UNIVERSITÀ DEGLI STUDI DI TRENTO

DIPARTIMENTO DI ECONOMIA

Importance Sampling for Sums of Lognormal Distributions, with Applications to Operational Risk

Marco Bee

Discussion Paper No. 28, 2007

The Discussion Paper series provides a means for circulating preliminary research results by staff of or visitors to the Department. Its purpose is to stimulate discussion prior to the publication of papers.

Requests for copies of Discussion Papers and address changes should be sent to:

Dott. Luciano Andreozzi
E.mail luciano.andreozzi@economia.unitn.it
Dipartimento di Economia
Università degli Studi di Trento
Via Inama 5
38100 TRENTO ITALIA

Importance Sampling for Sums of Lognormal Distributions, with Applications to Operational Risk

Marco Bee
*Department of Economics, University of Trento**

Abstract. In this paper we estimate tail probabilities for the sum of Lognormal distributions. We propose to use a defensive mixture, and develop a method of finding the optimal density via the EM algorithm; we also consider the technique which assumes the importance sampling density to belong to the same parametric family of the distribution of the random variables to be summed. Optimality is defined in terms of minimal Cross-Entropy. Several simulation experiments show that the defensive mixture has the best performance. Finally, we study the compound distribution framework, and present a real-data application concerning the Poisson-Lognormal compound distribution.

Keywords. Tail Probability, Importance Sampling, Cross-Entropy, Defensive Mixtures, Compound Distributions.

*Postal address: via Inama 5, 38100 Trento - Italy. Phone: +39-0461-882296. Fax: +39-0461-882222. E-mail: marco.bee@economia.unitn.it

1 Introduction

In the last three decades, Monte Carlo (MC) simulation has become a very popular tool. This success can be traced back to at least two reasons: first, the widespread availability of cheap computing power resulted in more feasible execution times; second, the development of models where no analytical solutions exist has forced researchers to resort to computer-intensive methodologies. The latter remark also applies to the case when only asymptotic approximations exist, and the small sample behavior of estimators or test statistics must be investigated by means of stochastic simulation.

In some cases, however, standard MC is not the most appropriate tool; this happens, for example, when we deal with rare events. Suppose to be interested in the estimation of the probability $p = P(X \geq c)$ for some "large" threshold c. The crude MC approach consists in simulating N observations $x_1, \ldots, x_N$ from the distribution of X and computing the estimate

$$\hat{p}^{MC} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{x_i \geq c\}}, \tag{1}$$

where $\mathbf{1}_A$ is the indicator function of the set A. If c is the α quantile of the distribution, we expect $\sum_{i=1}^{N} \mathbf{1}_{\{X_i \geq c\}}$ to contain only $N \cdot (1-\alpha)$ non-zero summands, so that, for very large values of α, N must be huge to get an estimator with good properties. Formally, this can be seen by means of the standard efficiency measure for estimators of rare events probabilities, namely the relative error $\tau =: \sqrt{\text{var}(\mathbf{1}_{\{X_i \geq c\}})}/E(\mathbf{1}_{\{X_i \geq c\}})$; see, for, example, Asmussen and Binswanger (1997). For the crude MC estimator the relative error diverges as $p \to 0$:

$$\tau = \frac{\sqrt{p(1-p)}}{p} \approx \frac{1}{\sqrt{p}} \overset{p \to 0}{\to} \infty.$$

The difficulty becomes even more relevant when we are interested in estimating the moments of the conditional distribution of X, as the following example shows.

Example 1. Consider the estimation of $\mu_c = E(X|X \geq c)$ when X has the standard Lognormal distribution, i.e. $X \sim \text{Logn}(0,1)$. The crude MC estimator is given by

$$\hat{\mu}_c^{MC} = \frac{1}{\#\{X_i \geq c\}} \sum_{i=1}^{N} X_i \mathbf{1}_{\{X_i \geq c\}}, \tag{2}$$

where $X_1, \ldots, X_N$ is a random sample from X. Figure 1a and 1b show the MC approximations of the variance and MSE of the estimator $\hat{\mu}_c^{MC}$ as a function of c, obtained by repeating $B = 500$ times the following two steps:

1. simulate $N = 1000000$ standard lognormal random numbers; notice that such a large value of N is necessary in order to get a reasonable estimate of the variance for the largest values of c considered in the example;

Figure 1b: Variance of the Monte Carlo estimator of E(X|X>c)

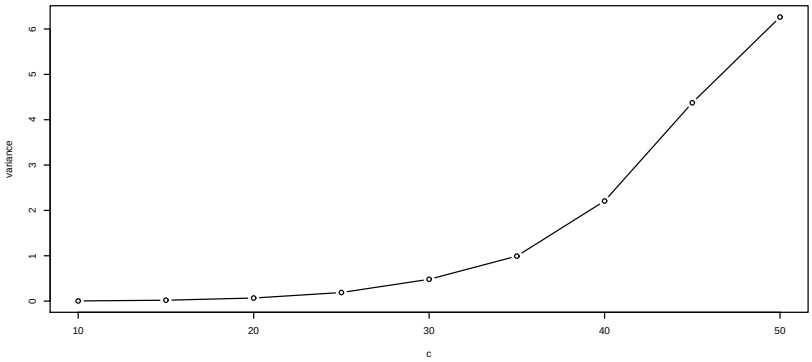
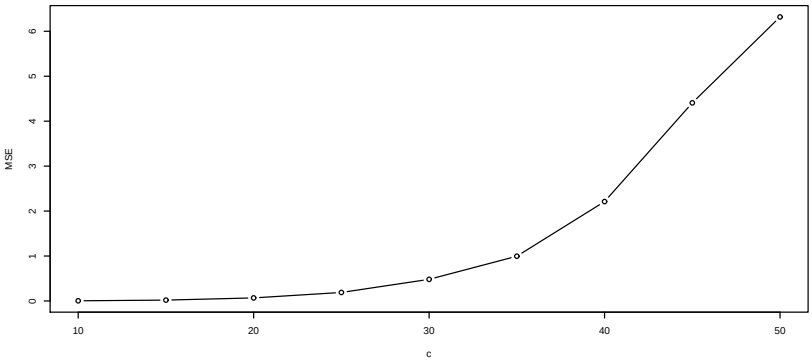


Figure 1a: MSE of the Monte Carlo estimator of E(X|X>c)



2. use (2) to compute the MC estimate of μ_c;

Then, we computed the mean and variance, respectively given by $\hat{\bar{\mu}}_c = \sum_{i=1}^{B} \hat{\mu}_c^{(i)}/B$ and $\text{var}(\hat{\mu}_c) = \sum_{i=1}^{B}(\hat{\mu}_c^{(i)} - \hat{\bar{\mu}}_c)^2/B$, where $\hat{\mu}_c^{(i)}$ is the i-th estimate ($i = 1, \ldots, B$) obtained in the simulation procedure; finally, the bias is equal to $b(\hat{\mu}_c) = \hat{\bar{\mu}}_c - \mu_c$. In the lognormal case we indeed know from elementary probability theory the true conditional expectation:

$$\mu_c = E(X|X \geq c) = e^{\mu+\sigma^2/2} \cdot \frac{1 - \Phi(\alpha - \sigma)}{1 - \Phi(\alpha)},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the density and the distribution function of the standard normal distribution and $\alpha = (\log(c) - \mu)/\sigma$. It follows that in this example $E(X|X \geq c) = \exp\{1/2\}(1 - \Phi(\log(c) - 1))/(1 - \Phi(\log(c)))$.

It can be seen that the MSE increases sharply as c gets larger; in section 3 we will show that the importance sampling estimator has much more favorable properties. ∎

Importance sampling (Hammersley and Handscomb 1964; see Casella and Robert 2004, sect. 3.3 or Glasserman 2003, sect. 4.6 for reviews) is a very powerful variance reduction technique. The setup is based on standard MC, and Importance Sampling (IS) shares indeed with it many desirable

properties; the main difference is that IS does not simulate observations from the probability measure of interest P, but from an instrumental probability measure $\tilde{P}$ which assigns "more weight" to the event of interest. Not surprisingly, the features of $\tilde{P}$ are of crucial importance, because they can result in either large efficiency gains or very bad estimators. Thus, IS is commonly considered one of the most effective but also most complex variance reduction techniques.

The theory of IS is well established for light-tailed distributions: the IS density is usually based on an exponential change of measure (more concretely, on the moment generating function), and efficiency results are known, at least asymptotically. On the other hand, when working with heavy-tailed distributions, no all-purpose recipe for finding the importance sampling probability measure is available, so that it is impossible to find general procedures of implementation; moreover, from the theoretical point of view, the lack of exponential moments makes it difficult to develop limit results. Given these premises, it is not surprising that the properties of the estimators must often be studied numerically.

In this article we will derive an IS technique for the computation of tail probabilities and expectations conditional on rare events when the distribution of interest is a sum of *iid* lognormal distributions. We will also extend the IS solution found in this setup to the estimation of the quantiles of a random sum, for example in a compound Poisson-Lognormal setup. This model is frequently employed in the analysis of the total loss distribution, in particular in actuarial and operational risk applications. In the latter field, IS can be of paramount importance, because the computation of risk measures requires (also for regulatory prescriptions contained in the New Basel Accord on Banking Supervision; see Basel Committee on Banking Supervision 2005) the estimation of extreme quantiles, for which crude MC suffers of the drawbacks shown in example 1 above.

The rest of the paper is organized as follows: section 2 reviews the basic methodology of importance sampling; section 3 gives some details about heavy-tailed distributions and derives the optimal parameters of the IS density for the estimation of the tail probability of a finite sum of lognormal distributions; section 4 gives the results of several simulation experiments aimed at verifying the properties of the estimators; section 5 applies the technique to a Poisson-Lognormal compound distribution and computes tail probabilities in an operational risk setup; section 6 concludes and discusses some directions for future research.

2 Importance Sampling

Most of the times, MC simulation is devoted to the computation of a definite integral. Consider a random variable defined on some probability space $(\Omega, \mathcal{F}, P)$ and assume that it is absolutely continuous with density $f(\cdot)$; moreover, let $h(\cdot)$ be a known function. In such a setup, suppose to be interested in evaluating the following integral:

$$\mu = E(h(X)) = \int_{\mathbb{R}} h(x) f(x) dx. \tag{3}$$

Sometimes, the density of X is too complex, and the analytical evaluation of (3) is difficult or even impossible. A readily available solution consists in using MC simulation, which proceeds as follows:

- Simulate $x_1, \ldots, x_N$ independently from $f(\cdot)$;
- Compute $\hat{\mu}^{MC} = (1/N) \sum_{i=1}^{N} h(x_i)$.

The convergence of the estimator $\hat{\mu}^{MC}$ follows from the strong law of large numbers, but in some cases the rate of decrease of the variance of the estimator is very low and the achievement of the desired precision level requires an extremely large N.

As was seen in the introduction, MC simulation is inefficient when dealing with rare events, because in such a case most simulated values do not contribute to the estimation of the quantity of interest and are just discarded. Intuitively, to obtain an improvement with respect to standard MC, we need to simulate from an instrumental distribution which "assigns a larger probability" (in a sense to be made more precise later) to the area where the events we are interested in tend to take place. We also need a "compensation" which weights the generated numbers according to the ratio between the true and instrumental probability distributions. This way of reasoning is the basis of importance sampling.

To formalize the methodology, note that (3) has the following alternative representation:

$$\mu = E_f(h(X)) = \int_{\mathcal{X}} h(x) \frac{f(x)}{g(x)} g(x) dx,$$

where $g(\cdot)$ is also a density. It follows that

$$\mu = E_g \left(\frac{h(X) f(X)}{g(X)} \right), \tag{4}$$

where the expectation is taken with respect to the density $g(\cdot)$. Equation (4) provides us with another method of simulating X:

Algorithm 1 *(Importance Sampling)*

- *Simulate* $x_1, \ldots, x_N$ *independently from* $g(\cdot)$*;*
- *Compute*

$$\hat{\mu}_{IS} = \frac{1}{N} \sum_{i=1}^{N} h(x_i) \frac{f(x_i)}{g(x_i)}. \tag{5}$$

The estimator $\hat{\mu}_{IS}$ is called an *importance sampling estimator* of μ and $g(\cdot)$ is the *importance sampling density*. The ratio $r(x) = f(x)/g(x)$, usually referred to as *likelihood ratio*, can be interpreted as a weight, so that (5) is a weighted mean. For this reason, importance sampling is also known as *weighted sampling*.

Unbiasedness, consistency and asymptotic normality of the estimator $\hat{\mu}_{IS}$ follow from the asymptotic theory of standard MC, under the only condition that the support of g includes the support of f. This has the interesting implication that the choice of g can be based on efficiency

criteria; in particular, we could seek a density g^* which is easy to simulate and is such that $\text{var}_{g^*}(\hat{\mu}_{IS}) = \min_{g \in \mathcal{G}} \text{var}_g(\hat{\mu}_{IS})$, where $\mathcal{G}$ is the set of all densities such that $\text{supp}(g) \supset \text{supp}(f)$.

The problem of finding the "best" choice of g, and therefore the best importance sampling estimator $\hat{\mu}_{IS}$, can be tackled by noting that a necessary condition for the variance of (5) to exist is that

$$E_g\left(\frac{h^2(X)f^2(X)}{g^2(X)}\right) = E_f\left(\frac{h^2(X)f(X)}{g(X)}\right) = \int_{I\!R} h^2(x)\frac{f^2(x)}{g(x)}dx < \infty.$$

As pointed out by Casella and Robert (2004, sect. 3.3.2), this result implies that we should analyze carefully the behavior of $r(X)$ in the tails of the distribution, because importance sampling distributions with unbounded likelihood ratio are likely to give estimators with infinite variance and/or widely varying weights.

Loosely speaking, "good" importance sampling distributions g should have thicker tails than f. Two useful sufficient conditions for the convergence of $E_f(h^2(X)f(X)/g(X))$, which considerably restrict the choice, have been proved by Geweke (1989):

$$\frac{f(x)}{g(x)} < M \ \forall x \in \mathcal{X} \text{ and } \text{var}_f(h) < \infty; \tag{6}$$

$$f(x) < F \text{ and } g(x) > \epsilon \ \forall x \in I\!R. \tag{7}$$

The function g that minimizes the variance of the estimator is defined by the following result (Rubinstein 1981):

Theorem 1 *The choice of g that minimizes the variance of the importance sampling estimator is*

$$g^*(x) = \frac{|h(x)|f(x)}{\int_{\mathcal{X}} |h(z)|f(z)dz}.$$

Proof. See, for example, Casella and Robert (2004, pag. 95). ∎

However, this result is of little help in applications as it requires the knowledge of $\int h(x)f(x)dx$, i.e. the integral we are interested in. In view of the above considerations, the obvious question is: how do we choose the importance sampling density? Ideally, we would like to find a standard procedure which works in all setups. Solutions of this type, mainly based on tilted densities (Ross 2006), are often available when working with light-tailed distributions. For heavy-tailed distributions the problem has to be solved differently, mostly on a case-by-case basis.

3 Importance Sampling for sums of Lognormals

Given K *iid* random variables $X_1, \ldots, X_K$, the rate at which $P(X_1 + \cdots + X_K > Kc) \to 0$ for $c > E(X)$ has been thoroughly investigated and the theory is known under the name of *Large Deviations*: see, for example, Durrett (1996, sect. 1.9). Essentially all the results assume the existence of the moment generating function of X; see Mikosch and Nagaev 1998 for a review of

the problems of the Large Deviations methodology in the heavy-tailed setup. Furthermore, the problem studied by Large Deviations is somewhat different from what is being investigated in this paper, because here we are concerned with fixed values of p and K, not with an asymptotic (as $p \to 0$) tail probability.

It is well known that, although the lognormal distribution has all the moments, its moment generating function does not exist. As a consequence, both the IS procedure based on tilted densities and the theory of large deviations are of little help in this case; thus, different tools are needed both to find an IS density and to investigate the speed of convergence.

In this paper we tackle the problem by means of *Defensive Mixtures* (DM; Hesterberg 1995; see also Davison and Hinkley 1997, pag. 457): the IS density is built as a mixture of the distribution of interest itself, denoted as $X \sim \text{Logn}(\mu, \sigma_1^2)$, and another lognormal distribution with a larger expected value: $X_1 \sim \text{Logn}(\mu_1, \sigma_2^2)$, with $\mu_1 = \mu + t$, $t \in \mathbb{R}^+$, so that it is given by

$$g(x) = \pi f(x; \mu, \sigma_1^2) + (1 - \pi) f_1(x; \mu + t, \sigma_2^2), \tag{8}$$

where f and f_1 are respectively the density of X and X_1. As pointed out by Hesterberg (1995), this approach has the advantage of providing us with weights bounded above by $1/\pi$; furthermore, it is quite clear that conditions (6) and (7) are satisfied. On the other hand, the main difficulty is that, in principle, three parameters (π, t and σ_2) have to be chosen according to some optimality criterion.

3.1 An asymptotically efficient approach

Before we focus on the Cross-Entropy minimization methodology, we look at the problem from a slightly different point of view, analyzing an asymptotically (as $p \to 0$) efficient methodology.

The Lognormal distribution is usually defined as a heavy-tailed distribution. A random variable is commonly said to be heavy-tailed if its moment generating function does not exist, but this is not a precise definition: not all the so-called heavy-tailed distributions have the same tail behavior, and this is of fundamental importance when estimating rare events probabilities. Thus we need to introduce two further characterizations: the Lognormal distribution belongs to the Maximum Domain of Attraction (MDA) of the Gumbel distribution and is subexponential. The first property is related to the limit law of normalized maxima. Referring the reader to Embrechts *et al.* (1997, sect. 3.3) for details, it turns out that only three possibilities exist: any *iid* sequence of random variables is in the MDA of the Frechet, the Weibull or the Gumbel distribution.

Roughly speaking, distributions in the Frechet MDA are "very heavy-tailed"; distributions in the Weibull MDA are light-tailed (in particular, their distribution functions have support bounded to the right); finally, the tail behavior of distributions in the Gumbel MDA is less clear-cut, ranging to moderately-heavy to light. The Lognormal distribution is moderately-heavy; see Embrechts *et al.* 1997, pag. 132-145.

Moreover, the Lognormal distribution belongs to the class of subexponential distributions (Embrechts *et al.* 1997, sect. 1.3.2 or Asmussen *et al.* 2000, sect. 2.1). We borrow from Asmussen *et al.* (2000, pag. 306) the intuitive description of the definition of subexponential distributions, which is also a fundamental feature of heavy-tailed distributions: a necessary condition for the the sum of N *iid* heavy-tailed random variables $X_1, \ldots, X_N$ to be large is that one summand is large.

For the sum of random variables belonging to the intersection of the subexponential class and of the MDA of the Gumbel distribution, Asmussen *et al.* (2000) derived some probabilistic results concerning optimality of IS estimation: they propose an IS density which satisfies the theoretical requirement of asymptotical efficiency (see e.g. Sadowsky 1993). The density is given by (Asmussen *et al.* 2000, pag. 310)

$$g_A(x) = \begin{cases} \frac{\eta}{x(\log(x))^2} & x \in (a, +\infty), \\ \gamma l(x) & x \in [0, a], \end{cases} \tag{9}$$

where $a > e$, $\eta > 0$, l is an arbitrary density on $[0, a]$ and γ is chosen so that $\int g_A(x)dx = 1$. This distribution is very heavy-tailed, at the extent that for most choices of the parameters there is a relatively large probability of obtaining huge values: for example, with $\eta = 0.25$ and $a = 2.8$, the probability of observing a value larger than 1000000 is more than 1%, and for larger values of η this probability is even higher; we ran some simulation experiments with 10000 replications and invariably observed at least a couple of overflows.

Unfortunately, as noted by Asmussen and Kroese (2006, pag. 550), the properties of the IS estimator obtained using (9) are rather poor for most parameter configurations. Where does the problem come from? Intuitively, the variance of $\hat{p}^{IS}$ based on g_A is large because the distribution of $r(X)$ is strongly concentrated around zero. As a result, the variance of $r(X)$ is small, but the IS estimator $\hat{p}^{IS}$ is almost entirely determined by the few weights which are away from zero, so that it is highly unstable.

More formally, fix $\epsilon > 0$ and let A_ϵ be the set of the x_i's such that the weights are larger than ϵ: $A_\epsilon =: \{x_i : r(x_i) > \epsilon\}$. Then $\hat{p} \approx (1/N) \sum_{x_i \in A_\epsilon} h(x_i) r(x_i)$ with $x_i \sim g_A$. The relative error is given by $\tau = \sqrt{p_{A_\epsilon}(1 - p_{A_\epsilon})}/p$, whose magnitude is ultimately determined by p_{A_ϵ}: if p_{A_ϵ} is smaller than p, the performance can be even worse than crude MC. The reason is that the IS density has a much heavier tail than the target density, so that the simulation produces many huge observations, for which both the likelihood ratio and the product $r(x)h(x)$ are essentially zero. Thus, the density (9) "puts more weight" than f on $[c, +\infty)$, but a non-negligible percentage of this weight is on the interval $[c_1, +\infty)$, where c_1 is such that $P(X > c_1) \approx 0$; this also explains why A_ϵ usually contains few observations.

In our opinion, in applications, an asymptotically bounded relative error is not the most relevant requirement of an IS estimator, because our aim consists in the estimation of tail probabilities for small but fixed values of p, so that the limiting behavior of the estimator as $p \to 0$ is less important than the properties of the estimator for the true value of p, namely the value of p to be estimated. In view of the preceding considerations we will abandon the criterion of asymptotical efficiency and

follow a different road.

3.2 Determining the optimal IS density

The most obvious approach to the selection of optimal values of the parameters in (8) would consist in choosing numerical values of these parameters which minimize the variance of $\hat{p}^{IS}$. As before, we assume to be interested in the estimation of p for some large threshold $c > E(X)$; putting $h(x) = \mathbf{1}_{\{x>c\}}(x)$, this probability can be written as

$$p = \int_c^\infty f(x)dx = \int_0^\infty h(x)f(x)dx = E_f(h(X)).$$

Notice that the optimal choice of the IS density is the same for $h(x) = \mathbf{1}_{\{x\geq c\}}$ and $h^*(x) = x\mathbf{1}_{\{x\geq c\}}$, which is such that $E(X|X > c) = \int_0^\infty h^*(x)f(x)dx$, so that it will be enough to develop the analysis with reference to $h(x)$; the same IS density can be used for the computation of the tail conditional expectation as well.

We have to find a density g such that we simulate N *iid* observations $x_1^*, \ldots, x_N^*$ from g and take their weighted average as an estimate of p, where the weight $r(x) = f(x)/g(x)$ is the likelihood ratio. Formally, the estimator is given by $\hat{p}^{IS} = (1/N)\sum_{i=1}^N h(X_i)r(X_i)$, where $X_i \sim g$.

To begin with, minimizing the variance of the estimator is equivalent to minimizing the expected value $E_f(r(X) \cap X \geq c)$: from the definition it is clear that $\hat{p}$ is the standard MC estimator of $E_g(h(X)r(X))$, and its variance is proportional to $E_g(h(X)^2r(X)^2)$. Minimizing the variance is therefore equivalent to minimizing the quantity

$$E_g(h(X)^2r(X)^2) = E_g(r(X)^2 \cap X \geq c) = E_f(r(X) \cap X \geq c). \tag{10}$$

However, there are at least two difficulties with this approach. First, evaluating (10) can be complicated because it depends on both π and t. Second, and more important, the random variable $(1/N)\sum_{i=1}^N r(X_i) = (1/N)\sum_{i=1}^N f(X_i)/g(X_i)$ with $X_i \sim g$ can have a very large variance and/or be too concentrated around zero. Thus, minimizing (10) is not enough, because it implies minimization of the variance of the estimator but can produce an estimator whose convergence to the true value is too slow for practical purposes; namely, there are cases where consistency is just formal and the estimator is biased for any reasonable sample size. Therefore, an analysis of the whole distribution of $r(X)$ is necessary: for example, if we use the IS density (9), the variance of $r(X)$ is small but the variance of $\hat{p}$ is not.

Let's now focus on the density (8); for simplicity, we put $\sigma_2^2 = \sigma_1^2 = \sigma^2$, although the EM algorithm presented below (see sect. 3.3.1) could easily be extended to the estimation of σ_2^2 as well.

The main reason for introducing such an IS density is that it guarantees bounded weights; this issue, which turns out to be crucial in the sum of Lognormals case, is indeed negligible when working with a single Lognormal distribution. To see why, take $\pi = 0$ in (8); some straightforward algebra

shows that

$$r(x) = \frac{\frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right\}}{\frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{\frac{-(\log(x)-(\mu+t))^2}{2\sigma^2}\right\}} = \exp\left\{\frac{t^2 - 2\mu t - 2t\log(x)}{2\sigma^2}\right\}.$$

This is a monotonically decreasing function, and $\lim_{x\to 0} r(x) = +\infty$. Thus, $r(x)$ is unbounded when x is small, but the only x_i's which contribute to $\hat{p}^{IS}$ are those belonging to the set $A_c = \{x_i : x_i \geq c\}$, so that $r(x)$ is bounded for $x \in A_c$, and the upper bound is $r(c)$. Thus, in the single Lognormal case, one can use an IS density of the form $\text{Logn}(\mu + t, \sigma^2)$ and minimize directly (10) with respect to t. However, in this setup one can just integrate numerically the density function, so that we do not present further details here.

Consider now the estimation of tail probabilities for the random variable $Y_k = \sum_{i=1}^k X_i$, where the X_i's are *iid* with $X_i \sim \text{Logn}(\mu, \sigma^2)$ $(i = 1, \ldots, k)$. Defining $h_k(y) = \mathbf{1}_{\{Y_k > c\}}(y)$, the probability $p_k = P(Y > c)$ is given by

$$p_k = E_{\boldsymbol{f}}(h_k(\boldsymbol{x})) = \int_{I\!R^k} h_k(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x},$$

where $\boldsymbol{f}$ is the joint density of $X_1, \ldots, X_k$, which by virtue of the hypothesis of independence is equal to $\boldsymbol{f}(x_1, \ldots, x_k) = \prod_{i=1}^k f(x_i)$. The IS approach to this problem is based on the fact that

$$p_k = E_{\boldsymbol{f}}(h(\boldsymbol{x})) = E_{\boldsymbol{g}}(h(\boldsymbol{x})r_k(\boldsymbol{x})),$$

where $r_k(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})/\boldsymbol{g}(\boldsymbol{x})$ and $\boldsymbol{g}$ is a k-variate density; the most obvious solution consists in taking $\boldsymbol{g}(x_1, \ldots, x_k) = \prod_{i=1}^k g(x_i)$. With these hypotheses, the likelihood ratio is equal to

$$r_k(\boldsymbol{x}) = \frac{\prod_{i=1}^k f(x_i)}{\prod_{i=1}^k g(x_i)}. \tag{11}$$

When implementing IS with $k > 1$, the distribution of $r_k(\boldsymbol{x})$ has a disturbing feature: the values of $r_k(\boldsymbol{x})$ which actually contribute to the estimator are the values $r_k(\boldsymbol{x}) = \prod_{i=1}^k f(x_i)/\prod_{i=1}^k g(x_i)$ with $x_1, \ldots, x_k$ such that $\sum_{i=1}^k x_i > c$. It is clear that $\sum_{i=1}^k x_i$ can be larger than c even though one or more of the x_i's is arbitrarily small; but this implies that, if $\pi = 0$, $r_k(\boldsymbol{x})$ is unbounded.

In addition to the aforementioned difficulties, the behavior of $r_k(\boldsymbol{X})$ as $k \to \infty$ is somewhat pathological, as summarized by the following theorem.

Theorem 2 *Under the hypothesis*

$$E_{\boldsymbol{g}}(|\log(\boldsymbol{f}(\boldsymbol{X})/\boldsymbol{g}(\boldsymbol{X}))|) < \infty \tag{12}$$

the following results hold true:

$$E_{\boldsymbol{g}}\left(\frac{\prod_{i=1}^k f(X_i)}{\prod_{i=1}^k g(X_i)}\right) = 1 \text{ for any } k \in \boldsymbol{N}; \tag{13}$$

$$\lim_{k\to\infty} \frac{\prod_{i=1}^k f(X_i)}{\prod_{i=1}^k g(X_i)} = 0 \text{ with } \tilde{P}\text{-probability 1.} \tag{14}$$

Proof. See Casella and Robert (2004, pag. 551) or Glasserman (2003, pag. 259).

The theorem has two implications. The first one, sometimes termed "weight degeneracy" (Casella and Robert 2004, pag. 552) consists in the fact that as k gets large the distribution of the weights becomes more and more skewed, with most weights close to zero; in the limit a huge sample size is needed to get a single non-zero weight. The variance is obviously minimized, but the estimator is useless because it is essentially computed with only one observation. For intermediate values of k the estimator is downward biased unless N is very large. On the other hand, as π decreases, the maximum of the likelihood ratio gets larger: when using a defensive mixture, it is easy to see that $\max_{\boldsymbol{x}} r_k(\boldsymbol{x}) = (1/\pi)^k$. The speed of convergence to zero is a decreasing function of π: as π gets small, there is an increasing probability of larger values of X, for which the likelihood ratio is small.

