Proceedings of Joint Conference of The Korean Data Analysis Society and Korean Data And Information Science Society, October 17-18, 2003, pp. 147-152

고차 일반화극치분포와 PMLE를 이용한 환율자료분석

정보윤 · 전유나[1] · 박정수[2]

초 록

본 논문에서는 일반화극치분포(GEV)와 r개의 순서통계량을 이용한 r-GEV를 기술하였다. 모수 μ, σ, k를 추정하기 위해 최우추정법(MLE)과 Penalized MLE(P-MLE) 방법을 적용해 보았다. 이 분포를 원/달러 환율자료에 적용하여 일종의 재정위기 분석을 실시하였다.

1. 서론

재정관련 자료(주가, 환율 등)에 대한 분석에서 기존에는 정규분포를 사용하였으나, 큰 위험(Extreme case)을 볼 경우의 손실량을 제대로 예측하는데 어려움이 있다. 따라서, Extreme event는 EV분포(예: GEV, GPD)를 근사적으로 따르기 때문에, 극치이론(EVT : Extreme Value Theory)을 이용하여 큰 위험을 예측하는 것이 옳다고 본다. 본 논문에서는 한국의 원/달러 환율 자료에 EVT를 (특히 r-GEV분포 : MLE방법과 P-MLE방법) 적용하여 큰 위험을 예측하는데 도움이 되는 통계량을 계산하는데 의의를 둔다.

극치이론을 이해하기 위해서 다음과 같이 관측치의 극대값을

$$X_{(1)} = \max[X_1, X_2, \cdots, X_{n-1}, X_n], \quad 1 \le i \le n$$

로 정의하자. 극치이론이라고 불리는 Fisher-Tippett 정리(Extreme Value Theorem)로부터 우리는 극대값 $X_{(1)}$의 분포가 자료의 수가 증가하면서 안정적인 극한분포(Limiting Distribution)로 수렴한다는 사실을 알고 있다. 이러한 극한분포는 다음과 같은 일반화된 극치분포(Generalized Extreme Value Distribution)로 요약될 수 있다(한상범, 1999).

$$F(x) = \exp\left(-\left[1 - k\left(\frac{x-\mu}{\sigma}\right)\right]^{1/k}\right)$$

여기서 μ: location 모수, σ : scale 모수, k : tail 모수

또한, 이 분포는

i) $k=0$인 경우 Gumbel Distribution(Fisher-Tippett types I)

---

[1]전남대학교 통계학과 대학원

[2]전남대학교 통계학과 교수

ii) $k<0$인 경우 Frechet Distribution(Fisher-Tippett types II)

iii) $k>0$인 경우 negative Weibull Distribution(Fisher-Tippett types III)

라 불린다.

본 논문에서는 한국 환율 극관측치(Extreme Observation)를 사용하여 GEV, r-orders GEV/MLE, r-orders GEV/P-MLE에 적합하였다. 데이터가 추출된 기간은 1997년 1월부터 2002년 4월까지이고, 구간을 4개월, 12개월을 설정하여 각각의 구간에서 $X_{(1)}, X_{(2)}, \cdots, X_{(7)}$까지를 추출하였다. 위와 같은 방법을 통하여 한국 환율 극관측치를 이용하여 하여 어떤 분포가 더 적합하였지 고려하여 본다. 또한 최종적으로 위와 같은 방법을 통하여 VaR에서 사용되는 분포인 표준분포(정규분포, 로그정규분포)보다는 GEV분포를 사용하기 위한 통계적 도구를 제공한다.

2장은 GEV분포, r-GEV분포, MLE방법, P-MLE방법과 또한 quantile의 계산과정을 개략적으로 소개하고 3장은 우리나라의 원/달러 환율을 실증분석한 결과를 제시한다. 또한 quantile의 또 다른 방법인 Return Value에 대하여 소개한다. 제 4장에서는 결론을 맺고 향후 연구과제에 대하여 논의한다

2. 연구 내용

1) 일반화 극치 분포(GEV분포)

극치 데이터에 적합하기 위한 GEV 분포의 누적확률분포함수 $F(x)$는 다음과 같이, $k\neq 0$의 경우에, (여기서, $y_i=(x_i-\mu)/\sigma$, $G_i=1-ky_i>0$),

$$F(x_i)=\exp\left(-G_i^{1/k}\right),$$

이고, $k=0$의 경우, $F(x_i)=\exp[-\exp(-y_i)]$ 이다. 위의 CDF를 이용하여 확률분포함수 $f(x)$를 다음과 같이 얻을 수 있다. $k\neq 0$의 경우일 때,

$$f(x) \;=\; \frac{1}{\sigma}G_i^{1/k-1}\exp\left(-G_i^{1/k}\right)$$

이고, $k=0$의 경우 $f(x_i)=\frac{1}{\sigma}\{\exp(-y_i)\}\exp\{-\exp[-y_i]\}$ 이다. 본 논문에서는 모수의 추정을 위하여 MLE 방법을 사용하였다. MLE방법은 우도함수($\ln L(x)$)를 최대로 만드는 모수를 찾는 방법이다. MLE의 방법을 이용하여 모수를 찾기 위한, GEV분포의 우도함수식($L(x)$)는 다음과 같이 계산할 수 있다. $k\neq 0$의 경우,

$$L(x;\alpha,\beta,\sigma,k)=\log\prod_{i=1}^{n}f(x_i) \;=-n\log\sigma+\left(\frac{1}{k}-1\right)\sum_{i=1}^{n}\log G_i-\sum_{i=1}^{n}G_i^{1/k} \tag{2.3}$$

$k=0$인 경우,

$$L(x;\alpha,\beta,\sigma)=\log\prod_{i=1}^{n}f(x)=-nlog\sigma-\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}\exp(-y_i)$$

위의 식을 통해서 계산된 최우추정량을 이용하여 백분위함수 x_p가 다음과 같이 유도된다. 백분위 함수는 $x_p=\mu+\frac{\sigma}{k}\{1-[-\ln(1-p)]^k\}$ 이다. 여기서, x_p의 분산은 delta method에 의해서 계산되고, 계산된 추정치와 분산을 이용하여 X_p의 95%의 C.I.(신뢰구간)은 다음과 같이 구해진다: $\widehat{x_p}\pm1.96*\sqrt{Var(\widehat{X_p})}$.

2) r orders-GEV (r-GEV)

r-GEV는 순서 통계량인 $X_{(r)}<X_{(r-1)}<\cdots<X_{(1)}$의 Joint Distribution이 n이 무한대로 가까워 갈 때 r-GEV 분포로 수렴한다는 사실을 알고 있다(Twan, 1987). r-GEV분포를 사용하는 이유는 1-GEV를 사용하는 것보다, r-GEV를 사용할 경우에 더욱 더 많은 정보를 이용할 수 있으므로 더 정확한 모수 추정이 가능하기 때문이다. 정확한 모수의 추정이란 모수들의 표준 오차가 줄어듬을 의미한다. 또한, 경제 데이터의 경우 연 최대 강수량과 같은 시간적 어떠한 경향으로 나타낼 수 없으므로, 잘못된 구간 선택의 위험을 줄이는 방법으로 사용하기 위하여 r을 1보다는 여러 개를 사용하는 것이다.

다음 식은 r-GEV분포의 주변 확률분포함수 $f(x_1,x_2,\cdots,x_r)$이다. $k\neq0$의 경우에,

$$f(x_1,x_2,\cdots,x_r;\mu,\sigma,k)=\sigma^{-r}g(x_1,x_2,\cdots,x_{r-1})\left[1-k\left(\frac{x_r-\mu}{\sigma}\right)\right]^{1/k-1}\cdot\exp\left\{-\left[1-k\left(\frac{x_r-\mu}{\sigma}\right)\right]^{1/k}\right\}$$

여기서, $g(x_1,x_2,\cdots,x_{r-1})=\prod_{j=1}^{r-1}\left[1-k\left(\frac{x_j-\mu}{\sigma}\right)\right]^{1/k-1}$. 그리고, 최대값 x_1에 대한 quantile을 계산하기 위하여, 주변 확률분포함수 $f(x_1)$을 계산하여야 한다. 이때, quantile의 추정을 위하여 사용되어질 추정치는 주어진 자료로부터 r-GEV를 적합시켜 얻어진 $\hat{\mu},\hat{\sigma},\hat{k}$ 이다. 먼저 주변 확률분포함수 $f(x_1)$를 계산하여 보면, $f(x_1)$는 다음과 같이 결합 밀도함수 $f(x_1,x_2,\cdots,x_r)$를 x_r부터 x_2까지 부분 적분을 함으로써 얻을 수 있는데, 이는 (구체적 계산 생략) $f(x_1)$의 주변 확률분포함수가 GEV분포의 확률분포함수와 같음을 알 수 있다. 위의 얻어진 결과로부터, r-GEV분포로부터 $\hat{\mu},\hat{\sigma},\hat{k}$ 를 추정하여 이를 1-GEV의 x_p에 적용 가능함을 알 수 있다.

3) Penalized MLE

앞에서 언급된 GEV에서 MLE방법은 표본의 크기가 크고, 또한 -0.5<k<0.5 있어야 모수를 추정할 때 적합하다고 이미 밝혀졌다. 그러나, 표본의 크기가 작은 경우 모수 k의 추정치의 편의(bias)가 매우 크고, k>1/2인 경우 MLE의 점근적인 효율성을 보증할 수 없을 뿐만 아니라, k<-1/2인 경우 변수의 분산이 무한대로 가까워짐을 선행연구를 통해서 이미 입증되었다. 위의 문제점을 개선하기 위하여 최근 연구에서는 다음과 같은 MLE의 계산식에 Penalty 함수($\pi(k)$)를 추가하여 P-MLE의 값을 최대로 만드는 모수를 추정하는 방법이 연구되고 있다(Martin & Steidinger(2000), Coles & Dikson(2001)).

$$\text{P-MLE: } -\log L(x;\mu,\sigma,k) - \log \text{Penalty Function}(\pi(k))$$

$$\text{Penalty Function}: \ \pi(k) = (0.5+k)^{u-1} \cdot (0.5-k)^{v-1}/B(u,v)$$

본 논문에서 사용되는 Penalty Function($\pi(k) = B(2.5, 2.5)$, $-0.5 \leq k \leq 0.5$)은 이미 시뮬레이션을 통하여 $\pi(k)$를 갖는 P-MLE가 MLE보다 X_p추정을 훨씬 잘한다고 보고되었다(박정수, 2002). r-GEV분포에서 Penalty Function로서 $\pi(k) = B(2.5, 2.5)$가 적절하다고 시뮬레이션의 결과에 의해 밝혀지지 않았지만, 위의 논문을 확장하여 r-GEV 분포에서도 $\pi(k)$함수로 $B(2.5, 2.5)$분포가 적절하다는 가정 하에 사용되었다.

3. 실증분석

실증분석에 사용된 자료는 1987년 1월 1일부터 2002년 4월 12일까지의 대한민국의 원/달러의 일별자료를 각각 x_t/x_{t-1}(일일 환율 최대 변동율)은 최대값으로 x_{t-1}/x_t (일일 환율 최소 변동율)은 최소값으로 정의하였다. 이들 자료를 각각 4개월, 12개월로 구분하여 각 구간에서 7번째까지의 큰 값을 이용하여 r-GEV/MLE, r-GEV/P-MLE분포에 적합하여 본다. 프로그램은 Coles가 개발한 GEV, r-GEV의 모수를 추정하는 프로그램을 수정하여 사용하였고 S-Plus, Excel을 사용하였다. 1997년 12월 이전 최대 수익률의 소폭 변화가 존재하지만 1997년 말 외환위기이후 최대 수익률의 큰 변화가 존재함을 알 수 있다(데이타 그림 생략).

위 방법을 이용하여 추출된 모수들은 12개월의 r=2일 때를 제외하고, r이 증가할수록 모수의 추정치들이 증가함을 알 수 있다(모수추정값 등의 표 생략). 또한, μ, σ의 s.e들은 r이 증가할수록 증가하는 반면, k의 s.e는 감소하는 것을 알 수 있다. s.e들의 합계를 보면, r이 증가하면 할수록 전체적으로 감소하는 경향을 나타내고 있다. 이는 μ, σ의 s.e는 작은 값을 갖는 반면, 우리가 관심을 갖고 있는 꼬리 영역을 나타내는 k의 s.e가 큰 값을 갖기 때문에, 전체적으로 s.e들의 합계가 줄어듬을 알 수 있다. r과 r-1의 지점에서 s.e합계가 안정적인 지점(stable point)이 되는 r을 선택하였다. MLE방법과 P-MLE방법을 비교하면, 전체적인 50%이상 모수와 모수들의 s.e의 값이 낮아 졌음을

알 수 있고, 모수 k의 값은 MLE로 추정되었을 경우 0.5보다 큰 반면에 P-MLE의 방법을 사용하였을 경우 0.5보다 작음을 알 수 있다. 다시 말하면, MLE를 사용하였을 경우 MLE의 점근성을 보장할 수 없지만, P-MLE를 사용하였을 경우 MLE의 점근성을 보장 할 수 있다.

<표 1> X_p(1-p Quantile)과 신뢰구간이다. MLE의 방법으로 계산된 X_p의 경우 r과 개월이 높아갈수록 높아가고 있고, P-MLE의 방법도 높아가고 있다. 그리고, 신뢰구간은 r과 개월이 증가할수록 감소한다.

<표 1> X_p의 추정치

		4개월		12개월	
		X_p	±95% C.I	X_p	±95% C.I
r-GEV	1순위	1.0545	*	1.0920	*
	2순위	1.0725	*	1.1087	*
	3순위	1.0970	*	1.1322	*
	4순위	1.1253	*	1.1538	*
Penalized - MLE	1순위	1.0425	0.1244	1.0655	0.1038
	2순위	1.0519	0.0941	1.0787	0.0898
	3순위	1.0572	0.0688	1.0904	0.0723
	4순위	1.0624	0.0553	1.0998	0.0597

<표 2>는 각 개월에 따른 일일 최대/최소 변동률에 따른 r-GEV분포에 MLE /P-MLE방법을 이용하여 추정된 모수 μ, σ, k를 이용하여 아래에 주어진 식에 대입하여 계산된 값이다. Return Value ($R_{n개월, k}$)는

$$R_{n,m} = F^{-1}_{\mu,\sigma,k}(1-1/m) = \mu + \frac{\sigma}{k}(1-[-\ln(1-1/m)]^{k})$$

Return Level($R_{n개월, k}$)이란 quantile의 대안적인 방법으로, n개월을 m개의 구간 내(n*m월)에 나타날 수 있는 최대 일일 증가비를 의미한다. Return Level의 하나의 예를 들면, 'IMF시기에 나타난 일일 최대 증가율이 몇 년에 한 번 나타날수 있는가'와 같은 질문의 답은 '몇년에 한번 IMF가 나타난다'를 의미한다. MLE의 방법으로 추정된 R-Value는 큰 차이를 갖고 있기 때문에 옳은 방법이라고 말할 수 없지만, P-MLE의 방법은 근사적으로 같은 값을 갖고 있기 때문에 r-GEV분포에 적용시킬 때에 MLE방법보다는 P-MLE를 적용시키는 것이 옳다고 본다.

실제로 IMF시기 내에서 1997년 12월 22일에 나타난 최대 증가량인 1.146은 몇 년에 한번 나타날 최대 증가량인가를 실제 대입한 결과 r-GEV/MLE(12개월, 4th를 이용)는 18년 만에 한번씩 나타날 수 있는 값이고, r-GEV/P-MLE(12개월, 4th를 이용)는 48년에 한번씩 나타날 수 있는 값이다. IMF는 일반인들이 말하기를 100년에 한번 나타나는 국가 경제 위기라고 한다. P-MLE방법으로 계산된 R-Value의 해석상의 의미는 'IMF라는 경제위기는 40년 ~ 48년 사이에 발생한다'고 해석 할 수 있다. 그러나, 여기서 주의해야 할 점은 본 논문에서 사용한 데이터는 IMF라는 경제위기를 포함한 자료이기 때문에 자료의 변동이 심하고 극치의 값들이 높다고 할 수 있다. 그러므로, P-MLE

방법으로 계산된 R-Value의 값은 낮게 나왔다고 볼 수 있다.

결론은 한국 환율 데이터에 MLE의 방법으로 r-GEV분포에 적합하는 것보다는 P-MLE방법으로 분포에 적합하는 것이 적절하다고 말할 수 있다.

<표 2> $R_{n,k}$의 추정치

		4개월		12개월	
		10년	20년	10년	20년
		$R_{4개월,30}$	$R_{4개월,60}$	$R_{12개월,10}$	$R_{12개월,20}$
r-GEV /MLE	1순위	1.0694	1.1034	1.0592	1.0920
	2순위	1.0930	1.1408	1.0732	1.1087
	3순위	1.1286	1.2062	1.0877	1.1322
	4순위	1.1699	1.2836	1.1008	1.1538
r-GEV /P-MLE	1순위	1.0511	1.0689	1.0480	1.0655
	2순위	1.0629	1.0861	1.0580	1.0787
	3순위	1.0699	1.0974	1.0660	1.0904
	4순위	1.0766	1.1076	1.0723	1.0998

4. 결론 및 제언

본 논문에서는 위험측정에 대한 기존의 표준모형보다는 대안적 방법으로서 고차 일반화극치분포모형과 이의 추정방법들을 이용하여, quantile과 Return Value로 우리나라의 엔/달러 외환시장을 대상으로 그 성과를 실증분석 하였다. r-GEV/P-MLE를 이용해서 한국 원/환율의 extreme의 경우를 분석 할 수 있었고, GEV < r-GEV/MLE < r-GEV/P-MLE순으로 모수와 quantile의 추정을 안정되게 할 수 있었다. 또한, GEV를 적용하기 위해서는 block간의 r개의 순서통계량 사이의 독립성이 가정되어야 하는데, Financial time series인 관계로 이 가정을 만족시키기가 좀 어렵다. 따라서 이러한 독립성 가정이 필요 없는 일반화 파레토분포(GPD)를 이용한 접근법이 더 바람직 할 것으로 생각한다. r을 체계적으로 선택할 기존의 방법(wang, 1996)을 실제로 적용하지 못하였다. 현재 본 연구자들은 GPD를 이용하여 어느 값 이상을 넘는 극단자료들에 대해 적합시키고 특히 VaR(value-at-risk)를 구하는 연구를 진행중이다.

참고문헌

[1] 한상범, 극치분포(Extreme Value Distribution)를 이용한 VaR(Value at Risk)의 추정 및 성과분석, mimeo, *한국증권연구원*, 1999.

[2] Martins. E. S., Stedinger. J. R., Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resources Research*, Vol 36, No. 3, pp. 737 ~ 744, 2000.

[3] Twan. J. A., An Extreme Value Theory Model for Dependent Observations , *Journal of Hydrology*, Vol 101, pp. 227 ~ 250, 1988.