

국제현물원유가의 일일 상승 및 하락율의 극단값 분석

윤석훈¹

¹수원대학교 통계정보학과

(2010년 9월 접수, 2010년 10월 채택)

요약

극단값 통계 분석의 도구로는 전통적인 연간 최대값 방법과 현대적인 분계점 방법, 그리고 분계점 방법을 개선한 변형체 등으로 분류할 수 있다. 연간 최대값 방법은 시계열자료의 연간 최대값들에 대하여 일반화극단값분포를 적합시키는 것이고, 분계점 방법은 충분히 큰 하나의 분계점을 넘어서는 초과값들의 초과여분들에 대하여 일반화과레토분포를 적합시키는 것이다. 분계점 방법의 한 변형체로서 본 논문에서는 분계점 방법에 추가적으로 초과값들의 전체 개수가 포아송분포를 따른다고 가정하는 포아송-GPD 방법을 다루고, 이를 1988.01.04부터 2009.12.31까지 수집된 서부텍사스산중질유의 현물가격 자료로부터 계산된 일일 상승율과 일일 하락율에 적용한다. 이에 따르면 일일 상승율과 일일 하락율의 분포는 정규분포와 달리 두터운 꼬리를 갖는 분포로 나타났는데, 이는 오늘날의 많은 금융 자료 분석에서 나타나는 일반적인 현상과 잘 부합하는 것이다.

주요어: 극단값이론, 포아송-GPD 방법, 국제현물원유가, 서부텍사스산중질유.

1. 서론

오늘날 국제 원유 가격이 우리나라 경제에 미치는 영향은 실로 막대하다고 할 수 있다. 통상 세계 3대 유종이라 함은 미국 서부텍사스 지역에서 생산되는 서부텍사스산중질유(WTI; West Texas Intermediate), 영국 북해 지역에서 생산되는 브렌트유(Brent oil), 중동 아랍에미리트에서 생산되는 두바이유(Dubai oil)를 일컫는데, 이 3대 유종이 국제 원유시장의 기준 유종으로 자리잡은 이유는 우선 생산량이 많고 또한 특정 생산자가 생산을 독점하지 않아 가격 형성이 투명하기 때문이다. 이 중에서 브렌트유는 WTI보다 유황 성분이 많고 질이 떨어져 보통 2~3달러 낮게 가격이 형성되지만, 두바이유보다는 품질이 뛰어나 2~3달러 정도 비싸게 거래된다. 이렇게 WTI의 가격이 가장 높은 것은 유황 함량이 낮아 유황 제거 작업을 할 때 비용이 작게 들고 또한 원유를 정제할 때 가격이 비싼 휘발유와 나프타 등의 고급 유류가 많이 생산되기 때문이다. WTI는 미주지역 시장에서 거래되는 모든 원유의 가격을 결정하는 기준 유종으로서 뉴욕상업거래소(NYMEX; New York Mercantile Exchange)에서 현물거래와 선물거래가 이루어지고, 브렌트유는 유럽과 아프리카에서 거래되는 원유 가격의 기준 유종으로서 런던국제석유거래소(IPE; International Petroleum Exchange of London)에서 주로 선물로 거래되며, 두바이유는 우리나라 원유 수입의 70% 이상을 차지하는 중동산 원유의 기준 유종으로서 중동권과 싱가포르에서 주로 현물로 거래되는 것이 특징이다. 보통 두바이유는 우리나라 시간으로 오후 6시 무렵에 현물 가격이 결정되기 때문에 오전 2시 30분과 오전 4시 30분경에 각각 가격이 결정되는 브렌트유와 WTI에 비해 하루 늦게 가격이 반영되어 가격 하락과 상승이 두 원유와 다르게 나타나기도 한다.

¹(445-743) 경기도 화성시 봉담읍 와우리 산2-2, 수원대학교 통계정보학과, 부교수. E-mail: syun@suwon.ac.kr

21세기에 들어와 2004년 초부터 2008년 중순까지 유가가 급등한 바 있는데, 이의 원인으로는 달러화의 하락과 미국 원유 재고량의 감소 그리고 이란 정세의 불안정 등을 꼽을 수 있다. 이 중 달러화 문제를 제외하면 정정 불안과 수급 불안정이라는 통상적인 분석이 가능한데, 세계석유회의에서 석유 수입국들이 석유 생산국들을 향해 원유 증산을 요구한 것도 이 같은 이유 때문이다. 그러나, 또한 급격한 유가 상승은 수요-공급의 문제가 아니라 석유에 대한 투기 때문이라는 주장이 끊임없이 제기되고 있는데, 미국의 에너지 전문가인 Engdahl (2008)에 따르면 석유 선물 온라인 장외시장에 대해 미국 선물거래위원회의 감시를 면제해준 2000년 상품선물현대화법에 의해 골드만삭스 같은 석유 거래 은행들의 고삐가 풀어졌고, 따라서 마음 놓고 가격조작과 투기가 가능한 환경이 만들어졌기 때문이라는 것이다. 그는 또 미국인들이 미국산 석유 상품을 미국내 컴퓨터를 사용하면서도 미국시장이 아닌 영국시장을 통해 거래할 수 있게 된 것도 투기를 부추긴 이유라면서 현재 유가의 60%는 투기 요소 때문이라고 주장하고 있다.

