UNIVERSITY
OF TRENTO - Italy
Department of Economics

Statistical analysis of the Lognormal-Pareto distribution using Probability Weighted Moments and Maximum Likelihood

Marco Bee

n. 08/2012

Statistical analysis of the Lognormal-Pareto distribution using Probability Weighted Moments and Maximum Likelihood

Marco Bee

Department of Economics, University of Trento

Abstract

This paper deals with the estimation of the lognormal-Pareto and the lognormal-Generalized Pareto distributions, for which a general result concerning asymptotic optimality of maximum likelihood estimation cannot be proved. We develop a method based on probability weighted moments, showing that it can be applied straightforwardly to the first distribution only. In the lognormal-Generalized Pareto case, we propose a mixed approach combining maximum likelihood and probability weighted moments. Extensive simulations analyze the relative efficiencies of the methods in various setups. Finally, the techniques are applied to two real datasets in the actuarial and operational risk management fields.

Keywords: Probability Weighted Moments; Mixed Estimation Method; Lognormal-Pareto Distribution; Loss Models.

1 Introduction

The lognormal and the Pareto distribution play a central role as probabilistic models for the distribution of various phenomena in different fields. In finance, the lognormal is the standard model for prices of financial assets since the development of the Black-Scholes (Black and Scholes, 1973) option pricing theory based on geometric Brownian motion. Both distributions have long been used as loss models in actuarial sciences (Klugman et al., 2004) and, more recently, in operational risk management (Panjer, 2006). Hydrological variables are also often modelled as lognormal random variables (Kroll and Stedinger, 1996).

The Pareto distribution has first been proposed as a model of income distribution (Pareto, 1896), but, as power-law relations have been discovered in several fields of the natural sciences (Sornette, 2004), its range of applicability has enlarged significantly in the last few decades. Empirically, the increased use of the Pareto distribution has also resulted from the inability of the lognormal to account for large-impact low-probability events, i.e. to model the right tail of the distribution. In finance, for example, empirical analysis has shown that log-returns are usually not normal and prices are not lognormal, mostly because of the presence of very large observations. This finding is supported theoretically by Extreme Value Theory (EVT), which shows that excesses over high thresholds follow a Generalized Pareto Distribution (GPD).

Referring the interested reader to Embrechts et al. (1997) for details, statistical inference within the EVT framework is mostly carried out using the Peaks Over Threshold (POT) method, which has often proved to be more effective (Embrechts et al., 1997; McNeil et al., 2005). Given n *iid* observations $x_1, \ldots, x_n$ from a random variable X with cumulative distribution function (cdf) F and some predefined large value $u \in \mathbb{R}^+$ in the support of F, the POT approach studies the excess distribution over the threshold u, namely $Y = X - u$. The conditional distribution function of Y is given by:

$$\mathbb{P}\Big(X - u \leq y \mid X > u\Big) = F_u(y) = \frac{F(y+u) - F(u)}{1 - F(u)}, \qquad 0 \leq y < x_0 - u \ , \tag{1}$$

where $x_0 \leq \infty$ is the right endpoint of F. The key result is the Balkema, de Haan and Pickands theorem (see, for example, McNeil et al., 2005, p. 277): under regularity conditions and for a certain class of underlying distributions, it is possible to find a function $\beta(u)$ such that the excess distribution converges to the GPD.

Another active field of research concerns the distribution of the tail of variables relevant for economic theory, such as trade bilateral flows, firm size and city size; see, e.g., Eeckhout (2004); Rozenfeld et al. (2011); Bee et al. (2011). Finally, numerous other phenomena in economics and finance have been found to follow heavy-tailed (often power-law) distributions; see Malevergne and Sornette (2006) and Gabaix (2009).

A message that frequently arises is that the lognormal distribution may be a model that fits well the body of the distribution but not the tail, where a Pareto distribution provides a better fit. Thus, it makes sense to consider a composite Lognormal-Pareto model, namely a mixture of a right-truncated lognormal and a Pareto. Denoting with $\Phi(\cdot)$ the standard normal distribution function, the density of the resulting distribution is

$$f(x) = r f_1(x; \mu, \sigma^2, \theta) + (1 - r) f_2(x; \theta, \alpha), \tag{2}$$

where

$$f_1(x; \mu, \sigma^2, \theta) = \frac{1}{\Phi\left(\frac{\log(\theta) - \mu}{\sigma}\right)} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x) - \mu}{\sigma}\right)^2} \mathbb{I}_{\{0 < x \leq \theta\}} \tag{3}$$

is the density of the Logn(μ, σ^2) distribution right-truncated at θ and

$$f_2(x; \theta, \alpha) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \mathbb{I}_{\{x>\theta\}}$$

is the Pareto density with parameters θ and α. Without further constraints, (2) has the disturbing feature of being discontinuous at θ. A solution to this drawback has been proposed by Scollnik (2007), who finds two versions of (2) that are respectively continuous and differentiable (see Sect. 2 for details). The constraint(s) needed to make the density continuous (differentiable) has the consequence of reducing the number of free parameters of the model: whereas (2) contains five free parameters, the continuous (differentiable) version contains four (three) free parameters because r (r and μ) are functions of the remaining parameters.

A more general approach uses the GPD instead of the Pareto in (2). This extension is important both for modeling purposes and because of the central role played by the GPD as the distribution of excesses above high thresholds. Similarly to the preceding case, the density is, in general, discontinuous; Scollnik (2007) gives continuous and differentiable versions of the density.

The composite model proposed above is just one of many hybrid distributions used in applications to model skewed populations. In particular, hydrologists routinely employ mixed exponential, exponential-GPD and hybrid Gamma-GPD distributions: see, for example, Li et al. (2012).

Finite mixture distributions are a well-known topic. The earliest and still most common mixture model has normal population components but, more recently, mixtures of several other distributions have been studied and employed in various fields; see McLachlan and Peel (2000) for a review. Maximum Likelihood Estimation (MLE) is, in general, rather complicated both from the computational point of view, because the estimators cannot be obtained in closed form, and from the inferential point of view, because the likelihood function is typically non-regular. The first difficulty has been overcome, in the majority of cases, by the introduction of the EM algorithm (Dempster et al., 1977). As for the statistical properties, conditions for MLEs to be consistent, efficient and asymptotically normal are given by Redner and Walker (1984); see also McLachlan and Peel (2000, Sect. 2.5). Unfortunately, as will be shown in Sect. 3, in general the lognormal-Pareto distribution does not satisfy these conditions.

From the computational point of view, the fact that r is a function of the remaining parameters makes the usual approach to MLE based on the EM algorithm unfeasible, because the likelihood cannot be decomposed in different functions that depend on different sets of parameters and can be maximized separately.

On the other hand, the differentiable model contains only three unknown parameters, so that moment-based estimation procedures are more convenient, because it is necessary to find only three theoretical moments. Within this class of estimators, the Probability Weighted Moments (PWM) method, first introduced by Landwehr et al. (1979) and subsequently employed in various theoretical and empirical settings, proved to be quite effective. It is worth noting that the method has been widely used for the estimation of the Generalized Extreme Value distribution and of the GPD, in particular in hydrology (see, e.g., Landwehr et al., 1979; Hosking et al., 1985; Hosking and Wallis, 1987; Morrison and Smith, 2002; Ailliot et al., 2011). Although not much is known concerning its efficiency (see Hosking, 1990 for a thorough discussion), PWMs turn out to be often more efficient than MLEs in small samples, even in regular cases, that is when MLEs are asymptotically efficient.

The main contribution of this paper consists in deriving the first three approximate PWMs of (2), necessary for estimating the three unknown parameters. Whereas PWMs for the lognormal-Pareto can be computed in a relatively straightforward way using a logarithmic transformation, the problem is more complicated when the true data-generating process is the lognormal-GPD. In the latter case, standard PWM-based estimation breaks down, and for this reason we develop a mixed approach based on the technique proposed by Morrison and Smith (2002) and Ailliot et al. (2011).

The rest of the paper is organized as follows. Section 2 reviews the versions of the lognormal-Pareto distribution and analyzes its statistical features. Section 3 treats MLE, considering separately the cases when θ is known and unknown. Section 4 investigates PWM estimation, deriving PWM-based estimators for the lognormal-Pareto and mixed estimators for the lognormal-GPD distribution. Section 5 studies the properties of the estimators by means of simulation experiments and uses them in two real-data applications. Finally, Section 6 concludes.

2 The composite model

2.1 The lognormal-Pareto distribution

The lognormal-Pareto distribution with density (2) is a finite mixture distribution, but unlike most commonly-used mixtures, the two populations do not overlap. Moreover, it is in general discontinuous at $x = \theta$. This last feature is undesirable for most modelling purposes. Hence, Scollnik (2007) formalized two restricted versions of (2) that are respectively continuous and differentiable.