Second, a naive optimization of (10) is unfeasible, because (i) two parameters have to be found (π and t), (ii) they also depend on k and c and (iii) numerical integration over a large dimensional space is problematic. There is, however, an alternative solution, based on the concept of Cross-Entropy.

3.3 The Cross-Entropy approach

It has already been said that the crucial aspect of IS consists in finding the optimal instrumental density. We can now be a little more precise, and add that this issue is twofold: one has first to choose the parametric form of the IS density, then to define an optimality criterion and use it for finding the parameters. In this section we propose the optimality criterion of minimum Cross-Entropy (CE) and apply it to IS densities belonging to (i) the DM class and (ii) the same parametric family (i.e., Lognormal) of the density of interest.

3.3.1 The defensive mixture approach

Hesterberg (1995) finds the numerical values of the parameters of the defensive mixture mostly on the basis of heuristic considerations. However, the parameters should be found by means of some optimality criterion; thus, in this section we develop a technique, based on the EM algorithm (Dempster *et al.* 1977), for determining the parameters. The parameters are optimal in the sense that they minimize the CE between the distribution of interest and the IS distribution, i.e. the defensive mixture.

The method of CE minimization was first proposed by Rubinstein (1997); see also Rubinstein and Kroese (2004) and Asmussen *et al.* (2005). Referring the interested reader to these references for details, we start by noting, from theorem 1 above, that the optimal IS probability measure $\tilde{P}$ should be "as similar as possible" to the original probability measure conditioned on the event of interest, which we define as $P^{(c)}$. How can we measure the discrepancy? The most commonly used approach consists in using the Cross-Entropy or *Kullback-Leibler distance* (Kullback 1968) between

the two distributions:

$$D(P^{(c)}, \tilde{P}) = E^{(c)} \log \frac{P^{(c)}}{\tilde{P}}. \tag{15}$$

In the present setup, the probability measure $\tilde{P}$ is restricted to be of the form (8) and is therefore identified by parameters π and t. When, as in this case, the probability measure $\tilde{P}$ is absolutely continuous with density $f_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ is a vector of parameters, it can be shown that the problem of minimizing $D(P^{(c)}, \tilde{P})$ is equivalent to

$$\max_{\boldsymbol{\theta}} E^{(c)} \left(\sum_{i=1}^{k} \log(f_{\boldsymbol{\theta}}(X_i)) \right).$$

It is now clear that there is a close relation between entropy minimization (namely, minimization of (15)) and likelihood maximization. The log-likelihood of k observations is indeed given by

$$\sum_{i=1}^{k} \log(f_{\boldsymbol{\theta}}(x_i)) = k \int \log(f_{\boldsymbol{\theta}}(x)) P_k(dx) = -kD(P_k, P_{\boldsymbol{\theta}}) + \text{const}, \tag{16}$$

where P_k is the empirical distribution. Comparing (15) to (16), it follows that maximum likelihood results can be translated into minimum CE results by replacing P_k with $P^{(c)}$. The second fundamental result we need in the following is presented in lemma 1 below.

Lemma 1 *Let $A(c) = \{X_1 + \cdots + X_k > c\}$, where X_i's are subexponential distributions with distribution function F and let $X_{(1)}, \ldots, X_{(k)}$ be the corresponding order statistics. Then*

$$\| P(X_{(1)}, \ldots, X_{(k)} \in \cdot | A(c)) \| - \underbrace{F \otimes \cdots \otimes F}_{k-1} \otimes P(X \in \cdot | X > c) \to 0.$$

Proof. See, for example, Asmussen 2000, lemma 5.6.

In words, the lemma means that $A(c)$ occurs if $k-1$ of the X_i's have distribution F and one has the conditional distribution of X given $X > c$.

How do these results combine to provide a method for determining the optimal parameters of the IS density? We explain this issue by focusing on our setup. Recalling that maximum likelihood estimation of the parameters of a Lognormal mixture is equivalent to maximum likelihood estimation of the parameters of a normal mixture, we can just consider the MLE's of the parameters π and $\mu_2 \stackrel{\text{def}}{=} \mu + t$ of a normal mixture (see, for example, Flury 1997, sect. 9.2):

$$\hat{\pi} = \frac{1}{k} \sum_{i=1}^{k} \pi_{1i} = \int_{-\infty}^{\infty} \pi_{1i} F_k(dx), \tag{17}$$

$$\hat{\mu}_2 = \frac{1}{k(1-\hat{\pi})} \sum_{i=1}^{k} \pi_{1i} x_i = \frac{1}{1-\hat{\pi}} \int_{-\infty}^{\infty} x \pi_{1i} F_k(dx), \tag{18}$$

where F_k is the empirical distribution function and π_{1x} is the so called posterior probability of x. Defining ϕ_{μ,σ^2} and ϕ_{μ_2,σ^2} respectively as the $N(\mu, \sigma^2)$ and the $N(\mu_2, \sigma^2)$ densities, π_{1x} is given by

$$\pi_{1x} = \frac{\pi \phi_{\mu,\sigma^2}(x)}{\pi \phi_{\mu,\sigma^2}(x) + (1-\pi) \phi_{\mu_2,\sigma^2}(x)}. \tag{19}$$

Putting $c^* = \log(c)$, the conditional density of $X|X > c^*$ is equal to

$$\phi^{(c^*)}_{\mu,\sigma^2}(x) = \frac{\phi_{\mu,\sigma^2}(x)}{1 - \Phi_{\mu,\sigma^2}(c^*)} \mathbf{1}_{\{x>c^*\}},$$

where Φ_{μ,σ^2} is the $N(\mu, \sigma^2)$ distribution function. Now the values of π and μ_2 (π^* and μ_2^*, say) which minimize entropy are given by (17) and (18) with F_k replaced by $\Phi^{(c)}_{\mu,\sigma^2}$. More precisely, from lemma 1 we know that $(k-1)$ observations have density ϕ_{μ,σ^2} and one has density $\phi^{(c)}_{\mu,\sigma^2}$, so that

$$\pi = \frac{k-1}{k} \int_0^\infty \pi_{1x} \phi_{\mu,\sigma^2}(x)dx + \frac{1}{k} \cdot \frac{1}{1 - \Phi_{\mu,\sigma^2}(c^*)} \int_{c^*}^\infty \pi_{1x} \phi_{\mu,\sigma^2}(x)dx, \tag{20}$$

$$\mu_2 = \frac{1}{1-\pi} \left(\frac{k-1}{k} \int_{-\infty}^\infty x\pi_{2x} \phi_{\mu,\sigma^2}(x)dx + \frac{1}{k} \cdot \frac{1}{1 - \Phi_{\mu,\sigma^2}(c^*)} \int_{c^*}^\infty x\pi_{2x} \phi_{\mu,\sigma^2}(x)dx \right). \tag{21}$$

Now the key to the solution of the system formed by (20) and (21) consists in noting that equations (19), (20) and (21) are the equations of the EM algorithm for maximum likelihood estimation of the parameters of a random variable X distributed as a two-population normal mixture with parameters (μ, σ^2) and (μ_2, σ^2) respectively, where $(k-1)$ observations are from the mixture itself and one observation is from the distribution of $X|X \geq c^*$. In particular (19) implements the E-step, (20) and (21) the M-step. Hence, in order to get the optimal values π^* and μ_2^*, we just have to iterate (19), (20) and (21) until convergence; note that the integrals in (20) and (21) have to be solved numerically at each iteration.

3.3.2 The standard Cross-Entropy approach

In this subsection we sketch the "standard" CE approach. By "standard CE approach" (Rubinstein and Kroese 2004, sect. 2.3) we mean that (i) the IS density is chosen in the same parametric family of the variable of interest and that (ii) the CE method is used to find the optimal tilting parameter: in the present setup this implies $g \sim \text{Logn}(\mu_2, \sigma^2)$, where $\mu_2 = \mu + t$ ($t \geq \mu$). Therefore, the optimal value of t is determined by minimizing the Kullback-Leibler distance, and in this case the analogy with maximum likelihood estimation provides us with a simple solution: given k observations $y_1, \ldots, y_k$, the MLE of μ is indeed given by:

$$\mu = \frac{1}{k} \sum_{i=1}^k x_i = \int_{-\infty}^\infty x F_k(dx),$$

where $x_i = \log(y_i)$ and F_k is the empirical distribution function. Following the same way of reasoning of the preceding subsection and using the well-known expression of the expected value of the truncated normal distribution, the optimal value of μ_2 can be obtained analytically:

$$\begin{aligned} \mu_2^* &= \frac{k-1}{k} \int_{-\infty}^\infty x\phi(x)dx + \frac{1}{k} \cdot \frac{1}{1 - \Phi(c)} \int_{(c^*-\mu)/\sigma}^\infty x\phi(x) = \\ &= \frac{k-1}{k}\mu + \frac{1}{k} \left(\mu + \sigma \frac{\phi((c^*-\mu)/\sigma)}{1 - \Phi((c^*-\mu)/\sigma)} \right), \end{aligned} \tag{22}$$

where ϕ and Φ are respectively the pdf and cdf of the standard normal random variable. The optimal value μ_2^* completely determines the IS density.

3.3.3 The Adaptive Cross-Entropy approach

As pointed out by Rubinstein and Kroese (2004, pag. 38), the standard CE approach (from now on ST) does not work well if the probability of the event of interest is too small (below 10^{-5}); to overcome this difficulty, Rubinstein and Kroese (2004, sect. 3.4) propose a multilevel algorithm which is based on a two-step adaptive procedure (also called "Adaptive Cross-Entropy" - ACE) where not only the parameter μ but also the threshold c is updated at each iteration. While referring the interested reader to Rubinstein and Kroese (2004) for theoretical properties of this technique, we will give some details about the implementation to our setup in the next section.

4 Some simulation experiments

In this section we focus on the estimation of tail probabilities; results concerning tail conditional expectations are completely analogous and are therefore not reported here. Consider a random variable $Y_k = \sum_{i=1}^{k} X_i$ with $X_i \sim \text{Logn}(0,1)$ and $k = 10$. Table 1 shows the optimal values of the parameters of the IS density in the three approaches (DM, ST and ACE) presented in the preceding section. In the first case the IS density is a defensive mixture and the optimal parameters π^* and t^*_{DM} are determined by means of the EM algorithm; in the remaining two cases the IS density is a lognormal density and the optimal value of t is found respectively using (22) and the multilevel algorithm proposed by Rubinstein and Kroese (2004, sect. 3.4). The latter algorithm was implemented with $N = 10000$ MC replications and a sample quantile equal to $1-\rho = 0.99$; different values of ρ produced almost identical results. According to remark 3.9 in Rubinstein and Kroese (2004, pag. 74), we decided to iterate the algorithm ten more times after the stopping criterion has been reached.

Table 1: Optimal parameter values in the three approaches with $k = 10$ and $X_i \sim \text{Logn}(0,1)$.

	π^*	t^*_{DM}	t^*_{ST}	t^*_{ACE}
$c = 65$	0.893	4.265	0.439	0.636
$c = 80$	0.895	4.489	0.459	0.540
$c = 100$	0.896	4.726	0.481	1.251
$c = 150$	0.898	5.148	0.520	1.150
$c = 200$	0.899	5.441	0.548	2.043
$c = 300$	0.899	5.850	0.587	2.287
$c = 400$	0.900	6.137	0.615	2.455
$c = 500$	0.900	6.359	0.637	2.366

From the table it can be seen that in the DM approach the optimal value of π remains almost constant as c increases, but the tail of the IS density becomes heavier because t_{DM} increases. As for

the two standard approaches, in the adaptive approach the parameter t_{ACE} is much more sensitive to c than in the standard CE approach.

As pointed out above, the performance of any IS estimator is related to the features of the distribution of $r_k(\boldsymbol{X})$. Table 2 gives some details about this distribution in the DM, ST and ACE approaches with $c \in \{65, 80, 100, 150, 200, 300, 400, 500\}$ and $k = 10$.

Table 2: Descriptive statistics about the distribution of $r_{k,DM}$, $r_{k,ST}$ and $r_{k,ACE}$ with $k = 10$; "max^c" is the maximum of the weights corresponding to observations such that $\sum_{i=1}^{k} x_i > c$, "$\bar{r}^c$" is the average of the observations such that $\sum_{i=1}^{k} x_i > c$, "f^c" is the proportion of the observations such that $\sum_{i=1}^{k} x_i > c$ with respect to the number of simulations.

	$c = 65$	$c = 80$	$c = 100$	$c = 150$	$c = 200$	$c = 300$	$c = 400$	$c = 500$
$\bar{r}_{DM}$	0.94	0.99	0.87	0.96	1.02	1.02	1.09	0.94
$\bar{r}_{ST}$	1.05	1.02	1.03	1.39	0.98	0.87	0.84	0.81
$\bar{r}_{ACE}$	0.87	0.17	1.67	0.20	0.31	$9 \cdot 10^{-3}$	$4 \cdot 10^{-5}$	$7 \cdot 10^{-5}$
$\max_{DM}$	3.10	3.04	3.00	2.94	2.91	2.89	2.88	2.88
$\max_{ST}$	26.16	37.49	57.15	460.03	114.28	34.47	26.02	55.20
$\max_{ACE}$	52.78	22.96	1112.88	65.41	129.47	4.97	0.04	0.03
$\bar{r}^c_{DM}$	10^{-3}	$3 \cdot 10^{-4}$	10^{-4}	10^{-5}	$2 \cdot 10^{-4}$	$2 \cdot 10^{-7}$	$4 \cdot 10^{-8}$	$8 \cdot 10^{-9}$
$\bar{r}^c_{ST}$	0.15	0.07	0.01	-	-	-	-	-
$\bar{r}^c_{ACE}$	0.11	$4 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	$2 \cdot 10^{-9}$	$4 \cdot 10^{-10}$	$8 \cdot 10^{-14}$	$7 \cdot 10^{-7}$	$4 \cdot 10^{-16}$
$\max^c_{DM}$	0.03	$9 \cdot 10^{-3}$	$7 \cdot 10^{-3}$	$2 \cdot 10^{-4}$	$4 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	$5 \cdot 10^{-7}$	10^{-7}
$\max^c_{ST}$	0.77	0.14	0.01	-	-	-	-	-
$\max^c_{ACE}$	1.47	0.03	$8 \cdot 10^{-4}$	$5 \cdot 10^{-4}$	$7 \cdot 10^{-9}$	$9 \cdot 10^{-10}$	10^{-12}	$2 \cdot 10^{-15}$
f^c_{DM}	.532	.509	.532	.512	.471	.471	.433	.471
f^c_{ST}	.007	.004	.002	.001	0	0	0	0
f^c_{ACE}	.017	.115	.047	.016	.004	.004	.033	.005

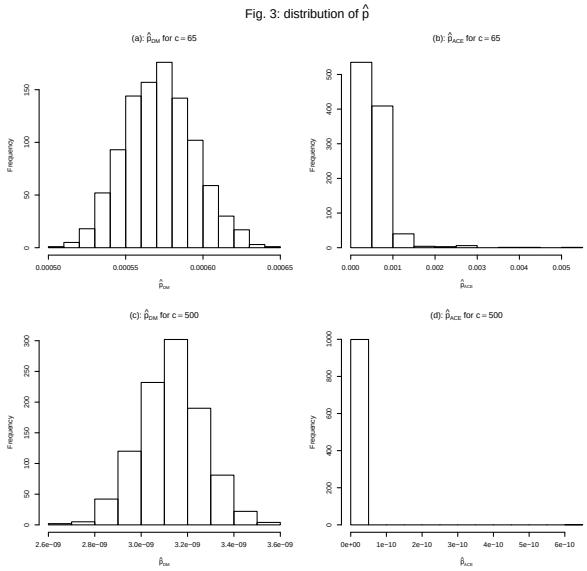
The results show quite clearly that the distribution of $r_{k,DM}$ has much more desirable properties than the distributions of $r_{k,ST}$ and $r_{k,ACE}$: in particular, as expected, in the ST approach the average of the weights remains approximately stable as c grows, but the number of observations exceeding c is extremely small, in particular for large c. Somewhat more surprisingly, we see that $r_{k,ACE}$ has quite poor properties as well: in particular, the average of the weights tends to zero as c increases, and the number of observations actually used for the computation of the tail probability becomes very small as c gets large; these features are typical of the weight degeneracy mentioned in section 3.2. The distribution of $r_{k,DM}$, on the other hand, maintains the same satisfactory features for any c. The different properties of the distributions of $r_{k,DM}$ and $r_{k,ACE}$ are clearly related to the fact that the weights in the DM case are bounded above, whereas in the ACE approach are not. As we are now going to see, these results have a strong impact on the properties of the resulting estimator.

In the same setup used above for the simulation of $r_k(\boldsymbol{X})$, we investigate the properties of the three estimators of $p = P(Y_k > c)$ with $N = 10000$. To assess the stability of the estimators, we repeated the simulation $B = 1000$ times and computed the MC estimate of the standard error of the estimator $\hat{\text{se}}(\hat{p}) = (1/B)\text{var}(p^{(i)})$, where $p^{(i)}$ $(i = 1, \ldots, B)$ is the estimate obtained at the i-th replication. We immediately notice that the ACE estimator is downward biased for the largest values of c; therefore, neither the standard deviation nor the relative error are good measures of performance, because both of them are only appropriate when the estimator is unbiased; as a consequence we estimated a version of the relative error (we will call it "MSE Relative Error") based on the MSE instead of the variance: $\tau_{MSE} \stackrel{\text{def}}{=} \sqrt{\text{MSE}(\mathbf{1}_{\{X_i \geq c\}})}/E(\mathbf{1}_{\{X_i \geq c\}})$. The problem with τ_{MSE} is that we do not know the true value of p and therefore, in principle, we can't compute the MSE; however, using here the conclusions drawn from the results in figure 3 (see below), $\hat{p}_{DM}$ seems to have essentially reached convergence for sample sizes larger than $100,000$. As the estimator is consistent, we computed $\hat{p}_{DM}$ for all c's with a sample size as large as $N = 5,000,000$ and computed τ_{MSE} treating the value so obtained as the true value. Table 3 displays the results.

Table 3: Some results from the simulation of the three estimators $\hat{p}_{DM}$, $\hat{p}_{ST}$ and $\hat{p}_{ACE}$ with $k = 10$.

	$c = 65$	$c = 80$	$c = 100$	$c = 150$
$\hat{p}_{DM}$	$5.71 \cdot 10^{-4}$	$1.74 \cdot 10^{-4}$	$4.89 \cdot 10^{-5}$	$4.84 \cdot 10^{-6}$
$\hat{p}_{ST}$	$5.71 \cdot 10^{-4}$	$1.75 \cdot 10^{-4}$	$4.93 \cdot 10^{-5}$	$4.86 \cdot 10^{-6}$
$\hat{p}_{ACE}$	$5.70 \cdot 10^{-4}$	$1.65 \cdot 10^{-4}$	$3.79 \cdot 10^{-5}$	$1.59 \cdot 10^{-6}$
$\hat{\sigma}_{\hat{p}_{DM}}$	$2.33 \cdot 10^{-5}$	$6.38 \cdot 10^{-6}$	$1.84 \cdot 10^{-6}$	$1.90 \cdot 10^{-7}$
$\hat{\sigma}_{\hat{p}_{ST}}$	$1.70 \cdot 10^{-4}$	$9.89 \cdot 10^{-5}$	$4.65 \cdot 10^{-5}$	$1.61 \cdot 10^{-5}$
$\hat{\sigma}_{\hat{p}_{ACE}}$	$2.38 \cdot 10^{-4}$	$1.90 \cdot 10^{-4}$	$6.46 \cdot 10^{-5}$	$7.22 \cdot 10^{-6}$
$\tau_{MSE,\hat{p}_{DM}}$	0.04	0.04	0.04	0.04
$\tau_{MSE,\hat{p}_{ST}}$	0.30	0.57	0.95	3.30
$\tau_{MSE,\hat{p}_{ACE}}$	0.42	1.09	1.34	1.63
	$c = 200$	$c = 300$	$c = 400$	$c = 500$
$\hat{p}_{DM}$	$9.12 \cdot 10^{-7}$	$7.98 \cdot 10^{-8}$	$1.32 \cdot 10^{-8}$	$3.13 \cdot 10^{-9}$
$\hat{p}_{ST}$	$7.97 \cdot 10^{-7}$	$6.58 \cdot 10^{-8}$	$7.82 \cdot 10^{-10}$	$3.87 \cdot 10^{-8}$
$\hat{p}_{ACE}$	$5.39 \cdot 10^{-7}$	$2.95 \cdot 10^{-9}$	$9.77 \cdot 10^{-12}$	$9.37 \cdot 10^{-14}$
$\hat{\sigma}_{\hat{p}_{DM}}$	$3.85 \cdot 10^{-8}$	$3.52 \cdot 10^{-9}$	$5.71 \cdot 10^{-10}$	$1.38 \cdot 10^{-10}$
$\hat{\sigma}_{\hat{p}_{ST}}$	$4.60 \cdot 10^{-6}$	$1.03 \cdot 10^{-6}$	$1.71 \cdot 10^{-8}$	$1.22 \cdot 10^{-6}$
$\hat{\sigma}_{\hat{p}_{ACE}}$	$6.61 \cdot 10^{-6}$	$6.02 \cdot 10^{-8}$	$1.50 \cdot 10^{-10}$	$1.26 \cdot 10^{-12}$
$\tau_{MSE,\hat{p}_{DM}}$	0.04	0.04	0.04	0.04
$\tau_{MSE,\hat{p}_{ST}}$	5.05	12.89	1.60	391.59
$\tau_{MSE,\hat{p}_{ACE}}$	7.26	1.22	1.00	1.00

Figure 2 shows the histograms of the simulated distributions of $\hat{p}_{DM}$ and $\hat{p}_{ACE}$ respectively for the cases $c = 65$ and $c = 500$; whereas the first estimator is approximately normal as expected from the theory of MC simulation, the latter has a very skewed distribution, in particular for $c = 500$.

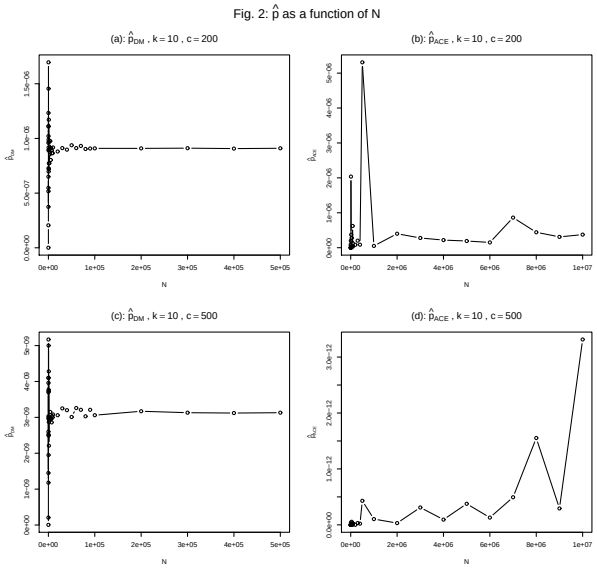


Before commenting the outcomes in table 3, we investigate the asymptotic properties of the estimators. Figure 3 shows the convergence of $\hat{p}_{DM}$ and $\hat{p}_{ACE}$ as a function of N for $k = 10$ and two different values of c, namely $c = 200$ and $c = 500$; notice also that the final sample size is $N_{ACE} = 10,000,000$ for ACE and only $N_{DM} = 500,000$ for DM, because $\hat{p}_{DM}$ remains essentially constant for $N_{DM} > 500,000$.

As expected, the DM approach clearly outperforms ACE, but there is a marked difference between the two cases: for $c = 200$, $\hat{p}_{ACE}$ shows an acceptable precision, although only for the largest sample sizes (the final values are $\hat{p}_{DM} = 9.1 \cdot 10^{-7}$ and $\hat{p}_{ACE} = 3.75 \cdot 10^{-7}$). On the other hand, for $c = 500$ $\hat{p}_{ACE}$ does not seem to reach convergence, even though it moves in the "right" direction as the sample size increases.

From table 3 and figure 3 we can derive some interesting conclusions. First, the DM approach is always preferable to ST and ACE, because it is approximately unbiased even for small sample sizes and has a much lower MSE Relative Error. Notice that the value of τ_{MSE} for $\hat{p}_{DM}$ remains stable up to the second decimal digit for all values of c, whereas τ_{MSE} deteriorates considerably for $\hat{p}_{ACE}$ when the probability of interest gets smaller: in particular, for $c = 500$ the ratio of the two MSE Relative Errors is approximately equal to 25.

The estimator $\hat{p}_{ACE}$ is approximately unbiased and has a low MSE Relative Error only for very large N; this fact has the obvious implication that computing time increases; moreover, it is not clear how large N should be. As for the ST approach, it only performs well for relatively large probabilities; for the largest values of c the tilting parameter is clearly too small. For these reasons, the DM approach seems to be preferable in all instances.



5 Computing tail probabilities in Operational Risk

Operational risk management (see Davis 2006 for an overview of problems and techniques) is usually defined as the area of risk management concerned with non-financial losses: it includes internal and external frauds, employment practices and workplace safety, clients, products, and business practices, damage to physical assets, business disruption and system failures, execution, delivery and process management. It has recently become more and more important, both because of the regulators' pressure and of the severity of losses.

Operational risk presents peculiar features with respect to market and credit risk; it follows that its measurement and management require different tools. In particular, the distribution of losses is mostly modeled directly because operational losses are not related to underlying financial factors. This characteristic has been the key to the development of a purely statistical approach which assumes a fully parametric model for the losses and estimates its parameters using historical observations.

The *Loss Distribution Approach* is the most advanced approach contemplated by the Basel II accord; it is based on the well known actuarial methodology which estimates the whole loss distribution for each business line by modeling separately the frequency and the severity of losses; see Embrechts *et al.* 1997 or Klugman *et al.* 1998 for details. The standard parametric model currently used in applications is based on the compound Poisson-Lognormal distribution. According to this hypothesis, the joint probability density function of $Y_K =: \sum_{i=1}^{K} X_i$ and K over a fixed time horizon T is given by:

$$f_{Y_K}(y,k) = P(K=k) \cdot f_{Y_K}(y),$$

where f_{Y_K} is the density of the sum of K *iid* lognormal random variables and $K \sim \text{Poisson}(\lambda)$ and $X_i \sim \text{Logn}(\mu, \sigma^2)$ model respectively the frequency and severity of losses. The marginal distribution of Y is the infinite mixture (often called compound) distribution

$$f_Y(y) = \sum_{i=0}^{\infty} P(K = i) \cdot f_{Y_i}(y), \tag{23}$$

The most common risk measure is the Value at Risk (VaR); the VaR at level α is the α quantile of (23), so that $1 - \alpha$ is the tail probability of L. However, (23) is not known in closed form; thus the only way of estimating quantiles relies on MC simulation. Moreover, the Basel II Accord prescribes large confidence levels (up to 99.9%), and crude MC encounters the problems mentioned in the preceding sections, and IS is likely to provide us with a much preferable solution.

The crude MC procedure commonly implemented is based on the following steps:

1. simulate a random number k^* from the Poisson(λ) distribution;
2. simulate k^* random numbers $x_1, \ldots, x_{k^*}$ from the Logn(μ, σ^2) distribution and compute $L = \sum_{i=1}^{k^*} x_i^*$.

Repeating B times (where B is a large number) steps 1. and 2. above we simulate the loss distribution; the VaR at confidence level α is finally given by the α quantile of the empirical distribution.

How can we apply IS to this problem? Recall the functional form of (23) and put $P(K = k) = q_k$ ($k = 0, 1, \ldots$). We have

$$P(Y > c) = \sum_{i=1}^{\infty} q_i \int_c^{\infty} f_{Y_i}(y) dy; \tag{24}$$

thus, we can apply to each summand the IS procedure developed above. The only problem is that we have to truncate the series; however, given the properties of the Poisson distribution, the series can usually be truncated after few terms. Obviously, the decision has to be made on the basis of the value of λ; for example, with $\lambda = 1$, $P(K \geq 7) \approx 8.3 \cdot 10^{-5}$; this implies that, if we simulate 10000 random numbers to estimate (24), we can stop when $K = 6$ or at most when $K = 7$, so that the sum will contain just 5 (respectively 6) summands (the first term, corresponding to $k = 0$, can also be discarded because the loss is zero).

In this application we reconsider the data used by Bee (2006) for estimating a Poisson-Lognormal model with truncated data; the estimated parameter values of the Poisson-Lognormal model were $\hat{\lambda} = 6.931$, $\hat{\mu} = 1.404$, $\hat{\sigma}^2 = 2.823$ and $N_T = 950.63$ (N_T is the estimated number of truncated data; see Bee 2006 for details).

First of all, we have to compute the optimal parameters of the IS densities. We do not give the numerical values of the optimal parameters for all c's and k's but consider that, for example, for $c = 1000$ in the *DM* approach π^* increases from 0.467 when $k = 2$ to 0.935 when $k = 20$ (as seen in sect. 3, the case $k = 1$ is special because $r_1(x)$ is bounded even for $\pi = 0$, so that there is no

need to use the DM approach: for $k = 1$ we got indeed $\pi \approx 6^{-10}$) and t_{DM} decreases from 7.35 for $k = 1$ to 6.76 for $k = 20$ (with $\lambda = 6.931$, the probability of a value larger than 20 is approximately equal to $1.25 \cdot 10^{-5}$; hence, we truncated the series (24) at $k = 20$); in the ST approach t_{CE} ranges from 5.95 for $k = 1$ to 0.297 when $k = 20$, and in the ACE approach from 5.93 to 0.390.

Finally, table 4 shows estimated tail probabilities and MSE relative errors obtained with the DM, ST, ACE and crude MC approaches with $N = 10000$; not surprisingly, and according to the results of the simulations of the preceding section, the first one has the best performance; notice, however, that the ST approach also works well if the threshold is not too high.

Table 4: Estimates and MSE relative errors for operational risk data.

	$c = 1000$€	$c = 3000$€	$c = 5000$€	$c = 7500$€	$c = 10000$€
$\hat{p}_{DM}$	0.00486	0.00033	$8.53 \cdot 10^{-5}$	$2.77 \cdot 10^{-5}$	$1.22 \cdot 10^{-5}$
$\tau_{MSE}(\hat{p}_{DM})$	0.012	0.013	0.014	0.015	0.015
$\hat{p}_{ST}$	0.00486	0.00033	$8.40 \cdot 10^{-5}$	$2.74 \cdot 10^{-5}$	$1.18 \cdot 10^{-5}$
$\tau_{MSE}(\hat{p}_{ST})$	0.043	0.155	0.306	0.560	0.534
$\hat{p}_{ACE}$	0.00484	0.00031	$8.92 \cdot 10^{-5}$	$2.40 \cdot 10^{-5}$	$9.37 \cdot 10^{-6}$
$\tau_{MSE}(\hat{p}_{ACE})$	0.050	0.335	1.226	1.409	1.346
$\hat{p}^{MC}$	0.00484	0.00033	$7.66 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	$1.17 \cdot 10^{-5}$
$\tau_{MSE}(\hat{p}^{MC})$	0.137	0.517	1.041	2.058	2.969

6 Conclusions

Estimating rare events probabilities by means of the standard MC method is in general very inefficient; on the other hand, when the functional form of the density of the variable of interest is not known, deterministic numerical approaches cannot be applied. Borrowing an idea first introduced by Hesterberg (1995), in this paper we have developed a mixture-based IS strategy for the estimation of tail events probabilities when the distribution of interest is the finite sum of Lognormal random variables and for the compound Poisson-Lognormal distribution. We solved the problem of finding the optimal values of the parameters of the mixture by means of the CE method: in particular, by exploiting the relationship between minimal CE and maximum likelihood, we showed that the parameters can be found using the EM algorithm. With the help of simulation experiments we verified that this technique works better than the standard and adaptive CE approaches. Finally, we applied the methodology to the computation of tail probabilities in operational risk.

Although the Lognormal distribution is by far the most common choice, in operational risk different distributional hypotheses for the severity of losses are sometimes used, according to the estimated tail heaviness: in particular the Gamma or the Generalized Pareto. While in the first case the implementation of IS should not be too difficult, as the Gamma distribution has the moment generating function, so that we can use the approach based on tilted densities, in the latter setup the problem would require some further research.

Finally, in the Poisson-Lognormal compound distribution, importance sampling could also be applied by leaving unchanged the parameters of the Lognormal and twisting the Poisson parameter, i.e. increasing the number of losses. This topic is somehow aside from the main object of interest of this paper, but should be studied for its possible relevance in applications.

References

Asmussen, S. (2000). *Ruin Probabilities*. London: World Scientific.

Asmussen, S. Binswanger, K. (1997). Simulation of ruin probabilities for subexponential claims. ASTIN Bulletin 27: 297-318.

Asmussen, S. Kroese, D.P. (2006). Improved algorithms for rare event simulation with heavy tails. Advances in Applied Probability 38: 545-558.

Asmussen, S. Binswanger, K. Højgaard, B. (2000). Rare events simulation for heavy-tailed distributions. Bernoulli 6: 303-322.

Asmussen, S. Kroese, D.P. Rubinstein, R.Y. (2005). Heavy tails, importance sampling and cross-entropy. Stochastic Models 21: 57-76.

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*. London: Risk Books.

Basel Committee on Banking Supervision (2005). *Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework*. www.bis.org.

Casella, G. Robert, C.P. (2004). *Monte Carlo Statistical Methods*, second edition. New York: Springer.

Davis, E. ed. (2006). *The Advanced Measurement Approach to Operational Risk*. London: Risk Books.

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

Davison, A.C. Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

Durrett, R. (1996). *Probability: Theory and Examples*. Belmont: Duxbury Press.

Embrechts, P. Klüppelberg, C. Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57: 1317-1340.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Hammersley, J.M. Handscomb, D.C. (1964). *Monte Carlo Methods*. London: Methuen.

Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. Technometrics 37: 185-194.

Klugman, S.A. Panjer, H.H. Willmot, G.E. (1998). *Loss Models - From Data to Decisions*. New York: Wiley.

Kullback, S. (1968). *Information Theory and Statistics*. New York: Wiley.

Mikosch, T. Nagaev, A.V. (1998). Large deviations of heavy-tailed sums with applications in insurance. Extremes 1: 81-110.

Ross, S.M., 2006. *Simulation*, fourth edition. San Diego: Academic Press.

Rubinstein, R.Y. (1981). *Simulation and the Monte Carlo Method*. New York: Wiley.

Rubinstein, R.Y. (1997). Optimization of computer simulation models with rare events. European Journal of Operational Research 99: 89-112.

Rubinstein, R.Y. Kroese, D.P. (2004). *The Cross-Entropy Method*. New York: Springer.

Sadowsky, J.S. (1993). On the optimality and stability of exponential twisting in Monte Carlo simulation. IEEE Transactions on Information Theory IT-39: 119-128.

Elenco dei papers del Dipartimento di Economia

2000.1 *A two-sector model of the effects of wage compression on unemployment and industry distribution of employment*, by Luigi Bonatti

2000.2 *From Kuwait to Kosovo: What have we learned? Reflections on globalization and peace,* by Roberto Tamborini

2000.3 *Metodo e valutazione in economia. Dall'apriorismo a Friedman ,* by Matteo Motterlini

2000.4 *Under tertiarisation and unemployment.* by Maurizio Pugno

2001.1 *Growth and Monetary Rules in a Model with Competitive Labor Markets,* by Luigi Bonatti.

2001.2 *Profit Versus Non-Profit Firms in the Service Sector: an Analysis of the Employment and Welfare Implications,* by Luigi Bonatti, Carlo Borzaga and Luigi Mittone.

2001.3 *Statistical Economic Approach to Mixed Stock-Flows Dynamic Models in Macroeconomics,* by Bernardo Maggi and Giuseppe Espa.

2001.4 *The monetary transmission mechanism in Italy: The credit channel and a missing ring,* by Riccardo Fiorentini and Roberto Tamborini.

2001.5 *Vat evasion: an experimental approach,* by Luigi Mittone

2001.6 *Decomposability and Modularity of Economic Interactions,* by Luigi Marengo, Corrado Pasquali and Marco Valente.

2001.7 *Unbalanced Growth and Women's Homework,* by Maurizio Pugno

2002.1 *The Underground Economy and the Underdevelopment Trap,* by Maria Rosaria Carillo and Maurizio Pugno.

2002.2 *Interregional Income Redistribution and Convergence in a Model with Perfect Capital Mobility and Unionized Labor Markets,* by Luigi Bonatti.

2002.3 *Firms' bankruptcy and turnover in a macroeconomy,* by Marco Bee, Giuseppe Espa and Roberto Tamborini.

2002.4 *One "monetary giant" with many "fiscal dwarfs": the efficiency of macroeconomic stabilization policies in the European Monetary Union,* by Roberto Tamborini.

2002.5 *The Boom that never was? Latin American Loans in London 1822-1825,* by Giorgio Fodor.

2002.6 *L'economia senza banditore di Axel Leijonhufvud: le 'forze oscure del tempo e dell'ignoranza' e la complessità del coordinamento,* by Elisabetta De Antoni.

2002.7 *Why is Trade between the European Union and the Transition Economies Vertical?,* by Hubert Gabrisch and Maria Luigia Segnana.

2003.1 *The service paradox and endogenous economic gorwth,* by Maurizio Pugno.

2003.2 *Mappe di probabilità di sito archeologico: un passo avanti,* di Giuseppe Espa, Roberto Benedetti, Anna De Meo e Salvatore Espa.
(*Probability maps of archaeological site location: one step beyond,* by Giuseppe Espa, Roberto Benedetti, Anna De Meo and Salvatore Espa).

2003.3 *The Long Swings in Economic Understianding,* by Axel Leijonhufvud.

2003.4 *Dinamica strutturale e occupazione nei servizi,* di Giulia Felice.

2003.5 *The Desirable Organizational Structure for Evolutionary Firms in Static Landscapes,* by Nicolás Garrido.

2003.6 *The Financial Markets and Wealth Effects on Consumption* An Experimental Analysis, by Matteo Ploner.

2003.7 *Essays on Computable Economics, Methodology and the Philosophy of Science,* by Kumaraswamy Velupillai.

2003.8 *Economics and the Complexity Vision: Chimerical Partners or Elysian Adventurers?*, by Kumaraswamy Velupillai.

2003.9 *Contratto d'area cooperativo contro il rischio sistemico di produzione in agricoltura,* di Luciano Pilati e Vasco Boatto.

2003.10 *Il contratto della docenza universitaria. Un problema multi-tasking*, di Roberto Tamborini.

2004.1 *Razionalità e motivazioni affettive: nuove idee dalla neurobiologia e psichiatria per la teoria economica?* di Maurizio Pugno.
(*Rationality and affective motivations: new ideas from neurobiology and psychiatry for economic theory?* by Maurizio Pugno.

2004.2 *The economic consequences of Mr. G. W. Bush's foreign policy. Can th US afford it?* by Roberto Tamborini

2004.3 *Fighting Poverty as a Worldwide Goal* by Rubens Ricupero

2004.4 *Commodity Prices and Debt Sustainability* by Christopher L. Gilbert and Alexandra Tabova

2004.5 *A Primer on the Tools and Concepts of Computable Economics* by K. Vela Velupillai

2004.6 *The Unreasonable Ineffectiveness of Mathematics in Economics* by Vela K. Velupillai

2004.7 *Hicksian Visions and Vignettes on (Non-Linear) Trade Cycle Theories* by Vela K. Velupillai

2004.8 *Trade, inequality and pro-poor growth: Two perspectives, one message?* By Gabriella Berloffa and Maria Luigia Segnana

2004.9 *Worker involvement in entrepreneurial nonprofit organizations. Toward a new assessment of workers? Perceived satisfaction and fairness* by Carlo Borzaga and Ermanno Tortia.

2004.10 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification* by Lorenzo Sacconi

2004.11 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity* by Lorenzo Sacconi

2004.12 *A Fuzzy Logic and Default Reasoning Model of Social Norm and Equilibrium Selection in Games under Unforeseen Contingencies* by Lorenzo Sacconi and Stefano Moretti

2004.13 *The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality* by Gianluca Grimalda and Lorenzo Sacconi

2005.1 *The happiness paradox: a formal explanation from psycho-economics* by Maurizio Pugno

2005.2 *Euro Bonds: in Search of Financial Spillovers* by Stefano Schiavo

2005.3 *On Maximum Likelihood Estimation of Operational Loss Distributions* by Marco Bee

2005.4 *An enclave-led model growth: the structural problem of informality persistence in Latin America* by Mario Cimoli, Annalisa Primi and Maurizio Pugno

2005.5 *A tree-based approach to forming strata in multipurpose business surveys,* Roberto Benedetti, Giuseppe Espa and Giovanni Lafratta.

2005.6 *Price Discovery in the Aluminium Market* by Isabel Figuerola-Ferretti and Christopher L. Gilbert.

2005.7 *How is Futures Trading Affected by the Move to a Computerized Trading System? Lessons from the LIFFE FTSE 100 Contract* by Christopher L. Gilbert and Herbert A. Rijken.

2005.8 *Can We Link Concessional Debt Service to Commodity Prices?* By Christopher L. Gilbert and Alexandra Tabova

2005.9 *On the feasibility and desirability of GDP-indexed concessional lending* by Alexandra Tabova.

2005.10 *Un modello finanziario di breve periodo per il settore statale italiano: l'analisi relativa al contesto pre-unione monetaria* by Bernardo Maggi e Giuseppe Espa.

2005.11 *Why does money matter? A structural analysis of monetary policy, credit and aggregate supply effects in Italy*, Giuliana Passamani and Roberto Tamborini.

2005.12 *Conformity and Reciprocity in the "Exclusion Game": an Experimental Investigation* by Lorenzo Sacconi and Marco Faillo.

2005.13 *The Foundations of Computable General Equilibrium Theory*, by K. Vela Velupillai.

2005.14 *The Impossibility of an Effective Theory of Policy in a Complex Economy*, by K. Vela Velupillai.

2005.15 *Morishima's Nonlinear Model of the Cycle: Simplifications and Generalizations,* by K. Vela Velupillai.

2005.16 Using and Producing *Ideas* in Computable Endogenous Growth, by K. Vela Velupillai.

2005.17 *From Planning to Mature: on the Determinants of Open Source Take Off* by Stefano Comino, Fabio M. Manenti and Maria Laura Parisi.

2005.18 *Capabilities, the self, and well-being: a research in psycho-economics,* by Maurizio Pugno.

2005.19 Fiscal and monetary policy, unfortunate events, and the SGP arithmetics. Evidence from a *growth-gap* model, by Edoardo Gaffeo, Giuliana Passamani and Roberto Tamborini

2005.20 *Semiparametric Evidence on the Long-Run Effects of Inflation on Growth,* by Andrea Vaona and Stefano Schiavo.

2006.1 *On the role of public policies supporting Free/Open Source Software. An European perspective,* by Stefano Comino, Fabio M. Manenti and Alessandro Rossi.

2006.2 *Back to Wicksell? In search of the foundations of practical monetary policy,* by Roberto Tamborini

2006.3 *The uses of the past,* by Axel Leijonhufvud

2006.4 *Worker Satisfaction and Perceived Fairness: Result of a Survey in Public, and Non-profit Organizations,* by Ermanno Tortia

2006.5 *Value Chain Analysis and Market Power in Commodity Processing with Application to the Cocoa and Coffee Sectors,* by Christopher L. Gilbert

2006.6 *Macroeconomic Fluctuations and the Firms' Rate of Growth Distribution: Evidence from UK and US Quoted Companies,* by Emiliano Santoro

2006.7 *Heterogeneity and Learning in Inflation Expectation Formation: An Empirical Assessment,* by Damjan Pfajfar and Emiliano Santoro

2006.8 Good *Law & Economics* needs suitable microeconomic models: the case against the application of standard agency models: the case against the application of standard agency models to the professions, by Lorenzo Sacconi

2006.9 *Monetary policy through the "credit-cost channel". Italy and Germany,* by Giuliana Passamani and Roberto Tamborini

2007.1 *The Asymptotic Loss Distribution in a Fat-Tailed Factor Model of Portfolio Credit Risk,* by Marco Bee

2007.2 *Sraffa?s Mathematical Economics – A Constructive Interpretation,* by Kumaraswamy Velupillai

2007.3 *Variations on the Theme of Conning in Mathematical Economics,* by Kumaraswamy Velupillai

2007.4 *Norm Compliance: the Contribution of Behavioral Economics Models,* by Marco Faillo and Lorenzo Sacconi

2007.5 *A class of spatial econometric methods in the empirical analysis of clusters of firms in the space,* by Giuseppe Arbia, Giuseppe Espa e Danny Quah.

2007.6 *Rescuing the LM (and the money market) in a modern Macro course,* by Roberto Tamborini.

2007.7 *Family, Partnerships, and Network: Reflections on the Strategies of the Salvadori Firm of Trento,* by Cinzia Lorandini.

2007.8 *I Verleger serici trentino-tirolesi nei rapporti tra Nord e Sud: un approccio prosopografico,* by Cinzia Lorandini.

2007.9 *A Framework for Cut-off Sampling in Business Survey Design,* by Marco Bee, Roberto Benedetti e Giuseppe Espa

2007.10 *Spatial Models for Flood Risk Assessment,* by Marco Bee, Roberto Benedetti e Giuseppe Espa

2007.11 *Inequality across cohorts of households:evidence from Italy*, by Gabriella Berloffa and Paola Villa

2007.12 *Cultural Relativism and Ideological Policy Makers in a Dynamic Model with Endogenous Preferences*, by Luigi Bonatti

2007.13 *Optimal Public Policy and Endogenous Preferences: an Application to an Economy with For-Profit and Non-Profit*, by Luigi Bonatti

2007.14 *Breaking the Stability Pact: Was it Predictable?*, by Luigi Bonatti and Annalisa Cristini.

2007.15 *Home Production, Labor Taxation and Trade Account*, by Luigi Bonatti.

2007.16 *The Interaction Between the Central Bank and a Monopoly Union Revisited: Does Greater Uncertainty about Monetary Policy Reduce Average Inflation?*, by Luigi Bonatti.

2007.17 *Complementary Research Strategies, First-Mover Advantage and the Inefficiency of Patents*, by Luigi Bonatti.

2007.18 *DualLicensing in Open Source Markets*, by Stefano Comino and Fabio M. Manenti.

2007.19 *Evolution of Preferences and Cross-Country Differences in Time Devoted to Market Work*, by Luigi Bonatti.

2007.20 *Aggregation of Regional Economic Time Series with Different Spatial Correlation Structures*, by Giuseppe Arbia, Marco Bee and Giuseppe Espa.

2007.21 *The Sustainable Enterprise. The multi-fiduciary perspective to the EU Sustainability Strategy*, by Giuseppe Danese.

2007.22 *Taming the Incomputable, Reconstructing the Nonconstructive and Deciding the Undecidable in Mathematical Economics*, by K. Vela Velupillai.

2007.23 *A Computable Economist's Perspective on Computational Complexity*, by K. Vela Velupillai.

2007.24 *Models for Non-Exclusive Multinomial Choice, with Application to Indonesian Rural Households*, by Christopher L. Gilbert and Francesca Modena.

2007.25 *Have we been Mugged? Market Power in the World Coffee Industry*, by Christopher L. Gilbert.

2007.26 *A Stochastic Complexity Perspective of Induction in Economics and Inference in Dynamics,* by K. Vela Velupillai.

2007.27 *Local Credit ad Territorial Development: General Aspects and the Italian Experience,* by Silvio Goglio.

2007.28 *Importance Sampling for Sums of Lognormal Distributions, with Applications to Operational Risk,* by Marco Bee.

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO