

# 원/달러 환율 투자 손실률에 대한 극단분위수 추정

윤석훈<sup>1,a</sup>

<sup>a</sup>수원대학교 통계정보학과

## 요약

금융자료에 극단값이론을 적용하는 것은 위험관리에서 중요한 최신 통계기법 중의 하나라고 할 수 있다. 극단값분석에서 전통적으로 사용해 오던 연간 최대값 방법은 시계열자료의 연간 최대값들에 대하여 일반화 극단값분포를 적합시키는 것이고, 최근 대안으로 널리 사용되고 있는 분계점 방법은 시계열자료 중 충분히 큰 하나의 분계점을 넘어서는 초과값들에 대하여 일반화과레토분포를 적합시키는 것이다. 그러나, 보다 실질적인 방법은 분계점을 넘어서는 초과값들을 하나의 점과정으로 해석하는 것인데, 즉 초과값들의 초과시점과 초과여분을 점근적으로 비동질 포아송과정을 갖는 하나의 2차원 점과정으로 간주하는 것이다. 본 논문에서는 이러한 2차원 비동질 포아송과정 모형을 1982.1.4부터 2008.12.31까지 수집된 원/달러 환율 시계열자료로부터 계산된 일별 환율 투자 손실률, 즉 일별 로그 손실률에 적용한다. 여기서 주된 관심은 10년 혹은 50년에 한번 정도 발생하는 대형 손실률 수준과 같은 극단분위수를 어떻게 추정하느냐 하는 것이다.

주요어어: 극단값이론, 2차원 점과정, 극단분위수 추정, 환율.

## 1. 서론

환율이란 일반적으로 우리 돈과 외국 돈을 바꿀 때 적용되는 교환 비율을 말하는데, 환율을 표시하는 방법으로는 자국통화표시환율과 외국통화표시환율을 들 수 있다. 예를 들어, 우리나라 원화(KRW)와 미국 달러화(USD)의 환율을  $KRW/USD = 1,000$ 으로 표시하였다면 이는 1달러를 받기 위해 1,000원을 지급해야 한다는 것으로서 이러한 표시환율을 자국통화표시환율 또는 지급환율이라고 부르고, 반면에  $USD/KRW = 0.001$ 로 표시하였다면 이는 1원으로 0.001달러를 받을 수 있다는 것으로서 이러한 표시환율을 외국통화표시환율 또는 수취환율이라고 부르는데, 본 논문에서의 환율은 지금부터 자국통화표시환율, 즉  $KRW/USD$ 를 뜻하는 것으로 약속하기로 한다.

환율은 국제수지, 국내물가, 내외금리차 등 경제여건에 따라 수시로 변동하지만, 반대로 환율 변동이 국내물가, 생산, 국제수지 등 국내경제에 영향을 미치기도 한다. 환율이 떨어져 원화가치가 절상되면 일반적으로 수출은 줄고 수입은 늘어나며 이에 따라 대외존도가 높은 우리나라에서는 경제성장이 둔화되고 경상수지도 악화되지만, 반면에 수입물가가 떨어져 국내물가가 하락하고 기업의 외채상환부담도 줄어드는 효과가 있다. 반대로 환율이 올라 원화가치가 절하되면 수출이 늘어나고 수입이 줄어들기 때문에 경상수지가 개선되지만 수입물가가 올라 국내물가가 또한 상승하는 부작용을 초래할 수도 있다. 이와 같이 환율 변동은 경제 여러 분야에 상반되는 영향을 미치게 되므로 환율은 가능한 실세가 잘 반영되어 안정적으로 유지되는 것이 바람직할 것이다.

우리나라의 환율제도는 해방이후 고정환율제도(외환시세의 변동을 전혀 인정하지 않고 고정시켜 놓은 환율제도: 1945.10-1964.05)를 유지해 오다가 1960년대에 이르러 단일변동환율제도(환율의 하한선을 기준환율로 설정해 놓고 약간의 변동을 허용하는 환율제도: 1964.05-1980.02)로 전환되었

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

다. 그러나 단일변동환율제도하에서 사실상 환율은 고정되어 있었고 이에 따라 경제 규모의 확대나 물가 상승 등과 같은 경제 환경 변화에 적절히 대응하기 어려웠으므로 1980년대 초 복수통화바스켓 환율제도(미국 달러화를 비롯한 주요 교역국 통화가치에 원화가치를 연동시키는 환율제도: 1980.02-1990.02.28)로 전환되었다. 이후 1980년대 후반에 국제수지 흑자의 확대에 의해 외국, 특히 미국으로부터 외환 조조의 비난을 받기 시작하고 금융과 경제의 국제화가 진전됨에 따라 다시 시장평균환율제도(전날 모든 외환은행이 국내 외환시장에서 거래한 원화와 달러화와의 환율을 거래량으로 가중평균하여 정하는 환율제도: 1990.03.02-1997.12.15)를 도입하였으나 일중환율변동폭은 제한하였다. 그러나, 1997년 하반기 외환위기가 발생한 이후 환율제도는 환율 질상 압력이 시장 메카니즘에 의해 단기기간에 잘 반영될 수 있도록 일중환율변동폭 제한을 폐지하여 자유변동환율제도(정부의 구조적인 개입이나 제한없이 환율이 외환시장에서 외화의 수요와 공급에 따라 자유롭게 결정되는 환율제도: 1997.12.16-현재)로 전환되어 오늘에 이르고 있다.

본 논문에서는 달러화에 투자하였을 경우 입게 되는 일별 환율 투자 손실을 분포의 극단분위수(extreme quantile)를 추정하는 방법으로 연간 최대값 방법, 분계점 방법, 2차원 포아송초과점과정 방법의 세가지 극단값 통계 모형을 소개하고 이를 1982.1.4부터 2008.12.31까지 수집된 일별 KRW/USD 환율 자료에 적용한다.

## 2. 극단분위수 추정 방법

모평균에 대한 추론시 추정량으로 모평균에 대한 정보를 가지고 있는 표본평균을 사용하고 대개의 경우 이에 대한 극한분포로 정규분포를 사용하는 것은 모분산이 유한할 경우 성립하는 중심극한정리의 덕분이다. 모분포의 극단분위수 추정은 모분포의 오른쪽 끝점(right end-point) 부근에서는 표본의 최대순서통계량을, 그리고 모분포의 왼쪽 끝점(left end-point) 부근에서는 표본의 최소순서통계량을 사용하는 것이 합리적일 것이다. 최소순서통계량은 표본자료에 -1을 곱하여 최대순서통계량으로 변환하여 생각할 수 있으므로 여기서는 최대순서통계량만을 고려하기로 한다. 그러나, 여기서 우리가 추정하고자 하는 극단분위수는 전형적으로 최대순서통계량보다도 훨씬 그 값이 더 큰 경우로서 최대순서통계량을 단순히 극단분위수의 추정량으로 사용하는 것은 매우 비효율적이 될 것이다. 이를 위해 극단값 분석에서는 모분포의 오른쪽 끝점 부근에서의 모분포 추정(또는 복원)의 한 방법으로 통상 모분포함수의 외삽법(extrapolation)을 사용하는데, 이것은 정규화된 최대순서통계량의 극한분포에 기초한다. 이 극한분포를 극단값분포(extreme value distribution)라고 부른다.

### 2.1. 연간 최대값 방법

이제, 모집단 분포함수(distribution function)  $F$ 로 부터 확률표본(random sample)  $X_1, \dots, X_n$ 이 주어졌다고 하자. 그러면, 전형적으로 적당한 정규화 상수  $a_n > 0$ 과  $b_n \in \mathbb{R}$ 이 존재하여  $n \rightarrow \infty$ 일 때

$$P(\max(X_1, \dots, X_n) \leq a_n x + b_n) = F^n(a_n x + b_n) \rightarrow G_\xi(x), \quad x \in \mathbb{R} \quad (2.1)$$

이 성립하게 되는데(참조: Fisher와 Tippett (1928), Gnedenko (1943)), 여기서  $\xi \in \mathbb{R}$ 를  $F$ 의 극단값지수(extreme value index) 또는 꼬리지수(tail index)라고 부르고,  $G_\xi$ 는 극단값분포함수(extreme value distribution function)로서

$$G_\xi(x) := \exp\left\{-\left(1 + \xi x\right)_+^{-1/\xi}\right\}, \quad x \in \mathbb{R}$$

로 정의된다. 여기서,  $y_+ := \max(y, 0)$ 이고,  $\xi = 0$ 인 경우는 항상  $\xi \rightarrow 0$ 일 때의 극한으로 해석한다.

식 (2.1)의  $G_\xi$ 에 추가적으로 위치모수(location parameter)  $\mu \in \mathbb{R}$ 와 척도모수(scale parameter)  $\sigma >$

0를 도입하여

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

로 놓으면,  $n$ 이 충분히 크고  $z$  역시 충분히 클 경우 식 (2.1)로부터 다음의 근사식을 얻게 된다:

$$P\{\max(X_1, \dots, X_n) \leq z\} = F^n(z) \approx H(z; \mu, \sigma, \xi). \quad (2.2)$$

여기서,  $H(x; \mu, \sigma, \xi)$ 를 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.

식 (2.2)로부터  $n$ 과  $z$ 가 충분히 크면 근사식  $F(z) \approx H^{1/n}(z; \mu, \sigma, \xi)$ 가 성립하므로, 이는 결국 모집단 분포함수  $F$ 의 오른쪽 끝점 부근의 값은  $GEV(\mu, \sigma, \xi)$ 로 근사시켜 외삽법을 사용할 수 있다는 것이다. 이에 대한 통계적 방법으로 소위 연간 최대값 방법(annual maximum method)을 들 수 있는데 구체적인 내용은 다음과 같다. 우선, 기간  $m$ 년 동안 서로 *iid*(independent and identically distributed)인 일별 관측치가 주어지고 이들의 공통 분포함수를  $F$ 라고 하며, 이들로 부터 계산되는  $m$ 개의 연간 최대값들을  $M_1, \dots, M_m$ 이라고 하면, 이들이  $GEV(\mu, \sigma, \xi)$ 로 부터의 확률표본이라고 간주하여 모수의 최대가능도추정량(MLE: maximum likelihood estimator)  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ 을 쉽게 계산할 수 있게 된다. 이로 부터 적합된  $H(x; \hat{\mu}, \hat{\sigma}, \hat{\xi})$ 를 얻을 수 있고 또한 이를 이용하여  $F$ 의 극단분위수에 대한 최대가능도추정량을 계산할 수 있게 된다. 즉, 예를 들어  $t$ 가 큰 값일 때

$$F(q_t) = 1 - \frac{1}{365.25t} \quad (2.3)$$

을 만족하는  $t$ 년 재발수준( $t$ -year return level)  $q_t$ 는 근사식

$$H(q_t; \mu, \sigma, \xi) \approx 1 - \frac{1}{t} \quad (2.4)$$

을 만족하므로  $q_t$ 의 최대가능도추정량  $\hat{q}_t$ 은  $H^{-1}(1 - 1/t; \hat{\mu}, \hat{\sigma}, \hat{\xi})$ 으로 계산될 수 있을 것이다. 여기서,  $H^{-1}$ 는  $H$ 의 분위수함수(quantile function)를 나타낸다.

## 2.2. 분계점 방법

적당한  $\xi \in \mathbb{R}$ 에 대하여 식 (2.1)이 성립할 때 Pickands (1975)는 충분히 큰 임의의 분계점(threshold)  $u$ 에 대하여 역시 다음의 근사식이 성립한다는 사실을 보였다:

$$P\{X_i - u \leq y | X_i > u\} = \frac{F(u+y) - F(u)}{1 - F(u)} \approx G(y; \phi, \xi), \quad y > 0. \quad (2.5)$$

여기서  $\phi > 0$ 는 분계점  $u$ 에 의존하는 척도모수이고, 일반화파레토분포함수(generalized Pareto distribution function)라고 불리는  $G(y; \phi, \xi)$ 는

$$G(y; \phi, \xi) := 1 + \log G_\xi \left( \frac{y}{\phi} \right) = 1 - \left( 1 + \frac{\xi y}{\phi} \right)_+^{-1/\xi}, \quad y > 0$$

로 정의되며 이 분포를 간단히 기호  $GPD(\phi, \xi)$ ,  $\phi > 0$ ,  $\xi \in \mathbb{R}$ 로 나타낸다.

식 (2.5)를 이용하여 모집단 분포함수  $F$ 의 극단분위수를 추정하는 통계적 방법을 소위 분계점 방법(threshold method)이라고 부르는데 구체적인 내용은 다음과 같다. 우선, 앞의 연간 최대값 방법에서

와 같이 기간  $m$ 년 동안 서로 iid인 일별 관측치  $X_1, \dots, X_n$ 이 주어져 있고 이들의 공통 분포함수를  $F$ 라고 하자. 분계점 방법에서는 일단 충분히 큰 하나의 분계점  $u$ 를 선택하여 이를 넘어서는 초과값들을 모두 이용하는 것인데,  $u$ 를 너무 크게 선택하면 초과값들의 개수가 작아져서 추정의 정밀도(precision)가 낮아지게 되고, 반면에  $u$ 를 너무 작게 선택하면 근사식 (2.5)의 성립을 보장할 수 없으므로 추정의 편향(bias)이 발생하게 된다. 분계점  $u$ 의 선택은 통상 전체 관측치의 상위 5% 미만이 초과값으로 선택되도록 정하는데 보통 다음의 방법을 이용한다.

하나의 확률변수  $Y$ 가  $GPD(\phi, \xi)$ 를 따르면  $\phi + \xi t > 0$ 인 모든  $t > 0$ 에 대하여  $Y - t|Y > t \sim GPD(\phi + \xi t, \xi)$ 가 성립하므로  $\xi < 1$ 일 경우

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

가 성립함을 쉽게 알 수 있는데, 이는 곧  $t$ 의 함수로서  $E(Y - t|Y > t)$ 는 기울기가  $\xi/(1 - \xi)$ 이고 절편이  $\phi/(1 - \xi)$ 인 직선을 의미한다는 것이다. 따라서, 분계점  $u$ 는 일별 관측치들의 MRL 그림(mean residual life plot), 즉 모든  $t > u$ 에 대하여  $t$ 와  $t$ 를 넘어서는 초과값들의 초과여분(excesses) 산술 평균값의 산점도가 근사적으로 직선을 따르는 최소의 값으로 선택할 수 있을 것이다.

이제, 이렇게 선택된 분계점  $u$ 에 대하여  $X_1, \dots, X_n$  중  $u$ 를 넘어서는 초과값들이  $N$ 개 관찰되고 이들을 각각  $Z_1, \dots, Z_N$ 이라고 하면, 이들의 초과여분, 즉  $Z_1 - u, \dots, Z_N - u$ 들이  $GPD(\phi, \xi)$ 로 부터의 확률표본이라고 간주하여 모수의 최대가능도추정량  $\hat{\phi}, \hat{\xi}$ 을 계산한다. 이로 부터 적합된  $G(y; \hat{\phi}, \hat{\xi})$ 을 얻고  $q_t > u$ 가 식 (2.3)을 만족하는  $t$ 년 재발수준이라고 하면 식 (2.5)로 부터  $q_t$ 는 근사식

$$\frac{1}{365.25t} = P\{X_i > u\} \times P\{X_i > q_t | X_i > u\} \approx p(1 - G(q_t - u; \phi, \xi))$$

를 만족하므로  $q_t$ 의 최대가능도추정량  $\hat{q}_t$ 은

$$u + G^{-1}\left(1 - \frac{1}{365.25tp}; \hat{\phi}, \hat{\xi}\right)$$

으로 계산될 수 있을 것이다. 여기서,  $p = P\{X_i > u\}$ 로  $p$ 는 추정량  $N/n$ 으로 대체하여 사용할 수 있는데,  $N$ 이 이항분포를 따르므로 이 추정량은 역시  $p$ 의 최대가능도추정량이 된다.

### 2.3. 2차원 포아송초과점과정 방법

모집단 분포함수  $F$ 로 부터의 확률표본  $X_1, \dots, X_n$ 이 주어졌을 때, 식 (2.1)이 성립하면 식 (2.2)가 역시 성립하고, 따라서 충분히 큰 하나의 분계점  $u$ 에 대하여  $n$ 이 충분히 크면, 근사식

$$n(1 - F(u)) \approx -\log H(u; \mu, \sigma, \xi) = \left\{1 + \frac{\xi(u - \mu)}{\sigma}\right\}_+^{-1/\xi}$$

가 성립하게 되는데, 하나의 초과점과정(exceedance point process)  $N_n = \{N_n(B), B \in \mathcal{B}((0, 1])\}$ 을

$$N_n(B) = \#\{i : i/n \in B, X_i > u, i = 1, \dots, n\}$$

으로 정의하면, 즉  $N_n(B)$ 를 분계점  $u$ 를 넘어서는 초과값  $X_i$  중  $i/n$ 가  $B$ 에 속하는 점들의 개수라고 정의하면,  $N_n$ 은  $n$ 이 충분히 클 때 점근적으로 강도(intensity)가  $-\log H(u; \mu, \sigma, \xi)$ 인 포아송과정(Poisson process)을 따르게 된다 (참조: Leadbetter 등 (1983), Resnick (1987)). 여기서,  $\mathcal{B}((0, 1])$ 은  $(0, 1]$  상의 Borel  $\sigma$ -field를 의미한다.

이 결과는 통계적 응용의 목적으로  $(0, 1] \times (u, \infty)$  상의 2차원 초과점과정으로 다음과 같이 확장하여 생각할 수 있다. 즉, Smith (1989)는  $D = (0, 1] \times (u, \infty)$  상의 2차원 초과점과정  $N = \{N(A), A \in \mathcal{B}(D)\}$ 를

$$N(A) = \#\{i : (i/n, X_i) \in A, i = 1, \dots, n\}$$

으로 정의한 후  $N$ 이 근사적으로 강도함수(intensity function)가

$$\lambda(t, y) = \frac{1}{\sigma} \left\{ 1 + \frac{\xi(y - \mu)}{\sigma} \right\}_+^{-1/\xi-1}, \quad (t, y) \in D \tag{2.6}$$

인 2차원 비동질 포아송과정(two-dimensional non-homogeneous Poisson process)을 따른다고 가정하였다. 이 가정에 따르면, 임의의  $z > u$ 에 대하여  $A = (0, 1] \times (z, \infty)$ 라고 놓을 때  $N(A)$ 가 근사적으로 평균이

$$\int_0^1 \int_z^\infty \lambda(t, y) dy dt = -\log H(z; \mu, \sigma, \xi)$$

인 포아송분포를 따르게 되므로

$$P\{\max(X_1, \dots, X_n) \leq z\} = P\{N(A) = 0\} \approx H(z; \mu, \sigma, \xi)$$

가 성립하고, 또한  $\phi = \sigma + \xi(u - \mu) > 0$ 라고 놓으면

$$P\{X_i > z | X_i > u\} = \frac{E(N(A))}{E(N(D))} \approx \frac{-\log H(z; \mu, \sigma, \xi)}{-\log H(u; \mu, \sigma, \xi)} = 1 - G(z - u; \phi, \xi)$$

가 성립하게 되어 앞의 근사식 (2.2)와 (2.5)가 동시에 만족됨을 알 수 있다.

따라서, 2차원 비동질 포아송과정으로서의 2차원 초과점과정의 고려는 앞의 연간 최대값 방법과 분계점 방법을 통합하는 잇점을 갖는다고 할 수 있을 것이다. 여기서, 이를 사용하는 통계적 방법을 2차원 포아송초과점과정 방법으로 부르기로 하는데, 구체적인 내용은 다음과 같다. 앞에서와 같이 기간  $m$ 년 동안 서로 iid인 일별 관측치  $X_1, \dots, X_n$ 이 주어지고 이들의 공통 분포함수를  $F$ 라고 하자. 우선, 앞의 분계점 방법에서와 같이 일별 관측치들의 MRL 그림을 이용하여 충분히 큰 하나의 분계점  $u$ 를 선택하고,  $D = (0, m] \times (u, \infty)$  상의 2차원 초과점과정  $N = \{N(A), A \in \mathcal{B}(D)\}$ 를

$$N(A) = \#\{i : (im/n, X_i) \in A, i = 1, \dots, n\}$$

으로 정의한 후  $N$ 이 식 (2.6)과 동일한 강도함수  $\lambda(t, y)$ ,  $(t, y) \in D$ 를 갖는 2차원 비동질 포아송과정을 따른다고 간주한다. 이제  $u$ 를 넘어서는  $N_m = N(D)$ 개의 초과값들이  $(T_1, Z_1), \dots, (T_{N_m}, Z_{N_m}) \in D$ 에서 관찰되었다고 하면, 이에 기초한 가능도함수(likelihood function)는

$$L(\mu, \sigma, \xi) = \exp \left[ -m \left\{ 1 + \frac{\xi(u - \mu)}{\sigma} \right\}_+^{-1/\xi} \right] \times \prod_{i=1}^{N_m} \left[ \frac{1}{\sigma} \left\{ 1 + \frac{\xi(Z_i - \mu)}{\sigma} \right\}_+^{-1/\xi-1} \right]$$

과 같이 주어지는데, 이를 이용하면 모수의 최대가능도추정량  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ 을 계산할 수 있게 된다. 따라서, 이로부터 적합된  $H(z; \hat{\mu}, \hat{\sigma}, \hat{\xi})$ 을 얻고  $q_t > u$ 가 식 (2.3)을 만족하는  $t$ 년 재발수준이라고 하면  $q_t$ 는 역시 근사식 (2.4)를 만족하므로  $q_t$ 의 최대가능도추정량  $\hat{q}_t$ 은  $H^{-1}(1 - 1/t; \hat{\mu}, \hat{\sigma}, \hat{\xi})$ 으로 계산될 수 있다.

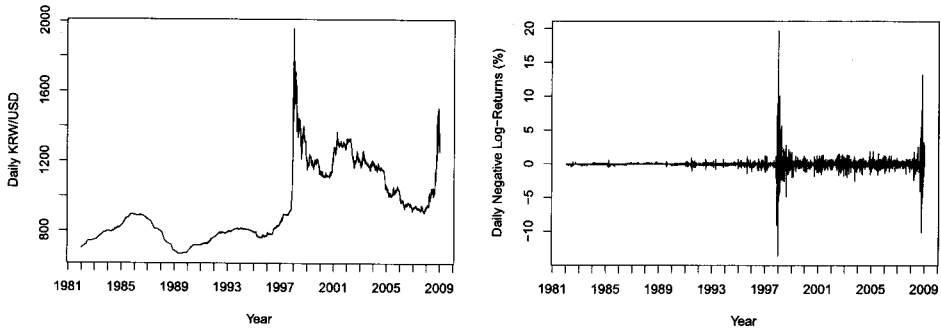


그림 1: 일별 KRW/USD 환율과 일별 로그 손실률(%)

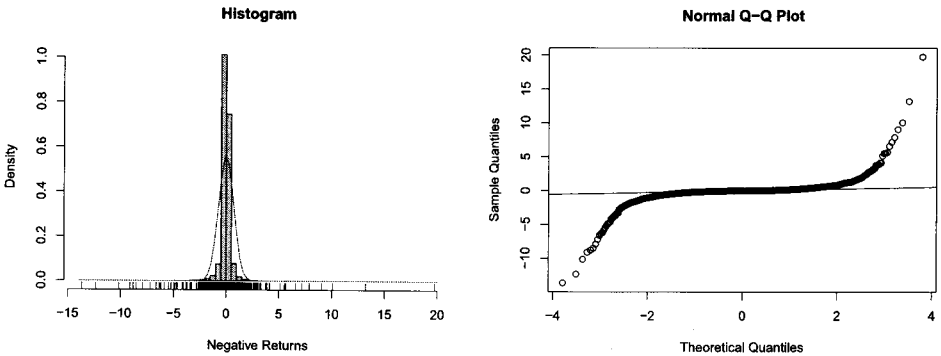


그림 2: 일별 손실률의 히스토그램과 정규분포 Q-Q 그림

### 3. 원/달러 환율 시계열자료에서의 대형 투자 손실률 추정

본 논문에서 다루고자 하는 자료는 대미 달러화 환율 KRW/USD 자료인데, 이 자료는 미국 뉴욕 연방준비은행이 수집하고 있는 일별 뉴욕 정오 전신환 매매환율 자료로서 누구나 다운로드가 가능하다 (참조: <http://www.federalreserve.gov/RELEASES/H10/Hist/default.htm>). 본 논문에서는 1982.1.4부터 2008.12.31까지 27년간의 자료를 분석하는데 총 관측치의 개수는 6,740개이고, 이 일별 KRW/USD 환율을 그림 1의 좌측에 나타내었다. 이 자료로부터  $t$ 일 로그 손실률(단위: %)(negative log-returns in percentage)  $X_t$ 를  $X_t = -100 \log(t$ 일 환율/ $t$ 일 전일 환율)과 같이 계산할 수 있는데, 그림 1의 우측은 이렇게 계산된 일별 로그 손실률  $\{X_t\}$ (자료 개수는 6,739개)를 나타낸다. 앞으로 환율의 로그 손실률을 간단히 손실률이라고 부르기로 하는데, 이는 우리가 달러화에 투자하였을 경우 입게 되는 손실률을 의미 하며 동시에 달러화에 대한 원화의 절상률에 해당한다. 손실률은 최고 19.759%(1997.12.26)와 두번째 13.222%(2008.10.30)를 제외하면 모두 10.08% 미만으로 나타났다. 그림 1에서 보듯, 좌측의 원시 환율 자료와 달러 우측의 손실률 자료  $\{X_t\}$ 는 자료간 상관성이 없어 보이는데, 여기서는  $\{X_t\}$ 가 iid로서 적당한  $\xi \in \mathbb{R}$ 가 존재하여 식 (2.1)이 성립한다고 가정한다.

일별 손실률의 분포가 정규분포를 따르는 지를 알아보기 위해 그림 2에 손실률의 히스토그램과 정규분포 Q-Q 그림(normal quantile-quantile plot)을 각각 그려 넣었다. 전체 손실률 자료의 평균과 표준 편차는 각각 -0.009와 0.714인데, 히스토그램에는  $N(-0.009, 0.714^2)$  분포의 확률밀도함수를 함께 그려 넣었다. 그림 2에서 보듯, 일별 손실률의 분포는 정규분포보다도 중심과 양쪽 꼬리 부분이 훨씬 두

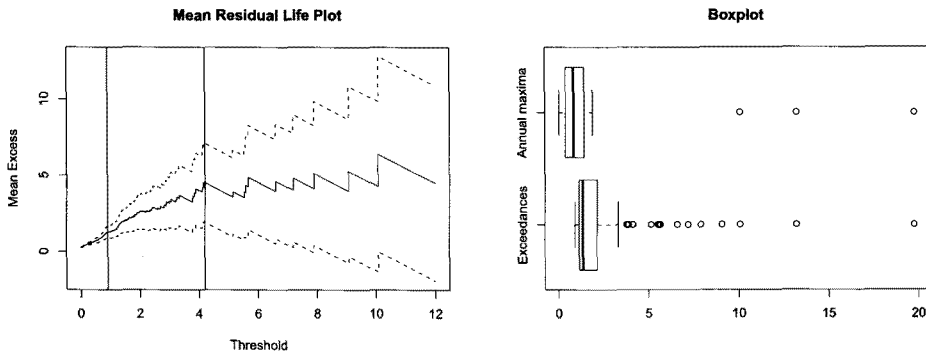


그림 3: 일별 손실률의 MRL 그림과 연간 최대값 및  $u = 0.9$  초과값들의 상자 그림

표 1: 일별 손실률의 연간 최대값들과  $u = 0.9$  초과값들의 기초 통계량

자료	표본크기	최소값	1사분위수	중앙값	평균값	3사분위수	최대값
연간 최대값	27	0.039	0.399	0.840	2.380	1.447	19.759
$u = 0.9$ 초과값	143	0.902	1.138	1.342	2.169	2.141	19.759

표 2: GEV, GPD, TPP 모형에서의 모수 추정 결과(SE: 표준오차, CI: 프로파일 로그 가능도에 기초하여 계산된 신뢰구간)

자료	모형	추정	$\mu$	$\sigma$	$\xi$	$\phi$
연간 최대값	GEV	MLE (SE)	0.599 (0.134)	0.617 (0.154)	0.772 (0.213)	
		95% CI	(0.374,0.919)	(0.386,1.062)	(0.427,1.290)	
$u = 0.9$ 초과값	GPD	MLE (SE)			0.537 (0.124)	0.627 (0.090)
		95% CI			(0.325,0.818)	(0.471,0.826)
$u = 0.9$ 초과값	TPP	MLE (SE)	2.589 (0.249)	1.533 (0.279)	0.537 (0.124)	
		95% CI	(2.173,3.195)	(1.120,2.324)	(0.324,0.818)	

더움을 알 수 있다.

이제, 2절에서 소개한 세가지 방법으로 일별 손실률 분포의 극단분위수를 추정해 보기로 한다. 우선, 연간 최대값 방법에서는 27년 동안의 연간 최대 손실률 27개에 대하여  $GEV(\mu, \sigma, \xi)$  모형을, 분계점 방법에서는 충분히 큰 하나의 분계점  $u$ 를 넘어서는 초과값들의 초과여부에 대하여  $GPD(\phi, \xi)$  모형을, 그리고 2차원 포아송초과점과정 방법에서는 분계점 방법에서와 동일한 분계점  $u$ 를 넘어서는 초과값들에 대하여 강도함수가 (2.6)으로 주어진 2차원 포아송과정(two-dimensional Poisson process)을 최대가능도추정법으로 적합시키는데, 지금부터 세번째 모형을 간단히  $TPP(\mu, \sigma, \xi)$  모형으로 칭하기로 한다. GPD 모형과 TPP 모형을 적합시키기 위해서는 적절한 하나의 분계점  $u$ 를 선택해야 하는데, 이를 위한 일별 손실률들의 MRL 그림을 그림 3의 좌측에 그려 넣었다. 그림을 보면  $t = 4.2$ 를 넘어서는 손실률 11개를 제외하면  $t$ 가 (0.9, 4.2)에서 움직일 때 MRL 그림이 양의 기울기를 갖는 직선 형태이므로, 여기서  $u = 0.9$ 를 분계점으로 선택하기로 한다. 그림 상으로 더 작은 분계점도 역시 가능할 것으로 생각되나, 분계점 선택의 추가적인 고려는 하지 않기로 한다.

표 1은 일별 손실률의 연간 최대값들과 분계점  $u = 0.9$ 를 넘어서는 초과값들에 대한 기초 통계량을 보여주며, 이들의 상자 그림은 그림 3의 우측에 그려 넣었다. 표와 그림에 의하면, 연간 최대값의 분포와 초과값의 분포는 모두 오른쪽으로 두터운 꼬리를 갖고 있으며 또한 오른쪽으로 매우 기울어져 있음을 알 수 있다. 특별히, 상자 그림에 의하면 연간 최대값들의 경우 상위 3개를 제외하면 초과값들에게서 보이는 큰 값들이 대부분 포함되어 있지 않는데, 이는 일별 손실률 분포의 오른쪽 꼬리에 대한 정보

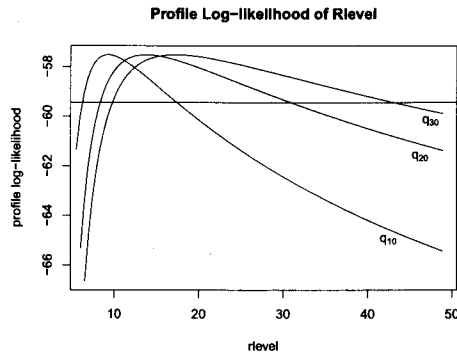


그림 4: TPP 모형에서  $q_{10}$ ,  $q_{20}$ ,  $q_{30}$ 에 대한 프로파일 로그 가능도

표 3: GEV, GPD, TPP 모형에서의  $t$ 년 재발수준  $q_t$  추정 결과(SE: 표준오차, CI: 프로파일 로그 가능도에 기초하여 계산된 신뢰구간)

모형	추정	$q_{10}$	$q_{20}$	$q_{30}$	$q_{50}$
GEV	MLE (SE)	4.339 (1.737)	7.713 (4.006)	10.695 (6.349)	16.049 (11.084)
	95% CI	(2.340,12.573)	(3.558,31.014)	(4.463,47.759)	(5.838,54.678)
GPD	MLE (SE)	11.787 (3.408)	17.217 (6.226)	21.414 (8.659)	28.173 (12.950)
	95% CI	(7.512,24.331)	(9.925,42.460)	(11.636,58.908)	(14.138,89.086)
TPP	MLE (SE)	9.287 (2.329)	13.792 (4.439)	17.289 (6.315)	22.910 (9.656)
	95% CI	(6.267,17.413)	(8.217,32.340)	(9.841,43.055)	(11.934,65.347)

를 가지고 있는 자료들이 누락된 것이므로 형상모수(shape parameter)  $\xi \in \mathbb{R}$ 의 추정시 과대추정과 과대 오차에 예상된다 할 수 있을 것이다.

표 2는 일별 손실률의 연간 최대값들에 대하여 GEV( $\mu, \sigma, \xi$ ) 모형을, 그리고  $u = 0.9$ 를 넘어서는 초과값들에 대하여 GPD( $\phi, \xi$ ) 모형과 TPP( $\mu, \sigma, \xi$ ) 모형을 수치적인 방법을 사용하여 최대가능도추정법으로 적합시킨 결과를 보여준다. 표 2에는 또한 프로파일 로그 가능도(profile log-likelihood)에 기초하여 계산된 95% 신뢰구간이 포함되어 있다. 표에 의하면 모수  $\xi$ 의 경우 예상대로 GPD 모형이나 TPP 모형보다 GEV 모형에서 과대추정과 과대오차가 발생하였음을 알 수 있다.

표 3은 앞의 세가지 모형 각각에 대해서 일별 손실률의 10년, 20년, 30년, 50년 재발수준에 대한 최대가능도추정의 결과를 보여준다. 표 3에는 역시 프로파일 로그 가능도에 기초하여 계산된 95% 신뢰구간이 포함되어 있는데, 이를 선호하는 이유는 특별히 재발수준과 같은 값을 추정할 때 가능도 함수의 모양이 전형적으로 비대칭적이어서 표준오차와 정규분포에 기초하여 계산된 신뢰구간은 비현실적이기 때문이다. 표에 의하면 재발수준은 GEV 모형에서 가장 작게 추정되었고, GPD 모형에서 가장 크게 추정되었으며, TPP 모형에서 중간 정도의 크기로 추정되었음을 알 수 있다.

그림 4는 TPP 모형에서 일별 손실률의 10년, 20년, 30년 재발수준에 대한 프로파일 로그 가능도 함수를 보여 주는데, 모두 비대칭적으로 생겼음을 알 수 있다. 여기서, 수평 참조선은 재발수준의 95% 신뢰구간을 계산하기 위한 절단선을 나타내는데, 이 값은 로그 가능도비 통계량(log-likelihood ratio statistic)의 점근분포인  $\chi^2(1)$  분포에 의하여 계산된 값이다.

#### 4. 결론

본 논문에서는 1982.1.4부터 2008.12.31까지 수집된 일별 KRW/USD 환율 자료를 사용하여 10년, 20년, 30년 혹은 50년에 한번 정도 발생하는 대형 환율 투자 손실률을 추정하였다. 추정 방법으로는 연



간 최대값 방법, 분계점 방법, 2차원 포아송초과점과정 방법의 세가지 극단값 통계 모형을 사용하였는데, 연간 최대값 방법에서는 꼬리지수  $\xi$ 가 과대평가되어 추정된 반면 재발수준은 과소평가되어 추정되었고, 분계점 방법에서는 재발수준이 과대평가되어 추정되었으나, 이러한 단점은 2차원 포아송초과점과정 방법을 사용하면서 다소 해결되었다. 꼬리지수  $\xi$ 의 MLE(SE)와 95% 신뢰구간은 2차원 포아송초과점과정 방법에서 0.537(0.124)과 (0.324,0.818)로 나타났는데, 이는 일별 손실률의 분포가 정규분포와 달리 두꺼운 꼬리를 갖는 분포라는 최근의 여러 주장들을 뒷받침하는 것이다.

### 참고 문헌

- 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.
- Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics*, **3**, 119–131.
- Resnick, S. (1987). *Extreme Values, Point Processes, and Regular Variation*, Springer, New York.
- 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.

2009년 7월 접수; 2009년 9월 채택

# Extreme Quantile Estimation of Losses in KRW/USD Exchange Rate

Seokhoon Yun<sup>1,a</sup>

<sup>a</sup>Department of Applied Statistics, University of Suwon

---

## Abstract

The application of extreme value theory to financial data is a fairly recent innovation. The classical annual maximum method is to fit the generalized extreme value distribution to the annual maxima of a data series. An alternative modern method, the so-called threshold method, is to fit the generalized Pareto distribution to the excesses over a high threshold from the data series. A more substantial variant is to take the point-process viewpoint of high-level exceedances. That is, the exceedance times and excess values of a high threshold are viewed as a two-dimensional point process whose limiting form is a non-homogeneous Poisson process. In this paper, we apply the two-dimensional non-homogeneous Poisson process model to daily losses, daily negative log-returns, in the data series of KRW/USD exchange rate, collected from January 4th, 1982 until December 31st, 2008. The main question is how to estimate extreme quantiles of losses such as the 10-year or 50-year return level.

**Keywords:** Extreme value theory, two-dimensional point process, extreme quantile estimation, exchange rate.

---

---

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