When only the continuity condition $f(\theta-) = f(\theta+)$ is imposed, the density is still given by (2), but r is equal to (Scollnik, 2007):

$$r(\mu, \sigma^2, \alpha, \theta) = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\theta^*)e^{\frac{1}{2}{\theta^*}^2}}{\sqrt{2\pi}\alpha\sigma\Phi(\theta^*)e^{\frac{1}{2}{\theta^*}^2} + 1}, \tag{4}$$

where $\theta^* = (\log(\theta) - \mu)/\sigma$. The notation emphasizes that, although $r(\mu, \sigma^2, \alpha, \theta)$ can be interpreted as a mixing weight, it is not a free parameter, but a function of μ, σ^2, α and θ.

The differentiability condition $f'(\theta-) = f'(\theta+)$ implies $\mu = \log(\theta) - \alpha\sigma^2$. When this is the case, r simplifies to

$$r(\sigma^2, \alpha) = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{1}{2}(\alpha\sigma)^2}}{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)e^{\frac{1}{2}(\alpha\sigma)^2}+1}. \tag{5}$$

Therefore, only σ^2, α and θ are free parameters, whereas μ is a function of σ^2, α and θ, and r is a function of σ^2 and α. From now on we will refer to (2) with no constraint as *first lognormal-Pareto model*, under the constraint (4) as *second lognormal-Pareto model* and subject to (5) as *third lognormal-Pareto model*.

2.2 The lognormal-Generalized Pareto distribution

As outlined in the introduction, the GPD is of crucial importance for modelling extreme observations. Thus, we consider a more general composite model where the first population remains truncated lognormal but the second is GPD. The density is given by

$$f(x) = rf_1(x; \mu, \sigma^2, \theta) + (1-r)f_3(x; \xi, \tau, \theta), \tag{6}$$

where f_1 is the truncated lognormal density (3) and f_3 is the density of the GPD with parameters ξ, τ and θ:

$$f_3(x; \xi, \tau, \theta) = \frac{1}{\tau}\left(1 + \xi\frac{x-\theta}{\tau}\right)^{-\left(\frac{1}{\xi}+1\right)} \mathbb{I}_{\{\theta < x\}}. \tag{7}$$

Similarly to the lognormal-Pareto case, (6) is in general discontinuous. Scollnik (2007) derives the condition for the density to be continuous:

$$r(\mu, \sigma^2, \xi, \theta) = \frac{\sqrt{2\pi}\xi^{-1}\sigma\Phi(\theta^*)e^{\frac{1}{2}\theta^{*2}}}{\sqrt{2\pi}\xi^{-1}\sigma\Phi(\theta^*)e^{\frac{1}{2}\theta^{*2}} + \tau\xi^{-1}}.$$

If we want (6) to be continuous and differentiable, we also have to impose a condition on μ:

$$\mu = \log(\theta) - \sigma^2\frac{(\theta-\tau)/\xi + \theta}{\tau/\xi}.$$

Note that the conditions above are different from Scollnik (2007) because we use a different parametrization. In the following we will just use the continuous and differentiable version of the model, which will be referred to as *fourth lognormal-Pareto model* because loss models are typically characterized by smooth densities. However, the simulation experiment of Sect. 5.1 has been performed also with the second lognormal-Pareto model, and the results are very similar.

3 Maximum Likelihood Estimation

In this section we study MLE of the distributions described in Sect. 2. For the sake of completeness, we first treat the unrealistic case where θ is known. Then we focus on the more relevant setup where all parameters are unknown. When θ is known, MLE is asymptotically efficient, but in the second and third model finding the estimators requires numerical maximization of the log-likelihood function and computing the asymptotic standard errors of the estimators is computationally prohibitive. When θ is unknown, asymptotic optimality of MLEs can no longer be proved.

3.1 When θ is known

When θ is known and no continuity or differentiability conditions at θ are imposed, population membership of each observation is known and r is a free parameter. The problem therefore reduces to separate estimation of (i) the binomial parameter r, (ii) the parameters (μ, σ^2) of a truncated Lognormal random variable X_1, (iii) the shape parameter α of a Pareto r.v. X_2. The Pareto distribution with known location parameter and the Binomial distribution are one-parameter exponential families of distributions. As for the truncated Lognormal, by taking

$$\boldsymbol{\gamma} = \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)', \ h(x) = \frac{1}{\sqrt{2\pi x}}, \ A(\boldsymbol{\gamma}) = \frac{\mu^2}{2\sigma^2} - \log\left(\frac{1}{\sigma}\Phi\left(\frac{\log(\theta) - \mu}{\sigma}\right)\right),$$
$$\boldsymbol{T}(x) = (T_1(x), T_2(x))' = \left(\log(x), -\frac{\log(x)^2}{2}\right)',$$

it is easily seen that it is a two-parameter exponential family with canonical parameter $\boldsymbol{\gamma}$ and density

$$f_X(x; \boldsymbol{\gamma}) = h(x) \exp\left(\sum_{i=1}^{2} \gamma_i T_i(x) - A(\boldsymbol{\gamma})\right).$$

Let now $x_1, \ldots, x_n$ be a random sample from (2), $\mathcal{X}^{(1)} = \{x_i : x_i \leq \theta\}$, $\mathcal{X}^{(2)} = \{x_i : x_i > \theta\}$, $n_1 = \#\mathcal{X}^{(1)}$ and $n_2 = n - n_1$. Then the MLEs of r and α are given by

$$\hat{r} = \frac{n_1}{n}; \quad \hat{\alpha} = \frac{n_2}{\sum_{\mathcal{X}^{(2)}} (\log(x_i) - \log(\theta))}.$$

As for μ and σ^2, they cannot be found in closed form, but can be computed numerically, for example by means of the EM algorithm (McLachlan and Krishnan, 2008, Sect. 2.8; Bee, 2006). Note that one has to assume that the lognormal distribution is truncated, not censored. If we incorrectly treated the distribution as censored, we would use for estimation of the lognormal parameters also the cardinality of $\mathcal{X}^{(2)}$, assuming implicitly that the observations in $\mathcal{X}^{(2)}$ are unknown but lognormally distributed, whereas in this setup they are actually Pareto(θ, α) distributed.

According to a general result concerning MLE in multiparameter exponential families (Lehmann and Casella, 1998, Example 6.3), all the estimators derived above are asymptotically efficient. Moreover, $\hat{r}$ has limiting variance equal to $r(1-r)/n$ and the finite sample distribution of $\hat{\alpha}$ is $2n_2\alpha/\hat{\alpha} \sim \chi^2_{2n_2-2}$ (Lehmann and Casella, 1998, Example 7.13).

3.1.1 Continuous density

The log-likelihood function of the second model is given by

$$l(\mu,\sigma^2,\alpha;\boldsymbol{x}) = \sum_{i=1}^{n} \log(\{\mathbb{I}_{\{0\le x_i\le\theta\}} r(\mu,\sigma^2,\alpha) L_1^{\theta}(\mu,\sigma^2;x_i) + +(1-r(\mu,\sigma^2,\alpha))\mathbb{I}_{\{x_i>\theta\}} L_2^{\theta}(\alpha;x_i)\}). \tag{8}$$

In (8), the dependence of r on θ is suppressed as θ is known; $L_1^{\theta}(\mu,\sigma^2;x)$ is the likelihood function of a right-truncated Lognormal random variable X_1 with parameters μ and σ^2, and θ is the truncation point; $L_2^{\theta}(\alpha;x)$ is the likelihood function of a Pareto random variable X_2 with parameters θ and α. Apart from a constant that does not depend on unknown parameters, (8) can be rewritten as

$$\begin{aligned} l(\mu,\sigma^2,\alpha;\boldsymbol{x}) &= \sum_{i=1}^{n_1} \log\{(r(\mu,\sigma^2,\alpha)L_1^{\theta}(\mu,\sigma^2;x_i^{(1)})\}) + \sum_{i=n_1+1}^{n} \log(\{(1-r(\mu,\sigma^2,\alpha))L_2^{\theta}(\alpha;x_i^{(2)})\}) \\ &= n_1\log(r(\mu,\sigma^2,\alpha)) + n_2\log(1-r(\mu,\sigma^2,\alpha)) + \sum_{i=1}^{n_1} l_1^{\theta}(\mu,\sigma^2;x_i^{(1)}) + \sum_{i=n_1+1}^{n} l_2^{\theta}(\alpha;x_i^{(2)}), \end{aligned}$$

where $l_1^{\theta}(\mu,\sigma^2;x) = \log(L_1^{\theta}(\mu,\sigma^2;x))$, $l_2^{\theta}(\alpha;x) = \log(L_2^{\theta}(\alpha;x))$ and the notation $x_i^{(k)}$ means that the i-th observation belongs to the k-th population ($k=1,2$). In this setup the log-likelihood function has to be maximized numerically, and MLEs are asymptotically efficient, since all the standard conditions (Lehmann and Casella, 1998, Chap. 6, Theorem 3.10) for efficiency of MLEs are satisfied: l_1 and l_2 belong to the exponential family, the first two summands are infinitely differentiable and their derivatives are bounded.

