

# Goodness of Fit and Independence Tests for Major 8 Companies of Korean Stock Market

Seungsik Min<sup>a,1</sup>

<sup>a</sup>Department of Natural Science, Korea Naval Academy

(Received November 9, 2015; Revised December 21, 2015; Accepted December 24, 2015)

---

## Abstract

In this paper, we investigated the major 8 companies of Korean stock market, and carried out the goodness of fit and independence tests. We found out the distributions of absolute returns are closed to compressed exponential distribution. The parameters are dominant that  $1 < \beta < 2$ , followed by  $\beta = 1$ (exponential distribution) and  $\beta = 2$ (normal distribution). Meanwhile, we assured that most of the absolute returns for major 8 companies have relevance to each other by chi-square independence test.

Keywords: distribution, absolute return, goodness of fit test, independence test, compressed exponential distribution

---

## 1. 서론

경기종합지수는 국내 경제의 각 부문을 대표하고, 경기대응성을 갖는 개별 경제지표들을 합성하여 만드는 대표적인 종합 경기판단지표이다 (Kim, 1998). 2015년 9월 기준 경기종합지수는 9개의 선행종합지수, 7개의 동행종합지수, 5개의 후행종합지수로 구성된다 (통계청). 이 중 종합주가지수는 가장 익숙한 값으로 실시간 변동성이 매우 크다. 지수를 구성하는 개별 주식의 매매는, 모기업 입장에서 자본금의 역할을 하고, 가계 입장에서 투자처, 또는 주권 행사의 수단이 되기 때문에 종합주가지수에 대한 수많은 연구가 진행되어 왔다. 하지만 효율적 시장 가설(efficient market hypothesis; EMH)에 의하면 주가는 시중의 정보를 흡수하기 때문에 투자자는 차익을 실현할 수 없다 (Fama, 1970). 바꾸어 말하면, 주가는 불확실한 외부 정보가 직접 반영되므로 임의 보행(random walk) 또는 브라운 운동(Brownian motion)을 따른다는 의미이다. 경험적 자료 분석에 근거한 많은 선행 연구들에서 임의 보행 가설(random walk hypothesis; RWH)이 지지된다 (Working, 1960; Fama, 1965; Niederhoffer와 Osborne, 1966). 이는 곧 주가 수익률의 분포가 정규분포로 수렴한다는 것을 의미한다. 그러나 최근 연구들에 의하면 주가지수는 기존의 임의 보행 가설에서 진화하여 프랙탈 브라운 운동(fractal Brownian motion), 멀티프랙탈 브라운 운동(multi-fractal Brownian motion) 등을 따르는 것으로 여겨지고 있다 (Panas와 Ninni, 2010). 따라서 주가 수익률 역시 정규분포뿐만 아니라 지수분포(exponential distribution), 또는 멱함

---

This research is an updated version of a part for the master's thesis of Seungsik Min (Department of Physics, KAIST, 2009).

<sup>1</sup>Department of Natural Science, Korea Naval Academy, 1 Jungwon-ro, Changwon-si, Gyeongsangnam-do 51704, Korea. E-mail: [fieldsm@gmail.com](mailto:fieldsm@gmail.com)

수분포(power-law distribution) 등 여러 분포함수들이 제시되고 있다 (Kaizoji와 Kaizoji, 2003; Hong 등, 2007). 나아가 주가 수익률은 확장된 지수 분포(stretched exponential distribution)를 따른다는 견해도 제기되고 있다 (Laherrere와 Sornette, 1999).

따라서 본 논문에서는 선행연구를 바탕으로 코스피 시장에 상장된 시가총액 상위 8개사 주가 수익률의 분포를 일원 분할표(1-way contingency table)에 기반한 카이제곱 적합도 검정 및 콜모고로프-스미르노프 검정을 통해 확인하고(일반적으로 연속형 확률분포 검정 시 Kolmogorov-Smirnov의 방법을 많이 사용하지만, 통계값을 일정 구역으로 분할(각 구역의 빈도수가 5 이하인 구역이 전체의 1/5 이하가 되도록 인위적으로)한 후,  $\chi^2$  적합도 검정을 시행하기도 한다. Hayter (2007, 466-467)는 저서에서 자동차의 브레이크 패드가 지수분포를 이루는지 여부와, 콘크리트 블록이 정규분포를 이루는지 여부를 적합도 검정하는 사례를 보였다.), 이들 수익률 간의 독립성을 이원 분할표(2-way contingency table)에 기반한 카이제곱 검정을 통해 확인하고자 한다.

## 2. 분석 데이터

### 2.1. 조사 대상

본 연구에서는 2008년 11월 28일 종가 기준 한국 유가증권시장에 상장된 시가 총액 상위 8개사 주가의 일일 데이터를 대상으로 하였다. 상위 8개사는 총 926개의 상장사 중 0.864%에 불과한 개수지만 시가 총액은 33.1%나 차지하고 있다. 이는 매우 편중된 구조로 상위 기업의 시장 교섭력이 비정상적으로 강함을 의미한다. 반면 상위 기업을 분석함으로써 유가증권시장 전체의 흐름을 추산하기 용이함을 의미하기도 한다 (Min, 2009).

### 2.2. 조사 기간

본 연구의 조사 기간은 2003년 4월 1일부터 2015년 3월 31일까지 2,976일이었다. 이는 8개사의 가격이 모두 제공되면서도 가격 변동폭이  $\pm 30\%$ 로 확대되기 이전 기간을 선정한 것이다(2015년 6월 15일부로 한국 유가증권시장의 가격 제한폭이  $\pm 30\%$ 로 확대됨). 한편 거시경제학에서는 4가지 대표적인 경기 순환 주기를 정의하는데, 이 중 설비 투자 변동, 기계 발명 등 개별 기술 혁신에 의한 순환 주기(Juglar cycle or business cycle)를 7~11년으로 보고 있으며, 보다 중기적으로 건축 경기 변동에 의한 순환 주기(Kuznets cycle)를 15~25년으로 보고 있다 (Korotayev와 Tsirel, 2010). 따라서 개별 기업을 대상으로 한, 연구에서의 조사 기간인 12년은 사업 순환 주기 전체를 포함하고 있어, 불황, 회복, 호황, 후퇴 등 특정 순환 주기에 의해 개별 기업의 주가 수익률이 확대 또는 축소 해석되는 효과를 줄일 수 있다.

Figure 2.1은 조사 대상 기업들 주가 수익률의 절대값을 나타낸 그림이다. 대부분의 기업에서 1,350일 근처에서 값이 크게 나타나는데, 이때가 2008년 9월 이후로 미국의 서브프라임 모기지 사태로 촉발된 세계적인 금융위기가 도래하던 시기이다. 또한 2,150일 전후로 약한 충격이 나타나는데, 이때는 2011년 12월경으로 유럽의 재정 위기로 불거지던 시기이다. 대부분의 기업들에서 주가 수익률 절대값이 동조화(synchronized)되는 경향이 있으나 SK텔레콤, KT&G 등 일부 기업은 어느 정도 독립성을 유지하고 있는 것으로 보인다.

## 3. 배경 이론

### 3.1. Goodness of fit test for one-way contingency table (Hayter, 2007)

각각의 구역이 미지의 확률  $p_1, \dots, p_k$ 를 가지는  $k$ 개의 구역을 고려하면 다음과 같은 귀무가설을 설정

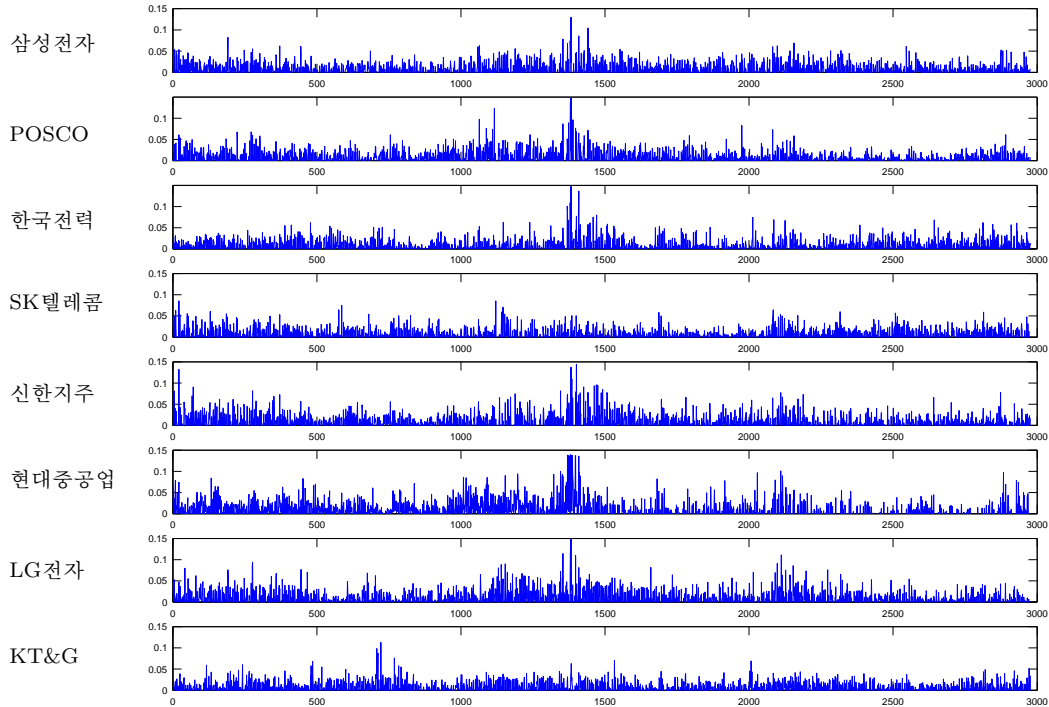


Figure 2.1. Absolute returns for 8 major companies in KOSPI.

할 수 있다.

$$H_0 : p_1 = p_i^*, \quad 1 \leq i \leq k.$$

이 때  $p_1^*, \dots, p_k^*$ 에 대한  $p$ -value는 다음의 카이제곱 검정 통계량(chi-square test statistic)으로부터 계산된다.

$$p\text{-value} = P(\chi_{k-3}^2 \geq X^2).$$

이 때

$$X^2 = \sum_{i=1}^k \frac{(x_i - e_i)^2}{e_i}, \quad \begin{array}{l} x_i \text{는 각각의 구역에서 관측된 빈도수로 } x_1 + \dots + x_k = n, \\ e_i = np_i^*, \\ \text{자유도} = \text{셀의 갯수} - 1 - \text{표본으로부터 추정된 모수의 갯수.} \end{array}$$

유의수준  $\alpha$ 로 검정하였을 때, 귀무가설은  $X^2 > \chi_{\alpha, k-3}^2$ 이면 기각된다. 분포 검정 시 평균과 분산이 주어지지 않고, 표본평균과 표본분산으로부터 추정하기 때문에 자유도는  $k - 3$ 이 된다.

### 3.2. Kolmogorov-Smirnov distribution test (Roussas, 1997)

독립이며 같은 분포를 따르는 확률변수  $X_i, i = 1, \dots, n$ 에 대해 경험적 분포(empirical distribution)는

다음과 같이 정의할 수 있다.

$$F_n(x) = \frac{1}{n} \{\text{number of } X_i \text{ such that } X_i < x\}.$$

이 때  $F_n(x)$ 의 수렴값  $F(x)$ 가  $F_0(x)$ 인지 여부는 다음의 귀무가설을 설정하여 확인할 수 있다.

$$H_0 : F = F_0.$$

이 때 콜모고로프 통계량  $K$ 를 아래와 같이 정의하면 누적 확률값을 도출해 낼 수 있다.

$$D_n = \sup |F_n(x) - F(x)|,$$

$$P(\sqrt{n}D_n \leq x) \rightarrow \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} \quad \text{as } n \rightarrow \infty.$$

따라서 충분히 큰  $n$ 에 대해 유의 확률  $\alpha$ 로 검정할 때,  $\sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2(\sqrt{n}D_n)^2} > 1 - \alpha$ 이면 귀무가설을 기각한다.

### 3.3. Compressed exponential distribution (McCauley, 2004)

주가 수익률에 대한 가장 전통적인 관점이 정규분포(Gaussian distribution)를 이룬다는 것이다. 하지만 근래에는 수익률 곡선이 보다 두터운 꼬리(heavy-tail)를 갖는 것으로 알려지고 있으며 지수분포(exponential distribution)를 이루는 것으로 분석되기도 한다. 이 같은 분포들을 일반화하여 압축된 지수 분포(compressed exponential distribution)라고 표현한다.

$$f_\beta(x) \sim \exp(-a|x|^\beta)$$

특히,  $\beta = 1$ 이면 지수분포,  $\beta = 2$ 이면 정규분포를 나타낸다. 만약  $\beta < 1$ 이면 확장된 지수 분포(stretched exponential distribution)라고 표현하며, 모든  $\beta$  범위에 대해 확장된 지수 분포로 정의하기도 한다.

### 3.4. Independence test for two-way contingency table (Hayter, 2007)

$r \times c$ 개로 이루어진 이원 분할표(contingency table)가 주어졌을 때 다음과 같은 귀무가설을 설정할 수 있다.

$$H_0 : \text{두 분포는 서로 독립적이다.}$$

마찬가지로 각 범주에 대한  $p$ -value는 다음의 카이제곱 검정 통계량을 사용한다.

$$p\text{-value} = P(\chi_{(r-1)(c-1)}^2 \geq X^2), \quad X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(x_{ij} - e_{ij})^2}{e_{ij}}.$$

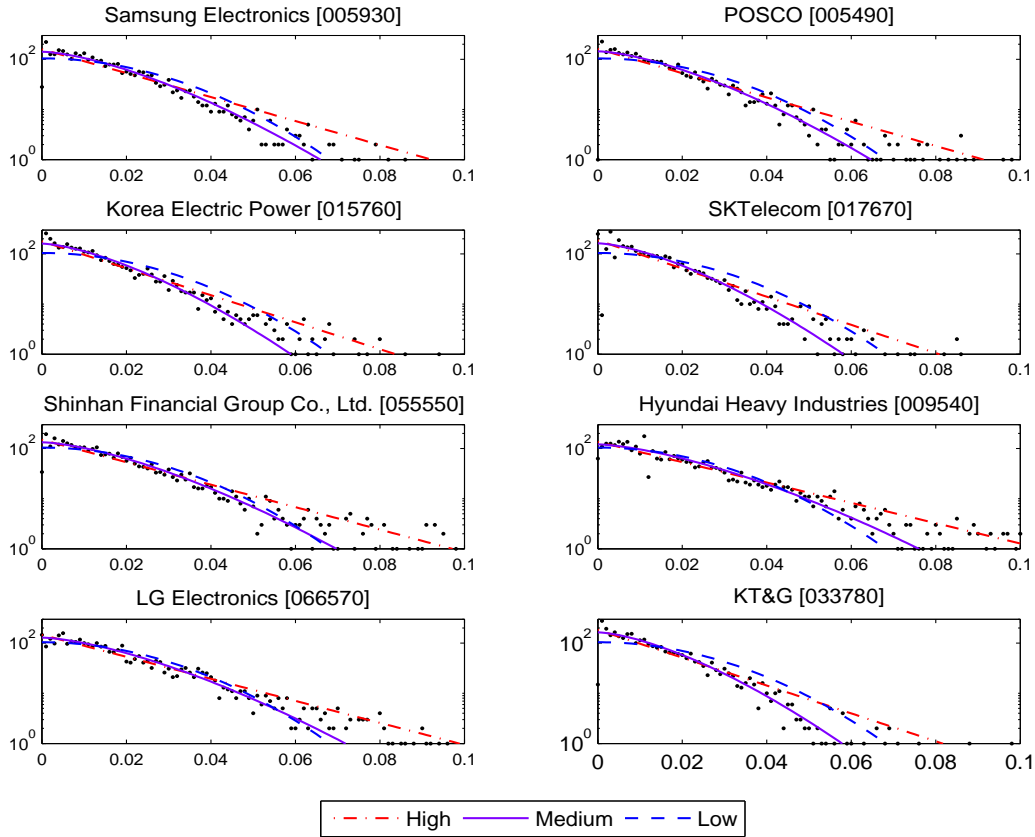
이 때  $x_{ij}$ 는 각 구역에서 관측된 빈도수이고,  $e_{ij}$ 는 그 기대값이다. 만일 두 범주가 독립적이라면 기대값은 다음과 같이 표현된다.

$$e_{ij} = \frac{x_{i \cdot} x_{\cdot j}}{n}.$$

이 때

$$x_{i \cdot} = \sum_{j=1}^c x_{ij}, \quad x_{\cdot j} = \sum_{i=1}^r x_{ij}$$

이다.



**Figure 5.1.** Log scaled frequencies of absolute price returns for 8 major companies in KOSPI, and their compressed exponential estimates for  $\beta = 1, 1.5, \text{ and } 2$ .

#### 4. 분석 방법

조사 기간 중 개별 기업들의 주가는 매일  $\pm 15\%$  이내로 변동한다. 이들 주가 수익률 절대값의 빈도수를 구하면 분포를 추정할 수 있다. 본 연구에서는 지수분포, 정규분포를 포함하는 압축된 지수분포(compressed exponential distribution)에 대해  $\chi^2$  통계량과 콜모고로프 통계량을 이용한 적합도 검정을 시행하였다.

한편 이들 분포들이 독립성을 띄는지 여부를  $\chi^2$  통계 검정을 통해 추정하였다. 이 때 각 구역을 세분화하면 빈도수 조건(각 셀의 빈도수 값이 5 이하인 경우가 20% 이하여야 함)이 만족되지 않는 경우가 많다. 따라서 본 조사에서는  $-15\% \sim -3\%$ ,  $3\% \sim 15\%$ 의 변동폭을 지니는 것을 각각의 구역으로 설정하고  $-3\% \sim 3\%$  사이에는 0.5% 단위로 12개의 구역을 설정하여 총  $14 \times 14 = 196$ 개 구역에 대해 독립성 검정을 시행하였다.

#### 5. 분석 결과

##### 5.1. Goodness of fit test results of the individual stock price return

Figure 5.1은 개별 주가 수익률 절대값의 빈도수를 로그 좌표로 그린 것이다. 구체적인 범위는 0~15%

**Table 5.1.** Goodness of fit tests for stock price returns of 8 major companies by chi-square test

Distribution	$\beta$	삼성전자	POSCO	한국전력	SK텔레콤	신한지주	현대중공업	LG전자	KT&G	
Exponential	1	179.5 (*)	125.7 (0.898)	105.8 (0.996)	132.4 (0.800)	104.8 (0.997)	98.9 (0.999)	100.0 (0.999)	125.2 (0.903)	
	1.1	116.3 (0.971)	71.8 (1.000)	58.3 (1.000)	82.0 (1.000)	59.0 (1.000)	59.9 (1.000)	57.7 (1.000)	73.7 (1.000)	
	1.2	71.3 (1.000)	40.1 (1.000)	35.8 (1.000)	47.9 (1.000)	32.1 (1.000)	<b>44.6</b> ( <b>1.000</b> )	<b>36.4</b> ( <b>1.000</b> )	37.0 (1.000)	
	1.3	38.6 (1.000)	22.7 (1.000)	<b>20.1</b> ( <b>1.000</b> )	28.6 (1.000)	<b>22.6</b> ( <b>1.000</b> )	53.6 (1.000)	37.3 (1.000)	16.9 (1.000)	
	1.4	21.1 (1.000)	<b>20.7</b> ( <b>1.000</b> )	30.9 (1.000)	<b>23.0</b> ( <b>1.000</b> )	32.5 (1.000)	48.1 (1.000)	41.2 (1.000)	<b>8.0</b> ( <b>1.000</b> )	
	Compressed exponential	1.5	<b>15.9</b> ( <b>1.000</b> )	23.7 (1.000)	62.9 (1.000)	33.9 (1.000)	52.9 (1.000)	52.1 (1.000)	50.2 (1.000)	13.4 (1.000)
		1.6	23.9 (1.000)	38.0 (1.000)	76.5 (1.000)	52.8 (1.000)	57.8 (1.000)	118.8 (0.958)	88.4 (1.000)	22.8 (1.000)
		1.7	38.5 (1.000)	94.3 (1.000)	216.9 (***)	71.2 (1.000)	110.4 (0.989)	117.5 (0.965)	111.2 (0.988)	28.3 (1.000)
		1.8	27.7 (1.000)	61.7 (1.000)	154.6 (0.318)	139.9 (0.648)	113.9 (0.980)	303.5 (***)	247.8 (***)	74.6 (1.000)
		1.9	93.6 (1.000)	139.2 (0.665)	378.8 (***)	56.8 (1.000)	288.3 (***)	256.5 (***)	195.2 (**)	81.7 (1.000)
Gaussian	2	251.8 (***)	135.3 (0.747)	131.0 (0.823)	89.2 (1.000)	100.8 (0.999)	214.8 (***)	388.9 (***)	124.5 (0.911)	

The values out of the parentheses are  $\chi^2$  statistics, and those in the parentheses are  $p$ -values.

\*:  $p$ -value < 0.05, \*\*:  $p$ -value < 0.01, \*\*\*:  $p$ -value < 0.001.

사이에 놓이는데, 이를 0.1% 크기로 150등분하여 각각의 구간에 놓이는 빈도수를 그래프로 나타낸 것이다. 우리가 검정에 사용한 분포는  $f_\beta(x) \sim \exp(-a|x|^\beta)$  형태의 압축된 지수 분포(compressed exponential distribution)이다. 그림에서 직선형 점선이  $\beta = 1$ 인 지수분포(exponential distribution)이고 곡선형 점선이  $\beta = 2$ 인 정규분포(Gaussian distribution)이다. 이 때 모든 경우에서 분포 추정 시 각각의 빈도수 총합을 유지하였다. 조사 기간이 2003년 4월 1일부터 2015년 3월 31일까지 2,976일이므로 0.1% 단위 구간에서 빈도수 총합은 2,976이 된다. 대부분의 실측값이 초반기에는 직선형을 유지하다가 값이 커지면서  $\beta = 1$ 인 경우의 적합선(fitted line) 아래에 놓이는 것을 확인할 수 있다. 반면,  $\beta = 2$ 인 경우와는 처음부터 모양에 차이가 있는 것을 확인할 수 있다. 따라서 우리는 일반적인  $\beta$  값에 대해 추가로 검정하였다(곡선형 실선). 그림에서는  $\beta = 1.5$ 인 경우를 대표적으로 표시하였는데, 대체적으로  $\beta$ 가 1, 또는 2인 경우보다 실측값에 더 가깝다는 것을 확인할 수 있다.

Table 5.1은 개별 주가 수익률 분포에 대한 카이제곱 적합도 검정 결과이다.  $\beta$  값에 따라 카이제곱 값에 큰 차이가 나는 것을 확인할 수 있다. 이 때 괄호 안의 값은 주어진 분포를 가정한 적합도 검정의  $p$ -value이다. 각 종목을 살펴보면  $1.2 \leq \beta \leq 1.5$  범위에서 카이제곱 값이 최대를 나타내므로, 분포함수는  $\beta = 1$ 인 지수분포나,  $\beta = 2$ 인 정규분포보다  $f_\beta(x) \sim \exp(-a|x|^\beta)$  ( $\beta \neq 1, 2$ )가 우세하게 지지된다. 한편, 유의도  $\alpha = 0.01$ 로 검정하였을 때,  $\beta = 1$ (exponential)인 경우는 8개사 모두에서 기각할 수 없는 반면,  $\beta = 2$ (Gaussian)인 경우는 3개사를 기각할 수 있으므로 주가 수익률은 정규분포보다는 지수분포에 가깝다는 것을 알 수 있다. 다만 앞서 언급하였듯이 값이 커질수록 실측값의 빈도수가 지수분포에 비해 급격히 감소하므로 수익률 절대값이 3%를 넘어서면서 지수분포와는 거리가 있는 것으로 여겨진다.

**Table 5.2.** Goodness of fit tests for stock price returns of 8 major companies by Kolmogorov-Smirnov test

Distribution	$\beta$	삼성전자	POSCO	한국전력	SK텔레콤	신한지주	현대중공업	LG전자	KT&G
Exponential	1	0.935 (0.346)	0.772 (0.591)	0.736 (0.651)	0.718 (0.680)	0.689 (0.729)	0.613 (0.847)	0.581 (0.888)	0.754 (0.621)
	1.1	0.719 (0.679)	0.642 (0.804)	0.717 (0.684)	0.551 (0.922)	0.483 (0.974)	0.416 (0.995)	0.448 (0.988)	0.675 (0.753)
	1.2	0.536 (0.936)	0.627 (0.827)	0.700 (0.712)	0.513 (0.955)	0.443 (0.990)	0.383 (0.999)	0.336 (1.000)	0.656 (0.782)
	1.3	0.491 (0.969)	0.613 (0.846)	0.685 (0.736)	0.479 (0.976)	0.430 (0.993)	0.363 (0.999)	0.274 (1.000)	0.640 (0.807)
	1.4	0.479 (0.976)	0.602 (0.862)	0.673 (0.756)	0.448 (0.988)	0.418 (0.995)	<b>0.352</b> <b>(1.000)</b>	<b>0.245</b> <b>(1.000)</b>	0.626 (0.828)
Compressed exponential	1.5	0.469 (0.980)	0.592 (0.874)	0.662 (0.774)	0.420 (0.994)	0.409 (0.996)	0.385 (0.998)	0.325 (1.000)	0.614 (0.845)
	1.6	0.460 (0.984)	0.584 (0.885)	0.652 (0.789)	0.395 (0.998)	0.400 (0.997)	0.486 (0.972)	0.430 (0.993)	0.603 (0.860)
	1.7	0.453 (0.987)	0.576 (0.894)	0.644 (0.802)	<b>0.377</b> <b>(0.999)</b>	<b>0.393</b> <b>(0.998)</b>	0.573 (0.897)	0.524 (0.946)	0.594 (0.872)
	1.8	0.446 (0.989)	0.569 (0.902)	<b>0.636</b> <b>(0.813)</b>	0.385 (0.998)	0.475 (0.978)	0.651 (0.791)	0.609 (0.852)	0.585 (0.883)
	1.9	0.440 (0.990)	<b>0.563</b> <b>(0.909)</b>	0.647 (0.797)	0.408 (0.996)	0.552 (0.921)	0.719 (0.680)	0.686 (0.734)	0.578 (0.892)
Gaussian	2	<b>0.434</b> <b>(0.992)</b>	0.605 (0.858)	0.719 (0.680)	0.476 (0.977)	0.621 (0.835)	0.782 (0.574)	0.756 (0.617)	<b>0.571</b> <b>(0.900)</b>

The values out of the parentheses are Kolmogorov statistics, and those in the parentheses are  $p$ -values.

한편, 지수분포가 정규분포보다 약간 우세하면서도 어떤 기업에서는 두 분포의 카이제곱 값들이 비슷하게 나타나는 반면, 모든 기업에서  $1.2 \leq \beta \leq 1.5$ 의 범위에서 카이제곱 값이 훨씬 낮게 나타나므로 지수 분포나 정규분포에 비해 압축된 지수분포가 우세하게 지지된다는 것이 입증된다.

Table 5.2는 개별 주가 수익률 분포에 대한 콜모고로프-스미르노프 검정 결과이다.  $\beta$  값에 따라 콜모고로프 통계량이 Table 5.1의 카이제곱 통계량에 비해 큰 차이를 보이지는 않는 것을 알 수 있다. 괄호 안의  $p$ -value도 일반적인 통계 검정에 비해 큰 차이를 보이지는 않는다. 바꾸어 말하면, 카이제곱 적합도 검정에 비해 콜모고로프-스미르노프 검정에서는 어떤 분포를 확실하게 기각할 수도 채택할 수도 없는 상황이 발생한다.  $1 \leq \beta \leq 3$ 인 범위에 대해 검정한 결과, 표에서 보이는 바와 같이  $1.4 \leq \beta \leq 2$ 인 범위에서 콜모고로프 통계량이 가장 적게( $p$ -value가 가장 크게) 나타났다. 이는 카이제곱 검정에 비해 대체적으로 높은 값이다. 하지만 각 종목 별로 보면 카이제곱 검정에서  $\beta$  값이 높은 종목은 콜모고로프-스미르노프 검정에서도 값이 높은 것을 알 수 있다. 따라서 두 검정을 절충하여 분포를 추정하는 것이 타당할 것으로 여겨진다. Table 5.1과 Table 5.2를 종합하여 볼 때, 한국 유가증권시장 상위 8개사의 주가 변동률 분포는  $1 < \beta < 2$ 의 범위를 갖는  $f_\beta(x) \sim \exp(-a|x|^\beta)$ 의 압축된 지수분포라는 것을 추정할 수 있으며, 개별 종목에 따라 그 지수는 달리 측정되는 것을 알 수 있다.

## 5.2. Independence test for absolute returns between 8 major stocks

Figure 5.2는 조사 대상 기업 8개에 대한 각각의 조합에서 독립성 검정을 시행한 결과이다. 오른쪽 상단 부분에 표시된 각각의 열지도(heatmap)는 각각 조합들의 이원분할표(2-way contingency table)에서

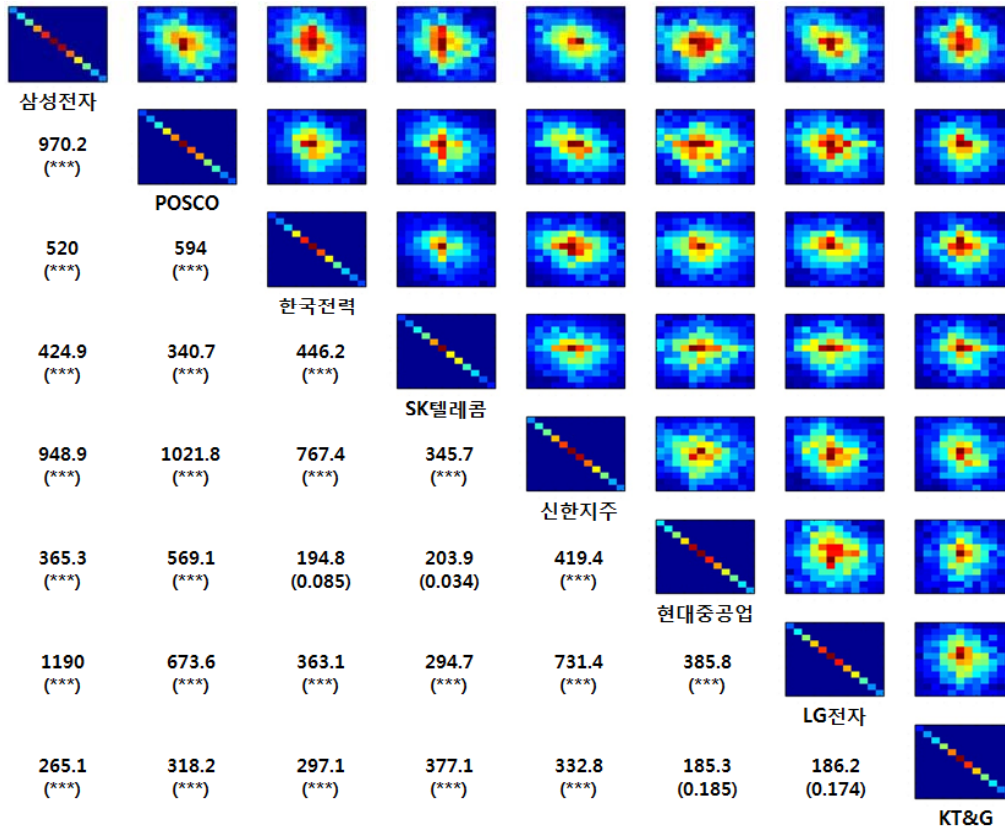


Figure 5.2. Independence test for absolute returns of 8 major companies: The upper diagonal figures are the heatmaps of 2-way contingency tables for all pairs of 8 companies, and the lower diagonal values are chi-square statistics and  $p$ -values of each pair (\*\*\*:  $p$ -value < 0.001).

나타나는 빈도들을 도식화한 것이다. 주로 가운데에 표시된 붉은색(어두운 색) 계열이 높은 빈도수를 나타내고, 가장자리에 표시된 푸른색(어두운 색) 계열이 낮은 빈도수를 나타낸다. 자명하게, 자기 자신과의 조합을 나타내는 열지도는 완전히 대각화되어(diagonalized) 독립성을 상실하고 있는 것을 확인할 수 있다. 한편, 왼쪽 하단 부분에 표시된 숫자들은 각각 조합들의 이원분할표를 바탕으로 독립성 검정을 시행한 결과이다. 괄호 밖에 표시된 숫자가 카이제곱 통계량이고 괄호 안에 표시된 숫자는  $p$ -value이다. 모든 조합에 대해 유의도  $\alpha = 0.01$ 로 검정한 결과 (한국전력, 현대중공업), (SK텔레콤, 현대중공업), (현대중공업, KT&G), (LG전자, KT&G)를 제외하고는 독립성을 보장할 수 없는 것으로 나타났다. 이로 보아 한국 유가증권시장의 30% 이상을 차지하는 상위 8개사 대부분의 주가 수익률은 서로 관계가 있는 것으로 추정할 수 있다. 이는 국내의 정세 등에 영향을 받아서 개별 주가들이 동시다발적으로 변동한 결과일 수도 있고, 서로의 주가들이 인과적으로 영향을 주고받은 결과일 수도 있다.

한편, 7개의 타 기업 중 3개사와 독립성을 유지하는 현대중공업과, 2개사와 독립성을 유지하는 KT&G의 분포를 Figure 5.1에서 살펴보면 타 기업들과 약간의 차별성을 지니는 것을 확인할 수 있다. 현대중공업은 수익률의 절대값이 큰 빈도수가 타 기업에 비해 많다. 그래서 Table 5.1과 Table 5.2에서도



확인되듯이 8개사 중  $\beta = 1$ 인 지수분포에 가장 근접하고 있음을 알 수 있다. 반면 Figure 5.1에서 KT&G는 5%이내의 낮은 값을 기록한 빈도수가 타 기업에 비해 많고  $\beta = 1.5$ 인 분포에 가장 근접하고 있음을 알 수 있다.

## 6. 결론 및 논의

2008년 11월 28일 종가 기준 한국 유가증권시장의 시가총액 상위 8개사인 삼성전자, POSCO, 한국전력, SK텔레콤, 신한지주, 현대중공업, LG전자, KT&G를 대상으로 2003년 4월 1일부터 2015년 3월 31일까지 2,976일 간의 주가 수익률 절대값(absolute return)을 분석한 결과 다음의 사실들을 알 수 있었다.

첫째, 개별 기업 주가 수익률의 분포를 추정하기 위해 적합도 검정(goodness of fit test)을 두 가지 방법으로 실시하였다. 일원 분할표(1-way contingency table)를 이용한 카이제곱 검정과 콜모고로프-스미르노프 검정을 실시하여 두 방법을 비교·분석하였다.  $1 \leq \beta \leq 3$ 의 범위에서 검정한 결과 주가 수익률의 분포는  $1 < \beta < 2$ 인 압축된 지수분포(compressed exponential distribution)가 지수분포( $\beta = 1$ ), 정규분포( $\beta = 2$ )보다 우세하게 나타났다. 결국 종합주가지수가 아닌 개별 주가 수익률은 정규분포에 비해 두터운 꼬리(heavy-tail)를 갖는 것으로 결론내릴 수 있으며, 이는 개별 기업 주가 수익률의 분산이 정규분포에 비해 크다는 것을 의미한다.

한편, 모든 개별 기업들의 조합(28가지)에 대해 주가 수익률 절대값의 독립성 분석(independence test)을 실시하였다. 적합도 검정과 비슷하게 분석을 위해 이원 분할표(2-way contingency table)를 이용한 카이제곱 검정 통계량을 사용하였다. 유의도 0.01로 독립성을 검증한 결과 28개 조합들 중 4개를 제외하고는 모두 독립성 가설을 기각하는 것으로 나타났다. 따라서 유가증권시장의 시가총액 30% 이상을 차지하는 상위 기업들의 주가 수익률은 서로 관계가 있는 것을 알 수 있다. 이는 국내외 요인에 의한 동시다발적인 변동, 또는 개별 주가 수익률의 직접적인 인과관계에 의한 것으로 추정된다.

## References

- Fama, E. F. (1965). The behavior of stock market prices, *The Journal of Business*, **38**, 34–105.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work, *Journal of Finance*, **25**, 383–417.
- Hayter, A. (2007). *Probability and Statistics for Engineers and Scientists*, 4th edition, Thomson Brooks/Cole, Belmont.
- Hong, B. H., Lee, K. E. and Lee, J. W. (2007). Power law of quiet time distribution in the Korean stock-market, *Physica A*, **377**, 576–582.
- Kaizoji, T. and Kaizoji, M. (2003). Empirical laws of a stock price index and a stochastic model, *Advances in Complex Systems*, **6**, 1–10.
- Kim, S. H. (1998). Comparison of methodologies for constructing composite indexes, *Journal of The Korean Official Statistics*, **3**, 21–60.
- Korotayev, A. V. and Tsirel, S. V. (2010). A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings, Juglar and Kitchin Cycles in Global Economic Development, and the 2008–2009 Economic Crisis, *Structure and Dynamics*, **4**, 3–57.
- Laherrere, J. and Sornette, D. (1999). Stretched exponential distributions in nature and economy: Fat tails with characteristic scales, *European Physical Journal B*, **2**, 525–539.
- McCauley, J. L. (2004). *Dynamics of Markets: Econophysics and Finance*, Cambridge University Press.
- Min, S. S. (2009). *Forecasting of KOSPI market trend using price and transaction volume* (Master's thesis), KAIST.

- Niederhoffer, V. and Osborne, M. F. M. (1966). Market making and reversal on the stock exchange, *Journal of the American Statistical Association*, **61**, 897–916.
- Panas, E. and Ninni, V. (2010). The distribution of London metal exchange prices: A test of the fractal market hypothesis, *European Research Studies*, **8**, 193–210.
- Roussas, G. G. (1997). *A Course in Mathematical Statistics*, 2nd edition, Academic Press, Chestnut Hill, MA.
- Working, H. (1960). Note on the correlation of first differences of averages in a random chain, *Econometrica*, **28**, 916–918.

# 한국 주식시장 상위 8개사에 대한 적합도 검정 및 독립성 검정

민승식<sup>a,1</sup>

<sup>a</sup>해군사관학교 이학과

(2015년 11월 9일 접수, 2015년 12월 21일 수정, 2015년 12월 24일 채택)

---

## 요약

본 논문에서는 한국 유가증권시장의 시가총액 상위 8개사 주가 수익률 절대값(absolute return)을 이용하여, 분포의 적합도 검정(goodness of fit test) 및 기업들 간의 독립성 검정(independence test)을 실시하였다. 검정 결과 개별 주가 수익률은 압축된 지수분포(compressed exponential distribution)를 이루는 것으로 나타났다. 이 때 파라미터는  $1 < \beta < 2$ 인 경우가  $\beta = 1$ (지수분포),  $\beta = 2$ (정규분포)보다 우세한 것으로 확인되었다. 한편 독립성 검정에서는 대부분의 기업들이 관련성을 지니고 있는 것으로 나타났다.

주요용어: 분포, 주가 수익률 절대값, 적합도 검정, 독립성 검정, 압축된 지수분포

---

<sup>1</sup>(51704) 경상남도 창원시 진해구 중원로 1, 해군사관학교 이학과. E-mail: fieldsmi@gmail.com