ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Measuring risk with ordinal variables

Silvia Figini
(Università di Pavia)

Paolo Giudici
(Università di Pavia)

 # 32 (02-13)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

February 2013

Measuring risk with ordinal variables

Silvia Figini, Paolo Giudici

Department of Economics and Management, University of Pavia

giudici@unipv.it

February 1, 2013

Abstract

In this paper we propose a novel approach to measure risks, when the data available are expressed in an ordinal scale.

As a result we obtain a new index of risk bounded between 0 and 1, that leads to a risk ordering that is consistent with a stochastic dominance approach.

The proposed measure, being non parametric, can be applied to a wide range of problems, where data are ordinal and where a point estimate of risk is needed. We also provide a method to calculate confidence intervals for the proposed risk measure, in a Bayesian non parametric framework.

In order to evaluate the actual performance of what we propose, we analyse a database provided by a telecommunication company, with the final aim of measuring operational risks, starting from a self-assessment questionnaire.

keywords Risk measurement, Ordinal variables, Operational risk

1 Introduction

The guidelines proposed by the Basel Committee on Banking Supervision for the banking sector (known as Basel 2) encourage banking institutions to use mathematical and statistical approaches for the computation of their capital charge covering operational risk.

On the other hand, non financial institutions are motivated to measure operational risk by the need of having under control the quality of their processes. In this framework operational risk measurement should be seen as a preventive diagnostics aimed at prioritising interventions and triggering actions.

Operational risk is usually classified in event types, according to the type of risk involved; and in business lines according to the sector of area of the company that is mostly affected by the risk events. In order to measure the operational risk, for each business lines and for a given class of event types, the standard approach is to derive the severity and the frequency distributions of the random losses associated with that risk, and to integrate them through a Monte Carlo simulation. The most employed approach to operational risk modelling has been developed for financial institutions and considers a quantitative approach to calculate the Value at Risk and derive the total economic capital required to protect an institution against possible losses (see e.g. Cruz, 2002, Alexander 2003, Giudici et al., 2009).

However, when data are available only at the ordinal scale, as it is common for non financial companies, a pure quantitative approach is not possible. In this paper we show that operational risk measurement is possible also in this case, using an ordinal approach. Related approaches can be found in the literature on customer satisfaction, for which we refer the reader to Cerchiello et al., 2010, and the references therein. The neasure we propose here is novel, and novel is the application of ordinal measures to operational risk. Of course, results from ordinal modelling of operational risk will have to be interpreted accordingly: as priority of intervention indicators, rather than measures of capital to be allocated. This is the main contribution of our paper, and its main novelty.

More precisely, we propose a novel methodology to measure operational losses, starting from ordinal loss random variables. As a result we obtain an ordinal measure of risk which respects stochastic dominance ordering.

Empirical evidences of the performance of what proposed are given on a real data set, that concerns the measurement of operational risk management for a telecommunication company.

The paper is structured as follows: Section 2 describes the requirements and the theoretical background of our methodology and Section 3 reports the empirical evidences obtained. Finally, Section 4 contains some conclusion remarks.

2 Proposal

In this section we present a general framework to measure risks, on the basis of ordinal variables. The framework will be presented, without loss of generality, with reference to operational risks.

Operational data for risk measurement are typically summarised in a matrix composed of I event types (the columns of the matrix) and J business lines (the rows of the matrix).

Let E_{ij} be a risk event, in the i-th event type $(i = 1, \ldots, I)$ and in the j-th business line $(j = 1, \ldots, J)$. For each combination of event type and business line, we have two different measures of risk: the frequency (how many risk events have appeared in that combination) and the severity (the mean loss of the events in that combination).

In the Basel 2 framework the severity is continuous random variable, while in the context of non financial companies the severity is generally expressed in an ordinal scale, characterised by S distinct levels ordered according to the corresponding magnitude (for example S=3, with H=high severity; M=medium severity and L=low severity). In this context, in order to summarise the frequency and the severity in a location measure, we may structure a loss contingency table, which counts, for each event type - business line and a given severity level, the absolute frequency. More formally, let n_{H11} be the number of times for which the first event type in the first business lines appears with high severity; n_{M11} the number of times for which the first event type appears in the first business line with medium severity and n_{L11} the number of times for which the first event appears in the first business line with low severity. In general, let n_{Hij}, n_{Mij} and n_{Lij} be the number of times for which high, medium or low severity occur for the event type $i = 1, \ldots, I$ in the $j = 1, \ldots, J$ business line.

The above counts can be represented in a contingency table composed of J rows, representing the business lines $(BL_1, \ldots, BL_J)$ and $I \times S$ columns, equivalent to the number of event types multiplied by the levels of severity S under analysis (in our running example three: high=H, medium=M and low=L). Each cell in Table 1 contains the frequency of a combination of business line (row) and event type*severity level (column). RM and CM are the marginal distributions and the sum across each of them is equal to N, which represents the total number of risk events observed.

The location measure we propose is based on the cumulative distribution function of each cell variable in the loss contingency table. The latter can be calculated from each relative frequency $p_{hij} = n_{hij}/N$. Let F_{hij} be the cumulative distribution function calculated in each cell of the loss contingency table.

On the basis of data structured as in Table 1, we propose to estimate the operational risk in each business line / event type combination through a Stochastic Dominance Index (SDI) measure which is defined as follows:

$$SDI_{ij} = \sum_{h=1}^{S} \frac{F_{hij}}{S} = \sum_{h=1}^{S} \sum_{l=1}^{h} p_{lij}, \tag{1}$$

for $i = 1, \ldots, I$, $j = 1, \ldots, J$ and F_{hij} for $h = 1, \ldots, K$ are the cumulative frequencies of each event type / business line combination. The SDI_{ij} is a novel proposal, in the context of risk assessment; it has been used in the context of customer satisfaction (see Cerchiello et al., 2010, and the references therein) and it appears a good approach when risk data are expressed in an ordinal scale. Note that $SDI_{ij} = 0$ when the risk event never appears; and $SDI_{ij} = 1$ when the risk event is concentrated only on values with highest severity.

Furthermore, it is easy to show that SDI_{ij} is bounded between 0 and 1. On the basis of the well known properties of the cumulative distribution function, the proposed index could be employed to compare events of interest and business lines, producing an ordering among risks. These results may be very useful for the data owner, to prioritise interventions, also in terms of improvement of the related process controls.

Furthermore, since our index is based on the cumulative distribution function, we are able to introduce an ordering criteria through the stochastic dominance approaches for ordinal variables, thus generating an order of preference on the set of $I \times J$ distribution functions involved in the analysis (see e.g. Kaur et al., 1994 and Shaked et al., 1994).

Stochastic dominance approaches are a good solution to compare the ordinal loss distribution looking at all loss distributions arising from the combinations between business lines and event types. In this way the decision maker can compare operational risks and prioritise interventions. We also remark that comparing loss distributions in terms of stochastic dominance is a good approach to take into account the whole distributions and not only particular location measures or quantiles, as done in standard approaches.

A further practical advantage of the introduced measure is that it can be used to combine risk measures in a simple way that, besides, may preserve stochastic dominance.

For example, in order to aggregate event type risks over different business lines, we can use the following equation, that expresses the overall business line risk as a geometric mean of the measure of risk associated with each event type in that business line:

$$SDI_j = (\prod_{i=1}^{I} SDI_{ij})^{1/I}. \qquad (2)$$

Similarly, in order to aggregate business line over different event type risks , we can use the following equation, that expresses the overall event type risk as a geometric mean of the measure of risk associated with each business line in that event type risk:

$$SDI_i = (\prod_{j=1}^{J} SDI_{ij})^{1/J}. \tag{3}$$

The above expressions show that risks over different units may be assumed to interact in a multiplicative way, being the units of an integrated system. Jean (1980) has shown that the geometric mean is a necessary condition to preserve stochastic dominance ranking when aggregating distribution functions. This because the geometric mean can be expressed as an arithmetic mean of logarithms and, since the logarithmic function is monotone, the corresponding geometric means are ordered. We remark that it is not true, in general, that the geometric mean is a sufficient condition to preserve stochastic dominance ranking.

However, there may be some specific distributions where a geometric mean ranking is not only necessary, but also sufficient to preserve stochastic dominance ranking, see for example Levy (1973) and Jean (1984).

The geometric mean has also a number of practical motivations. For example, many risk management problems require to combine multiplicatively the different components of risk; in operational risk the distribution of a financial loss is obtained multipling the frequency distribution with the severity distribution.

The relationship between the geometric mean and the stochastic dominance framework is central to provide a simple and mathematically coherent approach to the construction of effective operational risk measures.

Specifically, it can be demonstrated that the logarithm of the geometric mean is a coherent measure of risk (see e.g. Landenna, 1994 and Artzner et al. 1999).

So far we have discussed a methodology to derive a point estimate of the expected risk, through the SDI measure. As in risk management we are typically interested in percentile values, it is of interest to derive confidence interval risk measures for ordinal variables. One possible approach is to derive a Bayesian distribution for the SDI measures assuming that the $I \times J$ observed frequencies p_{hij}, $h = 1, \ldots, S$ independently follow a Multinomial Distribution with prior parameter $\theta_{ij} = \theta_{hij}, \quad h = 1, \ldots, S$ and that θ_{ij} follows a Dirichlet prior distribution with prior parameters $\alpha_{ij} = \alpha_{hji}, \quad h = 1, \ldots, S$. From standard prior to posterior analysis (see e.g. Bernardo and Smith, 1994), we know that the posterior distribution of θ_{ij} is a Dirichlet with parameters $\alpha^*_{ij} = \alpha_{hji} + n_{hij} \quad h = 1, \ldots, S$, from which confidence intervals on SDI can be derived approximately by means of Markov Chain Monte Carlo simulations (Robert and Casella, 2004). In the next Section we exemplify, by means of a practical application, how our proposed ordinal risk measure and the related confidence interval can be used and integrated in a complex problem of risk management, that involves the objective of ordering operational risk.

3 Application

This section describes the application of the proposed methodology to real data provided by Tadiran Telecom. Tadiran is a telecommunication company that installs telephone exchange systems (called Private Branch eXchange: PBX) and offers post-installation technical assistance for upgrading and problem resolution in hardware, software, interface, network communications, security. The service is offered to a wide range of customers/enterprises grouped in several business lines. The main objective is to estimate, for each business line and for each problem (event type) a measure of risk. The data set we use here is composed of observations on 1126 PBX systems. The data is collected by the Tadiran call centre and remote diagnostic capabilities embedded in the PBXs. A customer from a specific business line (i.e. Banking, Health, Defence, Cooperative etc) calls Tadiran to signal problems about hardware (H), software (ST), interface (I), network communications (NC) or security (S). The call centre operator inputs to the system a reference to the call listing the PBX number and the level of severity of the problem as reported by the customer (High, Medium, Low).

The structure of the collected data is a data base with problem reports listed by PBX. For each event type, the operator enters a level of severity (Low, Medium, High). A single customer could report more then one event in a time window. For example, a customer in the banking business line called Tadiran three times to report event type problems of hardware (severity=L), software (severity=M), interface (severity=L). Another customer in the same banking business line called Tadiran four times only to report event type problems of hardware (severity=H). Our database correspondingly updates the banking business line as follows: one more frequency count to the software event type, level Medium; one more frequency count to the interface event type, level Low and four more frequency counts for the hardware event type: one with level Low and four with level High. A similar classification can be repeated for all the 1126 observed reported PBX, leading to a complete cross classification of the risk events in seven business lines, five event types and three severity levels. On the basis of this cross classification we can calculate the corresponding 7×5 summary SDI measures according to equation (1). Such calculation is reported in Table 1.

Table 1 about here

On the basis of the results summarised in Table 1, it can be easily seen that the hardware event type is very important for the business lines Hotel, Industry and Banking; the software event type for Banking, and Cooperatives; the network communication event type is overall a problem common to the business lines considered; the interface event type is not a problem only for Defence.

An important advantage of the proposed SDI measure is that it can be easily aggregated over different dimensions, such as business lines and event types. To reach this objective, the geometric mean of different SDI measures, can be applied.

In Table 2 we have reported the results obtained applying the geometric mean on the SDI measures obtained in Table 2 thus obtaining an overall SDI measure of risk for each Business Line.

Table 2 about here

On the basis of Table 2 we can underline that Hotels is the Business Line with higher risk followed by Health, Industry and Banking. On the basis of the results reported in Table 1 we are not able to give a relative confidence interval for the SDI measure. In order to obtain confidence intervals for the SDI measure we have followed the Bayesian approach described in the previous section. In order to choose α_{hij} since the a priori knowledge for the data at hand is poor we have decided to use a uniform distribution, setting $\alpha_{hij} = 1$. In addition, we have performed sensitivity analysis, with alternative prior specifications, and the results do not substantially change. With the software R we have written a function which generates a sample from the posterior distribution of a multinomial likelihood with a Dirichlet prior. In order to obtain the posterior sample, and of course the posterior distribution of the SDI, we have implemented Gibbs Sampling MCMC algorithm with a number of iterations equal to 10000. The results are in Table 3. Table 3 reports, for the banking business line and all five event types,

the SDI value obtained using the Bayesian estimate.

Table 3 about here

We remark that in the Bayesian approach using the SDI and the relative confidence interval we are able to underline in a more clear way the uncertainty about the punctual estimation of the SDI for each business line and event type. More precisely, the length of the Bayesian confidence interval, provide us a measure of reliability: if the confidence interval is small, it means that almost surely the punctual estimation made on the basis of the data at hand is precise. From Table 3, for example, we derive that the largest confidence interval are for the event type Hardware in the Business Lines Computers and Defence and for the event type Interface in the Business Line Defence. The outcome of this analysis, provides to the data owner a clear picture to prioritise intervention on specific business lines and event types. We remark that our approach is rather general and it works well also in different area of application where the data are expressed in an ordinal scale as in self assessment questionnaire or with customer satisfaction data. In terms of computational aspects, our approach is easy to be implemented both in the classical and in the Bayesian framework.

Business Line	H	SW	NC	I	S
Banking	0.86	0.76	0.69	0.96	0.72
Computers	0.00	0.67	0.68	1.00	0.71
Cooperatives	0.73	0.74	0.69	1.00	0.75
Defence	0.00	0.67	0.67	0.00	0.83
Health	0.83	0.72	0.72	1.00	0.82
Hotels	1.00	0.73	0.72	1.00	0.85
Industry	1.00	0.67	0.68	1.00	0.69

Table 1: SDI measure for business lines and event types

Business Line	SDI Overall
Banking	0.79
Computers	0.13
Cooperatives	0.77
Defence	0.02
Health	0.81
Hotels	0.85
Industry	0.79

Table 2: Overall SDI measure for each business line

Business Line	H	SW	NC	I	S
Banking	0.67-0.93	0.71-0.79	0.66-0.72	0.73-0.97	0.67-0.75
Computers	0.40-0.92	0.57-0.75	0.63-0.72	0.60-0.97	0.59-0.79
Cooperatives	0.5-0.89	0.68-0.78	0.66-0.72	0.55-0.96	0.67-0.81
Defence	0.4-0.92	0.57-0.75	0.58-0.74	0.4-0.93	0.65-0.9
Health	0.68-0.91	0.68-0.75	0.68-0.75	0.83-0.99	0.74-0.86
Hotels	0.55-0.96	0.64-0.79	0.63-0.75	0.65-0.97	0.68-0.90
Industry	0.49-0.95	0.64-0.69	0.66-0.70	0.55-0.96	0.64-0.73

Table 3: Bayesian Confidence Interval for the SDI measure in each business line

4 Conclusions

In this paper we have proposed a novel measure of risk (SDI), valid when loss data are expressed in an ordinal scale, based on the cumulative distribution function of the frequencies at different severity levels. The proposed measure is coherent measure of risk, which preserves stochastic dominance ordering and, being non parametric, it can be applied to a wide range of problems.
A further aspect of our proposal, is that it allows to integrate different measures of risk (over different business lines/event types) through their geometric mean: Monte Carlo computations are not necessary, in order to merge the statistical random variables involved, as in the standard quantitative approach.
In the paper we have also provided a method to derive confidence intervals for the proposed SDI measure using Bayesian

theory and MCMC computations. The proposed methods have been applied on a real operational risk data base that concerns a telecom company. The obtained empirical evidences shows that our proposal may be an effective statistical method, aimed at prioritise interventions on the control system, so to effectively reduce the impact of risks, ex ante rather than a posteriori, as done by allocating capital in the standard quantitative approach.

5 Acknowledgments

We thank the referees for the comments that have improved the qaulity of the paper. We also thank the European Project MUSING: Multy industry semantic based business intelligence, and the Italian Ministry of Research project MISURA: Multivariate Statistical models for risk assesment, for financial and scientific support.

6 References

1. **Alexander C. (2003). Operational Risk: Regulation, Analysis and Management, Prentice Hall.**

2. **Artzner, P., Delbaen, J. and Heat, D. (1999): Coherent measures of risk, *Mathematical Finance*, vol. 9, pp. 203-228.**

3. **B Bernardo and A. Smith (1994). Bayesian Theory, Wiley, New York.**

4. **Cerchiello, P., Dequarti E., Giudici, P., Magni C. (2010): Scorecard models to evaluate perceived quality of academic teaching.*Statistica e Applicazioni*, vol. 2 pp. 145-156.**

5. **Cruz, M. (2002): Modeling, Measuring and Hedging Operational Risk, Wiley. In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.**

6. **Giudici, P. and Figini, S. (2009): Applied data mining for business and industry. Wiley, London.**

7. **Landenna, G. (1994): Fondamenti di statistica descrittiva (in italian), il Mulino.**

8. **Levy, H. (1973): Stochastic Dominance Among Log-Normal Prospects, *International Economic Review*, vol. 14 n. 3.**

9. **Jean, W.H. (1980): The geometric mean and stochastic dominance, *Journal of Finance*, vol. 35, pp. 151-158.**

10. Jean, W.H. (1984): The harmonic mean and other necessary conditions for stochastic dominance, *Journal of Finance*, vol. 39, pp. 527-534.

11. Kaur, A., B.L. Prakasa Rao, and H. Singh (1994): Testing for Second-Order Stochastic Dominance of Two Distributions, *Econometric Theory*, vol. 10, pp. 849 - 866.

12. Robert C. and G. Casella (2004). Monte Carlo Statistical Methods, (second edition), Springer-Verlag, New York.

13. Shaked, M. and Shanthikumar, G.J. (1994): Stochastic Orders and Their Applications, Academic Press, Boston.