3.1.2 Continuous and differentiable density

The log-likelihood function of the third model becomes

$$l(\sigma^2,\alpha;\boldsymbol{x}) = \sum_{i=1}^{n} \log\{\mathbb{I}_{\{0\le x_i\le\theta\}} r(\theta,\sigma^2,\alpha) L_1^{\theta}(\sigma^2,\alpha;\boldsymbol{x}) + (1-r(\theta,\sigma^2,\alpha))\mathbb{I}_{\{x_i>\theta\}} L_2^{\theta}(\alpha;\boldsymbol{x})\} \tag{9}$$

$$= \sum_{i=1}^{n_1} \log\{(r(\theta,\sigma^2,\alpha)L_1^{\theta}(\alpha,\sigma^2;x_i^{(1)})\}) + \sum_{i=n_1+1}^{n} \log(\{(1-r(\theta,\sigma^2,\alpha))L_2^{\theta}(\alpha;x_i^{(2)})\})$$

$$= n_1\log(r(\sigma^2,\alpha)) + n_2\log(1-r(\sigma^2,\alpha)) + \sum_{i=1}^{n_1} l_1^{\theta}(\sigma^2,\alpha;x_i^{(1)}) + \sum_{i=n_1+1}^{n} l_2^{\theta}(\alpha;x_i^{(2)}).$$

In (9), $L_1^{\theta}(\sigma^2,\alpha;x)$ is the likelihood function of a right-truncated Lognormal random variable with parameters $\log(\theta)-\alpha\sigma^2$ and σ^2. Similarly to Sect. 3.1.1, MLEs can only be found numerically and are asymptotically efficient.

3.2 The general case

When θ is unknown, standard MLE asymptotic theory (Lehmann and Casella, 1998, Chap. 6, Theorem 3.10) no longer applies, because (at least) the second derivative with respect to θ is not continuous. Alternative conditions not involving the second derivative have been developed by Daniels (1961): if the log-likelihood satisifes either Conditions I (Daniels, 1961, p. 152) or Conditions II (Daniels, 1961, p. 155), MLEs are asymptotically efficient. However, checking these conditions for the log-likelihood functions of the composite models defined so far is in general difficult. To see why, consider that Condition I (1) can be rephrased as follows (Daniels, 1961, formula (2.2)):
(1) $l(x,\theta)$ is continuous in θ throughout the parameter space Θ. Let θ_0 be the true parameter value. At every θ_0 there is a neighborhood such that for all θ, θ' in it,

$$\left|\frac{f(x;\theta)}{f(x;\theta')}-1\right| < B(x;\theta_0)|\theta-\theta'|, \tag{10}$$

where $\int B^2(x;\theta_0)f(x;\theta_0)dx<\infty$.

As for Condition II (1), it still requires (10), but with $\int B^3(x;\theta_0)f(x;\theta_0)dx<\infty$. Hence, considering that Condition I (1) nests Condition II (1), if the former is not satisfied, Condition II (1) is not satisfied either.

In our case, we need a function $B(x;\theta_0)$ such that (10) holds true and $\int B^2(x;\theta_0)f(x;\theta_0)dx<\infty$. Given that the Pareto distribution is heavy-tailed and the tail gets heavier as α decreases, it is difficult to find a function $B(x;\theta_0)$ such that not only (10) is satisfied but also the expected value of $B^2(x;\theta_0)$ exists for any $\alpha>0$. It may, however, be possible to find $B(x;\theta_0)$ satisfying the two conditions if α is large enough. In conclusion, MLEs may be asymptotically efficient in models where the true value of the Pareto shape parameter α is sufficiently large, but not for any value of α. This seems to be in line with Smith (1985), who shows that, for Pareto-type distributions, MLEs are asymptotically efficient when $\alpha \geq 2$.

3.3 The lognormal-GPD

The theory for the lognormal-GPD is essentially identical to the lognormal-Pareto case. When θ is known and no conditions are imposed, the GPD belongs to the exponential family and therefore separate MLE of the parameters of the lognormal and of the GPD gives asymptotically optimal estimators (see Smith, 1985, for the properties of the MLEs of the parameters of the GPD). MLEs cannot be obtained in closed form; under the continuity and differentiability conditions, they are asymptotically optimal. Analogously to the preceding case, when θ is unknown Daniel's (1961) conditions are in general not satisfied.

4 Probability Weighted Moments estimation

4.1 The lognormal-Pareto distribution

The PWM method is a refinement of the well-known method of moments. For a random variable X with CDF F, the general definition of PWMs is (Greenwood et al., 1979; see also Landwehr et al., 1979 and Hosking et al., 1985):

$$M_{p,r,s} = \mathrm{E}(X^p F(X)^r (1 - F(X))^s), \quad p, r, s = 0, 1, \ldots$$

PWMs are linear combinations of L-moments and therefore procedures based on PWMs and on L-moments are equivalent (Hosking, 1990).

Little is known about the efficiency properties of PWM estimators. However, even in cases where MLE is asymptotically efficient, the asymptotic loss of efficiency of PWM estimators with respect to MLE is typically quite small (Hosking, 1990). Moreover, they are sometimes more efficient in finite samples. Thus, it is clearly of interest to develop PWM-based statistical inference in a framework such as the present one, where asymptotic efficiency of MLEs does not hold in general.

As the third lognormal-Pareto model contains three unknown parameters, we have to compute the first three PWMs of the Lognormal-Pareto distribution and equate them to their empirical counterparts. We consider moments $M_{1,s,0} \stackrel{\text{def}}{=} \beta_s = \int x F^s(x) dFx = \mathrm{E}(XF^s(X))$, thus we compute β_0, β_1 and β_2.

Let us first recall two results that will be used in the following.

Result 1 *Let X be a normal r.v. right-truncated at $c \in \mathbb{R}$. The quantile function of X is given by $\Phi^{-1}_{\mu,\sigma^2,c}(x) = \Phi^{-1}_{\mu,\sigma^2}(\alpha x)$, where $\alpha = \Phi_{\mu,\sigma^2}(c)$ and Φ_{μ,σ^2} and Φ^{-1}_{μ,σ^2} are respectively the CDF and the quantile function of the $N(\mu, \sigma^2)$ distribution.*

Result 2 *For a continuous r.v. X, the moments β_s ($s = 0, 1, \dots$) can be written as*

$$\beta_s = \int_0^1 X(F) F^s dF, \tag{11}$$

where $X(F)$ is the quantile function of X evaluated at F.

For the estimation of the lognormal-Pareto distribution, it is more convenient to use a logarithmic transformation. It is easy to see that the distribution of $Y = \log(X)$ is a mixture of a right-truncated normal and of a translated exponential, with density

$$f_Y(y; \sigma^2, \alpha, \theta) = r\phi_{\mu,\sigma^2,\theta^\diamond}(y) + (1-r) f_{\theta^*}(y; \alpha), \tag{12}$$

where $\theta^\diamond = \log(\theta)$,

$$\phi_{\mu,\sigma^2,\theta^\diamond}(y) = \frac{\phi_{\mu,\sigma^2}(y)}{\Phi_{\mu,\sigma^2}(\theta^\diamond)} \mathbb{I}_{\{y \le \theta^\diamond\}} \tag{13}$$

is the density of the Normal distribution right-truncated at $\theta^\diamond$ and

$$f_{\theta^\diamond}(y; \alpha) = \frac{1}{\alpha} \exp\left\{-\frac{1}{\alpha}(y - \theta^\diamond)\right\} \mathbb{I}_{\{\theta^\diamond < y\}} \tag{14}$$

is the density of the translated exponential distribution, with translation parameter equal to $\theta^\diamond$. This transformation increases the generality of the PWM method, as β_s exists for any s for the normal-exponential mixture but not for the lognormal-Pareto distribution. For the random variable Y with density (12), the first three PWMs β_0, β_1 and β_2 can be written as

$$\begin{aligned}
\beta_0 &= \int_{\mathbb{R}} y\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + (1-r)\int_{\mathbb{R}} x f_{\theta^\diamond}(y)dy;\\
\beta_1 &= r^2 \int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + r(1-r)\int_{\mathbb{R}} xF_{\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy +\\
&\quad + r(1-r)\int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y) f_{\theta^\diamond}(y)dy + (1-r)^2 \int_{\mathbb{R}} xF_{\theta^\diamond}(y) f_{\theta^\diamond}(y)dy;\\
\beta_2 &= r^3 \int_{\mathbb{R}} y\Phi^2_{\mu,\sigma^2,\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + 2r^2(1-r)\int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y)F_{\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy +\\
&\quad + r(1-r)^2 \int_{\mathbb{R}} xF^2_{\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + r^2(1-r)\int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y) f_{\theta^\diamond}(y)dy +\\
&\quad + 2r(1-r)^2 \int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y)F_{\theta^\diamond}(y) f_{\theta^\diamond}(y)dy + (1-r)^3 \int_{\mathbb{R}} y\Phi^2_{\mu,\sigma^2,\theta^\diamond}(y)F_{\theta^\diamond}(y) f_{\theta^\diamond}(y)dy,
\end{aligned} \tag{15}$$