위와 같은 유가 급등의 구체적인 사회적 또는 구조적 원인을 더이상 다루지 않는 대신, 본 논문에서는 실제 일일국제현물원유가 자료에 근거하여 유가의 일일 상승율과 일일 하락율의 극단값을 분석한다. 이를 위해 극단값 통계 분석 방법으로 연간 최대값 방법, 분계점 방법, 포아송-GPD 방법을 2장에서 간단히 소개하고, 3장에서 이를 1988.01.04부터 2009.12.31까지 수집된 일일 WTI 현물가격 자료에 적용하여 얻은 분석 결과를 제공하며, 4장에서 간단히 결론을 설명한다.

2. 극단값이론의 요약

X_1, \dots, X_n 이 공통의 분포함수(distribution function) F 를 갖는 iid (independent and identically distributed) 확률변수들이라고 하고, $M_n = \max\{X_1, \dots, X_n\}$ 으로 나타내기로 하자. 극단값이론에서의 기본 가정은 적당한 정규화 상수 $a_n > 0$ 과 $b_n \in \mathbb{R}$ 이 존재하여 적당한 비퇴화확률변수(non-degenerate random variable) Z 에 대하여 $n \rightarrow \infty$ 일 때 다음의 분포수렴(convergence in distribution)이 성립한다는 것이다:

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Z. \quad (2.1)$$

식 (2.1)은 F 가 연속 분포(continuous distribution)이면 대부분 성립하게 되는데, Z 의 분포함수를 G 라고 할 때 Fisher와 Tippett (1928)과 Gnedenko (1943)는 식 (2.1)이 성립하면 G 는 다음의 세 가지 극단값분포함수(extreme value distribution function) 중의 어느 하나와 형태가 동일해야 된다는 사실을 보였다:

$$\begin{aligned} \text{Type I: } \Lambda(x) &= \exp(-e^{-x}), \quad x \in \mathbb{R}, \\ \text{Type II: } \Phi_\alpha(x) &= \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases} \\ \text{Type III: } \Psi_\alpha(x) &= \begin{cases} \exp\{-(-x)^\alpha\}, & x < 0, \\ 1, & x \geq 0, \end{cases} \end{aligned}$$

여기서, α 는 양수의 값을 취하는 모수이고, 두 분포함수 G_1 과 G_2 의 형태가 동일하다는 것은 적당한 상수 $a > 0$ 와 $b \in \mathbb{R}$ 가 존재하여 $G_1(x) = G_2(ax + b)$, $x \in \mathbb{R}$ 가 성립함을 의미한다. von Mises (1936)는 위 세 가지 극단값분포함수들을 묶어 하나의 일반화된 극단값분포함수의 형태인

$$G(x; \mu, \sigma, \xi) := \exp \left[- \left\{ 1 + \frac{\xi(x - \mu)}{\sigma} \right\}_+^{-\frac{1}{\xi}} \right], \quad x \in \mathbb{R}$$

을 처음으로 제안하였는데, 이를 소위 일반화극단값분포함수(generalized extreme value distribution function)라고 부르고 이 분포를 간단히 기호 $GEV(\mu, \sigma, \xi)$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\xi \in \mathbb{R}$ 로 나타낸다. 여기서, $y_+ := \max\{y, 0\}$ 이고, $\xi = 0$ 인 경우는 항상 $\xi \rightarrow 0$ 일 때의 극한으로 해석하여(즉, $G(x; \mu, \sigma, 0) = \exp\{-e^{-(x-\mu)/\sigma}\}$), $\xi = 0$ 은 Type I에, $\xi > 0$ 은 $\alpha = 1/\xi$ 인 Type II에, 그리고 $\xi < 0$ 은 $\alpha = -1/\xi$ 인 Type III에 각각 대응된다. 따라서, 앞으로는 식 (2.1)이 성립한다고 할 때 극한 변수 Z 의 분포함수는 적당한 $\mu \in \mathbb{R}$, $\sigma > 0$, $\xi \in \mathbb{R}$ 에 대해 $GEV(\mu, \sigma, \xi)$ 가 된다고 가정하기로 한다. 형상모수(shape parameter)인 ξ 를 F 의 극단값지수(extreme value index) 또는 꼬리지수(tail index)라고 부르는데, ξ 는 F 의 오른쪽 꼬리를 특징화하는데 사용되는 중요 모수로서, $\xi > 0$ 이면 두터운 꼬리(heavy tail), $\xi = 0$ 이면 중간형 꼬리(medium tail), $\xi < 0$ 이면 짧은 꼬리(short tail)를 갖는다고 종종 표현한다.

따라서, 식 (2.1)을 가정한다는 것은 결국 $n \rightarrow \infty$ 이고 $z \uparrow x_F$ 일 때(x_F 는 분포함수 F 의 오른쪽 끝점을 나타냄), 적당한 모수 $\mu \in \mathbb{R}$, $\sigma > 0$, $\xi \in \mathbb{R}$ 가 존재하여 근사식

$$F^n(z) = P\{M_n \leq z\} \approx G(z; \mu, \sigma, \xi) \tag{2.2}$$

가 성립한다는 것인데, 이로부터 F 의 오른쪽 끝점 x_F 의 부근에서의 분포함수 값을 G 를 이용하여 근사시키는 외삽법(extrapolation)이 가능함을 알 수 있다. 이에 대한 전통적인 통계적 극단값 분석 방법으로, 예를 들어 서로 iid인 일별 관측값들이 여러 해 동안 관찰되었다고 할 때 $GEV(\mu, \sigma, \xi)$ 를 연간 최대값들에게 적합시키는 방법이 있는데 이를 연간 최대값 방법(annual maximum method)이라고 부른다. 여기서는, F 가 일별 관측값들의 공통 분포함수가 되고 G 는 연간 최대값들의 공통 분포함수가 되는 것이므로, 예를 들어 t 년 재발수준(t -year return level), 즉 t 년에 한 번 정도 넘어서는 높은 수준을 q_t 라고 나타내면, $n = 365$ 이고 t 가 큰 값일 때,

$$G(q_t; \mu, \sigma, \xi) \approx F^n(q_t) = \left(1 - \frac{1}{nt}\right)^n \approx e^{-\frac{1}{t}} \approx 1 - \frac{1}{t}$$

이 성립한다고 할 수 있으므로 q_t 는 $G^{-1}(1 - 1/t; \mu, \sigma, \xi)$ 로 근사시킬 수 있고, 따라서 적합된 모수 값들을 이용하면 q_t 의 추정이 가능해진다. 여기서, G^{-1} 은 G 의 분위수함수(quantile function)를 나타낸다.

많은 역사적 극단값 분석에 연간 최대값 방법이 사용되어 왔으나 (Gumbel, 1958), 이 방법은 연간 최대값들만을 사용하고 F 의 꼬리에 대한 정보를 가지고 있는 다른 자료들을 사용하지 않기 때문에 추정치의 정확도가 떨어지고 오차도 증가하는 단점을 갖는다. 이에 대한 개선 방안으로 널리 쓰이고 있는 방법에는 분계점 방법(threshold method)이 있다 (Smith, 1989; Davison과 Smith, 1990). 이 방법에 따르면, 우선 충분히 큰 임의의 분계점 u ($< x_F$)를 선택한다. 이 경우, 적당한 $\xi \in \mathbb{R}$ 에 대하여 식 (2.1)이 성립하면, Pickands (1975)는 다음의 근사식이 역시 성립한다는 사실을 보였다:

$$P\{X_i - u \leq y | X_i > u\} = \frac{F(u+y) - F(u)}{1 - F(u)} \approx H(y; \phi, \xi) := 1 - \left(1 + \frac{\xi y}{\phi}\right)_+^{-\frac{1}{\xi}}, \quad y > 0,$$

여기서 정의된 H 를 일반화파레토분포함수(generalized Pareto distribution function)라고 부르고 이 분포를 간단히 $GPD(\phi, \xi)$, $\phi > 0$, $\xi \in \mathbb{R}$ 로 나타낸다. 이는 곧, 분계점 u 를 넘어서는 초과값(exceedance)들의 초과여분(excess)들이 근사적으로 $GPD(\phi, \xi)$ 를 따른다는 것이므로, 분계점 방법은 예를 들어 서로 iid인 일별 관측값들이 일정 기간 동안 관찰되었다고 할 때 $GPD(\phi, \xi)$ 를 초과값들의 초과여분들에게 적합시키는 방법을 의미하게 된다. 여기서, 분계점 u 를 선택하는 하나의 경험적 방법은 일별 관측값들의 MRL 그림(mean residual life plot), 즉 모든 $t > u$ 에 대하여 t 와 t 를 넘어서는 초과값들의 초과여분 산술평균값의 산점도가 근사적으로 직선을 따르는 최소의 u 값으로 선택하는 것이

다. 이는 왜냐하면 이론적으로 하나의 확률변수 Y 가 $\text{GPD}(\phi, \xi)$ 를 따르면 $\phi + \xi t > 0$ 인 모든 $t > 0$ 에 대하여 $Y - t|Y > t \sim \text{GPD}(\phi + \xi t, \xi)$ 가 성립하므로 $\xi < 1$ 일 경우

$$E(Y - t|Y > t) = \frac{\phi + \xi t}{1 - \xi}$$

가 성립하는데, 이는 곧 t 의 함수로서 $E(Y - t|Y > t)$ 는 기울기가 $\xi/(1 - \xi)$ 고 절편이 $\phi/(1 - \xi)$ 인 직선을 의미하는 것이기 때문이다. 분계점 방법에서는 t 년 재발수준 $q_t (> u)$ 가 $n = 365$ 라고 할 때 근사식

$$\frac{1}{nt} = P\{X_i > u\} \times P\{X_i > q_t | X_i > u\} \approx p\{1 - H(q_t - u; \phi, \xi)\}$$

를 만족하므로 $u + H^{-1}(1 - 1/(ntp); \phi, \xi)$ 로 근사시킬 수 있고, 따라서 적합한 모수값을 이용하면 q_t 의 추정이 역시 가능해진다. 여기서 $p = P\{X_i > u\}$ 의 추정은 전체 자료 중 분계점 u 를 넘어서는 초과값들의 상대 비율인 표본비율을 사용할 수 있다.

분계점 방법의 한 변형체는 위에 소개한 분계점 방법에 분계점을 넘어서는 초과값들의 개수가 포아송분포를 따른다고 가정하는 것인데, 이는 식 (2.1) 또는 식 (2.2)가 성립하면 충분히 큰 임의의 분계점 $u (< x_F)$ 에 대하여 $n \rightarrow \infty$ 일 때 근사식

$$n(1 - F(u)) \approx -\log G(u; \mu, \sigma, \xi)$$

가 성립하므로 X_1, \dots, X_n 중 분계점 u 를 넘어서는 초과값들의 개수를 N 으로 나타내면 N 은 근사적으로 평균이 $-\log G(u; \mu, \sigma, \xi)$ 인 포아송분포를 따르게 되기 때문이다. 이제, 역으로 N 이 적절한 $\lambda > 0$ 에 대하여 평균이 λ 인 포아송분포를 따르고, N 이 주어졌을 때 분계점 u 를 초과하는 초과값들의 초과여분 Y_1, \dots, Y_N 들이 서로 *iid*이고 $\text{GPD}(\phi, \xi)$, $\phi > 0$, $\xi \in \mathbb{R}$ 를 따르는 확률변수들이라고 가정하면 모든 $z > u$ 에 대하여

$$\begin{aligned} P\{M_n \leq z\} &= P\{\max\{Y_1, \dots, Y_N\} \leq z - u\} \\ &= P\{N = 0\} + \sum_{k=1}^{\infty} P\{N = k\} \times P\{Y_1 \leq z - u, \dots, Y_k \leq z - u | N = k\} \\ &= e^{-\lambda} + \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} H^k(z - u; \phi, \xi) \\ &= \exp[-\lambda\{1 - H(z - u; \phi, \xi)\}] \\ &= G(z; \mu, \sigma, \xi) \end{aligned}$$

가 성립함을 알 수 있다. 여기서,

$$\mu = u + \frac{\phi(\lambda^\xi - 1)}{\xi}, \quad \sigma = \phi\lambda^\xi \quad (2.3)$$

이다. 이는 곧, 여기서의 기본 가정이 근사식 (2.2)의 정당성을 담보한다는 것을 의미하고, 또한 여기서의 모수 λ , ϕ , ξ 와 연간 최대값 방법에서의 모수 μ , σ , ξ 가 서로 어떻게 연관되어 있는지를 설명하는 것이다. 이러한 통계적 극단값 분석 방법을 간단히 포아송-GPD 방법으로 부르기로 하는데, 구체적인 방법은 예를 들어 서로 *iid*인 일별 관측값들이 m 년 동안 관찰되었다고 할 때 분계점 u 를 넘어서는 전체 초과값들의 개수가 적절한 $\lambda > 0$ 에 대하여 평균이 $m\lambda$ 인 포아송분포를 따르고(따라서, λ 는 초과값들의 연간 평균 개수를 나타냄), 이들 초과값들의 초과여분들은 서로 *iid*이고 $\text{GPD}(\phi, \xi)$, $\phi > 0$, $\xi \in \mathbb{R}$ 를 따르는 확률변수들이라고 가정하여 이를 시계열 관측값들에게 적합시키는 것이다. 여기서 적합한 모수값

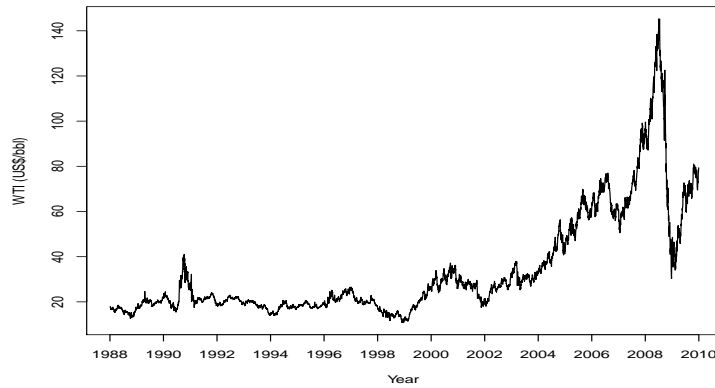


그림 3.1. WTI의 배럴당 일일 현물원유가

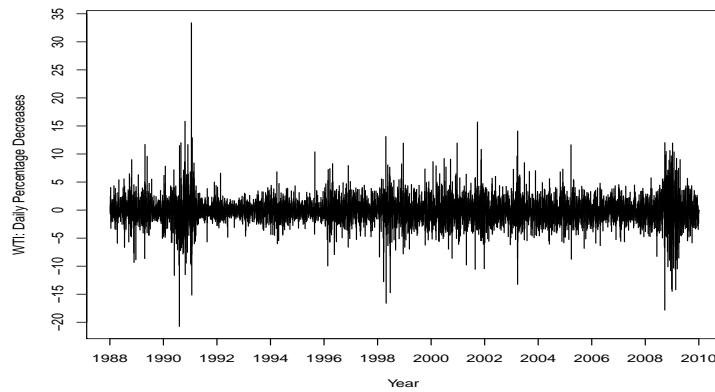


그림 3.2. WTI의 일일 하락율(%)

과 관계식 (2.3)을 이용하면 앞의 연간 최대값 방법에서의 모수값 추정이 역시 가능하므로 t 년 재발수준 q_t 의 추정은 연간 최대값 방법에서와 동일하게 수행될 수 있다. 포아송-GPD 방법은 윤석훈 (2009)에 소개된 2차원 포아송초과점과정 방법의 단순 형태로서 실제 자료 분석시 편리하게 사용될 수 있다.

3. 서부텍사스산증질유 국제현물가의 일일 상승율과 일일 하락율의 극단값 분석

본 논문에서 다루는 자료는 일일국제현물원유가 중 WTI의 배럴당 미달러화 표시 일일 현물가 자료로서(참조: <http://www.eia.doe.gov>), 극단값 통계분석은 1988.01.04부터 2009.12.31까지 총 22년간을 대상으로 하였다. 대상 기간 동안 결측치를 제외한 총 관측치의 개수는 5,551개인데 그림 3.1에 이를 나타내었다. 그림 3.1에서 보듯이 WTI의 가격은 2004년부터 꾸준히 상승하기 시작하여 2006년 말 잠시 하락한 후 2008.07.03에 배럴당 \$145.31의 최고가에 도달하기까지 급반등하였고 이후 2009년 초까지 다시 급락하는 등 21세기 초부터 급등락의 드라마틱한 양상을 보여주고 있다. 이 자료로부터 원유가의 t 일 상승율과 하락율(단위: %)을 각각

$$X_t = \pm 100 \times \left(\frac{t\text{일 원유가}}{t\text{일 전일 원유가}} - 1 \right)$$

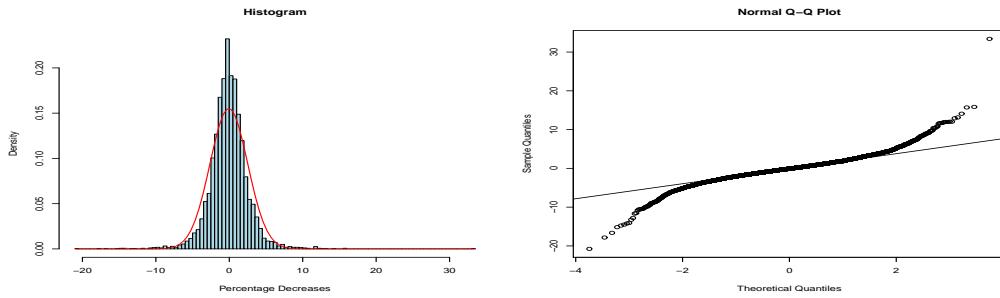


그림 3.3. WTI의 일일 하락율에 대한 히스토그램과 정규분포 Q-Q 그림

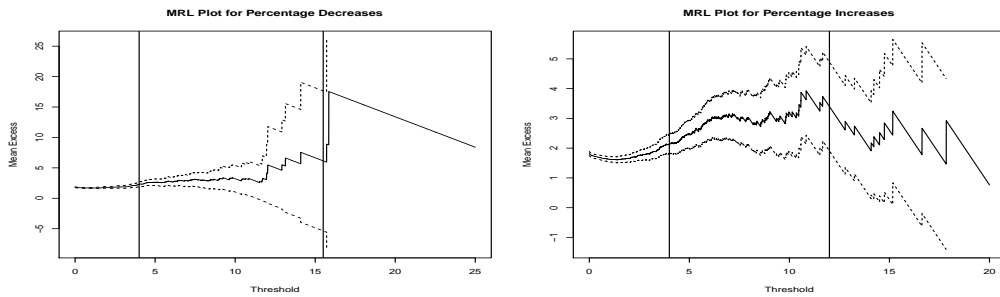


그림 3.4. WTI의 일일 하락율 MRL 그림과 일일 상승율 MRL 그림

과 같이 계산할 수 있는데, 그림 3.2는 이렇게 계산된 값 중 일일 하락율을 나타낸다. 그림 3.1과 달리 그림 3.2에서는 자료간 상관성이 없어 보이는데, 여기서는 $\{X_t\}$ 가 iid로서 적당한 $\xi \in \mathbb{R}$ 에 대하여 식 (2.1)이 만족된다고 가정한다. 하락율 1, 2위와 상승율 1, 2위는 각각 33.40%(1991.01.17), 15.85%(1990.10.22)와 20.77%(1990.08.06), 17.84%(2008.09.22)로 나타났는데, 참고적으로 1991.01.17에는 미국을 비롯한 33개국의 다국적군이, 원유 분쟁으로 인하여 1990.08.02 쿠웨이트를 무력 침공하고 점령한 이라크에 대해 즉각적인 철수를 요구하며 이라크를 공습한 날로 이때부터 걸프전이 시작되었고, 1990.08.06에는 유엔 안전보장이사회가 이라크에 대해 무역제재조치를 취하는 결의안을 통과시키고 이라크로 하여금 쿠웨이트로부터 철수할 것을 요구하였으나 이라크가 이를 거부한 날이었다.

그림 3.3은 일일 하락율의 히스토그램과 정규분포 Q-Q 그림(quantile-quantile plot)을 보여준다. 전체 일일 하락율 자료의 평균과 표준편차는 각각 -0.06 과 2.57 인데 히스토그램에는 $N(-0.06, 2.57^2)$ 분포의 확률밀도함수를 함께 그려 넣었다. 그림에서 보듯이 일일 하락율의 분포는 정규분포보다 양쪽 꼬리 부분이 훨씬 두터움을 알 수 있다. 이는 곧 일일 하락율(또는 상승율) 분포의 극단값지수가 0보다 클 것이라는 예상을 가능하게 한다.

이제, 2장에서 소개한 세 가지 방법을 사용하여 일일 하락율 및 상승율의 극단값 모형을 최대가능도추정법으로 적합시키고 t 년 재발수준을 역시 최대가능도추정법으로 추정해 보기로 한다. 먼저, 연간 최대값 방법에서는 22년 동안의 연간 최대 하락율(또는 상승율) 22개에 대하여 $GEV(\mu, \sigma, \xi)$ 모형을 적합시키고, 분계점 방법에서는 일일 하락율(또는 상승율) 자료에서 충분히 큰 하나의 분계점 u 를 넘어서는 초과값들의 초과여분에 대하여 $GPD(\phi, \xi)$ 모형을 적합시킨다. 마지막으로, 포아송-GPD 방법에서는 분계점 방법에서와 동일한 분계점 u 를 넘어서는 초과값들에 대하여 전체 초과값들의 개수에는 평균이 22λ 인 포아송분포를, 그리고 초과값들의 초과여분에는 $GPD(\phi, \xi)$ 모형을 적합시키고 난 후, 관계식 (2.3)을

표 3.1. 일일 하락율과 일일 상승율의 연간 최대값들과 $u = 4$ 초과값들의 기초 통계량(n : 표본크기, Q_1 : 1사분위수, Q_2 : 2사분위수, Q_3 : 3사분위수, \bar{x} : 평균)

구분	자료	n	최소값	Q_1	Q_2	\bar{x}	Q_3	최대값
하락율	연간 최대값	22	4.03	6.66	9.70	10.51	12.02	33.40
	$u = 4$ 초과값	232	4.01	4.46	5.43	6.28	7.06	33.40
상승율	연간 최대값	22	4.31	6.04	8.63	9.55	12.58	20.77
	$u = 4$ 초과값	244	4.01	4.51	5.21	6.14	6.65	20.77

표 3.2. 일일 하락율과 일일 상승율에 대한 GEV, GPD, P-GPD 모형에서의 모수 추정 결과(SE: 표준오차, CI: 프로파일 로그 가능도에 기초하여 계산된 신뢰구간)

구분	모형	추정	μ	σ	ξ	ϕ	λ
하락율	GEV($\xi = 0$)	MLE(SE)	8.07(0.87)	3.88(0.68)			
		95% CI	(6.32, 9.90)	(2.83, 5.62)			
	GPD	MLE(SE)			0.21(0.08)	1.82(0.19)	
		95% CI			(0.07, 0.40)	(1.47, 2.22)	
	P-GPD	MLE(SE)	9.51(0.51)	2.95(0.44)	0.21(0.08)	1.82(0.19)	10.55(0.69)
		95% CI			(0.07, 0.40)	(1.47, 2.22)	
상승율	GEV($\xi = 0$)	MLE(SE)	7.47(0.74)	3.29(0.59)			
		95% CI	(5.99, 9.04)	(2.37, 4.82)			
	GPD	MLE(SE)			0.23(0.08)	1.67(0.17)	
		95% CI			(0.08, 0.41)	(1.35, 2.04)	
	P-GPD	MLE(SE)	9.34(0.50)	2.88(0.44)	0.23(0.08)	1.67(0.17)	11.09(0.71)
		95% CI			(0.08, 0.41)	(1.35, 2.04)	

이용하여 GEV 모수 μ 와 σ 를 추정하는데, 따라서 이 마지막 방법을 지금부터는 간단히 P-GPD(λ, ϕ, ξ) 모형으로 부르기로 한다.

GPD 모형과 P-GPD 모형을 적합시키기 위해서는 충분히 큰 하나의 분계점 u 를 선택해야 하는데, 그림 3.4는 이를 위한 일일 하락율들의 MRL 그림과 일일 상승율들의 MRL 그림을 보여준다. 우선, 일일 하락율들의 MRL 그림에서 $t = 15.5$ 를 넘어서는 하락율 3개를 제외하면 t 가 (4, 15.5)에서 움직일 때 MRL 그림이 양의 기울기를 갖는 직선 형태이므로 $u = 4$ 를 분계점으로 선택하고, 일일 상승율들의 MRL 그림에서는 $t = 12$ 를 넘어서는 상승율 10개를 제외하면 t 가 (4, 12)에서 움직일 때 MRL 그림이 역시 양의 기울기를 갖는 직선 형태이므로 여기서도 $u = 4$ 를 분계점으로 선택한다. 이 경우 분계점의 초과율은 일일 하락율과 일일 상승율의 경우 각각 4.18%와 4.4%가 된다.

표 3.1은 일일 하락율과 일일 상승율의 연간 최대값들과 $u = 4$ 를 넘어서는 초과값들에 대한 기초 통계량을 보여준다. 표에 따르면, 연간 최대값의 분포와 초과값의 분포는 일일 하락율 및 일일 상승율에서 모두 오른쪽으로 두터운 꼬리를 갖고 있으며 또한 오른쪽으로 매우 기울어져 있음을 알 수 있다.

표 3.2는 일일 하락율과 일일 상승율에 대한 GEV, GPD, P-GPD 모형의 최대가능도추정법에 의한 적합 결과를 보여 준다. 우선, 연간 최대값들의 경우 GEV(μ, σ, ξ) 적합 결과 μ, σ, ξ 의 최대가능도추정치(표준오차)가 7.58(0.87), 3.43(0.71), 0.24(0.22)로 나타났고, 일일 상승율에서는 6.79(0.69), 2.56(0.64), 0.44(0.30)으로 나타났다. 여기서, 두 경우 모두 귀무가설 $H_0 : \xi = 0$ 에 대한 Wald 검정의 결과는 H_0 를 기각할 수 없다는 것이므로, 다시 $\xi = 0$ 으로 고정시키고 GEV($\mu, \sigma, \xi = 0$) 모형을 재적합시킨 결과 표 3.2의 내용이 얻어졌다. 다음으로, $u = 4$ 를 넘어서는 초과값들에 대해서는 일일 하락율과 일일 상승율 모두 GPD(ϕ, ξ) 모형과 P-GPD(λ, ϕ, ξ) 모형을 적합시킨 결과 표 3.2가 얻어졌다. 이로부터, 일일 하락율(또는 상승율) 분포의 오른쪽 꼬리가 매우 두터울 것이라는, 즉 극단값 지수 ξ 의 값

표 3.3. 일일 하락율과 일일 상승율에 대한 GEV, GPD, P-GPD 모형에서의 t 년 재발수준 q_t 의 추정 결과(SE: 표준오차, CI: 프로파일 로그 가능도에 기초하여 계산된 신뢰구간)

구분	모형	추정	q_{30}	q_{50}
하락율	GEV($\xi = 0$)	MLE(SE)	21.20(2.67)	23.21(3.01)
		95% CI	(16.95, 27.91)	(18.44, 30.78)
	GPD	MLE(SE)	26.28(5.58)	29.70(7.18)
		95% CI	(19.21, 45.83)	(20.82, 56.0)
	P-GPD	MLE(SE)	23.93(4.59)	27.16(6.0)
		95% CI	(17.96, 39.49)	(19.57, 48.37)
상승율	GEV($\xi = 0$)	MLE(SE)	18.62(2.34)	20.32(2.63)
		95% CI	(14.91, 24.48)	(16.16, 26.94)
	GPD	MLE(SE)	26.48(5.78)	30.17(7.52)
		95% CI	(19.02, 46.19)	(20.66, 56.77)
	P-GPD	MLE(SE)	24.0(4.74)	27.41(6.23)
		95% CI	(17.73, 39.64)	(19.37, 48.84)

이 0보다 클 것이라는 앞의 여러 탐색적 결과들과는 달리 연간 최대값 방법에서는 $\xi = 0$ 이라는 오도된 결과가 얻어졌음을 알 수 있다. 반면에, 분계점 방법이나 포아송-GPD 방법에서는 예상대로 $\xi > 0$ 이라는 결과가 얻어졌는데, 연간 최대값 방법에서의 이러한 잘못된 결과는 오른쪽 꼬리에 대한 정보를 가지고 있는 많은 자료 중 극히 일부분인 오직 연간 최대값들만을 사용하기 때문이다. 표 3.2에는 프로파일 로그 가능도(profile log-likelihood)에 기초하여 계산된 95% 신뢰구간이 포함되어 있는데, 이는 대체로 프로파일 로그 가능도 함수가 비대칭적으로 생긴 경우 사용되는 방법으로서, 예를 들어 일일 하락율의 GPD(ϕ, ξ) 모형 적합 결과 얻어진 ξ 의 최대가능도추정치(표준오차)인 0.21(0.08)로 ξ 의 95% 근사 신뢰구간을 구하면 $0.21 \pm 1.96 \times 0.08 \approx (0.05, 0.37)$ 이 되지만 프로파일 로그 가능도에 기초하여 계산된 ξ 의 95% 신뢰구간은 (0.07, 0.40)이 되어 다소 차이가 나는 것을 알 수 있다.

표 3.3은 일일 하락율과 일일 상승율에 대한 GEV, GPD, P-GPD 모형에 대해서 30년, 50년 재발수준에 대한 최대가능도추정법에 의한 추정 결과를 보여 준다. 여기서, GEV 모형은 표 3.2의 결과에 기초하여 GEV($\mu, \sigma, \xi = 0$) 모형으로 국한한 것이다. 표에는 역시 프로파일 로그 가능도에 기초하여 계산된 95% 신뢰구간이 포함되어 있다. 표에 의하면 일일 하락율과 일일 상승율 모두에서 GEV 모형은 분포의 오른쪽 꼬리가 두텁지 않고 중간형으로 잘못 판정된 것이기 때문에 재발수준 역시 매우 낮게 추정되었다. 이러한 문제점은 GPD 모형에서 해결되었으나, 여기서의 재발수준 추정치는 GEV 모형에 비해서 다소 과다하게 추정된 것 처럼 보이는데, P-GPD 모형에서는 추정치가 GPD 모형에 비해서 다시 낮아져 이러한 과다 추정 문제가 완화되었다고 볼 수 있다.

4. 결론

본 논문에서는 WTI 국제현물가의 일일 상승 및 하락율의 극단값 분석을 수행하였다. 극단값 통계 분석의 도구로는 연간 최대값 방법, 분계점 방법, 포아송-GPD 방법을 사용하였는데, 연간 최대값 방법에서는 극단값 지수 ξ 가 0이라는 잘못된 검정 결과가 유도되어 재발수준이 매우 과소 평가되어 추정되는 결과가 나타났다. 반면에 분계점 방법에서는 사전 탐색적 자료 분석 결과와 일치하는 $\xi > 0$ 의 결과가 도출되었으나 재발수준은 역으로 너무 높게 추정되는 현상이 발생되었다. 이러한 문제점들은 포아송-GPD 방법을 사용함으로써 다소 완화된 추정 결과가 얻어졌다. 이상으로 부터의 극단값 분석 결론은 국제현물원유가의 일일 상승 및 하락율의 분포는 정규분포와 달리 꼬리가 매우 두터운 분포로서 극

단 분위수의 추정시 정규분포를 가정하면 잘못된 추정 결과를 얻을 수 있으므로 주의해야 한다는 점이다. 이러한 결과는 오늘날 많은 금융 자료 분석에서 나타나는 현상, 즉 수익율 또는 손실율의 분포가 정규분포와 달리 두터운 꼬리를 갖는 분포라는 현상과 잘 부합하는 것이다.

참고문헌

- 윤석훈 (2009). 원/달러 환율 투자 손실률에 대한 극단분위수 추정, <한국통계학회논문집>, **16**, 803-812.
- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion), *Journal of the Royal Statistical Society, Series B*, **52**, 393-442.
- Engdahl, F. W. (2008). See http://www.engdahl.oilgeopolitics.net/Financial_Tsunami/Oil_Speculation/oil_speculation.HTM.
- Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**, 423-453.
- Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119-131.
- Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion), *Statistical Science*, **4**, 367-393.
- von Mises, R. (1936). La distribution de la plus grande de n valeurs, Reprinted in *Selected Papers II*, *American Mathematical Society*, Providence, R.I. (1954), 271-294.

Analysis of Extreme Values of Daily Percentage Increases and Decreases in Crude Oil Spot Prices

Seokhoon Yun¹

¹Department of Applied Statistics, University of Suwon

(Received September 2010; accepted October 2010)

Abstract

Tools for statistical analysis of extreme values include the classical annual maximum method, the modern threshold method and variants improving the second one. While the annual maximum method is to fit the generalized extreme value distribution to the annual maxima of a time series, the threshold method is to fit the generalized Pareto distribution to the excesses over a high threshold from the series. In this paper we deal with the Poisson-GPD method, a variant of the threshold method with a further assumption that the total number of exceedances follows the Poisson distribution, and apply it to the daily percentage increases and decreases computed from the spot prices of West Texas Intermediate, which were collected from January 4th, 1988 until December 31st, 2009. According to this analysis, the distribution of daily percentage increases as well as decreases turns out to have a heavy tail, unlike the normal distribution, which coincides well with the general phenomenon appearing in the analysis of lots of nowadays financial data.

Keywords: Extreme value theory, Poisson-GPD method, crude oil spot price, West Texas Intermediate.

¹Associate Professor, Department of Applied Statistics, University of Suwon, Suwon 445-743, Korea.
E-mail: syun@suwon.ac.kr