thursby and Thursby(1984), Chow(1988), Judge et al(1988), Gujarati(1995), Pindyck and Rubinfeld(1998), Darrat(1999), 조영경(1999), 황성혁·이정희(2000), 이정림(2001), 조성표·정재용(2001)등에서 근거를 찾을 수 있다.

3. 모형의 설계

시차분포모형을 사용하는데 있어서 근본적인 문제점은 시차변수들로 인해 많은 자유도가 소모되고 다중공선성 문제가 야기될 가능성이 있으며, 또한 추정할 모수의 수가 대부분의 경우 너무 많다는 것이다. 따라서 시차변수에 의한 관측값의 손실을 줄이고 추정할 계수의 수를 감소시키기 위해 시차계수 w_i 가 어떤 체계적인 패턴을 따른다는 제약을 가하여 이러한 문제점을 해결한다. 아래의 시차분포모형을 고려해보자.

$$Y_t = \alpha + \sum_{i=0}^n w_i X_{t-i} + \epsilon_t \quad i = 0, 1, \dots, n \quad (3.1)$$

식(3.1)에서 독립변수의 가중값 w_i 는 아래와 같이 차수가 p 인 다항시차의 형태를 갖는다 가정하며, 차수 p 는 통상 시차의 길이 n 보다 작게 설정된다.

$$w_i = \sum_{k=0}^p \lambda_k (i)^k \quad p \leq n \quad (3.2)$$

식(3.1)은 선형회귀모형의 모든 가정들을 만족시킨다고 가정하면 이 모형의 추정 결과 불편, 효율, 일치추정계수를 얻게 된다. 식(3.2)를 (3.1)에 대입하면 아래와 같다.

$$Y_t = \alpha + \sum_{i=0}^n \left[\sum_{k=0}^p \lambda_k (i)^k \right] X_{t-i} + \epsilon_t \quad (3.3)$$

(3.3)을 다시 쓰면

$$Y_t = \alpha + \sum_{k=0}^p \lambda_k Z_{tk} + \epsilon_t \quad (3.4)$$

가 되고, 이때

$$Z_{tk} = \sum_{i=0}^n (i)^k X_{t-i} \quad k = 0, 1, \dots, p \quad (3.5)$$

이다. 식 (3.4)는 다항시차전환법을 적용시킬 때 추정되는 시차분포모형이다.

여기서

$X \equiv (X_t, X_{t-1}, \dots, X_{t-n})$: 자료의 $[1 \times (n+1)]$ 행벡터

$w \equiv (w_0, w_1, \dots, w_i, \dots, w_n)'$: 추정 시차계수의 $[(n+1) \times 1]$ 열벡터

$\lambda \equiv (\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_p)'$: 다항식 계수의 $[(p+1) \times 1]$ 열벡터

$\psi_i \equiv (1, i, i^2, \dots, i^p)$: i 의 다항식 함수

와 같이 정의하면 (3.1)은 아래와 같이 쓸 수 있다.

$$Y_t = Xw \quad (3.6)$$

여기서 $w_i = \psi_i \lambda$ 이다. $\phi_j (j = 0, 1, 2, \dots, p)$ 를 w_i 각각의 다항식의 값이라고 하면

$$\phi_j = \psi_{ij} \lambda \quad (3.7)$$

가 된다. $\phi \equiv (\phi_0, \phi_1, \phi_2, \dots, \phi_p)'$ 가 $1 \times (p+1)$ 행벡터이고, $\Lambda = (\psi'_{i0}, \psi'_{i1}, \psi'_{i2}, \dots, \psi'_{ip})'$ 인 정칙행렬이라면

$$\phi = \Lambda \lambda \quad (3.8)$$

가 되어 이는

$$\lambda = \Lambda^{-1} \phi \quad (3.9)$$

이므로 다항식의 계수 λ 가 다항식의 값 ϕ 로부터 유도됨을 뜻한다.

따라서 매 기간 독립변수의 가중치는

$$w_i = \psi_i \lambda = \psi_i \Lambda^{-1} \phi \quad i = 0, 1, 2, \dots, n \quad (3.10)$$

이고, 이때 $\Pi \equiv (\psi_0', \psi_1', \psi_2', \dots, \psi_n')$ 이라고 정의하면

$$w = \Pi\lambda = \Pi\Lambda^{-1}\phi \quad (3.11)$$

이므로 식 (3.6)과 (3.11)에서

$$Y_t = X(\Pi\Lambda^{-1})\phi \quad (3.12)$$

와 같이 쓸 수 있다. 식 (3.12)의 $\Pi\Lambda^{-1}$ 을 가중행렬이라고 하며, X 벡터와 이 행렬을 결합시켜 인공변수

$$Z_t = X(\Pi\Lambda^{-1}) \quad (3.13)$$

을 정의하면 $(n+1)$ 개의 새로운 인공변수가 생성된다. 이는 X_t 에서 X_{t-n} 까지의 가중합을 나타내고 가중치들은 $(\Pi\Lambda^{-1})$ 의 행이 된다.

식 (3.13)을 (3.12)에 대입하면

$$Y_t = Z_t\phi \quad (3.14)$$

가 된다. 계수들은 $(\Pi\Lambda^{-1})$ 이 다항식의 차수와 시차의 길이에 의해 주어졌으므로

$$\hat{w} = (\Pi\Lambda^{-1})\hat{\phi}$$

에 의해 계산되며, 또한 \hat{w} 의 분산-공분산 행렬은

$$V(\hat{w}) = (\Pi\Lambda^{-1})V(\hat{\phi})(\Pi\Lambda^{-1})'$$

가 된다.

다항시차분포모형의 연구 논문이 드문 이유 중 하나는 모형의 설계에 있어서 다항식의 차수와 시차의 길이를 결정하는데 어려움이 따르기 때문이다. 현실적으로 주요 변수들이 국내의 주가변동에 영향을 미치는 시차의 길이와 다항식의 차수에 대해서는 어떠한 사전 정보도 존재하지 않는다. 본 논문에서는 국내 증시의 일일변동에 영향을 미치리라 예상되는 변수들의 모든 가능한 쌍의 n 과 p 의 조합을 고려하여 통계학적인 모형선택의 기준에 따라 최적모형을 선정한다.

종합주가지수의 변동에 영향을 미칠 수 있는 변수들은 국내 요인과 해외 요인으로 구분할 수 있다. 먼저 국내 요인을 보면 국내 증시가 거래소시장과 코스닥시장의 양대 시장관계로 정립되면서 동반 및 경쟁관계가 반복되고 있다. 따라서 코스닥지수를 설명변수로 고려한다. 실세금리가 하락하면 기대수익으로 인해 시중자금은 금융기관에서 주식시장으로 선회한다. 금리지표로는 국고채, 회사채, 물금리가 있으며, 추정과정에서 이들 각각을 또 하나의 설명변수로 고려해 본다. 해외 요인을 보면 국내 경기는 환율의 변동에 영향을 받으며 이는 다시 증시에 영향을 미친다. 환율로는 엔, 달러, 엔-달러환율이 각각 고려된다.

국내 증시는 미국 증시와 동조화 현상이 강하며 종합주가지수는 기간에 따라 다우존스 지수나 나스닥지수와 변갈아 가며 높은 상관관계를 보여왔다. 특히 외국인 투자자들은 국내 변수보다는 미국 증시의 변동에 따라 움직이는 것으로 알려져 있으므로 미국의 증시를 또 하나의 설명변수로 고려한다. 일간자료를 사용함에 따라 국내 증시와 미국 증시의 개장일이 공휴일 등으로 인해 일치하지 않는 날은 결측값이 생기지 않도록 미국 지수는 3기간

이동평균값을 취한다. 추정에는 2000년 10월의 증시안정대책 발표와 11월의 퇴출기업 발표 이후 2001년 1월 미국의 금리인하, 국내 금융시장안정대책 발표 등으로 종합주가지수와 다른 자료들이 안정세를 보였던 2001년 1분기의 일간자료를 사용한다.

모든 시장변수들의 차수와 시차의 조합을 고려한 결과 종합주가지수의 변동을 설명하기 위해 최종적으로 선정된 모형은 아래와 같으며, 코스닥, 다우존스, 나스닥지수만이 통계적으로 유의한 설명력을 갖는 변수로 나타났다.

$$\ln Y_t = \alpha_1 + \sum_{i=0}^m w_{1i} \ln X_{1t-i} + \sum_{j=0}^n w_{2j} \ln X_{2t-j} + \alpha_2 \ln X_{3t} + \epsilon_t \quad (3.15)$$

시차분포모형을 사용한 거의 대부분의 연구들은 한 가지 주요 변수의 시차관계에만 분석을 한정시켜 왔으나 본 연구에서는 국내 주식시장의 구조적 특징을 잘 반영하는 더욱 현실적인 모형을 구해보고자 한다. 식 (3.15)에서 Y 는 종합주가지수, X_1 은 코스닥지수, X_2 는 미국의 다우존스이고, X_3 은 나스닥지수이다. 모형에서 코스닥지수의 영향은 시차의 길이가 8이고 차수가 2, 또 다우존스지수는 시차의 길이가 12이고 차수는 역시 2인 다항시차구조를 가진 반면 나스닥지수의 변동은 당일에만 통계적으로 유의한 영향을 미치는 것으로 나타났다. 코스닥지수와 다우존스지수의 변동은 아래와 같은 시차효과를 가진다.

$$w_i = \lambda_0 + \lambda_1 i + \lambda_2 i^2 \quad (3.16)$$

식 (3.16)을 (3.15)에 대입하면 다음과 같다.

$$\begin{aligned} \ln Y_t = & \alpha_1 + \lambda_0 \sum_{i=0}^m \ln X_{1t-i} + \lambda_1 \sum_{i=0}^m i \cdot \ln X_{1t-i} + \lambda_2 \sum_{i=0}^m i^2 \cdot \ln X_{1t-i} \\ & + \delta_0 \sum_{j=0}^n \ln X_{2t-j} + \delta_1 \sum_{j=0}^n j \cdot \ln X_{2t-j} + \delta_2 \sum_{j=0}^n j^2 \cdot \ln X_{2t-j} \\ & + \alpha_2 \ln X_{3t} + \epsilon_t \end{aligned} \quad (3.17)$$

(3.17)을 다시 쓰면

$$\ln Y_t = \alpha_1 + \lambda_0 Z_{0t} + \lambda_1 Z_{1t} + \lambda_2 Z_{2t} + \delta_0 W_{0t} + \delta_1 W_{1t} + \delta_2 W_{2t} + \alpha_2 \ln X_{3t} + \epsilon_t \quad (3.18)$$

가 되고, 이때 인공변수 Z 와 W 는 아래와 같이 정의 된다.

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^8 \ln X_{1t-i}, & Z_{1t} &= \sum_{i=0}^8 i \cdot \ln X_{1t-i}, & Z_{2t} &= \sum_{i=0}^8 i^2 \cdot \ln X_{1t-i} \\ W_{0t} &= \sum_{j=0}^{12} \ln X_{2t-j}, & W_{1t} &= \sum_{j=0}^{12} j \cdot \ln X_{2t-j}, & W_{2t} &= \sum_{j=0}^{12} j^2 \cdot \ln X_{2t-j} \end{aligned}$$

4. 추정결과

모형 (3.18)의 최대우도 추정결과는 식(4.1)과 같고, 이를 원모형 (3.15)로 변환시키면 표 4.1과 같으며, ()의 값은 t 값이다. 각 인공변수 Z 는 사실상 같은 설명변수들로 구성되어 있으므로 다중공선성의 가능성이 존재한다. Gujarati는 다항시차분포모형에서 다중공선성이 존재하는 경우 인공변수의 계수들은 통계적으로 유의하지 않게 판별된다고 했다.

$$\begin{aligned} \ln Y_t = & -2.6898 + 0.1543Z_0 - 0.0685Z_1 + 0.0062Z_2 + 0.2173W_0 \\ & (-2.50) \quad (5.64) \quad (-3.57) \quad (2.66) \quad (5.27) \\ & - 0.0976W_1 + 0.0085W_2 + 0.1384 \ln X_{5t} \\ & (-5.35) \quad (5.42) \quad (7.04) \end{aligned} \tag{4.1}$$

$$e_t = \epsilon_t + 0.7050e_{t-3}$$

Reg $R^2 = 0.9844$	Total $R^2 = 0.9637$
D-W = 1.8809	RMSE = 0.01123

Total R^2 은 설명변수에 의한 구조식과 자기상관오차를 포함한 전체모형의 설명력을 나타내는 R^2 으로서 구조식을 구성하는 설명변수와 현재 및 과거의 오차들이 얼마나 예측력을 갖는가를 나타내는 척도로 사용된다. Reg R^2 은 오차가 자기상관을 갖는다는 성질을 이용하여 오차가 자기상관을 갖지 않는 형태로 변환된 모형의 R^2 이다.

표 4.1의 계수들을 보면 코스닥지수와 다우존스지수의 변동은 시간이 흐르면서 일정 기간씩 변갈아가며 종합주가지수에 긍정적·부정적 영향을 미치는 것으로 나타났다. 즉 이들이 상승하면 거래소시장도 4기간 동반상승장을 이루지만 곧 반발매수세로 인해 4~5기간 동안 하락세로 반전되며 이후 다시 몇 기간은 상승여파로 이어지는 현실적인 장세를 반영하고 있다. 코스닥지수와 다우존스지수의 변화는 시간이 흐를수록 영향력이 감소해가는 일반적인 기하시차의 형태를 띠는 것이 아니라 알려진 바와 같이 상승세와 하락세가 이어지는 다항시차의 형태를 그대로 보이고 있다. 표 4.1에서 X_1 과 X_2 의 계수와 식 (4.1)의 Z 와 W 의 계수들의 추이를 보면 하락세에서 상승세로 변화하는 전환점이 한 개인 2차시차구조를 가지고 있음을 알 수 있다.

코스닥과 다우존스가 1% 상승하면 종합주가지수에 단기적으로는 각각 0.15%, 0.22%의 상승효과를 주며, 장기적으로는 각각 0.20%, 0.77%의 순상승효과를 가져오게 된다. 한편 나스닥지수가 1% 변화시 종합주가지수는 0.14% 정도의 즉각적 변화만 보일뿐 그 외의 효과는 미미했다. 따라서 국내 주식시장은 국내, 해외 시장요인의 변동에 비탄력적이며, 또한 예상대로 코스닥, 다우존스, 나스닥시장과 단기적, 장기적으로 같은 방향으로 움직인다는 것을 알 수 있다.

거래소시장은 다우존스지수가 변화할 때 단·장기적으로 가장 큰 반응을 보이며, 또한 적응과정이 가장 긴 것으로 나타났다. 표 4.1에서 보듯이 코스닥지수의 변화가 거래소시장에 미치는 초기효과는 나스닥지수보다는 크지만 시간이 흐르면서 반발매수세로 인한 하락

세가 강하게 작용해 장기적 순상승효과는 그리 크지 않은 것으로 추정되었다. 이와는 달리 다우존스지수가 미치는 장기적 총영향은 상승세가 하락세를 압도해 초기영향의 3.5배를 초과함과 동시에 코스닥지수가 미치는 총효과의 3.8배에 이르는 것으로 나타났다. 그 결과 종합주가지수는 해외 주식시장의 변동에 단기적·장기적으로 더 많은 영향을 받는다고 결론 지을 수 있다.

표 4.1: 추정결과 및 시차계수의 변동추이

변수	α_1	$\ln X_{1t}$	$\ln X_{2t}$	$\ln X_{3t}$
시차				
0	-2.6898 (-2.50)	0.1542 (5.61)	0.2173 (5.26)	0.1384 (7.04)
1		0.0920 (7.53)	0.1282 (5.01)	
2		0.0423 (5.44)	0.0562 (3.94)	
3		0.0051 (0.43)	0.0014 (0.14)	
4		-0.0195 (-1.38)	-0.0363 (-2.99)	
5		-0.0316 (-2.60)	-0.0570 (-3.87)	
6		-0.0313 (-4.11)	-0.0606 (-3.96)	
7		-0.0184 (-1.63)	-0.0471 (-3.45)	
8		0.0069 (0.26)	-0.0166 (-1.52)	
9			0.0310 (2.64)	
10			0.0958 (4.73)	
11			0.1776 (5.21)	
12			0.2766 (5.35)	

추정된 모형이 증시의 변동상황을 얼마나 정확히 반영하는가를 보기 위한 모의실험 결과는 그림 4.1과 같다. 추정계수와 설명변수들의 실제값을 사용하여 구한 지수값은 모형의 RMSE값이 의미하듯이 실제의 주가지수를 잘 재생해내고 있으며, 추정기간 동안 지수의 장기적 동향을 큰 오차없이 추적하고 있음을 알 수 있다. 그림 4.1의 우측은 추정기간 이후 한 달 동안의 종합주가지수와 추정모형을 사용하여 계산한 예측값을 비교한 것이다. 2001년 4월은 추정기간에 포함되어 있지 않으므로 실제값과 예측값과의 오차는 추정기간에 비해 상대적으로 클 수 밖에 없으나 시장의 움직임을 전반적으로 잘 예측하고 있다.

그러나 실제값과 예측값간 오차의 크기만으로 모형의 예측능력을 평가할 수는 없다. 주식시장을 분석하는 경우가 좋은 예로서 추정된 모형을 사용하여 하루가 다르게 급변하는 증시동향을 예측하는데는 지수의값을 얼마나 정확히 예측하느냐는 양적인 기준보다 지수의 등락여부를 얼마나 정확히 예측하는가를 보는 질적인 평가기준이 더욱 중요하다 하겠다.

추정기간 동안의 모의실험값과 그 이후 한 달간의 예측값들을 실제값과 비교하면 모형은 추정기간 동안 주가지수의 등락방향을 66%나 예측하는 수준급 적중률을 보였으며, 또

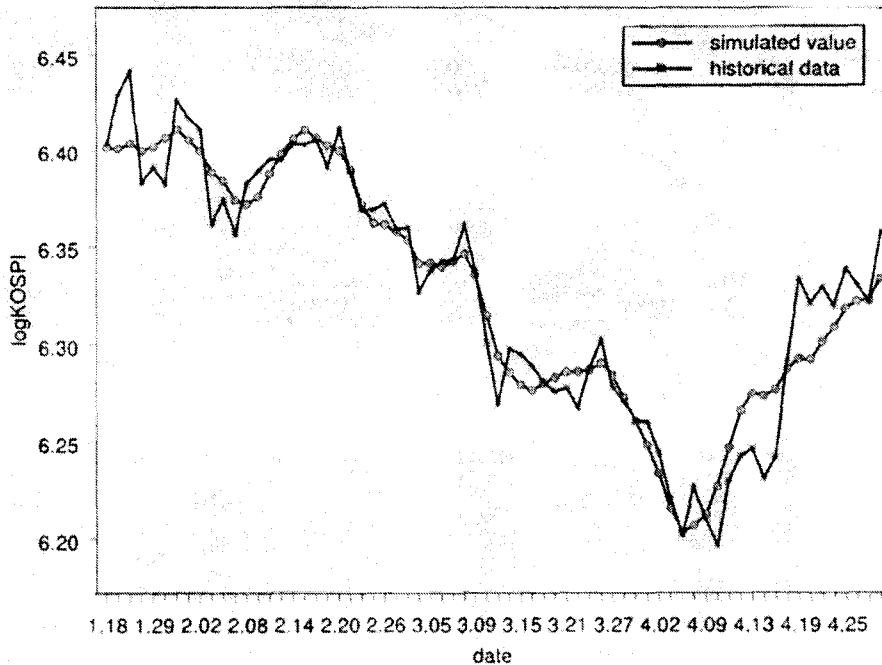


그림 4.1: 종합주가지수의 변동추이와 예측결과

한 추정기간을 벗어난 2001년 4월 한 달 동안은 20개의 개장일 중 15일의 등락방향이 예측되어 정확도가 75%로 높아졌다. 추정모형을 사용하여 추정기간 이후를 예측하는 경우 먼 미래로 갈수록 정확도는 떨어지게 된다. 따라서 4월 첫 15개 개장일만 고려해 본다면 15개 관측값 중 13개, 즉 87%의 놀라운 등락 예측능력을 보였다.

추가로 아래의 Theil계수를 계산하여 모형의 모의실험 및 예측능력을 평가해본다.

$$U = \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t^a - Y_t^b)^2} / \left[\sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t^a)^2} + \sqrt{\frac{1}{T} \sum_{t=1}^T (Y_t^b)^2} \right] \quad (4.2)$$

Y^a 는 모의실험으로 구한 종속변수의 재생값이고, Y^b 는 실제값, 또 T 는 모의실험 기간의 수이며, U 는 항상 0과 1사이의 값을 취한다. 만약 $U = 0$ 이라면 모든 기간에 대해 $Y^a = Y^b$ 가 되어 적합이 완벽한 반면 $U = 1$ 은 최악의 예측오차를 갖는 경우이다. 식(4.2)는 아래와 같이 분해될 수 있다.

$$\frac{1}{T} \sum (Y_t^a - Y_t^b)^2 = (\bar{Y}^a - \bar{Y}^b)^2 + 2(1 - \rho)\sigma_a\sigma_b + (\sigma_a - \sigma_b)^2 \quad (4.3)$$

여기서 \bar{Y}^a 와 σ_a 는 Y_t^a 의 평균과 표준편차, \bar{Y}^b 와 σ_b 는 Y_t^b 의 평균과 표준편차이며, ρ 는

이들의 상관계수이다.

식(4.3)에서 다음과 같은 통계량들을 정의할 수 있다.

$$U_1 = (\bar{Y}^a - \bar{Y}^b)^2 / \frac{1}{T} \sum (Y_t^a - Y_t^b)^2 \quad (4.4)$$

$$U_2 = (\sigma_a - \sigma_b)^2 / \frac{1}{T} \sum (Y_t^a - Y_t^b)^2 \quad (4.5)$$

$$U_3 = 2(1 - \rho)\sigma_a\sigma_b / \frac{1}{T} \sum (Y_t^a - Y_t^b)^2 \quad (4.6)$$

U_1, U_2, U_3 은 각각 편차, 분산, 공분산비율로서 식(4.3)에서 $U_1 + U_2 + U_3 = 1$ 이 되며, 모의 실험 오차의 발생 출처를 분류해 주는 수단으로 사용될 수 있다.

U_1 은 모의실험값과 실제값들간의 편차의 평균값을 측정해 주므로 체계적 오차를 나타내는 지표이다. 따라서 식(4.2)의 U 값에 상관없이 U_1 의 값은 0에 가까울수록 좋으며, 0.1 이나 0.2 만 되어도 체계적인 편차가 존재한다는 뜻으로 이 경우 모형을 수정해야 한다. U_2 는 관심의 대상이 되는 변수의 변동 정도를 복제할 수 있는 모형의 능력을 나타내주는 지표이다. 이 값이 크면 예컨대 실제값들은 진폭이 심하게 변동하지만 모의실험값들은 큰 변동이 없다는 뜻이므로 문제가 되며 역시 모형을 수정해야 한다. 마지막으로 U_3 은 평균값으로부터의 편차와 평균변동들이 고려된 이후의 잔여오차를 나타내므로 비체계적 오차를 측정해준다. 예측값들과 실제값들의 상관관계가 완벽하기를 기대하기는 어려우므로 이 부분의 오차는 별다른 문제가 되지 않는다. $U_1 \sim U_3$ 의 가장 이상적인 분포는 $U_1 = U_2 = 0$ 이고, $U_3 = 1$ 이라 할 수 있다.

표 4.2: Theil 계수의 분포

	추정기간	예측기간
U	0.000747	0.001847
U_1	0.002157	0.002037
U_2	0.026480	0.258267
U_3	0.992618	0.792211

모형의 추정결과로 추정기간과 예측기간에 대해 각각 식(4.2)와 (4.4) ~ (4.6)을 계산한 결과는 표 4.2 와 같다. 먼저 U 값은 거의 0에 해당하므로 모형이 구조적으로 정교하다는 것을 입증하고 있으며, $U_1 \sim U_3$ 의 합은 두 기간 각각 1에 가깝다는 것을 알 수 있다. U_1 의 값이 두 기간 모두 0에 가까우므로 체계적인 오차는 존재하지 않으며, U_2 와 U_3 의 값은 추정기간에 더욱 이상적인 분포를 보인다. 한편 예측기간의 U_2 값은 추정기간에 비해 다소 높은데 이는 실제값의 변동폭이 예측값에 비해 상대적으로 크다는 것을 의미하며 그림 4.1의 우측에 나타난 바와 같다. 따라서 추정기간과 예측기간으로 구분하여 모형의 수행능력을 비교해 본다면 양적인 기준으로는 추정기간에, 그러나 질적인 기준으로는 앞서 본 바와 같이

예측기간에 더 정확함을 알 수 있으며, 종합적으로 평가하면 U 통계값들은 모형의 우수함을 다시 입증해주고 있다.

5. 결론

서로 관련된 변수들 중 어느 변수에 변동이 발생할 때 다른 변수에 미치는 영향은 같은 기간 뿐 아니라 현실적으로는 여러 기간에 걸쳐 지속된다. 그러나 통상적인 통계모형으로 현실상황을 추정하는 경우 설명변수의 변동으로 인해 종속변수가 받는 영향은 그때 뿐이라는 제약이 가해지게 되어 모형이 현실을 제대로 반영하지 못한다. 실제의 시장동향을 설명하기 위해 추정되는 모형이 외생적 변동 후 시간이 흐름에 따라 변수들간의 관계가 변화해가는 과정까지 파악되도록 설계된다면 더욱 현실적이고도 다양한 분석이 가능해진다. 외생적 시장요인의 변동이 국내 주식시장에 미치는 영향은 실제로 여러 기간에 걸쳐 지속되며 조정되어 가므로 본 연구에서는 이러한 현실에 맞는 동태적 시차분포모형을 개발하였다.

추정 결과 모형은 외생적 충격에 따른 증시동향의 현실을 잘 반영하고 있으며, 주가지수는 알려진 바와 같이 국내 시장요인보다는 해외 증시의 변동에 장단기적으로 더 민감한 반응을 보이는 것으로 나타났다. 또한 거래소시장은 코스닥시장과는 상호보완관계, 그리고 해외 증시와는 상호동반관계에 있는 시장여건도 그대로 반영되고 있다. 주가지수는 코스닥시장과 해외시장에 변동이 발생하면 시간이 흐르면서 상승세와 하락세가 반복되는 동태적 영향을 받는다는 모형의 표기는 증시의 현실과 일치하며, 모의실험 결과에 의한 양적, 질적평가와 더불어 또 예측력에서도 모형의 정교함은 입증되고 있다.

참고문헌

- [1] 고광수 (1997). 기대수익률과 주가변동성의 관계 연구, <한국경영과학회지>, 22권 2호, 153-167.
- [2] 김서경 · 고광수 (2001). 주가지수와 주가지수 선물 관계의 일증거래 자료분석, <증권학회지>, 27집 4호, 101-137.
- [3] 김석용 · 이하일 (1998). KOSPI 200 선물을 이용한 동적헤징전략에 관한 실증적 연구, 증권학회 제2차 학술발표회.
- [4] 박상용 · 연강흠 (1994). 자본시장개방이 환율 · 주가 · 금리간의 상호연관성에 미치는 영향, <경영학연구>, 23권 4호, 47-79.
- [5] 박재환 (1998). 주식수익률 시계열의 ARIMA모형 설정 및 분석, <증권학회지>, 23집 6호, 187-210.
- [6] 박재환 (2000). 금리변수가 주식수익률에 미치는 영향, <증권학회지>, 26집 8호, 199-236.

- [7] 안태백 (1988). 신주발행이 주식수익률에 미치는 영향에 관한 연구, <경영학연구>, 18권 1호, 61-97.
- [8] 이일균 (1998). 주가시계열에 대한 확률미분방정식의 모수 추정과 자본시장의 운동법칙, 증권학회 제 2차 학술발표회.
- [9] 이일균 (2001). 시간의 가역성과 주가, <증권학회지>, 27집 3호, 71-99.
- [10] 이정도 · 김종호 (1996). 확률적 지배에 의한 포트폴리오보험전략의 비교우위 분석, <생산성논집>, 11권 1호, 7-38.
- [11] 이청림(2001). 경제·경영 변수간 장단기 동태적 시차관계의 설정, 충북대학교 통계학과 석사학위논문.
- [12] 조성표·정재용(2001). 연구개발지출의 다기간 이익효과 분석, <경영학 연구>, 30권 1호, 289-315.
- [13] 조영경(1999). 주요 거시경제변수와 KOSPI간의 실증적 영향도 분석, <경영학 연구>, 28권 3호, 589-608.
- [14] 조재범 · 김호일 (1998). 마르코프 국면전환모형을 이용한 KOSPI와 금리의 추이 분석, <한국통계학회논문집>, 5권 1호, 177-191.
- [15] 최정호 · 나종길 (2001). 부실기업 공시의 산업내 정보전이효과, <증권학회지>, 27집 9호, 261-299.
- [16] 허창수 (1988). 한국 주식시장과 미국 주식시장의 분산한계검증에 관한 비교연구, <경영학연구>, 18권 1호, 9-39.
- [17] 황성혁·이정희 (2000). 건강정보에 의한 육류수요의 분석, <농업경제연구>, 41집 1권, 61-79.
- [18] Chow, Gregory C. (1988). *Econometrics*, McGraw-Hill.
- [19] Cooper, J. Phillip (1972). Two Approaches to Polynomial Distributed Lags Estimation : An Expository Note and Comment, *The American Statistician*, 26, 32-35.
- [20] Darrat, Ali F. (1999). Are Financial Deepening and Economic Growth Causally Related? Another Look at the Evidence, *International Economic Journal*, vol.13 #3. 19-35.
- [21] Gujarati, Damodar N. (1995). *Basic Econometrics*, 3rd ed., McGraw Hill
- [22] Haynes, Stephen E. and Joe A. Stone (1983). Specification of Supply Behavior in International Trade, *Review of Economics and Statistics*, 626-632.
- [23] Judge, George G., R. Cater Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-

- Chao Lee (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd ed., John Wiley and Sons.
- [24] Leamer, Edward and Herman Leonard (1983). Reporting the Fragility of Regression Estimates, *Review of Economics and Statistics*, vol.65, 306-317.
- [25] Maddala, G. S. and A. S. Rao (1971). Maximum Likelihood Estimation of Solow's and Jorgensen's Distributed Lag Models, *Review of Economics and Statistics*, vol.53, 80-88.
- [26] Pagano, Marcello, and Michael J. Hartley (1981). On fitting Distributed Lag Models Subject to Polynomial Restrictions, *Journal of Econometrics*, vol.16 171-198.
- [27] Pindyck, Robert S. and Daniel L. Rubinfeld (1998). *Econometric Models and Economic Forecasts*, 4th ed., McGraw Hill.
- [28] Schmidt, P. and Roger N. Waud (1973). The Alman Lag Technique and the Monetary versus Fiscal Policy Debate, *Journal of American Statistical Association*, vol.68, 11-19.
- [29] Thursby, Jerry and Marie Thursby (1984). How Reliable are Simple, Single Equation Specification of Import Demand?, *Review of Economics and Statistics*, 120-128.
- [30] Warner, Dennis and Mordechai E. Kreinin (1983). Determinants of International Trade Flows, *Review of Economics and Statistics*, 65, 96-104.

[2002년 3월 접수, 2002년 11월 채택]

An Analysis for the Adjustment Process of Market Variations by the Formulation of Time Lag Structure

T.H.Kim ¹⁾ C.L.Lee ²⁾

ABSTRACT

Most of statistical data are generated by a set of dynamic, stochastic, and simultaneous relations. An important question is how to specify statistical models so that they are consistent with the dynamic feature of those data. A general hypothesis is that the lagged effect of a change in an explanatory variable is not felt all at once at a single point in time, but The impact is distributed over a number of future points in time. In other words, current control variables are determined by a function that can be reduced to a distributed lag function of past observations. It is possible to explain the relationship between variables in different points of time and to estimate the long-run impacts of a change in a variable on another if time lag series of explanatory variables are incorporated in the model specification. In this study, distributed lag structure is applied to the domestic stock market model to capture the dynamic response of the market by exogenous shocks. The Domestic market is found more responsive to the changes in foreign market factors both in the short and the long run.

Keywords: Time lags structure, Polynomial lag distribution, Artificial variable, Turning point forecast, Theil's inequality coefficient

1) Professor, Department of Statistics, Chungbuk National University

E-mail : thkim@trut.chungbuk.ac.kr

2) Former Graduate Student, Department of Statistics, Chungbuk National University

E-mail : clee@netian.com