where $F_\theta^\diamond$ is the density of the translated exponential distribution with translation parameter $\theta^\diamond$. As the two population densities of the Lognormal-Pareto distribution do not overlap, and therefore $\Phi_{\mu,\sigma^2,\theta^\diamond}(y) = 1$ for $y \geq \theta^\diamond$ and $F_{\theta^\diamond}(y) = 0$ for $y \leq \theta^\diamond$, the equations (15) simplify substantially. They are given by:

$$\beta_0 = r\left(\mu - \sigma\frac{\phi_{\mu,\sigma^2}(\theta^\diamond)}{\Phi_{\mu,\sigma^2}(\theta^\diamond)}\right) + (1-r)(\theta^\diamond + \alpha); \tag{16}$$

$$\begin{aligned}
\beta_1 &= r^2 \int_{\mathbb{R}} y\Phi_{\mu,\sigma^2,\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + \\
&+ r(1-r)(\theta^\diamond + \alpha) + (1-r)^2 \int_{\mathbb{R}} xF_{\theta^\diamond}(y)f_{\theta^\diamond}(y)dy; \\
\beta_2 &= r^3 \int_{\mathbb{R}} y\Phi^2_{\mu,\sigma^2,\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + r^2(1-r)(\theta^\diamond + \alpha) + \\
&+ 2r(1-r)^2 \int_{\mathbb{R}} xF_{\theta^\diamond}(y)f_{\theta^\diamond}(y)dy + (1-r)^3 \int_{\mathbb{R}} xF_{\theta^\diamond}(y)f_{\theta^\diamond}(y)dy = \\
&= r^3 \int_{\mathbb{R}} y\Phi^2_{\mu,\sigma^2,\theta^\diamond}(y)\phi_{\mu,\sigma^2,\theta^\diamond}(y)dy + r^2(1-r)(\theta^\diamond + \alpha) + \\
&+ (1-r)^2(1+r) \int_{\mathbb{R}} xF_{\theta^\diamond}(y)f_{\theta^\diamond}(y)dy.
\end{aligned}$$

In order to compute explicitly the remaining integrals, we use (11) and need an analytical approximation of the quantile function of the standard normal distribution. We use the approximation proposed by Joiner and Rosenblatt (1971) and employed by Kroll and Stedinger (1996):

$$\Phi^{-1}(p) = 5.05(p^{0.135} - (1-p)^{0.135}). \tag{17}$$

According to the results in Joiner and Rosenblatt (1971), (17) works quite well for $p \in [0.005, 0.995]$. In order to get some insight into this issue, we computed numerically the integrals above, with μ, σ and $\theta^\diamond$ equal to fixed arbitrary values, both with the normal quantile function and with (17). The results are equal to the second decimal digit.

Letting $\zeta = \Phi_{\mu,\sigma^2}(\theta^\diamond)$, we get

$$\begin{aligned}
\beta_1^a &= r^2 \int_0^1 (\mu + \sigma \cdot 5.05((y\zeta)^{0.135} - (1-(y\zeta)^{0.135})))ydy \\
&+ r(1-r)(\theta^\diamond + \alpha) + (1-r)^2 \int_0^1 (-\alpha^{-1}\log(1-y) + \theta^\diamond)ydy; \qquad (18) \\
\beta_2^a &= r^3 \int_0^1 (\mu + \sigma \cdot 5.05((y\zeta)^{0.135} - (1-(y\zeta)^{0.135})))y^2dy + r^2(1-r)(\theta^\diamond + \alpha) + \\
&+ (1-r)^2(1+r) \int_0^1 (-\alpha^{-1}\log(1-y) + \theta^\diamond)y^2dy, \qquad (19)
\end{aligned}$$

where the notation β_i^a $(i = 1, 2)$ emphasizes the fact that (18) are approximations of the β_is $(i = 1, 2)$ given in (16). All the integrals in (18) and (19) can be solved in closed form. Tedious but straightforward calculations give

$$\beta_1^a = r^2 \left(\frac{\mu}{2} + 2.3653\sigma \exp\{.135 \log(\zeta)\} - \frac{2.084\sigma(1.135\zeta^2 - 0.135\zeta - 1)(1-\zeta)^{0.135}}{\zeta^2} - \frac{2.084\sigma}{\zeta^2} \right)$$
$$+ r(1-r)(\theta^\diamond + \alpha) + (1-r)^2 \left(\frac{3}{4\alpha} + \frac{\theta^\diamond}{2} \right); \tag{20}$$

$$\beta_2^a = r^3 \left((1-\zeta)^{0.135} \left(\frac{1.3295\sigma}{\zeta^3} - 1.6108\sigma + \frac{0.1795\sigma}{\zeta^2} + \frac{0.1019\sigma}{\zeta} \right) + \frac{\mu}{3} + 1.6108\sigma\zeta^{0.135} - \frac{1.3295\sigma}{\zeta^3} \right)$$
$$+ r^2(1-r)(\theta^\diamond + \alpha) + (1-r)^2(1+r) \left(\frac{0.6111}{\alpha} + \frac{\theta^\diamond}{3} \right). \tag{21}$$

The approximate PWM estimators are finally obtained by equating (16), (20) and (21) to their estimators b_s. Unbiased estimators of β_s are given by (Landwehr et al., 1979):

$$b_s = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\cdots(i-s)}{(n-1)(n-2)\cdots(n-s)} X_{(i)}, \quad s = 0, 1, \ldots,$$

where $X_{(i)}$ is the i-th order statistic. The resulting system of nonlinear equations can be solved for σ^2, α and θ by means of standard numerical methods, and the solution $\tilde{\boldsymbol{\phi}} = (\tilde{\sigma}^2, \tilde{\alpha}, \tilde{\theta})$ is the vector of PWM estimators for the Lognormal-Pareto distribution. As outlined in Sect. 1, the estimators are asymptotically normal, but there is no formal result concerning efficiency (see Hosking, 1990).

4.2 The Lognormal-Generalized Pareto Distribution

When extending the method to the Lognormal-generalized Pareto distribution, several complications arise. The source of the difficulties is the impossibility of computing PWMs, both when using (6) and its logarithmic transformation. In the latter case it is straightforward to derive the distribution of the logarithm of the GPD and its quantile function, but the resulting integrals cannot be solved in closed form. On the other hand, when using the actual distribution (not the logarithm), the PWMs $\alpha_s = M_{1,0,s}$ of the GPD are easily computed (Hosking and Wallis, 1987) for $\xi < 1$ and it is possible to use the lognormal approximate quantile function $\exp\{5.05(p^{0.135} - (1-p)^{0.135})\}$. However, when plugging this function and the lognormal density into (11), the integral cannot be computed explicitly either. In both cases, the same problem is encountered using the definition of PWMs instead of Result 2.

The discussion above implies that PWMs can be computed for the mixture obtained by taking the logarithm of the first population (normal right-truncated at $\theta^\diamond$) and the second population as it is (GPD(ξ,τ,θ)), i.e. a two-population mixture of a normal distribution right-truncated at $\theta^\diamond$ and a GPD with scale parameter θ. More formally, we would estimate the distribution with density:

$$f(x;\sigma^2,\xi,\theta) = r\phi_{\mu,\sigma^2,\theta^\diamond}(x) + (1-r)f_3(x;\xi,\tau,\theta), \tag{22}$$

where $\phi_{\mu,\sigma^2,\theta^\diamond}$ is the Normal density (13) right-truncated at $\theta^\diamond$ and f_3 is given by (7). The support of (22) is therefore $(-\infty,\theta^\diamond]\cup[\theta,\infty)$. Unfortunately, this method is unfeasible unless θ is known, because the PWM estimators of (22) would be given by

$$a_s = \frac{1}{n}\sum_{i=1}^{n}\frac{(n-i)(n-i-1)\cdots(n-i-s+1)}{(n-1)(n-2)\cdots(n-s)}v_i,$$

where

$$v_i = \begin{cases}\log(x_{(i)}) & x_{(i)} < \theta;\\ x_{(i)} & x_{(i)} \geq \theta,\end{cases}$$

and would therefore depend on θ. To overcome this problem, we propose a variant of the mixed method introduced by Morrison and Smith (2002) and used by Ailliot et al. (2011) for fitting the Generalized Extreme Value Distribution. Consider first that, if θ were known, empirical PWMs of (22) would be immediately available. Thus, a possible solution consists in forming a profile log-likelihood that is a function of θ only. This log-likelihood is easily maximized numerically, as it is one-dimensional, and the remaining parameters are estimated via the PWM method.

Now the key to the use of the profile likelihood methodology is the existence of a 1-1 mapping between θ and $\boldsymbol{\eta} = (\sigma^2,\xi,\tau)'$. In the present setup, the mapping is defined for each partition of the observations. Suppose that, for some $1\leq j\leq n$, $\theta\in(x_{(j)},x_{(j+1)})$. This implies that, for all $\theta\in(x_{(j)},x_{(j+1)})$, the smallest j observations belong to the first population and the remaining observations to the second one. PWM estimators are uniquely defined for all θ in this interval, so that the maximum of the profile log-likelihood on $(x_{(j)},x_{(j+1)})$ is obtained by computing the empirical and theoretical PWMs a_s and α_s, equating them in order to get the PWM estimates $\tilde{\sigma}_j^2$, $\tilde{\xi}_j$ and $\tilde{\tau}_j$ as functions of θ, and maximizing this function with respect to θ, obtaining the profile MLE $\hat{\theta}_j$. The actual estimates are computed by repeating this procedure for each of the $n-p_1-p_2$ non-trivial partitions of $x_{(1)},\ldots,x_{(n)}$ and choosing the estimates obtained with the partition corresponding to the largest value of the profile log-likelihood. "Non-trivial" partitions are such that the i-th subset ($i=1,2$) contains at least p_i observations (where p_i is the number of unknown parameters of the i-th distribution), otherwise estimation is meaningless.

A step-by-step description of the first iteration of the algorithm follows:

- Partition the data in two subsets $\mathcal{A}_{11}=\{x_{(1)},\ldots,x_{(p_1)}\}$, $\mathcal{A}_{21}=\{x_{(p_1+1)},\ldots,x_{(n)}\}$;

- Use $\mathcal{A}_{11}$ and $\mathcal{A}_{21}$ to compute a_{s1} and α_{s1} ($s = 0, 1, 2$) of (22); solve for σ^2, ξ and τ the system

$$\begin{cases} a_{11} = \alpha_{11}; \\ a_{21} = \alpha_{21}; \\ a_{31} = \alpha_{31}. \end{cases}$$

  obtaining estimators $\tilde{\sigma}_1^2$, $\tilde{\xi}_1$ and $\tilde{\tau}_1$;

- Maximize the one-dimensional profile log-likelihood function $l(\theta; \tilde{\sigma}_1^2, \tilde{\xi}_1, \tilde{\tau}_1)$ with respect to θ, subject to the constraint $\theta \in [x_{(1)}, x_{(2)}]$. This function is the log-likelihood of the lognormal-GDP as a function of θ only, with $\tilde{\sigma}_1^2$, $\tilde{\xi}_1$ and $\tilde{\tau}_1$ equal to the PWM estimators just computed.

- Record the value $l(\tilde{\theta}_1; \tilde{\sigma}_1^2, \tilde{\xi}_1, \tilde{\tau}_1) = l_1$ of the maximized profile log-likelihood.

The four steps above are then repeated for all the remaining partitions, obtained by choosing θ between the next pair of observations: so, for example, the second iteration uses the partition $\mathcal{A}_{12} = \{x_{(1)}, \dots, x_{(p_1+1)}\}$, $\mathcal{A}_{22} = \{x_{(p_1+2)}, \dots, x_{(n)}\}$. Eventually, we have $n - p_1 - p_2$ values of the maximized profile log-likelihood. The estimators are $\tilde{\theta}_{j^*}$, $\tilde{\sigma}_{j^*}^2$, $\tilde{\xi}_{j^*}$ and $\tilde{\tau}_{j^*}$, where j^* is such that $l(\tilde{\theta}_{j^*}, \tilde{\sigma}_{j^*}^2, \tilde{\xi}_{j^*}, \tilde{\tau}_{j^*}) > l(\tilde{\theta}_j, \tilde{\sigma}_j^2, \tilde{\xi}_j, \tilde{\tau}_j)$ ($j \neq j^*$).

PWM estimators for the GPD distribution have been derived, for $\xi < 1$, by Hosking and Wallis (1987), who also find the asymptotic distribution and study via simulation the finite sample behavior. The asymptotic properties of the estimators based on the mixed method cannot be easily derived in this setup for the same reasons of the lognormal-Pareto case.

5 Simulation and Applications

5.1 Simulation

The main theoretical results available about the PWM-based estimators are unbiasedness and asymptotic normality (Landwehr et al., 1979; Hosking, 1990). As for MLEs, they are consistent (Lehmann and Casella, 1998, Chap. 6, Corollary 3.5), whereas it is difficult to check the conditions for asymptotic efficiency (see Sect. 3.2). Thus, it is very important to perform an extensive simulation study of the properties of the estimators. Considering that asymptotic optimality of MLEs cannot be proved in general, in this case the simulation study should be concerned with both small and large sample sizes. For the lognormal-Pareto distribution, the step-by-step description of the experiment is as follows.

- Simulate a random sample of size n from the lognormal-Pareto distribution with parameters σ^2, α and θ. We use the eight combinations of the following values of the parameters: $\sigma^2 \in \{0.25, 1\}$, $\alpha \in \{0.5, 1, 1.5, 2\}$, $\theta = 50$. As for the sample size, we employ values of $n \in \{20, 40, 60, 80, 100, 150, 200, 350, 500, 1000, 1500, 2000, 5000\}$.

- Estimate the parameters by means of the PWM-based method and by means of MLE;

For the lognormal-GPD, the experiment is identical, except for the values of the parameters, given by $\sigma^2 \in \{0.25, 1\}$, $\xi \in \{0.9, 0.75, 0.5, 0.3\}$, $\tau = 80$, $\theta = 50$.

In the first case, the number of replications is equal to 500, whereas for the lognormal-GPD, due to the heavy computational burden, it is equal to 100.

5.2 Lognormal-Pareto

Figures 1 and 2 show the Relative MSE (RelMSE) of the PWM and MLE estimators for the lognormal-Pareto model on a doubly-logarithmic scale. The RelMSE of a generic parameter δ is defined as $\text{RelMSE}(\delta) = \text{MSE}(\tilde{\delta})/\text{MSE}(\hat{\delta})$, so that a value larger (smaller) than 1 implies that MLE (PWM) is more efficient. Looking at the graps, a number of remarks are in order.

First, in general PWMs perform better for small sample sizes. However, this becomes more evident when α or/and r get larger, namely when the tail is less heavy or/and when the relative weight of the lognormal distribution is larger. As for the behavior of the estimators of the individual parameters, the RelMSE of the estimator of θ has the largest range, as PWMs are much better for small n and MLEs are preferable when n increases. On the other hand, the estimator of σ^2 has the smallest range and the estimator of α is in between. It is worth noting that, for $\alpha = 0.5$, $\hat{\sigma}^2$ is preferable even for the smallest sample sizes. Unreported results show that this is mainly due to the negative bias of $\tilde{\alpha}$, which is pronounced, for small n, when the Pareto tail is heavier. These outcomes are in line with Hosking and Wallis (1987).

Turning now to the lognormal-GPD, figures 3 and 4 are analogous to graphs 1 and 2. The results are qualitatively similar. PWMs are preferable for small sample sizes and/or small ξ, namely when the tail is lighter. It should also be noted that the RelMSE has a wider range of variation in the lognormal-GPD case, meaning that the performances of the estimators are sometimes very different. Moreover, when ξ is not too large and r is not too small, $\tilde{\theta}$ and $\tilde{\tau}$ are preferable to $\hat{\theta}$ and $\hat{\tau}$ even for the largest sample sizes.

In both models, MLEs have been computed numerically by means of the Nelder-Mead simplex algorithm. For small sample sizes, the maximization has sometimes failed. In these cases, we have also used both the BFGS and the Simulated Annealing algorithms, and tried a variety of starting

Figure 1: Relative MSE of PWMs and MLEs for the lognormal-Pareto distribution. Doubly logarithmic scale.

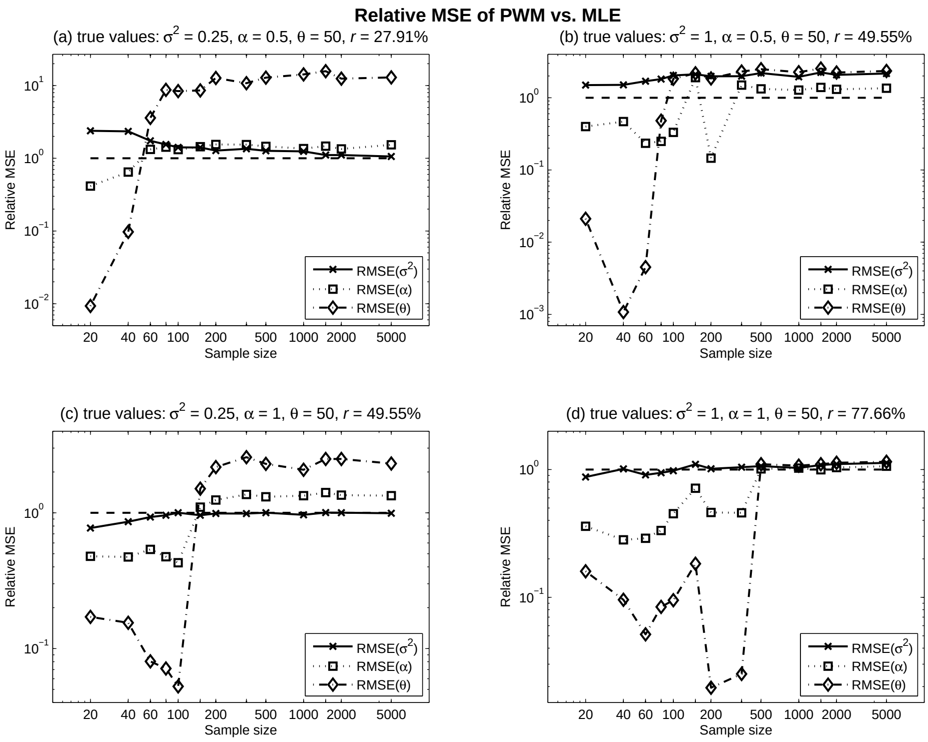


Figure 2: Relative MSE of PWMs and MLEs for the lognormal-Pareto distribution.

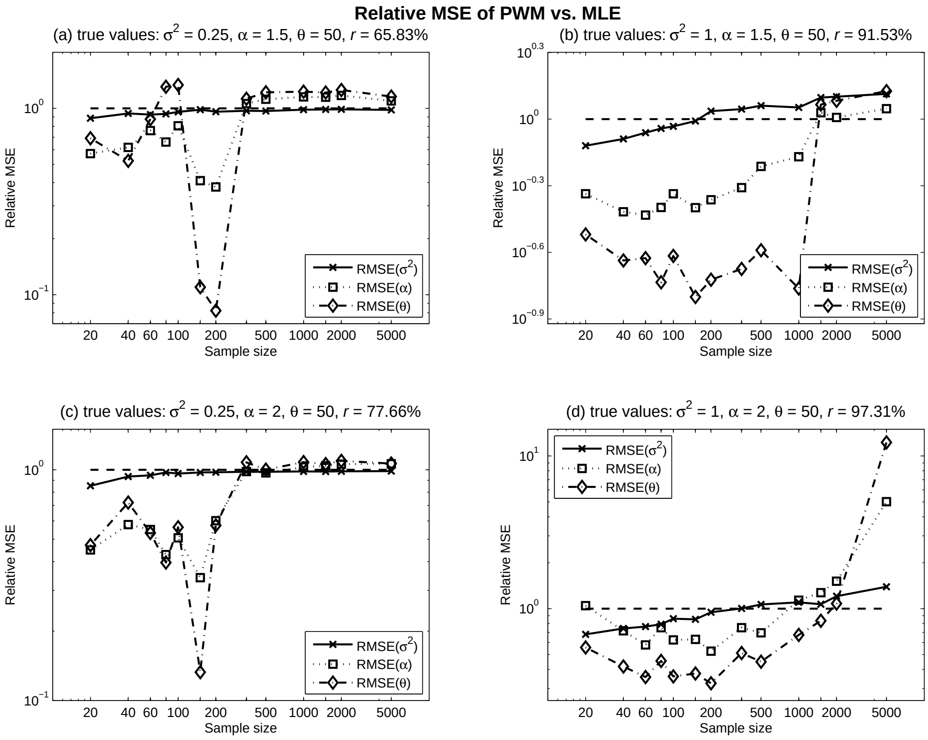


Figure 3: Relative MSE of PWMs and MLEs for the lognormal-GPD.

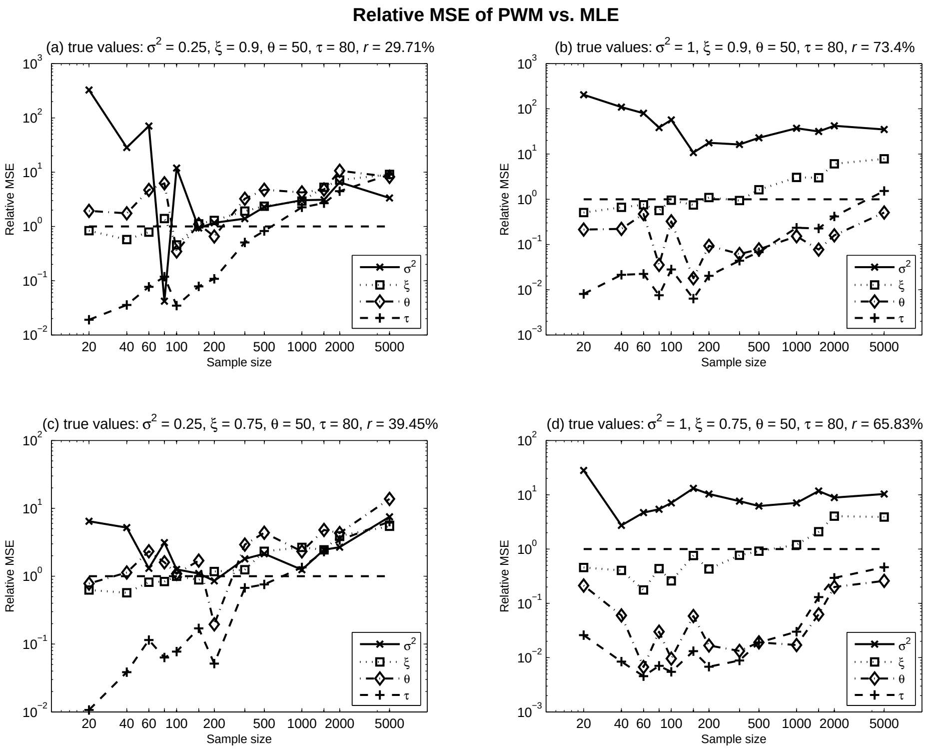


Figure 4: Relative MSE of PWMs and MLEs for the lognormal-GPD.

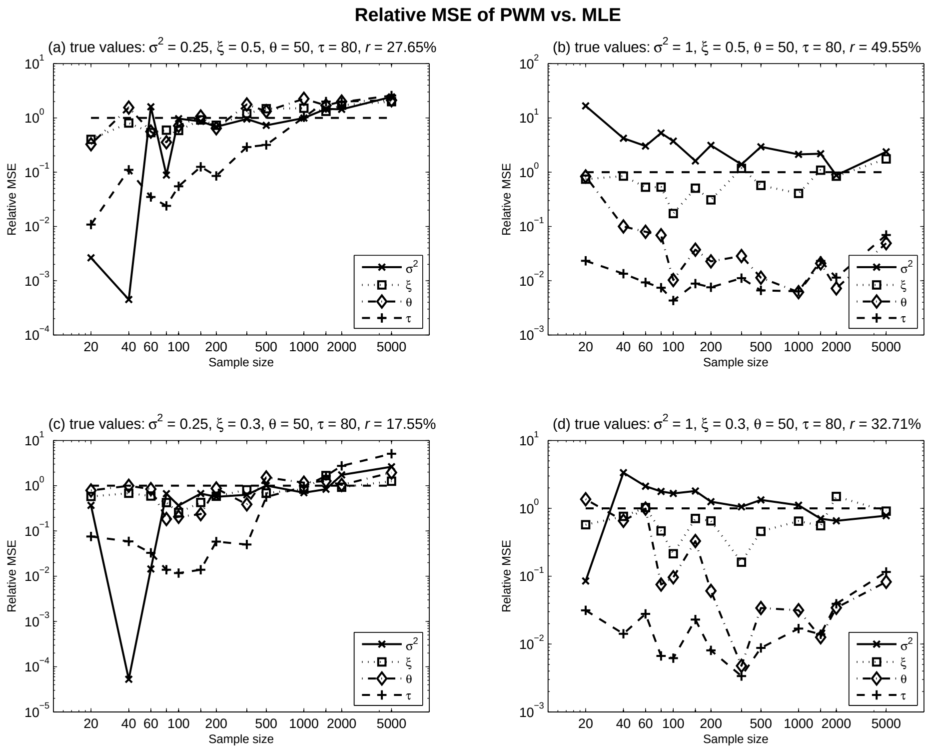
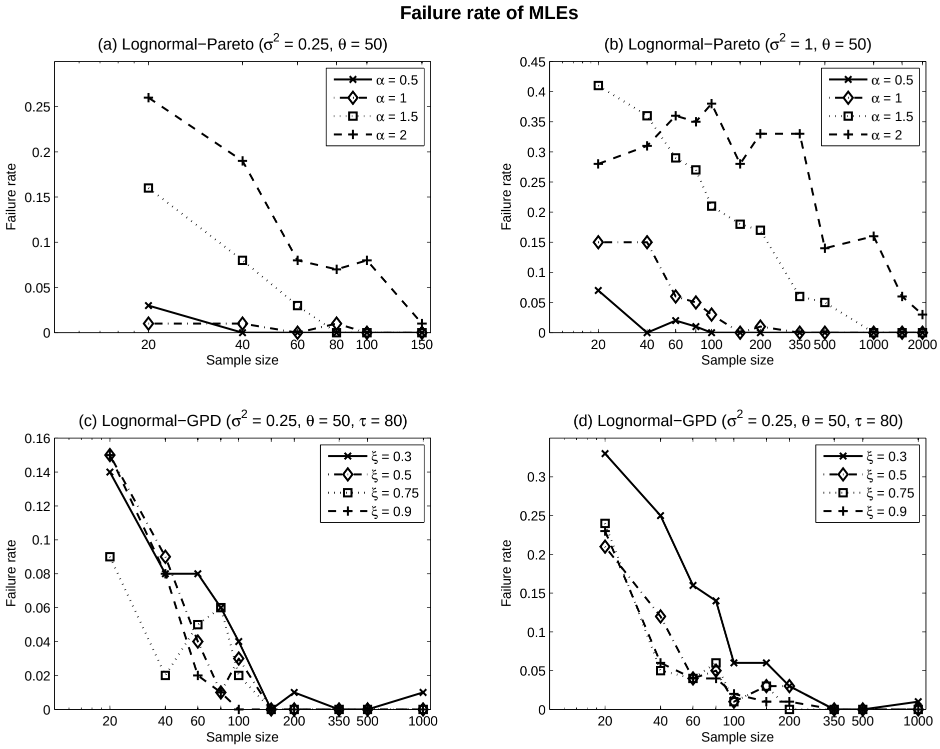


Figure 5: Failure rate of MLEs. In each panel, the largest sample size (n^*, say) reported on the x-axis is such that the failure rate is zero for all $n > n^*$, for all values of α (panels (a) and (b)) or ξ (panels (c) and (d)). The scale of the x-axis is logarithmic.



values, either equal to the sample moments and chosen randomly. In almost all cases (the percentage is approximately equal to 97%) we were unable to find a maximum. Thus, it is very likely that in the large majority of these cases the log-likelihood function does not have a local maximum. Again, similar results for the GPD have been obtained by Hosking and Wallis (1987). Fig. 5 shows the fraction of cases when MLE breaks down for the lognormal-Pareto distribution (panels (a) and (b)) and for the lognormal-GPD (panels (c) and (d)): it is clear that the problem disappears when the sample size increases.

5.3 Applications

5.3.1 Danish fire loss

The famous Danish fire loss dataset contains 2493 Danish fire insurance claims in millions of Danish Kroner (1985 prices) in the period 1980-1993 (Rytgaard, 1996). The data are now available in the R

Table 1: Estimation results for the Danish fire loss data.

Model	Estimates and standard errors	Log-lik	KS p-value
Lognormal	$\hat{\mu} = 0.672$, $\hat{\sigma}^2 = 0.537$	-4433.891	$< 10^{-5}$
	se$(\hat{\mu}) = 0.144$, se$(\hat{\sigma}^2) = 0.013$	-	-
Third PWM	$\hat{\sigma}^2 = 0.062$, $\hat{\alpha} = 1.390$, $\hat{\theta} = 1.354$	-	0.021
	se$(\hat{\sigma}^2) = 0.012$, se$(\hat{\alpha}) = 0.039$, se$(\hat{\theta}) = 0.069$	-	-
Third MLE	$\hat{\sigma}^2 = 0.039$, $\hat{\alpha} = 1.328$, $\hat{\theta} = 1.207$	-3865.864	0.013
	se$(\hat{\sigma}^2) = 0.013$, se$(\hat{\alpha}) = 0.040$, se$(\hat{\theta}) = 0.084$	-	-
Fourth PWM	$\hat{\sigma}^2 = 0.029$, $\hat{\xi} = 0.608$, $\hat{\theta} = 1.142$, $\hat{\tau} = 1.331$	-	$< 10^{-5}$
	se$(\hat{\sigma}^2) = 0.013$, se$(\hat{\xi}) = 0.080$, se$(\hat{\theta}) = 0.158$, se$(\hat{\tau}) = 0.517$	-	-
Fourth MLE	$\hat{\sigma}^2 = 0.033$, $\hat{\xi} = 0.640$, $\hat{\theta} = 1.145$, $\hat{\tau} = 0.965$	-3860.471	0.162
	se$(\hat{\sigma}^2) = 0.013$, se$(\hat{\xi}) = 0.041$, se$(\hat{\theta}) = 0.085$, se$(\hat{\tau}) = 0.033$	-	-

package SMPracticals. The different lognormal-Pareto and lognormal-GPD composite models have been fitted to these data via MLE by Scollnik (2007). However, the dataset he used is not exactly the same, as he only had the 2168 observations larger than 1, and simulated the remaining 324 observations. This explains why our results are slightly different.

According to the evidence in Scollnik (2007), the best models are the third and fourth lognormal-Pareto models. Thus, we fit these two distributions via both MLE and PWM; for comparison purposes, we also fit the lognormal distribution. Table 1 shows the estimated parameter values, the standard errors computed by means of non-parametric bootstrap, the maximized log-likelihood and the value of the two-sample Kolmogorov-Smirnov (KS) test. The KS test suggests that the best model is fourth MLE. However, both PWM and MLE lognormal-Pareto models perform similarly. On the other hand, the lognormal distribution is clearly inadequate. Fourth PWM also has a poor performance, mostly because of an inaccurate estimate of τ, clearly measured by the large standard error. Similar conclusions can be drawn from the left panel of Figure 6, which shows the Q-Q plots of the empirical distribution and the estimated distributions.

5.3.2 Operational risk data

The second application uses operational risk data. The data are the amounts (in Euros) of the losses recorded in a certain business line in a recent year in Banca Intesa Sanpaolo (Italy), and have been rescaled for confidentiality reasons. The sample size is $n = 224$. The results are reported in Table 2.

Note first that model fourth PWM has not been used, because ξ is larger than 1 and therefore PWMs do not exist. Second, the performance of the remaining four models, as measured by the

Figure 6: Q-Q plots. Empirical distribution and estimated distributions of Danish fire data (left panel) and of operational risk data (right panel).

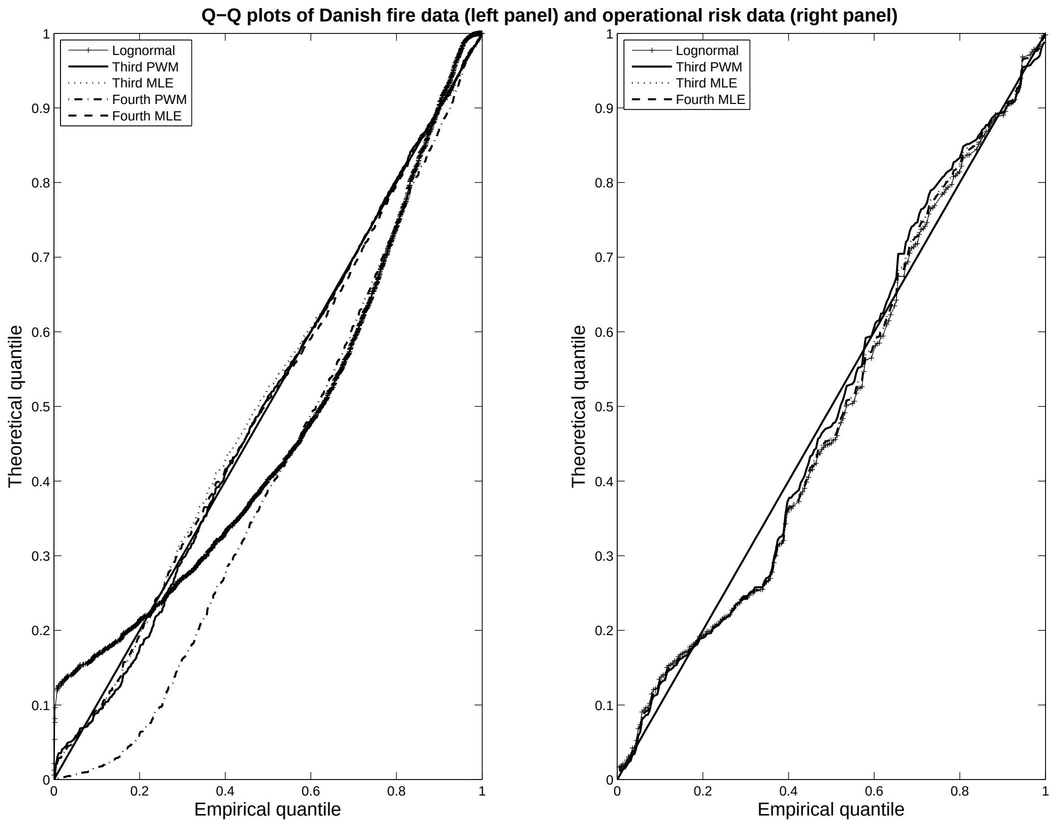


Table 2: Estimation results for operational risk data.

Model	Estimates and standard errors	Log-lik	KS
Lognormal	$\hat{\mu} = 9.916$, $\hat{\sigma}^2 = 4.292$	-2701.667	0.060
	$\text{se}(\hat{\mu}) = 2.777$, $\text{se}(\hat{\sigma}^2) = 0.381$	-	-
Third PWM	$\hat{\sigma}^2 = 3.405$, $\hat{\alpha} = 0.655$, $\hat{\theta} = 149354.26$	-	0.085
	$\text{se}(\hat{\sigma}^2) = 0.487$, $\text{se}(\hat{\alpha}) = 0.127$, $\text{se}(\hat{\theta}) = 1271.50$	-	-
Third MLE	$\hat{\sigma}^2 = 3.931$, $\hat{\alpha} = 0.839$, $\hat{\theta} = 510128.162$	-2701.657	0.060
	$\text{se}(\hat{\sigma}^2) = 0.994$, $\text{se}(\hat{\alpha}) = 2.269$, $\text{se}(\hat{\theta}) = 3222.98$	-	-
Fourth MLE	$\hat{\sigma}^2 = 3.981$, $\hat{\xi} = 1.069$, $\hat{\theta} = 242784.397$, $\hat{\tau} = 306371.659$	-2701.524	0.058
	$\text{se}(\hat{\sigma}^2) = 1.045$, $\text{se}(\hat{\xi}) = 0.654$, $\text{se}(\hat{\theta}) = 990.73$, $\text{se}(\hat{\tau}) = 728.32$	-	-

KS test and by the maximized log-likelihood (where available), is similar. The only model that performs a little better is the lognormal-Pareto with PWM estimators. Figure 6 (b) confirms these comments; in general, the fit is not as good as in the preceding application, mainly because of the smaller sample size.

