UNIVERSITA' DEGLI STUDI DI TRENTO - DIPARTIMENTO DI ECONOMIA

ON MAXIMUM LIKELIHOOD ESTIMATION OF OPERATIONAL LOSS DISTRIBUTIONS

Marco Bee

Discussion Paper No. 3, 2005

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. Stefano Comino
Dipartimento di Economia
Università degli Studi
Via Inama 5
38100 TRENTO ITALY

On Maximum Likelihood Estimation of Operational Loss Distributions

Marco Bee

University of Trento - Department of Economics
via Inama 5 - 38100 Trento, Italy
Phone +39-0461-882296, Fax +39-0461-882222
E-mail: mbee@economia.unitn.it

ABSTRACT

This paper develops a likelihood-based methodology to estimate loss distributions and compute Capital at Risk in risk management applications. In particular, we deal with the problem of estimating severity distributions with censored and truncated operational losses, for which numerical maximization of the likelihood function by means of standard optimization tools may be difficult. We show that, under the standard hypothesis of lognormal severity, maximum likelihood estimation can be performed by means of the EM algorithm. We derive the relevant equations of the algorithm and apply it to operational loss data. Finally, a simulation study shows that, in this setup, the EM algorithm has more desirable properties than both the BFGS algorithm and the Nelder-Mead simplex algorithm.

KEY WORDS: Truncated distribution, Loss model, EM algorithm, Capital at Risk.

JEL Classification: C13, C15, C21, C63.

Mathematics Subject Classification: 62F10, 62F40, 65C60.

1 Introduction

The New Basel Capital Accord has started to treat operational risk as an autonomous type of risk, with its own tools and methods. Nevertheless, it is not yet easy to formulate a clear-cut definition of operational risk, and there is a considerable debate going on about the precise bounds of the concept: the commonly used definition is "the risk of direct or indirect loss resulting from inadequate or failed internal process, people and systems or from external events" (Basel Committee on Banking Supervision 2001). According to Crouhy *et al.* (2000), 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; legal risk is usually included as well, whereas reputational and strategic risks are not. Therefore, operational risk is non-financial, so that it is not restricted to the financial sector, and it was indeed first measured by firms other than banks and financial institutions, like companies with complex IT and/or production processes (e.g. car, energy).

Because of the regulators' pressure, operational risk has recently received increasing attention in the financial industry. The augmented emphasis was also spurred by the growing relevance of this source of risk: it is clear that, as both financial products and IT processes become more sophisticated, failures tend to become more frequent and their consequences more severe.

Operational risk presents peculiar features with respect to market and credit risk and, as a consequence, its measurement and management require different tools. In particular, the distribution of losses has to be modeled directly, because, unlike typical financial losses, operational losses are not related to underlying financial factors. The relevant literature dates back to the analysis developed by non-life insurance companies, where losses are represented by claims; see Klugman *et al.* (1998) for a thorough analysis of these methods, which represent an important starting point for the measurement of operational risk.

The Basel II accord contemplates four methodologies, in increasing order of mathematical sophistication (see Nyström and Skoglund 2002 for a review); in this paper we will concentrate on the most sophisticated one, the so called Loss Distribution Approach (LDA), which computes the total loss amount for each business line by modelling separately the frequency and the severity of losses. The typical probabilistic models proposed to describe the severity are the Lognormal,

Gamma and Weibull distribution, or possibly the Generalised Pareto in case we are particularly concerned with tail events. Among them, attention has recently focused mainly on the Lognormal distribution, which guarantees a higher degree of analytical tractability. As for the frequency, we will follow the common approach of modelling it as a Poisson distribution; this is a rather strong hypothesis, in particular because it implies equality of mean and variance, a feature which is not always supported empirically. Despite of this remark, the Poisson-Lognormal model is now considered the standard one; see Cruz (2003) for an extensive analysis of the stochastic models used in the LDA approach.

In this setup, there are at least two reasons why statistical analysis is quite challenging. First, irrespectively of the distribution chosen for the severity, statistical inference about its parameters can't be performed along the familiar lines, because losses below a given threshold are discarded: using statistical terminology, data are left-truncated, and this makes inferential procedures more complicated.

Second, for measuring risk one has to determine the aggregate loss, as given by the combination of frequency and severity: this distribution is the convolution of k independent continuous r.v.'s, where k is the realization of a discrete random variable. The resulting distribution is usually very difficult to work with, so that, in order to compute loss and risk measures, simulation techniques are called for.

In this paper we focus on the problem of maximum likelihood estimation of the aggregate loss distribution with censored and truncated observations. The estimation of truncated distributions is considerably more difficult than the estimation of the corresponding plain (not truncated) distributions, and even the numerical maximization of the likelihood function is not straightforward. We show that the EM algorithm provides a solution that is both theoretically well founded and easily implementable; in addition, it has a feature which is very important in this setup: it produces an estimate of the number of truncated observations, which is essential for the estimation of the parameter of the frequency distribution.

For the sake of completeness it is worth mentioning an approach which has recently been put forward in response to the scarcity of data; the idea consists in combining the historical data with the expert knowledge gathered from internal managers, and is implemented by means of Bayesian methods. However, this methodology will not be used in this paper, so that we refer the interested reader to Cruz (2003, chap. 10).

The rest of this paper is organized as follows. In section 2 we describe the data generating process of operational losses; in section 3 we derive the EM algorithm to find the maximum likelihood estimators of the parameters in the case of both censored and truncated lognormal losses; section 4 uses the algorithm to measure risk with real data; moreover, the EM algorithm is compared to the BFGS and Nelder-Mead algorithms by means of simulation experiments; section 5 discusses the results and gives a review of the problems open to future research and of their connection with the tools developed in this paper.

2 The model

The most relevant operational risk measure is the Unexpected Loss, which is computed as the $(1-\alpha)$-th quantile minus the expected value of the loss distribution over a specified time horizon. The Basel II accord defines it as the Capital at Risk (CaR).

In order to describe formally the mathematical approach to CaR measurement, we first define the loss distribution. The total loss over a predetermined time horizon is given by the random sum

$$S = \sum_{i=1}^{K} W_i, \tag{1}$$

where K is a random variable with a counting distribution and $W_1, \ldots, W_K$ are *iid* continuous positive random variables, also independent from K. The corresponding cdf is

$$\begin{aligned} F_S(w) &= P(S \leq w) \\ &= \sum_{k=0}^{\infty} p_k P(S \leq w | K = k) \\ &= \sum_{k=0}^{\infty} p_k F_W^{*k}(w), \end{aligned}$$

where $p_k = P(K = k)$ and $F_W^{*k}(w)$ is the k-fold convolution of the cdf of W; the distribution of S is commonly known as a compound distribution. In the following we will assume that K and W are respectively Poisson and Lognormal.

The LDA approach can be formulated, more generally, in continuous time. In this setup the loss process $S(t),\ 0 \leq t \leq T$ is a compound counting process,

represented as

$$S(t) = \sum_{k=1}^{K(t)} W_k, \qquad 0 \leq t \leq T, \tag{2}$$

where $K(t)$ is a counting process and $W_1, \ldots, W_K$ are *iid* continuous r.v.'s. The pair $\{t_k, W_k\}_{k \in \mathbf{Z}}$, where t_k are the points of jump and $\mathbf{Z}$ is the set of integers, is usually called a marked point process. If we assume that $K(t)$ is a Poisson process with parameter ν, (2) is called a Poisson compound process; from the basic properties of the Poisson process (see, for example, Durrett 1996, pag. 145), if we fix $t^* \in [0, T]$, the compound counting process (2) reduces to the compound distribution (1), where the parameter of the Poisson distribution is $\lambda = \nu \cdot t^*$.

The joint distribution $f_{K,W}(k, w)$ of the two sources of randomness K and W in (1) is quite intricate, so that not only it is analytically intractable (although in some cases a recursive computation is possible: see Panjer 1981), but also a direct application of Monte Carlo simulation is not straightforward. The latter, however, can be performed by decomposing the joint distribution of K and S: $f_{K,S}(k, s) = f_{S|K}(s|k) \cdot f_K(k)$. Conditionally on $K = k$, the model for the aggregated loss is given by the random variable

$$(S|K = k) = \sum_{i=1}^{k} W_i,$$

which may or may not be analytically tractable depending on the distribution of W, but can easily be simulated as long as we can simulate K and W. If $W \sim Logn(\mu, \sigma^2)$, and conditionally on $K \sim P(\lambda)$, this distribution is the convolution of k lognormal r.v.'s, and random number generation from $f_{K,S}(k, s)$ can be accomplished by performing the following steps:

(i) simulate a random number k from the $P(\lambda)$ distribution;

(ii) simulate k random numbers $w_1, \ldots, w_k$ from the $Logn(\mu, \sigma^2)$ distribution and compute $s = \sum_{i=1}^{k} w_i$.

Repeating B times these two steps, we get a random sample of size B from the density $f_{K,S}(k, s)$. This is the approach commonly adopted by banks to compute CaR, which is given by the $(1 - \alpha)$-th quantile of the empirical distribution minus its mean.

From the point of view of statistical inference, the parameters of the frequency and of the severity distribution can be estimated separately either by maximum

likelihood or by percentile matching. Whereas in principle this is straightforward, a major difficulty is given by the fact that loss data are usually left-censored or, more frequently, left-truncated; according to whether data are truncated or censored, specific inferential procedures are needed. Despite of the fact that the data generating processes of truncated and censored data are different, the two terms are sometimes incorrectly treated as synonymous. For this reason, before tackling the problem we recall the appropriate definitions (Klugman *et al.* 1998, pag. 132).

Definition. *Data are said to be censored when the number of observations that fall in a given set is known, but the specific values of the observations are unknown; data are said to be censored from below when the set is all numbers less than a specific value.*

Definition. *Data are said to be truncated when observations that fall in a given set are excluded; data are said to be truncated from below when the set is all numbers less than a specific value.*

3 The estimation procedure

In financial models, estimation of parameters tends to be considered as a background matter. Whereas in pricing models this issue is perhaps less relevant, because they can often be calibrated to market data, rigorous estimation is essential for the computation of risk measures.

The derivation of MLE's is based on two steps: finding the functional form of the likelihood function and maximizing it. To give a motivation for the later developments of the paper, and referring to the next sections for details, we anticipate here the likelihood function corresponding to the density of the severity distribution when the observations are left-censored:

$$L(\mu, \sigma^2; c, \mathbf{y}) = const \cdot [\Phi(c^*)]^{N_C} \cdot \prod_{i=1}^{N_U} L(\mu, \sigma^2; c, y_i),$$

where y_i $(i = 1, \ldots, N_U)$ is the logarithm of the uncensored observations, $L(\mu, \sigma^2; c, y_i)$ is the normal likelihood function, N_U and N_C are respectively the number of uncensored and censored observations, c is the censoring threshold, $c^* = (c - \mu)/\sigma$ and *const* is a constant that does not depend on unknown parameters. Maximizing

this function is rather complicated: analytical differentiation is impossible because of the presence of the normal cdf, and numerical differentiation is computationally heavy. For this reason, standard gradient-based optimization methods are likely to give unstable results, and it is preferable to introduce algorithms which are better suited for the solution of this problem.

3.1 The EM algorithm

The EM algorithm is an iterative method for computing maximum likelihood estimators in presence of missing data. Typically, the algorithm is particularly convenient in cases when, if the missing data were known, estimation would be straightforward. The seminal paper concerning the EM algorithm is Dempster *et al.* (1977); a thorough review of both theory and applications is given by McLachlan and Krishnan (1996).

Roughly speaking, the intuition behind the algorithm consists in maximizing, instead of the likelihood based on the observed data, an hypothetical likelihood based not only on the observed data but also on a "proper replacement" of the missing data. The maximization of this function, called complete likelihood, is often straightforward.

More formally, let $\mathbf{Y}$ be the random vector of the observed data; let $g(\mathbf{y};\boldsymbol{\theta})$ and $l(\boldsymbol{\theta})$ be respectively its density and log-likelihood function, where $\boldsymbol{\theta}$ is the vector of all parameters. Let then $\mathbf{Z}$ be the vector containing the missing data, whose elements are unobservable; finally, let $\mathbf{X} = (\mathbf{Y}', \mathbf{Z}')'$ be the complete-data vector, with density and log-likelihood denoted respectively by $g_c(\mathbf{x};\boldsymbol{\theta})$ and $l_c(\boldsymbol{\theta})$.

The first step of the algorithm (called E-step, where E stands for Expectation) consists in computing the conditional expectation of the complete log-likelihood function $l_c(\boldsymbol{\theta})$, given the current value of $\boldsymbol{\theta}$ and the observed sample $\mathbf{y}$.

The second step (M-step, where M stands for Maximization) maximizes, with respect to $\boldsymbol{\theta}$, the conditional expectation of the complete log-likelihood computed in the E-step.

The structure of the algorithm can be described as follows: let $\boldsymbol{\theta}^{(0)}$ be the initial value of the parameter vector; the E-step consists in computing

$$Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(0)}) = E_{\theta^{(0)}}\{l_c(\boldsymbol{\theta})|\mathbf{y}\}. \tag{3}$$

Then $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(0)})$ is maximized with respect to $\boldsymbol{\theta}$, namely $\boldsymbol{\theta}^{(1)}$ is chosen such that

$$Q(\boldsymbol{\theta}^{(1)}; \boldsymbol{\theta}^{(0)}) = \max_{\theta} Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(0)}) \tag{4}$$

The procedure is then repeated, i.e. (3) and (4) are recomputed using $\boldsymbol{\theta}^{(1)}$ in place of $\boldsymbol{\theta}^{(0)}$. The two steps are iterated until some convergence criterion is met.

Referring the interested reader to McLachlan and Krishnan (1996, chap. 3) for details, we would like to mention some very desirable properties of the EM algorithm. First, it increases the likelihood at each iteration. Second, it is hardly sensitive to the choice of the starting values: it is indeed often the case that it converges to the global maximum irrespectively of the initialization; even with ill-behaved likelihood functions, only in case of very bad initializations the algorithm fails to converge. The only drawback is that convergence is rather slow: the rate of convergence of the algorithm can be shown to be inversely related to the amount of missing data (this point can be made more precise by introducing the concept of missing information; see McLachlan and Krishnan 1996, sect. 3.9). However, it will be seen in section 4 that the large number of iterations is frequently more than compensated by the small time required by a single iteration.

The actual form of (3) and (4) depends on the specific problem at hand and on the assumptions on the distribution of the data. In the next two sections we will develop the algorithm respectively for left-censored and left-truncated lognormal data.

3.2 Censored data

The problem of maximum likelihood estimation with censored data is more commonly encountered in situations where the observations are right-censored, as for example in survival analysis. A typical example is the case when some individuals are still alive at the end of the experiment: they constitute the censored observations, and all we know about them is their number and the value of the threshold. Loss data are instead left-censored or left-truncated; we first work out the details for the left-censored case.

Let $W_1, \ldots, W_N$ be a sample of independent losses from a lognormal population with parameters μ and σ^2; as usual with lognormally distributed observations, one can resort to the transformation $Y_i = \log(W_i) \sim N(\mu, \sigma^2)$ and restrict

attention to the normal case. Let N_U and $N_C = N - N_U$ be respectively the number of uncensored and censored observations, and let $\mathbf{y} = (y_1, \ldots, y_{N_U})'$ be the vector of uncensored observations. Similarly, let $\mathbf{x}$ be the missing data, i.e. $\mathbf{x} = (x_1, \ldots, x_{N_C})'$; their numerical values are unobservable, but their number N_C is known. Finally, let $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$ be the complete data. Notice that we can also write $\mathbf{y} = \{z_i : z_i \geq c\}$, and $\mathbf{x} = \{z_i : z_i < c\}$, $i = 1, \ldots, N$, where c is the logarithm of the truncation threshold.

In order to find the MLE's of the parameters, we have to write down the observed and complete likelihood functions; the complete likelihood function is given by

$$L_c(\mu, \sigma^2; \mathbf{z}) = \prod_{i=1}^{N} L(\mu, \sigma^2; z_i), \qquad \mu \in \mathbf{R},\ \sigma^2 \in \mathbf{R}^+,$$

where $L(\mu, \sigma^2; z_i)$ is the normal likelihood function evaluated at z_i. Let now $c^* = (c - \mu)/\sigma$; the observed likelihood (Maddala 1983, pag. 5) can be written as:

$$L_{obs}(\mu, \sigma^2; c, \mathbf{y}) = const \cdot \prod_{i=1}^{N_U} L(\mu, \sigma^2; c, y_i) \cdot [P(Y \leq c)]^{N_C} =$$

$$= const \cdot \prod_{i=1}^{N_U} L(\mu, \sigma^2; c, y_i) \cdot [\Phi(c^*)]^{N_C}, \qquad \mu \in \mathbf{R},\ \sigma^2 \in \mathbf{R}^+, \tag{5}$$

where $L(\mu, \sigma^2; c, y_i)$ is the normal density, considered as a function of μ and σ^2, so that

$$L_{obs}(\mu, \sigma^2; c, \mathbf{y}) = const \cdot \prod_{i=1}^{N_U} \left\{ \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y_i - \mu}{\sigma})^2} \right\} \cdot [\Phi(c^*)]^{N_C}. \tag{6}$$

The E-step is given by the conditional expectation of the complete log-likelihood function given the observed data and the current numerical values of the parameters. Putting $\tau = \sigma^2$ to simplify notation, at the t-th iteration we have

$$E[L_c(\mu, \tau; \mathbf{z}) | \mathbf{y}, \mu^{(t)}, \tau^{(t)})] = \prod_{i=1}^{N_U} L(\mu, \tau; y_i) \cdot \prod_{i=1}^{N_C} E[L(\mu, \tau; c, x_i | y_i, \mu^{(t)}, \tau^{(t)})].$$

Now the distribution of the missing data X_i, $i = 1, \ldots, N_C$, is right-truncated normal, because the X_i are normal and smaller than c, so that they have density

$$f_X(x) = \frac{1}{\Phi(c^*)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x - \mu}{\sigma})^2}, \qquad x \leq c.$$

The computation of the conditional expectation $E[L(\mu, \tau; c, x_i | y_i, \mu^{(t)}, \tau^{(t)})]$ is simplified by the fact that the complete log-likelihood function is linear in X and X^2; thus, it amounts to computing $E(X|\mu^{(t)}, \tau^{(t)}, y_1, \ldots, y_{N_U})$ and $E(X^2|\mu^{(t)}, \tau^{(t)}, y_1, \ldots, y_{N_U})$. These two quantities are given by the well known formulas for the expected value and the variance of the right-truncated normal distribution. Putting $\alpha(c^*) = \phi(c^*)/\Phi(c^*)$, we get:

$$E(X|\boldsymbol{\theta}^{(t)}, \mathbf{y}) = \mu^{(t)} - \sigma^{(t)} \cdot \alpha(c^{*(t)}); \tag{7}$$

$$E(X^2|\boldsymbol{\theta}^{(t)}, \mathbf{y}) = \tau^{(t)} \cdot [1 - c^{*(t)} \cdot \alpha(c^{*(t)}) - (\alpha(c^{*(t)}))^2] + [E(X|\boldsymbol{\theta}^{(t)}, \mathbf{y})]^2. \tag{8}$$

As for the M-step, we just have to use the formulas of the MLE's of the normal distribution, with missing data replaced by (7) and (8):

$$\mu^{(t+1)} = \frac{1}{N} \left(\sum_{i=1}^{N_U} y_i + N_C \cdot E(X|\boldsymbol{\theta}^{(t)}, \mathbf{y}) \right); \tag{9}$$

$$\tau^{(t+1)} = \frac{1}{N} \left(\sum_{i=1}^{N_U} y_i^2 + N_C \cdot E(X^2|\boldsymbol{\theta}^{(t)}, \mathbf{y}) \right) - (\mu^{(t+1)})^2. \tag{10}$$

Finally, MLE's of the parameters are given by iterating (7), (8), (9) and (10) until convergence.

3.3 Truncated data

The situation presented in the preceding section is not the one usually encountered when dealing with loss data. Such data are indeed typically truncated, namely losses below a given threshold are discarded; deriving the relevant equations of the algorithm in this case is slightly more tricky, but the practical implementation remains as easy as with censored observations. The following calculations are based on the way of reasoning followed by McLachlan and Krishnan (1996, sect. 2.8) in a model with grouped and truncated data.

The setup needs to be modified as follows. Let $W_1, \ldots, W_{N_U}$ be a sample of independent losses from a truncated lognormal population with parameters μ and σ^2; again, we use the transformation $Y_i = \log(W_i) \sim N(\mu, \sigma^2)$ and restrict attention to the normal case. N_U is now the number of untruncated observations,

and $\mathbf{y} = (y_1, \ldots, y_{N_U})'$ is the vector containing these data. We also know that N_T observations are smaller than c and have been discarded, but N_T is unknown. It follows that the missing data are now given by $\{N_T, \mathbf{x}\}$, where N_T is the number of truncated observations and $\mathbf{x} = (x_1, \ldots, x_{N_T})'$. As before, let finally $\mathbf{z} = (\mathbf{y}', N_T, \mathbf{x}')'$ represent the complete data, where $\mathbf{y} = \{z_i : z_i \geq c\}$, $N_T = \#\{z_i : z_i < c\}$ and $\mathbf{x} = \{z_i : z_i < c\}$, $i = 1, \ldots, N$.

First of all, we have to write down the complete and observed likelihood functions. The complete likelihood function is as in section 3.2:

$$L_c(\mu, \sigma^2; \mathbf{z}) = \prod_{i=1}^{N} L(\mu, \sigma^2; z_i), \qquad \mu \in \mathbf{R},\ \sigma^2 \in \mathbf{R}^+, \tag{11}$$

where $L(\mu, \sigma^2; z_i)$ is the normal likelihood function evaluated at z_i. The observed likelihood is given by

$$L_{obs}(\mu, \sigma^2; c, \mathbf{Y}) = \prod_{i=1}^{N_U} L_{trunc}(\mu, \sigma^2; c, y_i) =$$

$$= \prod_{i=1}^{N_U} \left\{ \frac{1}{1 - \Phi(c^*)} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{y_i - \mu}{\sigma})^2} \right\}, \qquad \mu \in \mathbf{R},\ \sigma^2 \in \mathbf{R}^+. \tag{12}$$

In order to implement the EM algorithm, it is convenient to rewrite (11) using the fact that the joint density of Y, N_T and X can be decomposed as

$$f(y, n_T, x) = f(y) \cdot f(n_T|y) \cdot f(x|n_T, y).$$

To compute the conditional expectation of the complete log-likelihood function, we need to determine the distribution of $N_T|Y$ and of $(X|Y, N_T)$.

Denoting with Z the r.v. having the complete-data distribution, i.e. $Z \sim N(\mu, \sigma^2)$, it can be shown (McLachlan and Krishnan 1996, pag. 76-77) that $N_T|Y$ has the negative binomial distribution with parameters N_U and π, where $\pi = P(Z > c) = (1 - \Phi(c^*))$. The conditional expectation is thus given by $E(N_T|Y) = N_U \cdot (1 - \pi)/\pi$.

Concerning $(X|Y, N_T)$, the quantities $E[X|Y, N_T]$ and $E[X^2|Y, N_T]$ can be computed as in the preceding section so that, to sum up, the E-step is given by

$$E(N_T|Y) \stackrel{def}{=} \hat{N}_T = N_U \cdot (1 - \pi)/\pi;$$

$$E(X|\boldsymbol{\theta}^{(t)}, Y) = \mu^{(t)} - \sigma^{(t)} \cdot \alpha(\boldsymbol{\theta}^{(t)}); \tag{13}$$

$$E(X^2|\boldsymbol{\theta}^{(t)}, Y) = \tau^{(t)} \cdot [1 - c^{*(t)} \cdot \alpha(\boldsymbol{\theta}^{(t)}) - (\alpha(\boldsymbol{\theta}^{(t)}))^2] + [E(X|\boldsymbol{\theta}^{(t)}, Y)]^2. \tag{14}$$

As for the M-step, the equations are analogous to the censored case, with N_C replaced by $N_T^{(t)}$:

$$\mu^{(t+1)} = \frac{1}{N_U + N_T^{(t)}} \left(\sum_{i=1}^{N_U} y_i + N_T^{(t)} \cdot E(X|\boldsymbol{\theta}^{(t)}, \mathbf{y}) \right); \tag{15}$$

$$\tau^{(t+1)} = \frac{1}{N_U + N_T^{(t)}} \left(\sum_{i=1}^{N_U} y_i^2 + N_T^{(t)} \cdot E(X^2|\boldsymbol{\theta}^{(t)}, \mathbf{y}) \right) - (\mu^{(t+1)})^2. \tag{16}$$

These equations are then iterated until convergence to get the MLE's.

In conclusion, it can be seen that the EM algorithm is easily implemented in both the censored and truncated case.

As usual, the problem can also be tackled by means of standard optimization tools. However, gradient-based algorithms (for example, Newton-Raphson, Scoring, BFGS), may not be the best solution in this case: we have already mentioned that the first and second derivative of the log-likelihood function cannot be obtained in closed form for (6) and (12). Moreover, the computational burden associated to the usual alternative, consisting in computing numerically the derivatives, is not negligible: this fact, besides slowing the algorithm, is likely to cause unstable results, in particular as concerns the Hessian matrix.

Another possibility consists of using a direct search method, as for example the Nelder-Mead simplex algorithm. The algorithms in this family do not require any gradient information, with the consequence of being usually less efficient: as pointed out by Press *et al.* (1992, sect. 10.4) they may be most useful when the aim consists in "get something working quickly". A detailed comparison of the EM, BFGS and NM algorithms will be performed in section 4.

3.4 Bootstrap for standard errors

The computation of standard errors in the EM framework is somewhat involved (McLachlan and Krishnan 1996, chap. 4), so that we prefer to assess them using the nonparametric bootstrap.

Roughly speaking, the nonparametric bootstrap is based on sampling with replacement from the observed data; more formally, this means sampling from the empirical distribution function F_N, which puts probability $1/N$ on each observation. Denoting with $\boldsymbol{\Psi}$ the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$, the actual implementation consists of the following steps:

(*i*) generate a sample (called *bootstrap sample*) from the empirical distribution of the data at hand;

(*ii*) use it to compute the MLE's of the parameters;

(*iii*) repeat steps (*i*) and (*ii*) a large number of times B and evaluate $\hat{\boldsymbol{\Psi}}^B = (1/B)\sum_{i=1}^{B}(\hat{\boldsymbol{\theta}}^{(i)} - \hat{\bar{\boldsymbol{\theta}}}^{(B)})(\hat{\boldsymbol{\theta}}^{(i)} - \hat{\bar{\boldsymbol{\theta}}}^{(B)})'$, where $\hat{\boldsymbol{\theta}}^{(i)}$ is the MLE of $\boldsymbol{\theta}$ obtained at step (*ii*) with the i-th bootstrap sample and $\hat{\bar{\boldsymbol{\theta}}}^{(B)} = (1/B)\sum_{i=1}^{B}\boldsymbol{\theta}^{(i)}$ is the mean of the bootstrap distribution of $\hat{\boldsymbol{\theta}}$.

It has been shown by Efron (1982) that $\hat{\boldsymbol{\Psi}}^B$ is, under fairly general conditions, a consistent estimator of $\boldsymbol{\Psi}$. See Davison and Hinkley (1997) for further details.

It is worth noting that the empirical distribution function is different for censored and truncated data. Whereas in the latter case it is just the pdf which puts probability mass $1/N_U$ over any of the untruncated observations, in the former it is the pdf which puts mass $1/N_U$ over any of the N_U observations and mass N_C/N over the set of censored observations; thus, for generating a bootstrap sample (i.e. for performing step (i) above), the following steps have to be repeated N times:

(a) simulate a random number from the $U(0,1)$ distribution; if it is smaller than N_C/N, increment by one the cardinality of the set of censored observations; otherwise, go to step (b);

(b) with probability $1/N_U$ choose one of the uncensored observations.

This implies that the cardinality of the set of censored observations changes at each bootstrap replication.

4 Application and simulation

4.1 A real-data application

In this section we apply the methodology developed so far to some operational losses recorded in Banca Intesa in a recent year. For confidentiality reasons, we mention neither the names of the business lines nor the year of the losses, and we have rescaled the data multiplying them by a constant. We consider data for two business lines, which will be called A and B. Table 1 shows the estimates of the parameters obtained with the EM algorithm and the standard errors computed by means of the nonparametric bootstrap.

Table 1. Univariate estimates and standard errors - Truncated data

	μ	σ^2	N_T
Business line A	3.078	4.895	52.271
$(N = 154)$	(0.50)	(1.19)	-
Business line B	1.898	2.106	234.38
$(N = 322)$	(0.44)	(0.62)	-

In order to apply the algorithm for the censored case, we used the same data, treating now the estimated number $\hat{N}_T$ of truncated observations (rounded to the nearest integer) as the true number of censored observations. Results are shown in table 2.

Table 2. Univariate estimates and standard errors - Censored data

	μ	σ^2
Business line A	3.082	4.884
	(0.16)	(0.59)
Business line B	1.898	2.106
	(0.06)	(0.13)

For comparison purposes, we also applied the Broyden-Fletcher-Goldfarb-Shanno (BFGS) and the Nelder-Mead (NM) simplex method to maximize the log-likelihood function. The BFGS algorithm is probably the most popular method in the class of quasi-Newton algorithms. Its main strength is fast convergence (it converges at a quadratic rate); however, the initial guess must be sufficiently good, or it does not converge; moreover, numerical differentiation is usually needed, and this may be a source of instability. The NM algorithm, on the other hand, is entirely based on function evaluations, so that it does not require derivatives; for this reason its use is mainly recommended for the maximization of highly discontinuous functions. For details about both algorithms we refer the interested reader to Press *et al.* (1992, chap. 10) and to the references therein.

Not surprisingly, BFGS and NM give the same point estimates as EM; the interest lies mostly in the analysis of the conditions under which the algorithms

converge and of the convergence times. Whereas we delay to the next section the investigation of the first issue, here we give some details about the second one: in table 3 we show the times and the number of iterations required to reach convergence in both the truncated and censored case.

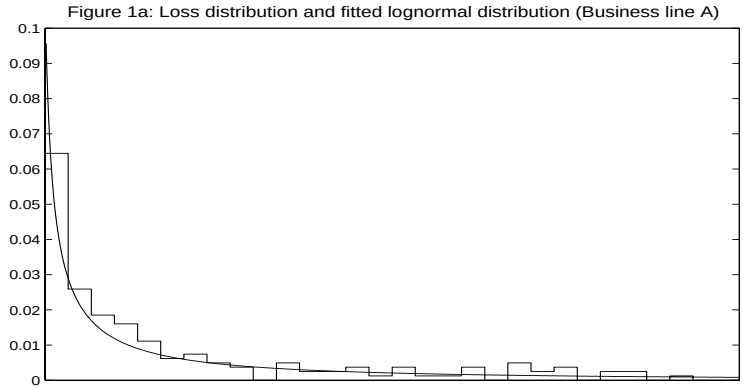
Table 3. Convergence times (in seconds) and number of iterations (nit) for EM, BFGS and NM algorithms in both the truncated (T) and censored (C) case. These results were obtained on a Pentium 4 PC, 2.66 GHz, 512 MB RAM, with the Matlab^{TM} programming language.

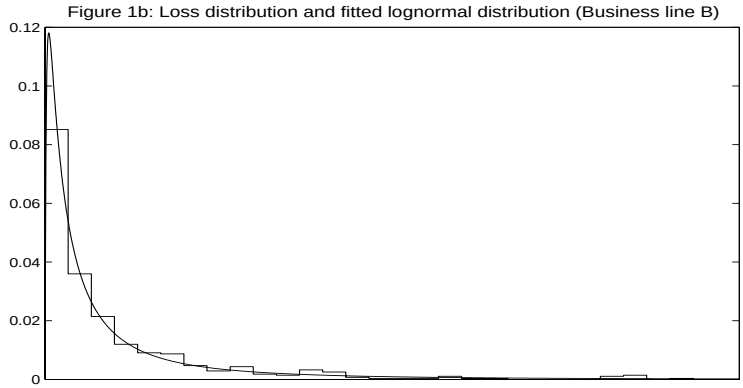
	Business line A				Business line B			
	t_T	nit_T	t_C	nit_C	t_T	nit_T	t_C	nit_C
EM	0.094	246	0.016	27	0.296	657	0.015	44
$BFGS$	0.078	14	0.109	16	0.078	13	0.093	14
NM	0.125	77	0.141	80	0.125	75	0.156	82

According to the results, two remarks are in order. First, a single iteration of the EM algorithm is much faster than an iteration of both the BFGS and the NM algorithm; this can be attributed to the fact that the E and M steps are computationally easier than function evaluation in this setup. Second, the performance of the NM algorithm deteriorates considerably in the censored case, because the function is more complicated, so that each function evaluation requires more time; similarly, derivative computation becomes heavier in the censored case, so that an analogous remark holds for BFGS, although the worsening is less pronounced. On the other hand, when switching from the truncated to the censored case, the performance of the EM algorithm improves, because not only it is not affected by the computational burden associated to the evaluation of the log-likelihood function (which is not used in the maximization process), but also gains from the larger informational content given by the knowledge of the cardinality of the set of censored observations. This gain is also reflected by the reduction of the number of iterations necessary for the EM algorithm to converge; this number remains approximately the same for BFGS and NM.

These outcomes confirm a rule of thumb of optimization: the choice of the algorithm should be guided by the problem at hand. When maximizing a likelihood function, consideration of the probabilistic aspects of the problem can often provide us with a much simpler solution with respect to a "brute force" approach based on deterministic numerical methods; see Casella and Robert (2000, pag. 18-19),

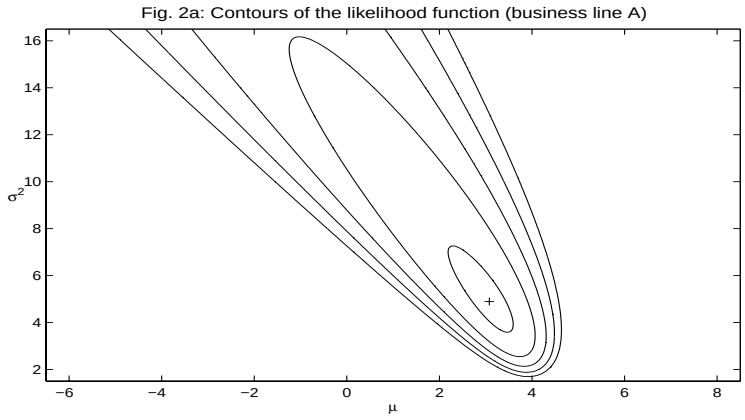
for more details about this point. In general, we notice that the EM algorithm can be very convenient in cases when the likelihood function is complicated but the E and M step are easy. Histograms of data with superimposed estimated densities are shown in figure 1a and 1b. The histogram is built using $\hat{N}_T$ as the frequency of the first bin $[0, c]$.

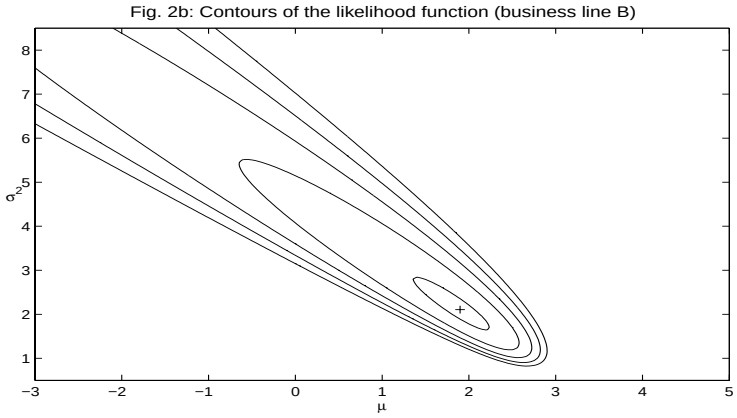




From the graphs it can be seen that the tails of the empirical distribution are slightly fatter than the estimated lognormal distribution, in particular for business line A, but the overall fit is rather good.

Figure 2 shows contour plots of the likelihood functions for the two business lines; the dot marks the maximum found by the algorithms.





Finally, CaR measures are computed by means of the procedure outlined in section 2 (with 10,000 Monte Carlo replications); the results are displayed in table 4.

Table 4. Daily Capital at Risk in Euros at the 99.9% level, based on 10,000 Monte Carlo replications

	Daily CaR
Business line A	$1,989,604$
Business line B	$393,918$

The parameters ν, μ and σ^2 have been estimated using one year of data; for the simulation, thanks to the properties of the Poisson process and the invaiance of MLE's, we used a Poisson distribution with parameter $\hat{\lambda} = \hat{\nu}/250$ to obtain daily measures (assuming 250 working days per year).

The fact that data are truncated has an impact on the estimation of the parameter ν as well: the MLE of ν is indeed the number of losses observed in one year, but with truncated data we do not know the number of losses smaller than c. However, the EM algorithm gives an estimate of this number, $\hat{N}_T$, so that the MLE of ν is $\hat{\nu} = \hat{N}_T + N_U$. It is worth noting that when using the BFGS and NM algorithm no estimate of N_T is available and thus maximum likelihood estimation of λ is impossible.

4.2 Simulation

It is well known that MLE's are consistent and asymptotically normal. However, the asymptotic theory of maximum likelihood estimation does not give any indication about the behavior of MLE's in small samples. Considering the relevance of the issue, the aim of this section consists in studying the rate of convergence of the estimators. Given the sample size N, the simulation experiment proceeds as follows:

(i) simulate a random sample of size N from the censored (truncated) normal distribution with the numerical parameter values estimated above $\hat{\mu}_C$ ($\hat{\mu}_T$) and $\hat{\sigma}^2_C$ ($\hat{\sigma}^2_T$);

(ii) estimate the parameters of the censored (truncated) normal distribution;

(iii) repeat the first two steps a large number of times M and compute mean and standard deviation of the results.

We perform the experiment for sample sizes $N = 20$, 40, 70, 100, 154, 200, 400 in case of business line A and $N = 20$, 40, 70, 100, 150, 200, 322, 400 for business line B; the results are shown in table 5 and 6.

Table 5. Simulation-based estimates and standard errors (Business Line A)

	$\hat{\mu}_T$	$\hat{\sigma}^2_T$	$\hat{\mu}_C$	$\hat{\sigma}^2_C$
$N = 20$	2.70 (1.74)	5.52 (4.18)	3.05 (0.52)	4.90 (2.13)
$N = 40$	2.92 (1.07)	5.18 (2.64)	3.08 (0.36)	4.84 (1.34)
$N = 70$	2.93 (0.85)	5.19 (2.12)	3.07 (0.27)	4.89 (1.01)
$N = 100$	3.00 (0.61)	5.04 (1.56)	3.08 (0.24)	4.89 (0.84)
$N = 154$	3.04 (0.47)	4.94 (1.19)	3.07 (0.19)	4.87 (0.69)
$N = 200$	3.04 (0.42)	4.98 (1.01)	3.08 (1.10)	4.90 (0.59)
$N = 400$	3.06 (0.28)	4.91 (0.73)	3.07 (0.12)	4.88 (0.42)

Table 6. Simulation-based estimates and standard errors (Business Line B)

	$\hat{\mu}_T$	$\hat{\sigma}_T^2$	$\hat{\mu}_C$	$\hat{\sigma}_C^2$
$N = 20$	1.62 (1.36)	2.34 (1.80)	1.88 (0.38)	2.08 (1.01)
$N = 40$	1.74 (1.02)	2.24 (1.31)	1.89 (0.26)	2.08 (0.66)
$N = 70$	1.78 (0.77)	2.23 (1.03)	1.89 (0.20)	2.12 (0.51)
$N = 100$	1.77 (0.67)	2.24 (0.87)	1.89 (0.16)	2.09 (0.44)
$N = 150$	1.80 (0.54)	2.20 (0.71)	1.89 (0.14)	2.10 (0.36)
$N = 200$	1.87 (1.44)	2.11 (0.57)	1.89 (0.12)	2.11 (0.31)
$N = 322$	1.87 (0.33)	2.14 (0.44)	1.89 (0.09)	2.11 (0.25)
$N = 400$	1.86 (0.28)	2.15 (0.39)	1.89 (0.08)	2.12 (0.22)

The main lesson to be learned from the simulations is that the "informational gap" associated with the switch from the censored to the truncated case is reflected by the increased sample size necessary to get stable results. In the truncated case the variability of the estimators is much higher, and the overall conclusion is that in this case at least 100 observations are needed to perform a reliable analysis.

In the censored case, when using the BFGS or NM algorithm to maximize the likelihood function (5), there is another numerical problem: the term $[\Phi(c^*)]^{N_C}$ becomes very small when N_C is large, and this can cause underflow problems. To analyze this issue, we performed an experiment with simulated data: for given N, we generated N standard normal random numbers and, for increasing values of the threshold c, we treated the numbers smaller than c as censored values and used the BFGS and NM algorithms to estimate the parameters of the distribution. In table 7 we give, for each sample size N, the maximum value of the fraction of censored data $\phi_N = N_C/N$ for which the algorithms converge (for larger values, the log-likelihood is a too large negative number and the algorithm breaks down).

Table 7. Maximum percentage $\phi_N = N_C/N$ of censored data for which the BFGS and NM algorithms converge.

N	ϕ_N
400	88.5%
500	73.6%
600	57.5%
700	47.3%
800	42.3%
900	36.4%
1000	28.8%
1200	23.4%
1500	17.6%
2000	11.9%
3000	7.6%
4000	5.7%
5000	4.0%

It can be seen that, for example, when $N = 3000$, with more than 7.6% of censored data the algorithm aborts. Again, this problem does not affect the EM algorithm (except for the fact that the numerical value of the maximized log-likelihood function cannot be computed), because it does not use the likelihood function.

All the results shown here were obtained using the sample mean and variance of the untruncated (uncensored) observations as starting values. A quick experiment was done to check the behavior of the algorithms with different initial guesses. To this aim, the algorithm was started with several different (randomly chosen) initializations: not surprisingly, in case of bad starting values EM and NM converged, while BFGS broke down.

5 Discussion

In this paper we have derived the EM algorithm for maximum likelihood estimation of lognormal severity distributions with censored and truncated data. Using real operational losses data, we computed estimates of the parameters, standard errors and Capital at Risk. A comparison was done with respect to the BFGS algorithm and the Nead-Melder simplex algorithm, and it turned out that the EM algorithm has several advantages: (i) it is faster than BFGS and NM when, as is the case here, the functional form of the likelihood function is complicated

but, thanks to the use of the probabilistic aspects of the problem, the E and M steps are simple; this advantage is particularly evident in the censored case (ii) unlike BFGS and NM, in the truncated case it gives an estimate of the number of truncated observations, which is essential for the estimation of the frequency distribution and for the computation of risk measures; (iii) in the censored case, when the sample size is large and the fraction of censored data is not sufficiently small, the BFGS and NM algorithms break down; in such a setup the EM algorithm converges normally. Finally, an advantage of both EM and NM is that they converge even for "very bad" starting points, whereas BFGS fails to converge if not initialized properly.

The field of operational risk represents a real challenge for the statistician, and many problems are still open to future research. First, the appropriateness of the lognormal model should not be taken for granted: it has often been noted that the severity distribution is heavy-tailed; in this case, Extreme Value Theory has sometimes been proposed (see Embrechts *et al.* 2003 for a thorough analysis). Second, the extension of the techniques to the multivariate case is still at a pioneering stage. Given the analytical intractability of the univariate distributions, a multidimensional model should likely be based on copulas (see Nyström and Skoglund 2002 and Frachot *et al.* 2001 for different proposals about this point); although in some cases losses in different business lines may be independent, we do not agree with the extreme point of view by Roehr (2002), who claims that correlation is spurious and losses are largely independent. Third, the empirical evidence in favor of the Poisson assumption is weak, because frequency data are typically characterized by overdispersion with respect to the Poisson distribution. This feature can possibly be accommodated by the negative binomial distribution; in case of large overdispersion, Poisson mixtures are a more flexible model. Finally, in CaR computation, it would be important to find more efficient simulation procedures: the high confidence levels requested by the regulators have the consequence that the variance of the quantile obtained by crude Monte Carlo is very large. Application of variance reduction techniques to this problem would be extremely useful.

Acknowledgements

I would like to thank Riccardo Cateni for providing the data and Stefano Benati and Giuseppe Espa for helpful comments on an earlier draft of the paper.

References

Basel Committee on Banking Supervision (2001), *Consultative Document - Operational Risk*, Bank for International Settlements, `www.bis.org/publ/bcbsca07.pdf`.

Crouhy, C., Galai, D. and Mark, R. (2000), *Risk Management*, New York, McGraw-Hill.

Cruz, M.G. (2002), *Modeling, Measuring and Hedging Operational Risk*, New York, Wiley.

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

Dempster, A.P., Laird, N.M. e 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.

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

Embrechts, P., Furrer, H., Kaufmann, R. (2003), "Quantifying Regulatory Capital for Operational Risk", *Derivatives Use, Trading & Regulation*, 9, 217-233.

Frachot, A., Georges, P. and Roncalli, T. (2001), "Loss Distribution Approach for Operational Risk", *working paper*, Credit Lyonnais.

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

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press.

McLachlan, G.J. and Krishnan, T. (1996), *The EM Algorithm and Extensions*, New York, Wiley.

Nyström, K. and Skoglund, J. (2002), "Quantitative Operational Risk Management", *working paper*, Swedbank.

Panjer, H.H. (1981), "Recursive Evaluation of Compound Distributions", *Astin Bulletin*, 12, 22-26.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992), *Numerical Recipes in C*, second edition, Cambridge, Cambridge University Press.

Robert, C.P. and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York, Springer.

Roehr, A. (2002), "Modelling Operational Losses", *Algo Research Quarterly*, 5, 2002.

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

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO