한 국 통 계 학 회 논 문 집
2009, 16권, 3호, 463–477

극단치 분포의 모수 추정방법 비교 연구(회귀 분석법을 기준으로)

우지용[a], 김명석[1,b]

[a]한국 채권평가, [b]서강대학교 경영학과

요 약

극단치 분포의 모수 추정방법으로 최우추정법, 확률가중적률법, 회귀분석법은 기존 연구에서 활발하게 적용되어져 왔다. 그러나 이들 세 가지 추정방법 가운데, 회귀분석법의 우수성은 엄격하게 평가되어진 적이 없다. 본 논문에서는 몬테칼로 시뮬레이션을 통하여 Generalized Extreme Value(GEV) 분포와 Generalized Pareto(GP) 분포의 모수 추정에 회귀분석법 및 다른 추정방법을 적용하여 비교 연구한다. 시뮬레이션 결과, 표본의 크기가 작은 경우 회귀분석법은 GEV 분포의 위치모수 추정시 편의 측면과 효율성 측면에서 다른 방법보다 우수한 경향을 나타내었다. GP 분포의 규모모수 추정시에는 표본의 크기가 작을 경우 회귀분석법이 다른 방법보다 작은 편의를 나타내었다. 회귀분석법은 표본의 크기가 작거나 적당히 큰 경우에도 GEV 분포나 GP 분포의 형태모수 추정시에 형태모수의 값이 −0.4일 경우, 다른 방법보다 우수한 경향을 나타내었다.

주요용어: 몬테칼로 시뮬레이션, 최우추정법, 확률가중적률법, 회귀분석법, Generalized Extreme Value 분포, Generalized Pareto 분포.

1. 서론

과거에 극단치 분포는 매년 발생하는 폭풍으로 인한 홍수의 연간 최대치 등의 자연현상을 모델링하는데 주로 사용되어져 왔다. 최근에는 경영 경제 자료의 극단치 연구에 대한 관심 고조와 더불어 보다 다양한 분야에서 활발하게 이용되어져 왔다. 극단치 분포의 인기와 더불어 극단치 분포의 모수를 추정하는 방법에 대한 관심도도 높아졌다. 많은 문헌 연구에서 다양한 극단치 분포에 대하여 여러 가지 모수 추정법이 이용되어져 왔다. Longin (1996)은 미국 주식 시장의 주어진 일정기간의 극단치 수익률을 Gumbel 분포와 Frechet 분포를 적용하여 분석하였다. 그들은 각 분포의 모수 추정 및 모델적합성 검정에 최우추정법과 회귀분석법을 이용하였다. Longin (2000)은 미국 주가 지수의 극단치 자료를 이용한 Value at Risk(VaR) 계산에 Generalized Extreme Value(GEV) 분포를 적용하였고, GEV 분포의 모수를 최우추정법을 이용하여 추정하였다. Bali (2003)는 미국 국채 수익률 자료를 이용한 VaR 계산에 다양한 극단치 분포를 적용하였다. GEV 분포, Gumbel 분포, Generalized Pareto(GP) 분포 및 Exponential 분포의 모수 추정시 최우추정법과 회귀분석법을 이용하였다. Gettinby 등 (2006)은 미국, 영국, 일본의 주식시장의 일정 기간 단위별 최소 수익률에 대한 Gumbel 분포, Weibull 분포, Frechet 분포, GEV 분포, GP 분포 및 Generalized Logistic(GLO) 분포의 모델적합성을 확률가중적률법을 이용하여 연구하였다. 여러 논문을 통하여 극단치 분포의 모수를 추정하는데 주로 많이 사용된 방법은 최우추정법, 확률가중적률법, 회귀분석법으로 요약된다.

이러한 방법들이 실제 자료에 응용되기 위해서는 시뮬레이션 연구를 통하여 각 방법의 정확성 및 장단점이 우선적으로 검증되어야 할 것이다. 최우추정법과 확률가중적률법의 정확성은 많은 논문에

이 연구는 2008년도 서강대학교 교내연구비 지원에 의한 연구임(과제번호: 200810048.01).
이 논문은 첫 번째 저자의 서강대학교 석사학위 논문을 바탕으로 작성되었음.
[1] 교신저자: (121-742) 서울시 마포구 신수동 1번지, 서강대학교 경영학과, 조교수.
E-mail: myungkim@sogang.ac.kr

서 연구되었다. Hosking 등 (1985)는 다양한 경우의 시뮬레이션을 통하여 GEV 분포의 모수추정시 확률가중적률법을 최우추정법과 비교 연구하였다. 그들에 의하면, 소표본 ($n = 15, 25$)의 경우 확률가중적률법에 의한 root mean square errors(RMSE)가 최우추정법에 의한 것보다 종종 더 작은 값을 기록한다. Hosking과 Wallis (1987)는 GP 분포의 모수 추정에 있어서 시뮬레이션을 통하여 확률가중적률법과 최우추정법 및 적률법을 비교하였다. 비교 결과 표본의 수가 500 이하일 경우 확률가중적률법이 다른 방법보다 불편의 및 효율성 측면에서 많은 경우 우수함을 입증하였다. Haktanir과 Bozduman (1995)은 GEV 분포를 포함한 다양한 분포의 모수 추정에 있어서 확률가중적률법을 이용할 경우 불편의 방법 및 여러 가지 plotting-position 방법에 따른 모수추정의 정확성을 시뮬레이션을 이용하여 비교 연구하였다. 그들은 box plot을 이용하여 GEV 분포의 모수추정시 Landwehr 등 (1979a)이 제시한 불편의 방법이 plotting-position을 이용한 다른 방법들 ($(i = 0.35)/n$, Weibull, Cunane) 보다 근소한 차이로 우수함을 보였다. Singh와 Ahmad (2004)는 시뮬레이션을 이용하여 GP 분포의 모수 추정에 적률법, 확률가중적률법, 최우추정법, 엔트로피추정법을 비교했다. 그들에 의하면, 표본의 크기와 모수의 종류 및 범위에 따라서 모수추정법의 우월이 불규칙하게 나타난다. 단, 확률가중적률법의 경우 정확성이 가장 우수하지는 않지만, 어느 정도 규칙성을 갖는 정확성을 보이므로 이 방법의 사용을 권고했다. Kim (2007)은 극단치 분포의 일종인 GLO 분포의 모수 추정시 불편의 방법 및 다양한 plotting-position에 근거한 확률가중적률법과 최우추정법을 시뮬레이션을 이용하여 비교 연구하였다. 그에 의하면 대체로 최우추정법이 확률가중적률법보다 우수하나 소표본의 경우 ($n = 20$) 형태모수가 0에 근사할 경우에는 형태모수 추정시 확률가중적률법이 보다 우수하였다. 대체로 불편의 방법이 다른 방법들보다 위치모수와 규모모수 추정시 더 적은 편의 값을 제공하였고, $(i - 0.35)/n$은 소표본하의 특정범위의 형태모수 추정시 효율성 측면에서 우수함을 보였다.

한편, 회귀분석법은 위에서 언급한 바와 같이 기존 연구에서 중요하게 응용되는데도 불구하고 시뮬레이션을 통하여 이 방법의 정확성을 검증한 논문은 찾아 볼 수가 없었다. 따라서, 본 연구에서는 몬테칼로로 시뮬레이션을 이용하여 회귀분석법을 이용한 GEV 분포와 GP 분포의 모수 추정의 정확도를 최우추정법 및 확률가중적률법과 비교하여 살펴보았다. 표본의 수는 15에서 500에 이르기까지 소표본에서 대표본을 골고루 고려하였다. 두 개의 분포 및 여러 종류의 표본크기에 대하여 세 가지 추정방법을 이용하여 모수를 추정하였다. 1000번을 반복하여 각 추정방법의 정확성을 불편의 정도와 RMSE를 이용하여 살펴보았다. 시뮬레이션 연구결과 GEV 분포의 위치모수 추정의 경우 소표본에서 회귀분석법이 불편의와 RMSE 측면에서 모두 우수한 경향이 나타났다. GEV 분포의 규모모수의 경우, RMSE 측면에서는 최우추정법이 가장 우수하였다. GP 분포의 규모모수의 경우 불편의 측면에서 소표본일 경우 회귀분석법이 우수하였다. GEV 분포와 GP 분포의 형태모수 추정의 경우, RMSE 측면에서 소표본에서는 확률가중적률법이 대체로 가장 우수하였고, 대표본일수록 최우추정법이 우수한 경향을 보였다. 불편의 측면에서는 위치모수 값이 -0.4인 경우, 회귀분석법이 소표본 및 일부 대표본에서 우수한 경향이 나타났다.

본 논문의 나머지 부분은 아래와 같이 구성되어있다. 제 2장에서는 분석에 사용할 대표적인 극단치 분포인 GEV 분포와 GP 분포의 확률함수 및 추정할 모수에 대하여 설명한다. 본 논문에서 고려할 세 가지 모수 추정방법인 최우추정법, 확률가중적률법, 회귀분석법을 제 3장에서 소개한다. 다양한 표본크기와 모수범위를 이용하여 얻어진 모수추정의 시뮬레이션 결과가 제 4장에서 논의되며, 마지막 결론이 제 5장에서 언급된다.

2. 극단치 분포

극단치 분포는 최대값 또는 최소값을 자료로 이용하여 만든 분포로 크게 두 가지 접근방법에 의하여 구분된다. 일정 구간에서의 극단치를 자료로 이용하는 방법과 특정값 이상 또는 이하의 극단치를

이용하는 방법이 있다. 예를 들자면, 첫 번째 경우, 수년치의 일별 주식 수익률을 기본 자료로 하되 월별 최대 또는 최소 주식 수익률을 자료로 하여 그 분포를 만드는 경우이다. 이 경우 일정구간으로서 월이 사용된다. 두 번째 경우는 역시 일별 주식 수익률을 이용하되 수익률이 10% 이상 (최대값)이거나 −15% 이하 (최소값)인 주식 수익률을 자료로 이용하여 모델하는 경우이다. 첫 번째 경우에 해당되는 대표적인 모델로 GEV 분포와 GLO 분포가 있다. GEV 분포는 Frechet 분포, Gumbel 분포, Weibull 분포를 포함하는 보다 일반화된 모델로 알려져 있다. GP 분포는 두 번째 경우에 적용하는 모델로 Pareto 분포, Exponential 분포, Uniform 분포를 포함한다. GEV 분포 및 GP 분포의 누적분포함수 및 확률밀도함수에 관한 설명은 아래와 같다.

2.1. GEV 분포

일정구간의 최대값을 이용한 GEV 분포는 Jenkinson (1955)에 의해 제안되었다. 최대값을 자료로 이용한 GEV 분포의 누적분포함수는 다음과 같다.

$$F(x) = \begin{cases} \exp\left[-\left\{1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right], & \xi \neq 0, \\ \exp\left[\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], & \xi = 0. \end{cases} \tag{2.1}$$

이때, x의 범위는 다음과 같다.

$$\begin{cases} -\infty < x \leq \mu - \dfrac{\sigma}{\xi}, & \xi < 0, \\ -\infty < x < \infty, & \xi = 0, \\ \mu - \dfrac{\sigma}{\xi} \leq x < \infty, & \xi > 0, \end{cases} \tag{2.2}$$

여기서, ξ는 형태모수, μ는 위치모수, σ는 규모모수를 나타낸다. Jenkinson (1955)에 의하면, 형태모수 ξ가 양수면, Frechet 분포, 음수면 Weibull 분포, 영이면 Gumbel 분포을 얻을 수 있다. 세 개의 분포는 극단치를 자료로 하는 분포로 Gnedenko (1943)에 의해 제시되었다. GEV 분포는 위의 세 개의 분포를 모두 포함하여 하나의 분포로 일반화하여 표현한 분포이다. 형태모수 ξ는 tail index라고도 불리는데, 기본 자료 분포의 꼬리 부분의 가중치를 반영한다. 일반적으로 GEV 분포의 형태모수의 범위는 −0.5에서 0.5 사이의 값이 이론적으로나 경험적으로 적절하다고 알려져 있다 (Hosking 등, 1985). 최대값을 이용한 GEV 분포의 확률밀도함수는 위의 누적분포함수를 미분하여 다음과 같이 구할 수 있다.

$$f(x) = \begin{cases} \dfrac{1}{\sigma}\left\{1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}-1} \exp\left[-\left\{1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right], & \xi \neq 0, \\ \exp\left(-\dfrac{x-\mu}{\sigma}\right)\exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], & \xi = 0. \end{cases} \tag{2.3}$$

GEV 분포의 분위수(quantile)함수는 누적분포함수 식 (2.1)의 역함수로 다음과 같다.

$$x(F) = \begin{cases} \mu + \dfrac{\sigma}{\xi}\left[\{-\log(F)\}^{-\xi} - 1\right], & \xi \neq 0, \\ \mu - \sigma\log(-\log F), & \xi = 0. \end{cases} \tag{2.4}$$

식 (2.4)를 이용하여 GEV 분포의 랜덤표본을 생성할 수 있다.

2.2. GP 분포

특정값 이상의 극단치를 이용한 GP 분포는 Pickands (1975)에 의하여 소개되었다. Hosking과 Wallis (1987)에 의하면, GP 분포의 누적분포함수는 다음과 같다.

$$F(x) = \begin{cases} 1 - \left(1 + \dfrac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x}{\sigma}\right), & \xi = 0. \end{cases} \tag{2.5}$$

이때, x의 범위는 다음과 같다.

$$\begin{cases} 0 \leq x < -\dfrac{\sigma}{\xi}, & \xi < 0, \\ 0 \leq x < \infty, & \xi \geq 0, \end{cases} \tag{2.6}$$

여기서, ξ는 형태모수, σ는 규모모수를 나타낸다. 일반적으로 형태모수의 적절한 범위는 −0.5에서 0.5 사이의 값으로 알려져 있다 (Hosking과 Wallis, 1987). 형태모수 ξ가 양수면, Pareto 분포, 음수면 Uniform 분포, 영이면 Exponential 분포를 얻을 수 있다. GP 분포는 세 개 분포를 하나의 분포로 일반화한 형태이다. Hosking과 Wallis (1987)는 확률변수 X가 GP 분포를 따르면 X가 특정값 t보다 클 때, 확률변수 $X - t$도 같은 형태모수 값을 갖는 GP 분포를 따른다. GP 분포의 확률밀도함수는 GP 분포를 미분하여 다음과 같이 얻어진다.

$$f(x) = \begin{cases} \dfrac{1}{\sigma}\left(1 + \dfrac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0, \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{x}{\sigma}\right), & \xi = 0. \end{cases} \tag{2.7}$$

GP 분포의 분위수(quantile) 함수는 식 (2.5)의 역함수로 다음과 같이 구할 수 있다.

$$x(F) = \begin{cases} \dfrac{\sigma}{\xi}\left\{(1 - F)^{-\xi} - 1\right\}, & \xi \neq 0, \\ -\sigma \log(1 - F), & \xi = 0. \end{cases} \tag{2.8}$$

식 (2.8)를 이용하여 GP 분포의 랜덤표본을 생성할 수 있다.

3. 모수추정법

이 장에서는 최우추정법, 확률가중적률법 및 회귀분석법의 내용을 정리한다. 최우추정법에 의한 모수 추정치는 일반적으로 불편의 추정치이며 근사적으로 최소 분산값을 갖는 정규분포를 따르는 등 좋은 성질을 가지고 있다고 알려져 있다. 표본자료의 크기가 충분히 클 때 가장 효율적인 추정치를 얻을 수 있으며 다른 모수 추정방법에 대하여 추정치의 효율성을 비교하는데 기준으로 사용된다. 그러나, 표본의 수가 작거나 모수가 특정 범위에 있을 경우, 확률가중적률법이 특정 극단치 분포의 모수 추정시에는 최우추정법보다 우수하다는 연구 결과가 있다. 앞서 언급한 바와 같이 회귀분석법은 실제 경영 경제 자료에 다양하게 응용되어지고 있음에도 불구하고 이 방법의 정확성에 관한 연구가 저자가 알기로는 없다.

3.1. 최우추정법

최우추정치는 로그우도함수를 최대화시키는 해로서 확률밀도함수가 $f(x)$일 때 로그우도함수는 다음과 같이 얻어진다.

$$\ln L = \sum_{i=1}^{n} f(x_i)$$

GEV 분포의 로그우도함수는 식 (2.3)에 의해 다음과 같다.

$$\ln L = -n \ln \sigma - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^{n} \ln \left\{1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right\} - \sum_{i=1}^{n} \left\{1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}$$

Jenkinson (1969, 제 5장)에 따르면 최우추정치는 반복 알고리듬을 이용하여 주어진 로그우도함수를 최대화시켜서 얻어진다. Prescott와 Walden (1980)은 최우추정법 계산시 조건 $\xi > -0.5$의 필요성을 주장하였다.

GP 분포의 로그우도함수는 식 (2.7)에 의해 다음과 같다.

$$\ln L = -n \ln \sigma - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^{n} \ln \left(1 + \frac{\xi x_i}{\sigma}\right).$$

Hosking과 Wallis (1987)에 의하면 $\xi < -0.5$일 경우 GP 분포의 분산은 무한대가 된다. GP 분포 모수의 최우추정법은 Davison (1984)와 Smith (1985)에 의하여 심화 연구되었다.

최우추정법은 확률가중적률법과 달리 비선형 방정식으로 표현되므로 모수 추정치를 얻을 수 없는 경우가 많다. 따라서 해를 얻기 위하여 Newton-Raphson법 또는 비선형 최적화 알고리즘과 같은 수치기법이 필요하며, 초기치에 따라 해를 구하지 못하는 어려움이 따르기도 한다. 또한 표본의 크기가 작을 때 일반적으로 잘 추정되어지지 않는다는 단점이 있다.

3.2. 확률가중적률법

$F(x) = P(X \le x)$를 분포함수로 가지는 확률변수 X의 확률가중적률은 다음과 같다 (Greenwood 등, 1979; Hosking, 1990).

$$M_{p,r,s} + E\left[X^p \{F(X)\}^r \{1 - F(X)\}^s\right].$$

GEV 분포와 GP 분포의 모수 추정시에는 다음의 확률가중적률이 이용되어진다.

$$\alpha_s = M_{1,0,s}$$
$$\beta_r = M_{1,r,0,}$$

이때 α_s와 β_r를 추정하는 방법은 불편의 방법과 plotting-position 방법이 있다. α_s와 β_r의 불편의 추정량은 각각 $\hat{\alpha}_s$와 $\hat{\beta}_r$로 다음과 같다 (Landwehr 등, 1979a).

$$\hat{\alpha}_s = \frac{1}{n} \sum_{j=1}^{n} \frac{(n-j)(n-j-1)\cdots(n-j-s+1)}{(n-1)(n-2)\cdots(n-s)} x_j$$

$$\hat{\beta}_r = \frac{1}{n} \sum_{j=1}^{n} \frac{(j-1)(j-2)\cdots(j-r)}{(n-1)(n-2)\cdots(n-r)} x_j$$

여기서 x_j는 $x_1 \le x_2 \le \cdots \le x_n$인 순서통계량 표본이다. α_s와 β_r의 plotting-position 추정량은 각각 $\hat{\alpha}_s(p_{j,n})$와 $\hat{\beta}_r(p_{j,n})$로 다음과 같다 (Landwehr 등, 1979a).

$$\hat{\alpha}_s(p_{j,n}) = \frac{1}{n}\sum_{j=1}^{n}(1-p_{j,n})^s x_j$$

$$\hat{\beta}_r(p_{j,n}) = \frac{1}{n}\sum_{j=1}^{n} p_{j,n}^r x_j$$

이때 $p_{j,n}$을 plotting-position이라 하고, 일반적으로 $p_{j,n} = (j+\gamma)/(n+\delta)$의 형태를 따른다. γ와 δ는 상수로 Landwehr 등 (1979b)은 $\gamma = -0.35$, $\delta = 0.0$을 추천했다.

불편의 추정량 $\hat{\beta}_r$을 이용하여 GEV 분포의 모수를 다음의 관계식을 이용하여 추정할 수 있다.

$$\hat{\xi} = -7.8590c - 2.9554c^2, \quad c = \frac{2\hat{\beta}_1 - \hat{\beta}_0}{3\hat{\beta}_2 - \hat{\beta}_0} - \frac{\log 2}{\log 3}$$

$$\hat{\sigma} = \frac{\left(\hat{\beta}_0 - 2\hat{\beta}_1\right)\hat{\xi}}{\Gamma\left(1-\hat{\xi}\right)\left(1-2^{\hat{\xi}}\right)}$$

$$\hat{\mu} = \hat{\beta}_0 - \frac{\hat{\sigma}\left\{\Gamma\left(1-\hat{\xi}\right)-1\right\}}{\hat{\xi}}.$$

한편, 불편의 추정량 $\hat{\alpha}_s$을 이용하여 GP 분포의 모수를 아래의 관계식을 이용하여 추정할 수 있다 (Hosking과 Wallis 1987).

$$\hat{\xi} = 2 - \frac{\hat{\alpha}_0}{\hat{\alpha}_0 - 2\hat{\alpha}_1}$$

$$\hat{\sigma} = \frac{2\hat{\alpha}_0\hat{\alpha}_1}{\hat{\alpha}_0 - 2\hat{\alpha}_1}$$

Plotting-position 추정량을 이용할 경우, $\hat{\alpha}_s$와 $\hat{\beta}_r$의 위치에 각각 $\hat{\alpha}_s(p_{j,n})$와 $\hat{\beta}_r(p_{j,n})$를 대신 사용한다.

3.3. 회귀분석법

회귀분석법은 Gumbel (1958)에 의하여 제시되었다. 회귀분석법은 순서통계량을 이용한다. x_j가 $x_1 \le x_2 \le \cdots \le x_n$인 순서통계량 표본일 때 j번째 순서 통계량의 확률밀도함수는 다음과 같다.

$$f_{x_j}(x) = \frac{x^{j-1}(1-x)^{n-j}}{B(j,\, n-j+1)}; \quad 0 \le x \le 1,$$

여기서, $B(j,\, n-j+1)$는 베타함수이다. 이때 x_j의 모평균은 다음과 같이 구해진다.

$$\mu_j = \int_0^1 \frac{x^j(1-x)^{n-j}}{B(j,\, n-j+1)}dx = \frac{j}{n+1}.$$

위의 내용을 최대값들의 순서 통계량에 적용하여 극단치 분포 모수 추정에 이용할 수 있다. 최대값들 중 j번째 순서 통계량의 누적확률함수의 값은 0과 1사이에 놓인 하나의 확률변수가 된다. 회귀분석방법은 이러한 순서화된 누적확률함수의 기대값을 이론적 평균치에 일치시킴으로써 얻어진다.

예를 들어서, $\xi \neq 0$인 경우를 살펴보자. GEV 분포의 경우 식 (2.1)을 이용하여 최대값들 중 j번째 순서 통계량, $X_{\max,j}$의 누적확률함수의 기대값을 다음과 같이 이의 이론적 평균치에 일치시킨다.

$$E\left[\exp\left[-\left\{1+\xi\left(\frac{X_{\max,j}-\mu}{\sigma}\right)\right\}^{-\frac{1}{\xi}}\right]\right]=\frac{j}{n+1}. \tag{3.1}$$

GP 분포는 식 (2.5)에 의해 다음 식을 이용하여 모수를 추정한다.

$$E\left\{1-\left(1+\frac{\xi X_{\max,j}}{\sigma}\right)^{-\frac{1}{\xi}}\right\}=\frac{j}{n+1}. \tag{3.2}$$

위의 두 식은 양쪽에 로그함수를 두 번 취하고 에러항을 추가함으로 추정 가능한 함수의 형태가 된다. 각 모델의 모수 추정은 비선형 회귀 방정식의 잔차들의 제곱합을 최소화시킴으로써 얻어진다.

4. 시뮬레이션 결과

세 가지 모수 추정방법을 비교하기 위해서 컴퓨터 시뮬레이션을 실시하였다. 시뮬레이션 표본크기 n을 15, 25, 50, 100, 200, 500로 한다. GEV 분포의 위치모수 μ 추정시에는 위치모수 값을 −10, −5, −1, −0.2, 0.2, 1, 5, 10으로 한다. 이때 형태모수 ξ와 규모모수 σ는 편의상 모두 0.2로 고정시킨다. 각각의 표본크기 n과 형태모수 μ의 조합에 대해 GEV 분포의 분위수 함수 식 (2.4)를 이용하여 Inverse Transformation Method 방식으로 식 (2.2) 범위의 랜덤표본을 생성할 수 있다. 각각의 표본에 대해 최우추정법, 확률가중적률법, 회귀분석법을 이용하여 모수를 추정한다. 확률가중적률법의 경우 Hosking 등 (1985)와 Hosking과 Wallis (1987)를 따라 Landwehr 등 (1979b)가 제시한 $\gamma = -0.35$, $\delta = 0.0$인 plotting-position에 근거한 확률가중적률법을 이용했다. GEV 분포의 경우, Haktanir과 Bozduman (1995)에 의하면 불편의 방법이 plotting-position $(i-0.35)/n$보다 근소하게 우수하다고 하였으나, box plot 상의 그 차이가 매우 미미하여 본 연구에서는 큰 무리 없이 $(i-0.35)/n$을 이용한다. 이와 같은 방법을 1000번 시행하여 모수 추정치의 RMSE와 편의정도를 계산하여 모수추정의 정확성을 평가할 수 있다. 비슷한 방법으로 GEV 분포와 GP 분포의 규모모수 σ 추정시에는 규모모수 값을 0.2, 0.4, 0.6, 1, 3, 5로 정한다. 이때 GEV 분포의 경우 위치모수 μ는 0으로 고정하고, 형태모수 ξ는 GEV 분포와 GP 분포 모두 0.2로 고정시킨다. GP 분포에 의한 랜덤표본 생성시에는 함수 식 (2.8)을 이용하여 식 (2.6) 범위의 랜덤표본을 생성할 수 있다. 형태모수 ξ 추정시 형태모수 값은 −0.4, −0.2, 0, 0.2, 0.4로 한다. 이때 GEV 분포의 위치모수 μ를 0, 규모모수 σ를 0.2로 고정하고, GP 분포의 규모모수 σ를 0.2로 고정한다. 위에서 사용된 모수의 범위는 기존 연구에서 고려되어졌던 범위이다. GEV 분포의 위치모수와 규모모수 및 GP 분포의 규모모수의 범위는 Kim (2007)의 GLO 분포의 위치모수 및 규모모수 추정을 위해 사용한 범위 값을 적용한 것이다. 언급하였던 바와 같이 GLO 분포도 극단치 분포의 일종이며, 세 개의 모수(위치모수, 규모모수, 형태모수)를 갖고 있다. 두 분포의 형태모수의 범위는 Hosking과 Wallis (1987)와 Hosking 등 (1985)의 논문에서 사용된 형태모수의 범위를 적용한 것이다. 이들에 의하면 이론적으로나 경험적으로 GEV 분포나 GP 분포의 형태모수는 −0.5에서 0.5 사이에 존재한다.

시뮬레이션 결과는 아래의 표에 잘 정리되어져 있다. 표 1부터 3에서는 GEV 분포의 위치모수, 규모모수, 형태모수에 관한 세 가지 추정법의 정확성 정도를 비교하였다. 세 가지 추정법인 최우추정법, 확률가중적률법, 회귀분석법은 각기 ML, PWM, Reg로 표시하였고, 이들 방법의 정확성은 편의정도와 RMSE (괄호안)를 이용하여 평가하였다. 표 1은 GEV 분포의 위치모수 추정에 관한 것으로, n = 15, 25와 같은 소표본에서 회귀분석법이 불편의 정도에서 우수한 경향을 보였다. 표본의 크기가 큰 경우에는 최우추정법이 전반적으로 우수한 경향을 나타내었다. RMSE 측면에서 볼 경우, 역시 소

표 1: GEV 분포 위치모수(μ)추정의 편의와 RMSE(괄호안); ML은 최우추정법, PWM은 plotting-position 확률가중적률법, Reg는 회귀분석법을 나타낸다. 위치모수 μ를 -10부터 10로 하고 규모모수 σ를 0.2, 형태모수 ξ를 0.2로 한다. 표본크기를 15, 25, 50, 100, 200, 500으로 하고, 반복은 1000으로 한다.

n	method	μ							
		−10.0	−5.0	−1.0	−0.2	0.2	1.0	5.0	10.0
15	ML	0.0082	0.0060	0.0119	0.0083	0.0080	0.0059	0.0045	0.0054
		(0.0655)	(0.0640)	(0.0637)	(0.0644)	(0.0659)	(0.0626)	(0.0617)	(0.0617)
	PWM	0.1369	0.0967	0.0283	0.0128	0.0079	−0.0042	−0.0540	−0.1078
		(0.1621)	(0.1109)	(0.0638)	(0.0596)	(0.0599)	(0.0585)	(0.0799)	(0.1234)
	Reg	−0.0029	−0.0040	0.0007	−0.0033	−0.0029	−0.0048	−0.0056	−0.0049
		(0.0617)	(0.0607)	(0.0595)	(0.0598)	(0.0613)	(0.0594)	(0.0586)	(0.0586)
25	ML	0.0061	0.0038	0.0026	0.0060	0.0050	0.0031	0.0029	0.0031
		(0.0490)	(0.0468)	(0.0481)	(0.0497)	(0.0476)	(0.0484)	(0.0473)	(0.0468)
	PWM	0.1168	0.0523	0.0136	0.0099	0.0057	−0.0031	−0.0349	−0.0717
		(0.1250)	(0.0674)	(0.0478)	(0.0481)	(0.0453)	(0.0467)	(0.0580)	(0.0857)
	Reg	−0.0033	−0.0052	−0.0060	−0.0026	−0.0036	−0.0062	−0.0061	−0.0059
		(0.0479)	(0.0461)	(0.0474)	(0.0486)	(0.0465)	(0.0475)	(0.0467)	(0.0462)
50	ML	0.0012	0.0000	0.0023	0.0014	0.0010	0.0024	0.0009	0.0023
		(0.0312)	(0.0349)	(0.0337)	(0.0330)	(0.0331)	(0.0330)	(0.0337)	(0.0327)
	PWM	0.0500	0.0239	0.0081	0.0039	0.0016	−0.0005	−0.0188	−0.0373
		(0.0579)	(0.0414)	(0.0341)	(0.0327)	(0.0326)	(0.0324)	(0.0388)	(0.0501)
	Reg	−0.0055	−0.0066	−0.0041	−0.0048	−0.0057	−0.0041	−0.0062	−0.0043
		(0.0318)	(0.0356)	(0.0341)	(0.0336)	(0.0333)	(0.0332)	(0.0346)	(0.0330)
100	ML	0.0005	0.0010	0.0004	0.0014	0.0014	0.0007	0.0017	0.0010
		(0.0226)	(0.0226)	(0.0240)	(0.0236)	(0.0233)	(0.0227)	(0.0229)	(0.0234)
	PWM	0.0241	0.0128	0.0033	0.0025	0.0019	−0.0005	−0.0084	−0.0196
		(0.0328)	(0.0259)	(0.0245)	(0.0238)	(0.0232)	(0.0227)	(0.0245)	(0.0307)
	Reg	−0.0042	−0.0038	−0.0047	−0.0037	−0.0035	−0.0043	−0.0034	−0.0041
		(0.0235)	(0.0235)	(0.0253)	(0.0245)	(0.0237)	(0.0234)	(0.0236)	(0.0242)
200	ML	0.0002	−0.0004	0.0009	0.0009	0.0006	−0.0001	0.0004	0.0010
		(0.0159)	(0.0156)	(0.0164)	(0.0160)	(0.0160)	(0.0158)	(0.0156)	(0.0158)
	PWM	0.0118	0.0057	0.0025	0.0015	0.0008	−0.0009	−0.0046	−0.0095
		(0.0198)	(0.0167)	(0.0166)	(0.0160)	(0.0160)	(0.0162)	(0.0166)	(0.0187)
	Reg	−0.0036	−0.0038	−0.0027	−0.0028	−0.0030	−0.0037	−0.0032	−0.0026
		(0.0169)	(0.0166)	(0.0170)	(0.0165)	(0.0166)	(0.0170)	(0.0165)	(0.0165)
500	ML	0.0004	0.0005	0.0002	0.0002	0.0002	0.0001	0.0001	0.0003
		(0.0101)	(0.0104)	(0.0102)	(0.0103)	(0.0098)	(0.0100)	(0.0100)	(0.0103)
	PWM	0.0050	0.0030	0.0008	0.0005	0.0003	−0.0003	−0.0020	−0.0039
		(0.0113)	(0.0108)	(0.0104)	(0.0105)	(0.0102)	(0.0101)	(0.0105)	(0.0112)
	Reg	−0.0019	−0.0015	−0.0020	−0.0019	−0.0019	−0.0023	−0.0021	−0.0018
		(0.0105)	(0.0108)	(0.0108)	(0.0109)	(0.0105)	(0.0107)	(0.0108)	(0.0109)

표본에서는 회귀분석법이 우수한 경향을 보이는 반면, 그 외의 표본에서는 최우추정법과 확률가중적률법이 대등하게 우수한 경향을 나타내었다. 불편의 측면에서 보았을 때, 전반적으로 회귀분석법의 경우 under estimate이고, 다른 두 추정 방법은 over estimate하는 경향을 나타내었다. 표 2는 GEV 분포의 규모모수 추정에 관한 것이다. 불편의 측면에서는 어떤 특정 추정 방법론이 우수하다는 결론을 찾기 힘들었다. 다만, 표본의 크기가 큰 경우에도 회귀분석법이 우수한 경향이 발견되기도 하였다. 전반적으로 최우추정법과 확률가중적률법은 일관되게 under estimate 경향을 나타내었다. RMSE 측면에서는 최우추정법이 전반적으로 우수한 경향을 나타내었다. 표 3은 GEV 분포의 형태모수 추정 결과가 보

표 2: GEV 분포 규모모수(σ)추정의 편의와 RMSE(괄호안); ML은 최우추정법, PWM은 plotting-position 확률가중적률법, Reg는 회귀분석법을 나타낸다 규모모수 σ를 0.2부터 5로 하고 위치모수 μ를 0, 형태모수 ξ를 0.2로 한다. 표본크기를 15, 25, 50, 100, 200, 500으로 하고, 반복은 1000으로 한다.

n	method	σ					
		0.2	0.4	0.6	1.0	3.0	5.0
15	ML	−0.0170 (0.0532)	−0.0266 (0.1032)	−0.0347 (0.1550)	−0.0784 (0.2619)	−0.2132 (0.8123)	−0.3048 (1.2786)
	PWM	−0.0170 (0.0523)	−0.0275 (0.1040)	−0.0335 (0.1548)	−0.0739 (0.2615)	−0.2282 (0.7958)	−0.3419 (1.2508)
	Reg	0.0195 (0.0662)	0.0469 (0.1350)	0.0828 (0.2131)	0.1109 (0.3303)	0.3300 (1.0433)	0.6116 (1.6485)
25	ML	−0.0093 (0.0390)	−0.0164 (0.0769)	−0.0261 (0.1128)	−0.0317 (0.1910)	−0.1147 (0.5703)	−0.1989 (0.9673)
	PWM	−0.0090 (0.0405)	−0.0162 (0.0803)	−0.0263 (0.1182)	−0.0300 (0.1974)	−0.1197 (0.5946)	−0.1984 (1.0169)
	Reg	0.0137 (0.0466)	0.0306 (0.0942)	0.0429 (0.1376)	0.0826 (0.2332)	0.2400 (0.7091)	0.3965 (1.2165)
50	ML	−0.0048 (0.0271)	−0.0108 (0.0519)	−0.0144 (0.0802)	−0.0245 (0.1274)	−0.0970 (0.4042)	−0.0931 (0.6518)
	PWM	−0.0048 (0.0291)	−0.0108 (0.0549)	−0.0150 (0.0868)	−0.0216 (0.1383)	−0.0868 (0.4371)	−0.0959 (0.6976)
	Reg	0.0064 (0.0309)	0.0125 (0.0600)	0.0188 (0.0904)	0.0369 (0.1495)	0.0826 (0.4646)	0.1989 (0.7708)
100	ML	−0.0025 (0.0174)	−0.0071 (0.0371)	−0.0070 (0.0548)	−0.0082 (0.0961)	−0.0426 (0.2798)	−0.0721 (0.4660)
	PWM	−0.0028 (0.0190)	−0.0063 (0.0411)	−0.0061 (0.0592)	−0.0078 (0.1044)	−0.0417 (0.3081)	−0.0748 (0.5089)
	Reg	0.0021 (0.0192)	0.0039 (0.0426)	0.0085 (0.0620)	0.0177 (0.1104)	0.0300 (0.3163)	0.0415 (0.5089)
200	ML	−0.0005 (0.0129)	−0.0022 (0.0265)	−0.0056 (0.0375)	−0.0050 (0.0646)	−0.0100 (0.1958)	−0.0372 (0.3175)
	PWM	−0.0004 (0.0139)	−0.0028 (0.0290)	−0.0060 (0.0420)	−0.0074 (0.0719)	−0.0102 (0.2140)	−0.0459 (0.3465)
	Reg	0.0010 (0.0143)	0.0004 (0.0298)	−0.0010 (0.0433)	0.0013 (0.0727)	0.0141 (0.2219)	−0.0053 (0.3558)
500	ML	−.0009 (0.0079)	−0.0007 (0.0155)	−0.0030 (0.0242)	−0.0026 (0.0409)	−0.0096 (0.1175)	−0.0091 (0.2104)
	PWM	−0.0013 (0.0088)	−0.0011 (0.0175)	−0.0028 (0.0267)	−0.0032 (0.0460)	−0.0176 (0.1327)	−0.0145 (0.2272)
	Reg	−0.0015 (0.0093)	−0.0014 (0.0186)	−0.0033 (0.0284)	−0.0043 (0.0483)	−0.0197 (0.1408)	−0.0207 (0.2338)

고되어 있다. 표본 크기 $n = 15, 25, 50, 100$에서 형태모수가 −0.4인 경우, 불편의 측면에서 회귀분석법이 우수한 결과를 보였다. 그 외의 표본에서는 최우추정법과 확률가중적률법이 대등하게 우수한 경향을 보였다. 확률가중적률법의 경우 모수가 양수일 경우 under estimate하는 경향이 나타났고, 음수일 경우에 over estimate하는 경향이 나타났다. RMSE 측면에서는 소표본의 경우, 확률가중적률법이 우수한 경향을 보였고, 대표본에서는 최우추정법이 우수한 경향을 나타내었다.

GP 분포의 결과는 표 4와 5에 정리되어 있다. 표 4는 GP 분포의 규모모수 추정 결과를 포함하고 있다. 표본의 크기가 $n = 15, 25$처럼 작을 경우, 회귀추정법이 불편의 측면에서 우수한 경향이 나타났

표 3: GEV 분포 형태모수(ξ)추정의 편의와 RMSE(괄호안), ML은 최우추정법, PWM은 plotting-position 확률가중적률법, Reg는 회귀분석법을 나타낸다. 형태모수 ξ를 −0.4부터 0.4로 하고 위치모수 μ를 0, 규모모수 σ를 0.2로 한다. 표본크기를 15, 25, 50, 100, 200, 500으로 하고, 반복은 1000으로 한다.

n	method	ξ				
		−0.4	−0.2	0.0	0.2	0.4
15	ML	−0.0976 (0.3244)	−0.0743 (0.3149)	−0.0401 (0.3171)	0.0137 (0.3420)	0.0289 (0.3435)
	PWM	0.1203 (0.2213)	0.0712 (0.1931)	0.0215 (0.1871)	−0.0199 (0.2005)	−0.1085 (0.2273)
	Reg	−0.0473 (0.3084)	−0.0201 (0.2951)	0.0216 (0.3246)	0.0741 (0.3736)	0.0949 (0.3817)
25	ML	−0.0541 (0.1927)	−0.0306 (0.1815)	−0.0171 (0.2010)	0.0027 (0.2211)	0.0237 (0.2381)
	PWM	0.0787 (0.1683)	0.0514 (0.1448)	0.0139 (0.1439)	−0.0258 (0.1595)	−0.0723 (0.1862)
	Reg	−0.0091 (0.1892)	0.0171 (0.1835)	0.0333 (0.2173)	0.0594 (0.2364)	0.0898 (0.2803)
50	ML	−0.0268 (0.1098)	−0.0181 (0.1088)	−0.0111 (0.1189)	−0.0065 (0.1290)	0.0052 (0.1459)
	PWM	0.0388 (0.1113)	0.0243 (0.1025)	0.0038 (0.1040)	−0.0151 (0.1198)	−0.0517 (0.1487)
	Reg	0.0082 (0.1178)	0.0187 (0.1203)	0.0312 (0.1374)	0.0475 (0.1583)	0.0599 (0.1867)
100	ML	−0.0209 (0.0692)	−0.0138 (0.0719)	−0.0029 (0.0796)	0.0009 (0.0890)	0.0011 (0.0959)
	PWM	0.0155 (0.0777)	0.0103 (0.0708)	0.0048 (0.0756)	−0.0058 (0.0937)	−0.0319 (0.1156)
	Reg	0.0050 (0.0740)	0.0133 (0.0816)	0.0315 (0.0988)	0.0420 (0.1143)	0.0433 (0.1297)
200	ML	−0.0087 (0.0419)	−0.0043 (0.0452)	−0.0046 (0.0522)	0.0010 (0.0583)	0.0037 (0.0686)
	PWM	0.0093 (0.0547)	0.0079 (0.0506)	−0.0004 (0.0540)	−0.0021 (0.0643)	−0.0168 (0.0878)
	Reg	0.0114 (0.0491)	0.0178 (0.0551)	0.0212 (0.0663)	0.0321 (0.0799)	0.0351 (0.0940)
500	ML	−0.0050 (0.0252)	−0.0031 (0.0278)	−0.0002 (0.0316)	0.0014 (0.0373)	−0.0004 (0.0408)
	PWM	0.0037 (0.0339)	0.0013 (0.0320)	0.0017 (0.0333)	−0.0001 (0.0426)	−0.0083 (0.0606)
	Reg	0.0052 (0.0339)	0.0121 (0.0357)	0.0165 (0.0429)	0.0198 (0.0520)	0.0206 (0.0586)

다. 그 외의 경우에서는 확률가중적률법이 우수한 경향을 나타내었다. 최우추정법과 확률가중적률법은 전반적으로 over estimate하는 경향을 나타내었다. RMSE 측면에서는 확률가중적률법이 압도적으로 우수한 경향을 나타내었다. 표 5에는 GP 분포의 형태모수 추정 결과가 나타나 있다. GEV 분포의 형태모수 추정에서 얻은 결과와 비슷한 경향이 발생되었다. 회귀분석법이 소표본에서 불편의 측면에서 우수한 결과를 보였다. 전반적으로 최우추정법과 확률가중적률법은 일관되게 under estimate 경향을 나타내었다. 회귀분석법은 형태모수가 −0.4일 때 under-estimate 경향을 나타냈고, 나머지 경우에서는 over-estimate 경향을 나타냈다. 한편, RMSE 측면에서는 소표본에서는 확률가중적률법이 우수

표 4: GP 분포 규모모수(σ)추정의 편의와 RMSE(괄호안), ML은 최우추정법, PWM은 plotting-position 확률가중적률법, Reg는 회귀분석법을 나타낸다. 규모모수 σ를 0.2부터 5로 하고 형태모수 ξ를 0.2로 한다. 표본크기를 15, 25, 50, 100, 200, 500으로 하고, 반복은 1000으로 한다.

n	method	σ					
		0.2	0.4	0.6	1.0	3.0	5.0
15	ML	0.0588 (0.1530)	0.1161 (0.2945)	0.1830 (0.4589)	0.3182 (0.8017)	0.8961 (2.2786)	1.4408 (3.7912)
	PWM	0.0260 (0.0994)	0.0604 (0.2067)	0.0804 (0.2897)	0.1494 (0.4994)	0.4014 (1.4339)	0.6701 (2.5079)
	Reg	0.0093 (0.1074)	0.0194 (0.2112)	0.0292 (0.3146)	0.0674 (0.5525)	0.1385 (1.5608)	0.2061 (2.7150)
25	ML	0.0236 (0.0836)	0.0571 (0.1758)	0.0810 (0.2519)	0.1594 (0.4601)	0.4723 (1.3180)	0.5688 (2.0293)
	PWM	0.0135 (0.0687)	0.0329 (0.1413)	0.0473 (0.2059)	0.0970 (0.3572)	0.2842 (1.0659)	0.2740 (1.6594)
	Reg	−0.0037 (0.0733)	−0.0024 (0.1521)	−0.0067 (0.2150)	0.0118 (0.3768)	0.0350 (1.1230)	−0.1305 (1.8168)
50	ML	0.0098 (0.0506)	0.0226 (0.0985)	0.0303 (0.1465)	0.0795 (0.2660)	0.1586 (0.7825)	0.3068 (1.2374)
	PWM	0.0058 (0.0464)	0.0146 (0.0919)	0.0196 (0.1346)	0.0545 (0.2385)	0.0887 (0.6977)	0.2099 (1.1272)
	Reg	−0.0084 (0.0502)	−0.0144 (0.0981)	−0.0242 (0.1450)	−0.0132 (0.2577)	−0.0975 (0.7733)	−0.1405 (1.2023)
100	ML	0.0056 (0.0340)	0.0116 (0.0670)	0.0215 (0.1017)	0.0267 (0.1703)	0.0772 (0.4950)	0.1801 (0.8400)
	PWM	0.0041 (0.0328)	0.0083 (0.0642)	0.0160 (0.0972)	0.0186 (0.1640)	0.0545 (0.4801)	0.1374 (0.8068)
	Reg	−0.0082 (0.0365)	−0.0139 (0.0723)	−0.0163 (0.1061)	−0.0374 (0.1811)	−0.1213 (0.5333)	−0.1395 (0.8873)
200	ML	0.0032 (0.0220)	0.0033 (0.0450)	0.0055 (0.0663)	0.0151 (0.1165)	0.0223 (0.3322)	0.0383 (0.5717)
	PWM	0.0026 (0.0218)	0.0023 (0.0446)	0.0033 (0.0658)	0.0116 (0.1160)	0.0083 (0.3270)	0.0221 (0.5619)
	Reg	−0.0067 (0.0249)	−0.0161 (0.0515)	−0.0226 (0.0758)	−0.0329 (0.1302)	−0.1143 (0.3879)	−0.1948 (0.6642)
500	ML	0.0015 (0.0137)	0.0019 (0.0289)	0.0048 (0.0416)	0.0040 (0.0718)	0.0254 (0.2219)	0.0265 (0.3468)
	PWM	0.0013 (0.0138)	0.0014 (0.0288)	0.0041 (0.0416)	0.0019 (0.0718)	0.0195 (0.2184)	0.0227 (0.3488)
	Reg	−0.0049 (0.0164)	−0.0104 (0.0351)	−0.0143 (0.0500)	−0.0247 (0.0873)	−0.0602 (0.2609)	−0.1373 (0.4156)

한 반면, 대표본에서는 최우추정법이 우수하였다.

마지막으로 한 분의 심사위원께서 제시한 바와 같이 모수 추정의 정확도를 두 분포에 대하여 비교해 보았다. GEV 분포와 GP 분포의 형태모수 및 규모모수에 대한 추정결과를 동일 표본수와 동일 모수값에서 비교해 본 결과, 추정방법에 상관없이 RMSE 측면에서 GEV 분포에서의 모수 추정의 정확성이 GP 분포에서 보다 높은 경향을 나타내었다. 불편의 측면에서 보았을 때, 소표본에서 규모모수 추정시 회귀분석법을 사용할 경우에는 GP 분포에서의 정확도가 GEV 분포에서 보다 높았으나, 그 외의 경우에서는 GEV 분포에서 더 좋은 결과를 가져왔다.

표 5: GP 분포 형태모수(ξ)추정의 편의와 RMSE(괄호안); ML은 최우추정법, PWM은 plotting-position 확률가중적률법, Reg는 회귀분석법을 나타낸다. 형태모수 ξ를 −0.4부터 0.4로 하고 규모모수 σ를 0.2로 한다. 표본크기를 15, 25, 50, 100, 200, 500으로 하고, 반복은 1000으로 한다.

n	method	ξ				
		−0.4	−0.2	0.0	0.2	0.4
15	ML	−0.3539 (0.5648)	−0.3002 (0.5326)	−0.2551 (0.5101)	−0.2253 0.5144)	−0.2024 (0.5224)
	PWM	−0.0946 (0.4295)	−0.0948 (0.3762)	−0.1046 (0.3418)	−0.1384 (0.3456)	−0.1777 (0.3536)
	Reg	−0.0453 (0.4691)	0.0357 (0.4457)	0.0929 (0.4466)	0.1255 (0.5081)	0.1904 (0.5672)
25	ML	−0.1707 (0.3391)	−0.1521 (0.3210)	−0.1249 (0.3069)	−0.1115 (0.3249)	−0.1286 (0.3532)
	PWM	−0.0287 (0.3128)	−0.0453 (0.2644)	−0.0573 (0.2558)	−0.0755 (0.2483)	−0.1341 (0.2663)
	Reg	−0.0057 (0.3240)	0.0406 (0.3112)	0.0903 (0.3218)	0.1319 (0.3652)	0.1274 (0.3949)
50	ML	−0.0861 (0.1724)	−0.0592 (0.1690)	−0.0655 (0.1777)	−0.0505 (0.1919)	−0.0452 (0.2137)
	PWM	−0.0324 (0.2174)	−0.0147 (0.1846)	−0.0335 (0.1636)	−0.0416 (0.1712)	−0.0646 (0.1834)
	Reg	−0.0308 (0.2166)	0.0446 (0.1923)	0.0670 (0.1995)	0.0984 (0.2306)	0.1232 (0.2795)
100	ML	−0.0363 (0.0986)	−0.0341 (0.0999)	−0.0351 (0.1182)	−0.0181 (0.1340)	−0.0194 (0.1467)
	PWM	−0.0053 (0.1388)	−0.0081 (0.1210)	−0.0192 (0.1197)	−0.0158 (0.1241)	−0.0358 (0.1354)
	Reg	−0.0083 (0.1401)	0.0317 (0.1175)	0.0468 (0.1344)	0.0794 (0.1699)	0.0888 (0.1959)
200	ML	−0.0231 (0.0647)	−0.0173 (0.0678)	−0.0108 (0.0783)	−0.0159 (0.0928)	−0.0113 (0.1002)
	PWM	−0.0080 (0.1019)	−0.0054 (0.0873)	−0.0030 (0.0850)	−0.0150 (0.0898)	−0.0223 (0.0975)
	Reg	−0.0152 (0.0990)	0.0221 (0.0834)	0.0422 (0.0958)	0.0498 (0.1181)	0.0596 (0.1318)
500	ML	−0.0100 (0.0365)	−0.0060 (0.0410)	−0.0070 (0.0471)	−0.0043 (0.0549)	−0.0056 (0.0628)
	PWM	−0.0026 (0.0657)	−0.0039 (0.0572)	−0.0038 (0.0519)	−0.0043 (0.0548)	−0.0118 (0.0655)
	Reg	−0.0089 (0.0628)	0.0173 (0.0524)	0.0233 (0.0577)	0.0326 (0.0725)	0.0337 (0.0835)

지금까지의 시뮬레이션 연구의 중요한 결과를 아래와 같이 요약 정리할 수 있다.

1. GEV 분포의 위치모수 추정에서는 불편의 측면과 RMSE 측면 모두에서 소표본 ($n = 15, 25$)일 경우 회귀분석법이 많은 경우 우수한 것으로 나타났으나, 그 외의 표본의 경우 확률가중적률법과 최우추정법이 우수하였다. 불편의 측면에서, 회귀분석법의 경우 under estimate이고, 다른 두 추정 방법은 over estimate하는 경향을 나타내었다.

2. GEV 분포의 규모모수 추정의 경우, RMSE 측면에서 최우추정법이 단연 우수함을 보였다. 최우추정법과 확률가중적률법은 일관되게 under estimate 경향을 나타내었다.

3. GP 분포의 규모모수 추정의 경우, 불편의 측면에서 소표본 (n = 15, 25)일 경우 회귀분석법이 우수하였고, 그 외에서는 확률가중적률법이 우수하였다. 최우추정법과 확률가중적률법은 전반적으로 over estimate하는 경향을 나타내었다. RMSE 측면에서는 확률가중적률법이 전반적으로 우수하였다.

4. GEV 분포나 GP 분포의 형태모수 추정의 경우, 불편의 측면에서 평가하면 소표본 또는 일부 대표본 (n = 15, 25, 50, 100)에서 형태모수가 −0.4일 때 회귀분석법이 가장 우수한 경향이 나타났다. GEV 분포의 경우 확률가중적률법은 모수가 양수일 경우 under estimate하는 경향이 나타났고, 음수일 경우에 over estimate하는 경향이 나타났다. GP 분포의 경우 최우추정법과 확률가중적률법은 일관되게 under estimate 경향을 나타내었다.

5. GEV 분포나 GP 분포의 형태모수 추정의 경우, RMSE 측면에서 평가하면 확률가중적률법 (소표본의 경우)과 최우추정법 (대표본의 경우)이 대체로 우수한 경향이 나타났다.

6. 두 분포의 형태모수 및 규모모수에 대한 추정결과를 동일 표본수와 동일 모수 값에서 비교해 본 결과, 추정방법에 상관없이 RMSE 측면에서 GEV 분포에서의 모수 추정의 정확성이 GP 분포에서 보다 높은 경향을 나타내었다.

5. 결론

본 논문에서는 GEV 분포와 GP 분포의 모수에 대하여 세 가지 추정 방법(회귀분석법, 확률가중적률법, 최우추정법)의 정확성을 몬테칼로 시뮬레이션 기법을 이용하여 다양한 표본 크기에 대하여 살펴보았다. 확률가중적률법 및 최우추정법에 대하여서는 기존에 많은 연구 결과가 나와 있으나, 회귀분석법에 대하여서는 저자가 알기로 전혀 없었다. 따라서, 회귀분석법을 중심으로 세 가지 추정 방법을 비교하였다. 비교한 결과 GEV 분포와 GP 분포의 형태모수 추정에 있어서는 소표본 및 일부 대표본에서 형태모수 값이 −0.5에 근사할 경우, 회귀분석법이 우수함을 발견할 수 있었다. GEV 분포의 위치 모수를 추정할 경우에는 소표본일 경우 불편의 및 RMSE 측면 모두에서 회귀분석법이 우수한 경향이 나타났다. 한편 GP 분포의 규모모수 추정시에는 회귀분석법이 소표본에서 불편의 측면에서 우수하였다. 요컨대, 회귀분석법은 특정 표본크기 조건하에서 GEV 분포나 GP 분포의 특정 모수 추정시에 불편의 측면이나 효율성 측면에서 최우추정법 또는 확률가중적률법 보다 우수한 결과를 가져 올 수 있음을 시뮬레이션 연구를 통하여 보였다.

감사의 글

이승천 편집위원장님과 두 분의 심사위원님께 감사의 말씀을 드립니다. 두 분 심사위원님들의 수정 요청사항은 본 논문이 보다 좋은 논문으로 만들어지는데 많은 도움이 되었습니다.

참고 문헌

Bali, T. G. (2003). An extreme value approach to estimating volatility and value at risk, *Journal of Business*, **76**, 83–108.

Davison, A. C. (1984). Modelling excesses over high thresholds, with an application, In *Statistical Extremes and Applications*, 461–482, Springer, Vimeiro.

Gettinby, G. D., Sinclair, C. D., Power, D. M. and Brown, R. A. (2006). An analysis of the distribution of extremes in indices of share returns in the US, UK and Japan from 1963 to 2000, *International Journal of Finance and Economics*, **11**, 97–113.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**, 423–453.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, **15**, 1049–1054.

Gumbel, E. J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

Haktanir, T. and Bozduman, A. (1995). A study of sensitivity of the probability-weighted moments method on the choice of the plotting position formula, *Journal of Hydrology*, **168**, 265–281.

Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalised extreme value distribution by the method of probability weighted moments, *Technometrics*, **27**, 251–261.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society, Series B*, **52**, 105–124.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quarterly Journal of the Royal Meteorological Society*, **81**, 158–171.

Jenkinson, A. F. (1969). Statistics of extremes of maximum floods. WMO Technical Note, 98. World Meteorological Organization, Geneva, 183–228.

Kim, M. S. (2007). On the effects of plotting positions to the probability weighted moments method for the generalized logistic distribution, *The Korean Communications in Statistics*, **14**, 561–576.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979a). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resources Research*, **15**, 1055–1064.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979b). Estimation of parameters and quantiles of Wakeby distributions, *Water Resources Research*, **15**, 1361–1379.

Longin, F. M. (1996). The asymptotic distribution of extreme stock market returns, *Journal of Business*, **69**, 383–408.

Login, F. M. (2000). From value at risk to stress testing: The extreme value approach, *Journal of Banking and Finance*, **24**, 1097–1130.

Pickands, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 119–131.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika*, **67**, 723–724.

Singh, V. P. and Ahmad, M. (2004). A comparative evaluation of the estimators of the three parameter generalized Pareto distribution, *Journal of Statistical Computation & Simulation*, **74**, 91–106

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

2009년 2월 접수; 2009년 3월 채택

Comparison Study of Parameter Estimation Methods for Some Extreme Value Distributions (Focused on the Regression Method)

Ji-Yong Woo[a], Myung Suk Kim[1,b]

[a]Korea Bond Pricing & Korea Ratings Co.
[b]College of Business Administration, Sogang University

Abstract

Parameter estimation methods such as maximum likelihood estimation method, probability weighted moments method, regression method have been popularly applied to various extreme value models in numerous literature. Among three methods above, the performance of regression method has not been rigorously investigated yet. In this paper the regression method is compared with the other methods via Monte Carlo simulation studies for estimation of parameters of the Generalized Extreme Value(GEV) distribution and the Generalized Pareto(GP) distribution. Our simulation results indicate that the regression method tends to outperform other methods under small samples by providing smaller biases and root mean square errors for estimation of location parameter of the GEV model. For the scale parameter estimation of the GP model under small samples, the regression method tends to report smaller biases than the other methods. The regression method tends to be superior to other methods for the shape parameter estimation of the GEV model and GP model when the shape parameter is −0.4 under small and moderately large samples.

Keywords: Generalized extreme value distribution, generalized Pareto distribution, maximum likelihood estimation method, Monte Carlo simulation, probability weighted moments method, regression method.

This work was supported by the Sogang University Research Grant of 2008 (200810048.01).
This paper is based on the first author's M.S. thesis at the Sogang University.
[1] Corresponding author: Assistant Professor, College of Business Administration, Sogang University, Shinsu-dong, Mapo-gu, Seoul 121-742, Korea. E-mail: myungkim@sogang.ac.kr