Despite the similar goodness of fit, the estimated parameters of the composite models are quite different. In particular, third PWM finds a Pareto tail containing 44 observations (as there are 44 observations larger than 149 354.26), whereas third MLE and fourth MLE find a Pareto tail containing respectively just 2 and 27 observations. In conclusion, third PWM seems to be preferable, and modelling the largest observations by means of the Pareto distribution provides a slightly better fit. Finally, the standard errors of the third PWM estimator of σ^2 is a little larger than the corresponding lognormal standard error, but the standard errors of the estimators of σ^2 and, more markedly, of α, are smaller when using PWM than when using MLE.

6 Conclusion

In this paper we have studied parameter estimation for the lognormal-Pareto and the lognormal-Generalized Pareto distributions. These models are mixture distributions, but have several non-standard features that jeopardize most typical properties of MLEs. In particular, except when the threshold is known, which is a case of limited practical interest, not only standard (Lehmann and Casella, 1998, Chap. 6, Theorem 3.10) but also Daniels' (1961) conditions for asymptotic efficiency of MLEs are in general not satisfied, so that nothing can be said about asymptotic optimality.

For both models, we propose an approach based on PWMs. We obtain approximate theoretical PWMs for the lognormal-Pareto distribution; as for the lognormal-GPD, we have to resort to the mixed method that combines MLE and PWM. In both models, an advantage of PWM-based methods is that, unlike MLEs, they can be computed even for the smallest sample sizes. On the other hand, in the lognomal-GPD, their applicability is restricted to the case where the shape parameter is smaller than 1.

We find, by means of simulations, that PWMs are in general preferable to MLEs for small to intermediate sample sizes. We are also able to identify situations when the individual parameters should be estimated by a specific method. In particular, for the lognormal-Pareto distribution, PWM estimators of α and θ are always preferable for $n < 50$ and almost always preferable for $n < 200$. For the lognormal-GPD, similar considerations hold true for ξ, θ and τ, but, when ξ is small, PWMs of θ and τ are better also for the largest sample sizes. On the other hand, in both models, the MLE estimator of σ^2 is mostly more precise; when α is small (or ξ is large), this happens not only for large sample size, but also when n is small.

Acknowledgements. The author would like to thank D.P.M. Scollnik for a useful discussion about the data and the method used in the Danish fire loss example and an anonymous referee for valuable comments that greatly helped to improve the contents of this paper.

References

Ailliot, P., Thompson, C., Thomson, P., 2011. Mixed methods for fitting the GEV distribution. Water Resources Research 47, W05551.

Bee, M., 2006. Estimating the parameters in the loss distribution approach: How can we deal with truncated data? In: Davis, E. (Ed.), The Advanced Measurement Approach to Operational Risk. Risk Books, pp. 123–144.

Bee, M., Riccaboni, M., Schiavo, S., Aug 2011. Pareto versus lognormal: A maximum entropy test. Physical Review E 84, 026104.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Daniels, H., 1961. The asymptotic efficiency of a maximum likelihood estimator. In: Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. Vol. 1. University of California press, pp. 151–163.

Dempster, N., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B 39, 1–38.

Eeckhout, J., 2004. Gibrat's law for (all) cities. American Economic Review 94 (5), 1429–51.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer.

Gabaix, X., 2009. Power laws in economics and finance. Annual Review of Economics 1, 255–93.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., 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.

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

Hosking, J. R. M., Wallis, 1987. Parameter and quantile estimation for the generalized pareto distribution. Te 29, 339–349.

Hosking, J. R. M., Wallis, J. R., Wood, E. F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27, 251–261.

Joiner, B. L., Rosenblatt, J. R., 1971. Some properties of the range in samples from Tukey's symmetric lambda distributions. Journal of the American Statistical Association 66, 394–399.

Klugman, S. A., Panjer, H. H., Willmot, G. E., 2004. Loss models: from data to decisions, 2nd Edition. Wiley.

Kroll, C. N., Stedinger, J. R., 1996. Estimation of moments and quantiles using censored data. Water Resources Research 32, 1005–1012.

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

Lehmann, E. L., Casella, G., 1998. Theory of Point Estimation, 2nd Edition. Springer.

Li, C., Singh, V., Mishra, A., 2012. Simulation of the entire range of daily precipitation using a hybrid probability distribution. Water Resources Research 48, 1–17.

Malevergne, Y., Sornette, D., 2006. Extreme financial risks: from dependence to risk management. Springer.

McLachlan, G., Krishnan, T., 2008. The EM Algorithm and Extensions, 2nd Edition. Wiley.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, Tools. Princeton University Press.

Morrison, J. E., Smith, J. A., 2002. Stochastic modeling of flood peaks usng the generalized extreme value distribution. Water Resources Research 38, 1305.

Panjer, H. H., 2006. Operational risk modeling analytics. Wiley.

Pareto, V., 1896. Cours d'economie politique. F. Rouge.

Redner, R., Walker, H., 1984. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review 26, 195–239.

Rozenfeld, H., Rybski, D., Gabaix, X., Makse, H., 2011. The area and population of cities: New insights from a different perspective on cities. American Economic Review 101, 2205–2225.

Rytgaard, M., 1996. Simulation experiments on the mean residual life function $m(x)$. In: Proceedings of the XXVII ASTIN Colloquium, vol. 1. pp. 59–81.

Scollnik, D., 2007. On composite lognormal-Pareto models. Scandinavian Actuarial Journal 1, 20–33.

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

Sornette, D., 2004. Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: concepts and tools. Springer Verlag.

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

2011.1 *Leaving home and housing prices. The Experience of Italian youth emancipation,* Francesca Modena e Concetta Rondinelli.

2011.2 *Pareto versus lognormal: a maximum entropy test,* Marco Bee, Massimo Riccaboni e Stefano Schiavo.

2011.3 *Does a virtuous circle between social capital and CSR exist? A "network of games" model and some empirical evidence,* Giacomo Degli Antoni e Lorenzo Sacconi

2011.4 *The new rules of the Stability and Growth Pact. Threats from heterogeneity and interdependence,* Roberto Tamborini

2011.5 *Chinese reserves accumulation and US monetary policy: Will China go on buying US financial assets?* Luigi Bonatti e Andrea Fracasso

2011.6 *Taking Keller seriously: trade and distance in international R&D spillovers*, Andrea Fracasso e Giuseppe Vittucci

2011.7 *Exchange Rate Exposure under Liquidity Constraints,* Sarah Guillou e Stefano Schiavo

2011.8 *Global Networks of Trade and Bits,* Massimo Riccaboni, Alessandro Rossi e Stefano Schiavo

2011.9 *Satisfaction with Creativity: A Study of Organizational Characteristics and Individual Motivations*, Silvia Sacchetti e Ermanno Tortia

2011.10 *Do Monetary Incentives and Chained Questions Affect the Validity of Risk Estimates Elicited via the Exchangeability Method? An Experimental Investigation*, Simone Cerroni, Sandra Notaro e W. Douglass Shaw

2011.11 *Measuring (in)security in the event of unemployment: are we forgetting someone?* Gabriella Berloffa e Francesca Modena

2011.12 *The firm as a common. The case of accumulation and use of common resources in mutual benefit organizations,* Ermanno C. Tortia

2012.1 *The implications of the elimination of the multi-fibre arrangement for small remote island economies: A network analysis,* Shamnaaz B. Sufrauj

2012.2 Post Mortem *Examination of the International Financia Network,* Matteo Chinazzi, Giorgio Fagiolo, Javier A. Reyes, Stefano Schiavo

2012.3 *International R&D spillovers, absorptive capacity and relative backwardness: a panel smooth transition regression model*, Andrea Fracasso and Giuseppe Vittucci Marzetti

2012.4 *The Impact of Financialization on the WTI Market*, Sergio Galli Lazzeri

2012.5 *Business change in Italian regions. A spatial shift-share approach to plant-level data*, Giuseppe Espa, Danila Filipponi, Diego Giuliani and Davide Piacentino

2012.6 *A Trick of the (Pareto) Tail*, Marco Bee, Massimo Riccaboni and Stefano Schiavo

2012.7 *A Germans' dilemma: save the euro or preserve their socio-economic model*, Luigi Bonatti and Andrea Fracasso

2012.8 *Statistical analysis of the Lognormal-Pareto distribution using Probability Weighted Moments and Maximum Likelihood*, Marco Bee

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO