



**HAL**  
open science

# Statistical Indicators for the Optimal Prediction of Failure Times of Reliability Stochastic Systems: A Rational Expectations-Like Approach

Jørgen Vitting Andersen, Roy Cerqueti, Jessica Riccioni

► **To cite this version:**

Jørgen Vitting Andersen, Roy Cerqueti, Jessica Riccioni. Statistical Indicators for the Optimal Prediction of Failure Times of Reliability Stochastic Systems: A Rational Expectations-Like Approach. IEEE Transactions on Reliability, In press, 10.2139/ssrn.4201520 . halshs-03891600

**HAL Id: halshs-03891600**

**<https://shs.hal.science/halshs-03891600>**

Submitted on 26 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Statistical indicators for the optimal prediction of failure times  
of reliability stochastic systems: a rational expectations-like  
approach.

Jorgen-Vitting Andersen

Université Paris 1 Pantheon-Sorbonne – CNRS, Centre d’Economie de la Sorbonne  
Maison des Sciences Economiques, 106-112 Boulevard de l’Hôpital, 75647, Paris Cedex 13 , France.

E-mail: Jorgen-Vitting.Andersen@univ-paris1.fr

Roy Cerqueti

Sapienza University of Rome – Department of Social and Economic Sciences

Piazzale Aldo Moro, 5 – 00185, Rome, Italy

Tel.: +39 06 4991 0511

E-mail: roy.cerqueti@uniroma1.it

and

London South Bank University – School of Business

London SE1 0AA, UK

and

Université d’Angers – GRANEM

SFR CONFLUENCES, F-49000, Angers, France

Jessica Riccioni\*

Sapienza University of Rome – Department of Social and Economic Sciences

Piazzale Aldo Moro, 5 – 00185, Rome, Italy

Tel.: +39 06 4991 0511

E-mail: jessica.riccioni@uniroma1.it

June 23, 2022

## Abstract

The estimation of the optimal failure time of stochastic systems is the aim of this paper. We are in the field of reliability theory, and we merge the Bayesian-like method of estimating the failure time of the so-called  $k$ -out-of- $n$  systems with rational expectations-like expectations, thus obtaining a new forecasting method and innovative and original results. To do this, we simulate the path of two sets of systems from zero time to failure time, using a series of statistical measures that allow us to compare the simulation procedures and understand which is the best predictive combination.

**Keywords:**  $k$ -out-of- $n$  systems; optimal failure forecasting; reliability theory; rational expectations.

## 1 Introduction

This paper arises from the need to deal with the uncertainty about the failure time of stochastic systems.

Our aim is devoted to estimating the optimal failure time of a set of stochastic systems by exploiting the potential of some statistical measures through a rational expectations approach. The fusion of the framework of rational expectations with that of reliability theory with a special focus on  $k$ -out-of- $n$  systems will be interesting and innovative.

Among the possible existing approaches for the prediction of failure times of stochastic systems, we propose the analysis of the path of failures of the components of reliability systems. The estimation takes place with a Bayesian computational method – to be intended here that we provide an estimation conditioned to the past information on the randomness collected during the evolution of the simulated systems.

In our study, we collect all the information related to a set of systems that we call *catalog*, which allows the Bayesian estimate of the average failure time of our simulated "real" stochastic systems. To determine the failure of every single system, the number and importance of the components are fundamental; in fact, the systems' reliability depends on its components' reliability. We are talking about  $k$ -out-of- $n$  systems.

Our study's innovation concerns applying rational expectations to reliability systems. A clarification word on the meaning of rational expectations in our context is in order. Indeed, rational expectations come from the economic area, where they provide a response to the [57]'s original statement, saying that it is *often necessary to make sensible predictions about the way expectations would change when either the amount of available information or the structure of the system is changed*. The original idea of Muth finds applications in macroeconomics and financial markets theory. In this

---

\*Corresponding author.

paper, as we will see, we take for us the grounding idea of rational expectations and apply it to the context of reliability theory. Specifically, through rational expectations, we can predict failure times on the basis of the knowledge about the past experience of our stochastic systems and condition the results on the information provided by the system – with interacting elements with unknown random lifetimes – thus obtaining a continuous gain on prediction performance.

A peculiarity of our model is called *failed component allocation rule* and acts on the weight distribution of the components once they have failed: the weight of the failed component is redistributed on the weights of the components still active within the system.

So we propose a method by using a Bayesian approach to compare  $k$ -out-of- $n$  real systems (called *in-vivo*) and  $k$ -out-of- $n$  cataloged systems obtained through an extensive simulation procedure which have components characterized by a dependence due to the structure of the systems. We deal with systems composed of homogeneous and heterogeneous components to perform the optimal failure prediction of stochastic systems with interconnected components by exploiting all available information efficiently through rational expectations to process it and make forecasts.

Usually, the methods used in the literature to issue forecasts on the failure times of stochastic systems through scenario analyses using the reliability of its components exploit what happens to real systems. In this regard, refer to [45], [73], [31] and [88]) for an in-depth study on this field.

Here, we introduce a new methodology that takes its cue from the approach introduced by [2, 80] devised on the basis of a method called in the literature "RTP" (reverse tracing of precursors) for the prediction of earthquakes proposed by [43] and [78]. We estimate the average failure times of real systems (which represent the goal of our investigation) on the basis of the information collected on the *catalog* systems. The information that is a priori recorded concerns all the weights of the components at each lifetime of a very large number of simulated systems, which will represent the starting point for the prediction of *in-vivo* systems. However, unlike Andersen and Sornette, we insert an interaction into our systems so that the failure of a component affects the rest of the still-active system.

The information collected at each specific time and for each simulated system concerns the weights of the components (adjusted from time to time according to the reallocation rule) and the final failure time. This cataloged set will then be compared with what happens in real systems to get our prediction.

These systems allow us to calculate rational expectations conditioned by all the possible observations available. In our approach, we record the failure data of a very large number of systems and provide the results of different scenarios on our information set.

We seek to provide the most reliable estimate of the failure time of a stochastic system whose components are interconnected through rational expectations.

To do this, we assess the reliability of our stochastic systems on the basis of several statistical in-

dicators calculated on the realizations of the weights, and we also take into consideration some initial weight distributions to obtain comparisons both between the various cases studied and concerning what is contained in the existing literature.

We want to find which combination of initial distribution associated with each statistical indicator allows us to obtain the highest predictive power through all the information available on our systems. Regarding the statistical indicators selected, our choice was based on those with the most informative power: variance, skewness, kurtosis, Gini coefficient and Shannon entropy. About the initial distributions selected for the analysis, we generate the weights of the components from different types of distribution to investigate whether the results are consistent with the literature (see [48], [77], [4], [25], [84], and [92]). In this direction, we compared five different initial distributions: the uniform distribution and four types of Beta distribution, assigning values to the parameters to obtain cases of symmetry and asymmetry.

We intend to analyze how much the functioning of the system and its probability of failure depend on the level of the statistical value under study. We link the *catalog* systems (the levels of the statistical indicators of the weight components of stochastic systems) with the residual failure times. An average residual failure time is assigned to each observed value. We then calculate the errors with respect to the residual failure times of the statistical indicator values of the *in-vivo* systems. A comparative analysis of the error distributions will allow us to understand how to obtain the best prediction. Time remains implicit in this approach. We use rational expectations by conditioning them to all available information, and we represent errors by taking away the time from the final visualization of the results and focusing on the values of the statistical indicators and the errors associated with each value. It will be useful to understand how the statistical indicators behave individually and how the levels of errors vary according to the size class they belong to.

A large number of analyses are then built up to simulate the system's failure, and we can identify the connections among the configurations and the lifetime of the system. In so doing, we derive a probability distribution for the random time of failure of the system conditioned on the observations.

## 2 Literature review

In this section, we illustrate the literature that underlies our work. Many methods have been developed over time, and several applications within the Bayesian approach are used to estimate the failure time of our stochastic systems, which is opposed to the probabilistic one that we will mention later.

Several papers have been published in the field of operational research that deal with reliability systems by applying loss functions. In this regard, [10] calculated reliability Bayesian estimators of the mean lifetime using asymmetric loss functions on the research line of life testing problems.

[11] dealt with a priori information on the values that the parameters can assume in the context of life testing techniques. In his paper [15] approached a reliability estimation problem by providing a solution that contains the prior information in the Bayesian theory and the loss function through the exponential reliability model. [24] discussed a comparison between new unbiased estimators for the parameter in the exponential distribution and the corresponding Bayesian estimators derived. [85] considered the properties of optimal forecasts under asymmetric loss from a theoretical perspective introducing a Linex loss function in the context of real estate assessment. [90] discussed properties of estimation and prediction procedures based on Varian's asymmetric Linex loss function. He delved into examples on normal  $n$ -mean problem, on least squares estimator of multiple regression coefficients, on a linear combination of regression coefficients. In engineering design, [32] developed a reliability-based optimization method using Bayesian statistics for problems with incomplete information.

We can find many other Bayesian applications in reliability analysis that deal with multilevel heterogeneous data sets. See [7], [51], [52], [53] and [54].

In literature, as opposed to the Bayesian method, we find the probabilistic approach, which differs from the previous one for studying the probabilities of time in which a system could fail. In these models, the use of the time-conditioned reliability function plays a fundamental role and calculates the probability of failure throughout the system's life. The contributions in this field by [6], [23], [42], [65], [62], [59], [61] and [64] are very interesting.

Although our approach falls within the Bayesian type, in our study, we also dealt with coherent systems with dependent components, representing a focal point of the probabilistic approach. In fact, the systems to which we have applied our method do not contain null components, and their importance grows as the number of failed components increases. If all the components are working, the system works; if all the components stop working, the system fails. In a coherent system, all its components are relevant, and its reliability is directly related to the reliability of its components.

Among the most relevant exponents in the literature, we find Navarro and his coauthors, who developed several papers on coherent systems containing dependent components.

For example, [62] compared systems by considering the properties of their components. They developed theorems and proofs showing that properties of coherent systems could be selected on the basis of the properties of these systems and mixtures. In [63], they extended the classical signature-based mixture representation to the systems with exchangeable components. In contrast, [59] determined ordering properties for coherent systems by considering the system reliability function as a distorted function of the common component reliability function. [61] discussed about coherent systems with dependent or independent components. They demonstrated sufficient conditions on the components and lifetimes and the number of components to improve the reliability of the whole system according to different stochastic orders. The examples included support their

study applied to the optimal random allocation of components in series and parallel systems. In the same context, we also find [23], where the authors provided results for systems of two components in series and systems of two components in parallel. They showed that the lifetime could be improved by replacing these components with certain mixtures of duplicated (dependent) components. [33] studied two groups of systems with the same dependence and composition by comparing the failure characteristics of the components no longer active in the system with the system's lifetime. A similar comparison approach between two coherent systems sharing the same components was written by [89]. These last two papers used a method very similar to ours.

We, therefore, introduce at this point another theme that characterizes our paper:  $k$ -out-of- $n$ . The  $k$ -out-of- $n$  systems are those in which a system of  $n$  components requires at least  $k$  components to operate. These systems are included in the system class called coherent systems. The stochastic systems we simulate fit exactly into this type of approach. In the literature, many researchers have studied  $k$ -out-of- $n$  systems in which the failure process of each component depends on the conditions in its operating environment, and many developments have been achieved in this field of research for the evaluation of the reliability of these systems. These systems have been introduced and extensively investigated in the literature by many researchers, see [35], [14], [8], [39], [55], [58], [38], [67], [69], [72], [76], [95] and [87] to deepen this really interesting topic. On the basis of the dependence of the components, we can divide the  $k$ -out-of- $n$  systems into two groups: those with homogeneous components that have the same degrees of importance in terms of the failure of the system and others with heterogeneous components and random weights, in which each component have different relevance and make a different contribution to the system reliability.

Many scholars have studied  $k$ -out-of- $n$  systems, where the failure process of each component depends on the conditions in its operating environment.

Applications of reliability analysis to  $k$ -out-of- $n$  systems with homogeneous components (see [30], [13], [56] and [20]) are less frequent. Instead, we find many more studies and generalizations in the context of  $k$ -out-of- $n$  systems with heterogeneous components such as those introduced by [67], [49], [86], [28], [26, 27], [81], [91, 92], [79] and [93].

Some special cases are the series and the parallel systems. In the first case, we suppose  $k=n$  and in the second case  $k=1$ . See [68], [19] and [21]. Another special case regards the consecutive  $k$ -out-of- $n$  systems, in which the spatial configuration of the systems plays a key role in assessing its failure. Indeed, in this case, the failure occurs if only if  $k$  consecutive components have failed (which includes both parallel and series systems). We refer to e.g. [17], [46] and [60]. A new extension of the  $k$ -out-of- $n$  systems was proposed by [16]. The components are divided into two groups, important and standard, and the system's failure is conditioned on the number of failed components in each of the two sets. This recent extension is very useful for financial applications.  $K$ -out-of- $n$  systems can be binary or multi-state. The binary system models are the most used in reliability theory. So, the

system can also be in two possible states, working or failed. In addition to binary systems, we find the multi-state  $k$ -out-of- $n$  systems. In this specific case, both the system and its components can reach more than two states. These models are more flexible, and the system and its components could be completely failed, partially failed, fully operational, partially operational (see, e.g. [37, 36], [96] and [82]).

The expansion of this field of research has involved many disciplines and, over the years, have been devised methodological studies, theoretical models and applications that have inspired our paper. Our systems can be considered of  $k$ -out-of- $n$  type: they fail when some of their components fail (the failure rule will be introduced later).

To face our problem, we include an additional fundamental element to our study: rational expectations.

The reliability of a system is linked to its ability not to fail within a certain time. There is, therefore, a relationship with the time, evaluated on the basis of the information the system provides us with. This allows us to estimate the behaviour and the impact on system reliability. Depending on the information stored by the system, we can evaluate its reliability by following and predicting its failure over time.

In modern economic theory, agents are assumed to adjust their expectations to anticipate future decisions. Over the years, in the study of economic behaviour, two main hypotheses have been put forward: adaptive expectations and rational expectations.

Adaptive expectations assume that people form their expectations by giving importance to past events when predicting future outcomes. They played a key role in the 1960s and 1970s (see [29] and [18]).

Instead, with rational expectations, people make their decisions (which on average will be correct) based on all the best information available (see for example [57], [50], [74], [75] and [9]).

It is evident that the application of rational expectations becomes important in many areas; in fact, it can be applied in macroeconomic models, microeconomic situations and a wide selection of other models.

Starting from the general equilibrium, the insights of rational expectations can be deduced analytically. From the equation of supply and demand and the assumptions of rational expectation, we obtain:

$$p_t^* = \mathbb{E}[p_t | \mathbb{I}_{t-1}] \tag{1}$$

where  $p_t^*$  is the price expectation,  $\mathbb{E}$  is the expected value operator,  $p_t$  is the price at time  $t$ , and  $\mathbb{I}_{t-1}$  is all the information available at time  $t - 1$ .

Many authors have studied rational expectations over the years in several fields.



For example, [34] estimated and tested parameters of economic agents' dynamic objective functions within nonlinear rational expectations models. Their model did not require a priori assumptions but only the specification of the economic context in which to apply it.

In the economic and financial field, [12] examined rational expectations in the financial markets by studying speculative bubbles. In contrast, [22] studied the relationship between the efficient market hypothesis and rational expectations.

In the field of operations research, there is a paper by [5] on discrete-time domains. They generalized Cagan's model of hyperinflation. They changed the idea of the time horizon of the original model without considering only the time intervals represented by the integers. The time-domain of their generalized model, therefore, includes all the real numbers of the considered time horizon.

So, we compare two sets of  $k$ -out-of- $n$  systems (*in-vivo* and *catalog*) through a large number of simulations by applying a Bayesian approach and exploiting rational expectations to obtain optimal predictions of the failure times of these systems.

### 3 The reliability model

We now briefly present the reliability context we deal with. Some detail related to this model can be found in [3].

This theoretical setting for the evolution of the system lifetime connects the rational expectations and the reliability of stochastic systems.

We consider the weighted  $k$ -out-of- $n$  type *system*  $\mathbf{S}$  as a unified entity composed by  $n$  individual interconnected *components* denoted by  $C_1, \dots, C_n$ . The system fails when some of its components fail.

The *state* of  $\mathbf{S}$  and of the  $j$ -th component  $C_j$  changes over time with  $t \geq 0$  and it is a binary quantity – i.e., the system can work or, alternatively, have failed (or be not active) at time  $t$ . At the starting point (time  $t = 0$ ), the system and the components are naturally assumed to work (to be active). We label the active components by  $A$  and the not active ones by  $NA$ . We define a *configuration* as a vector  $\gamma \in \{A, NA\}^n$  describing the states of the components of the system. The configuration at time  $t = 0$  is denoted by *initial configuration*.

If we maintain the same labels used for the components for active and not active systems ( $A$  and  $NA$ , respectively), then the arguments above can be formalized by introducing the so-called *structure function*  $\phi : \{A, NA\}^n \rightarrow \{A, NA\}$ .

At time  $t$ , the generic configuration is denoted by  $\gamma(t)$ . Configurations at a given time can be clustered in two groups: in the first one, we collect the configurations leading to the system's failure, while in the second group, we have the configurations whose occurrence is associated with a still

active system. As said before, we assume that the initial configuration belongs to the second group.

We introduce the *system lifetime* as:

$$\mathcal{T} := \inf\{t \geq 0 \mid \phi(\gamma(t)) = 0\}. \quad (2)$$

As in the mainstream reliability theory, we assume that the components are weighted, and the system's failure depends on the aggregate weight of the active components. In this respect, the weights capture the evidence that the different components of the system are not homogeneous in terms of their relevance. Furthermore, the relevance of the components is assumed to vary over time according to the changes in the status of the system's components.

We enter the details.

For each  $j = 1, \dots, n$  and  $t \geq 0$ , the *relative importance of the component*  $C_j$  over the entire system at time  $t$  is measured through  $\alpha_j(t)$ , where  $\alpha_j : [0, +\infty) \rightarrow [0, 1]$  and  $\sum_{j=1}^n \alpha_j(t) = 1$ , for each  $t$ .

For each  $t \geq 0$ , we collect the  $\alpha(t)$ 's in a time-varying vector  $\mathbf{a}(t) = (\alpha_j(t))_j$ , where

$$\mathbf{a} : [0, +\infty) \rightarrow [0, 1]^n \quad \text{such that} \quad t \mapsto \mathbf{a}(t). \quad (3)$$

If a component is not active at time  $t$ , then its relevance for the system is null. Moreover, each active component of the system has positive relative relevance (weight).

$$\alpha_j(t) = 0 \Leftrightarrow j \text{ is in status } NA. \quad (4)$$

Condition (4) is useful in that it allows to describe the status of the components of the system directly through the  $\alpha$ 's.

Once a component fails, its weight is reallocated to the survived components proportionally to their weights. This is what we call *failed component allocation rule*.

A rational expectations approach is used to compute the expectation of the random time in which the system fails. Specifically, we will compute the expected value of  $\mathcal{T}$  conditioned to the specific values of the weights  $\mathbf{a}(t)$  measured at time  $t$ . We denote by *RE* the rational expectations of time  $\mathcal{T}$  given all the possible configurations of the system. Specifically, for  $t \geq 0$ , we set

$$RE_t = \left\{ \mathbb{E}[\mathcal{T} \mid \bar{\mathbf{a}}(t)] : \bar{\mathbf{a}}(t) \in [0, 1]^n \right\}, \quad (5)$$

where  $\mathbb{E}$  is the expected value operator.

Formula (5) provides the expected value of the lifetime of  $\mathbf{S}(t)$  in correspondence to any system configuration.

## 4 Simulation procedure

This section aims to test the worthiness of the rational expectation estimation of the failure time of the system through an extensive set of numerical experiments. To this aim, we first provide a

detailed description of the numerical procedure with the steps of the algorithm; then, we give the parameter set used in the experiments.

The proportional *failed component allocation rule* and the *failure rule* give a clear intuition of the presence of stochastic dependence among the lifetimes of the system components.

The system is assumed to fail the first time in which the number of failed components becomes higher than the one of the active components.

## 4.1 Overview of the simulation procedure

To pursue our aim, the configurations of the systems will be opportunely synthesized through statistical indicators, pointing to central tendency, variability, shape and concentration information content of the distribution of the weights of the components (the  $\alpha$ 's).

Specifically, we will deal with: (i) the variance to evaluate the variability of the configurations and, therefore the dispersion of the weights; (ii) the skewness, to provide a measure of the symmetry of the configurations; (iii) the kurtosis, to study if the weights are heavy-tailed or light-tailed with respect to a normal distribution; (iv) the Gini coefficient, to evaluate the inequality within the configurations; (v) the Shannon entropy, to estimate the degree of complexity of information and the mutability of the configurations.

As seen above, the simulation procedure is characterized by two sets of systems: one (the *catalog* systems) to record the information and calculate rational expectations, and another (the *in-vivo* systems) to compare the failure times of real systems with those predicted. The prediction errors are obtained by comparing the *catalog* systems and the *in-vivo* systems.

Our attention is focused on all the statistical indicators values that represent the synthetic measures of our systems.

Time remains implicit, and, in the final representations of the errors committed in the prediction of failure times, we will directly relate the errors to the values of the statistical indicators.

## 4.2 Detailed steps of the numerical procedure

As a premise, we briefly give the ground of the numerical algorithm.

1. First, we create a large set of synthetic reliability systems with different initial configurations. Then, we build an iterative algorithm to observe the system's stepwise evolution and its components' failure. Such a procedure is stopped at the failure of the system. After the failure of a system, we store at each step the weights of the components and the system's failure time. This set of synthetic systems form the *catalog*, and the generic system of the *catalog* is denoted by  $(\mathbf{S}^C)_k$ . The *catalog* is the information set for checking the validity of the estimation in (5).

We denote by  $K\_simulations$  the number of simulated systems that will be recorded in the *information set* – the *catalog* – so that the generic *information set* system will be denoted by  $\mathbf{S}_k^C$ , with  $k = 1, \dots, K\_simulations$ .

For each system  $\mathbf{S}_k^C$  we collect the information by one of the five statistical indicators (i)-(v) presented above related to the specific configuration  $\bar{\mathbf{a}}(t) = (\alpha_1(t), \dots, \alpha_n(t))$  – where the reference to  $t$  indicates that the configuration is considered at time  $t$ . Here, we introduce the failure time  $\mathcal{FT}_k$  of the related system – where the index  $k$  points to the  $k$ -th system  $\mathbf{S}_k^C$  of the *catalog*; the residual failure time  $\mathcal{T}_k(t)$  is conditioned to  $\bar{\mathbf{a}}(t)$ , and depends on  $k$  and  $t$ ; the number of active components for system  $k$  at time  $t$  is  $n_k(t)$ . We will follow the systems from the initial time until their failure with the time  $t = 0, 1, 2, \dots, \mathcal{FT}_k$ .

By introducing opportune notation, we will collect for each  $k$ -th system:

- All the times until the failure  $\mathcal{FT}_k$  with  $t = 0, 1, 2, \dots, \mathcal{FT}_k$ .
- For each time  $t$ :
  - We record a set of  $\alpha$ 's of the configurations  $\bar{\mathbf{a}}$  that depends on  $k$ ,  $t$  and  $n_k(t)$  (all the weights of the  $\mathbf{S}^C$ 's);
  - Each set of the  $\alpha$ 's is associated with one value  $\bar{\star}_C$  of the statistical indicator represented by  $\star$ , with  $\star$  =variance, skewness, kurtosis, Gini coefficient and Shannon entropy. Any observed value  $\bar{\star}_C$  depends on  $k$  and  $t$ , while  $n_k(t)$  intervenes directly in the computation of the considered statistical indicator at time  $t$  and for system  $k$ . The values are rounded to 3 decimal places.
- All the failure times  $\mathcal{FT}$ 's of systems  $\mathbf{S}^C$ 's in a vector that depends on the  $k$ -th system  $\mathbf{S}_k^C$  and indicated with  $\mathcal{FT}_k$ , as said above.
- From the  $\mathcal{FT}$ 's we derive all the residual failure times  $\mathcal{T}$ 's in a matrix  $\mathcal{T}_C$ , that depends on the  $k$ -th system  $\mathbf{S}_k^C$  and the observed value  $\bar{\star}_C$  ( $\mathcal{T}_k(\bar{\star}_C)$ ). The residual failure time measures the time until the failure of each  $\bar{\star}_C$ . Specifically:

$$\mathcal{T}_k(\bar{\star}_C) = \mathcal{FT}_k - t, \forall t = 0, \dots, \mathcal{FT}, \quad (6)$$

2. Second, we are in the position of constructing the rational expectations in (5) on the basis of the  $\mathbf{S}^C$ 's.

At this aim, we build an application that links each observed value  $\bar{\star}_C$  of the statistical indicator represented by  $\star$ , with  $\star$  =variance, skewness, kurtosis, Gini coefficient and Shannon entropy, with a set of residual failure times contained in  $\mathcal{T}_C$ . We do it for all the *catalog* systems  $\mathbf{S}^C$ 's. We compute the arithmetic mean of each set of residual failure times in correspondence of the  $\bar{\star}_C$ 's, namely  $\mathbb{E}[\mathcal{T}_{RE}(\bar{\star}_C)]$ . In so doing, we obtain the rational expectations of formula (5) associated to each  $\bar{\star}_C$ .

3. Third, we create a second set of systems. We implement the same procedure used for the *catalog* and check if any weight vector is associated with the failure time obtained in the *catalog*. This second set of systems is called the *in-vivo*. We denote by  $X\_simulations$  the number of simulated *in-vivo* systems, and the generic *in-vivo* system is  $\mathbf{S}_x^V$ , with  $x = 1, \dots, X\_simulations$ .

In this case, by introducing suitable notation, we will record for each  $x$ -th system:

- All the times until the failure  $\mathcal{FT}_x$  with  $t = 0, 1, 2, \dots, \mathcal{FT}_x$ , where  $\mathcal{FT}_x$  is the failure time of the related system – the index  $x$  points to the  $x$ -th *in-vivo* system  $\mathbf{S}_x^V$ .
- For each time  $t$ :
  - We record a set of  $\alpha$ 's of the observations  $\bar{\mathbf{a}}$  that depends on  $x$ ,  $t$  and  $n_x(t)$  (all the weights of the  $\mathbf{S}^V$ 's), where  $n_x(t)$  is the number of active components.
  - Each set of the  $\alpha$ 's is associated with one value  $\bar{\star}_V$  of the statistical indicator represented by  $\star$ , with  $\star$  =variance, skewness, kurtosis, Gini coefficient and Shannon entropy. Any observed value  $\bar{\star}_V$  depends on  $x$  and  $t$ . The values are rounded to 3 decimal places.
- All the failure times  $\mathcal{FT}$ 's of systems  $\mathbf{S}^V$ 's, for each  $x$ , in a vector that depends on the  $x$ -th system  $\mathbf{S}_x^V$  (indicated with  $\mathcal{FT}_x$ ).
- From the  $\mathcal{FT}$ 's, all the residual failure times  $\mathcal{T}$ 's in a matrix  $\mathcal{T}_V$ , that depends on the  $x$ -th system  $\mathbf{S}_x^V$  and the observed value  $\bar{\star}_V$  ( $\mathcal{T}_x(\bar{\star}_V)$ ). Every value  $\bar{\star}_V$  is related to a residual failure time  $\mathcal{T}_x(\bar{\star}_V)$ .

Here, we consider all the values  $\bar{\star}_V$ 's. These values are different a priori from the  $\bar{\star}_C$ 's found in the first part of the simulation procedure.

4. Now we compare the observed values of the *in-vivo* systems  $\bar{\star}_V$ 's with the observed values of the *catalog* systems  $\bar{\star}_C$ 's when *in-vivo* and *catalog* systems show the “same configuration”.

We therefore need to replace every  $\bar{\star}_V$  that has never been observed in the *catalog* with the value  $\bar{\star}_C$  which represents the “same configuration”. The “same configuration” in the *catalog* is the observed value  $\bar{\star}_C$  which is closest to the value  $\bar{\star}_V$  we are replacing.

Specifically, we compute the minimum distance between each  $\bar{\star}_V$  and the values  $\bar{\star}_C$ 's.

Once we find the closest configuration of  $\bar{\star}_V$  among the  $\bar{\star}_C$ 's, we build a function that assigns to each  $\bar{\star}_C$  – considered the “same configuration” – the residual failure times contained in  $\mathcal{T}_V$ , namely  $\mathcal{T}_x(\bar{\star}_C)$ , linked to the  $\bar{\star}_V$ 's that have been replaced.

Then, we consider, for each  $\bar{\star}_C$ , the error between the rational expectations and all the residual failure times of the in vivo systems assigned to the specific  $\bar{\star}_C$ . Specifically,

$$E_{\bar{x}_{RE}}(\bar{x}_C) = |\mathbb{E}[\mathcal{T}_{RE}(\bar{x}_C)] - \mathcal{T}_x(\bar{x}_C)|, \quad (7)$$

where  $\mathcal{T}_x(\bar{x}_C)$  is the residual failure time of the *in vivo* systems assigned to the  $\bar{x}_C$ 's to replace rational expectations.

We create, for each  $\bar{x}_C$ , a distribution of errors between the expected residual times calculated with the rational expectations and the residual failure times of the *in-vivo* systems, according to (7).

5. For comparison purposes, we use the distribution of the errors of step 4 in rational expectations prediction for the computation of the quantiles  $q = 10\%$ ,  $q = 50\%$  and  $q = 90\%$  of the error distribution.

In this way, we quantify the impact of the error prediction comparing  $E_{\bar{x}_{RE.10}}$ ,  $E_{\bar{x}_{RE.50}}$  and  $E_{\bar{x}_{RE.90}}$  with the benchmark  $E_B$  for the different analysis carried out in this work (variance, skewness, kurtosis, Gini coefficient and Shannon entropy). The fifth part consists of the following steps:

- Create the distribution of the quantiles  $q = 10\%$ ,  $q = 50\%$  and  $q = 90\%$  of the errors distribution in rational expectations prediction ( $E_{\bar{x}_{RE}}$ ) for the  $\bar{x}_C$ 's. We need to construct statistics over predictions to assess their quality. We decide to focus our attention on fixed quantiles at level  $q$  of the distributions of the errors:  $q = 10\%$ ,  $q = 50\%$  and  $q = 90\%$ , i.e. the  $10^{th}$ ,  $50^{th}$  and the  $90^{th}$  quantile. Let us consider errors in ascending order so that the  $10^{th}$  quantile represents the smallest errors in our forecast of residual failure times. Instead, the  $90^{th}$  quantile identifies the biggest errors in predicting the residual failure times of stochastic systems. The  $50^{th}$  quantile gives us the median of errors, so the central errors of the distribution.  $E_{RE.10}$ ,  $E_{RE.50}$  and  $E_{RE.90}$  are, respectively, the  $10^{th}$  (Error 1),  $50^{th}$  (Error 2) and the  $90^{th}$  (Error 3) of the distribution of the errors. These quantiles take into account all the information recorded in the *catalog* over time.
- To obtain great comprehensibility in the interpretation of the results proceed by clustering the  $\bar{x}_C$ 's (in ascending order) in equal class intervals. The class width is 50, except for the last class, which includes the values excluded from the previous classes.
- Create the benchmark errors given by the errors made without the use of rational expectations:  $E_B$ .
- To understand how rational expectations allow us to obtain the optimal prediction of residual failure times using the information stored in the past, it is necessary to examine the comparison between  $E_{\bar{x}_{RE.10}}$ ,  $E_{\bar{x}_{RE.50}}$ ,  $E_{\bar{x}_{RE.90}}$  and  $E_B$  of the several indicators considered in the analysis.

### 4.3 Parameter set

The parameters are set as follows:

Parameter	Value
<i>number of components</i>	10
<i>K_simulations</i>	10.000
<i>X_simulations</i>	10.000

The initial distributions under study were selected to provide five possible cases:

- Uniform distribution in (0,1) type;
- Beta distribution type, which has support in (0,1), composed of two positive shape parameters, denoted by  $\gamma$  and  $\beta$  that represent the exponents of the random variable and control the shape of the distribution. Depending on the value assigned to the two parameters, we will get several distributions from which we can randomly extract our weights.
  - $\gamma = 1$  and  $\beta = 3$  that is an asymmetric distribution on the left;
  - $\gamma = \beta = 0.5$  that is a symmetric distribution focused on extremes;
  - $\gamma = \beta = 2$  that is a symmetric distribution concentrated on the central values of the distribution;
  - $\gamma = 1$  and  $\beta = 0.5$  that is an asymmetric distribution on the right.

## 5 Results and discussion

In this section, we present the study's results and the related discussion. The paragraph will be divided into two blocks: in the first part of the discussion, we will investigate the relationship between the statistical indicators' values and error levels; in the second part, we will go into the composition of errors by studying their frequency.

We now focus our attention on the various statistical measures under analysis.

We start from the variance. Fig. 1 shows the prediction errors (in absolute value) for the analysis of the variance considering all the different five types of initial weight distribution. Magenta fixed line is the Benchmark. Red line, blue line and light blue line, blue line and red line are respectively Errors 1, Errors 2 and Errors 3.

Variance is a good estimator of the dispersion; it represents the average volatility of the weights of each system around the mean.

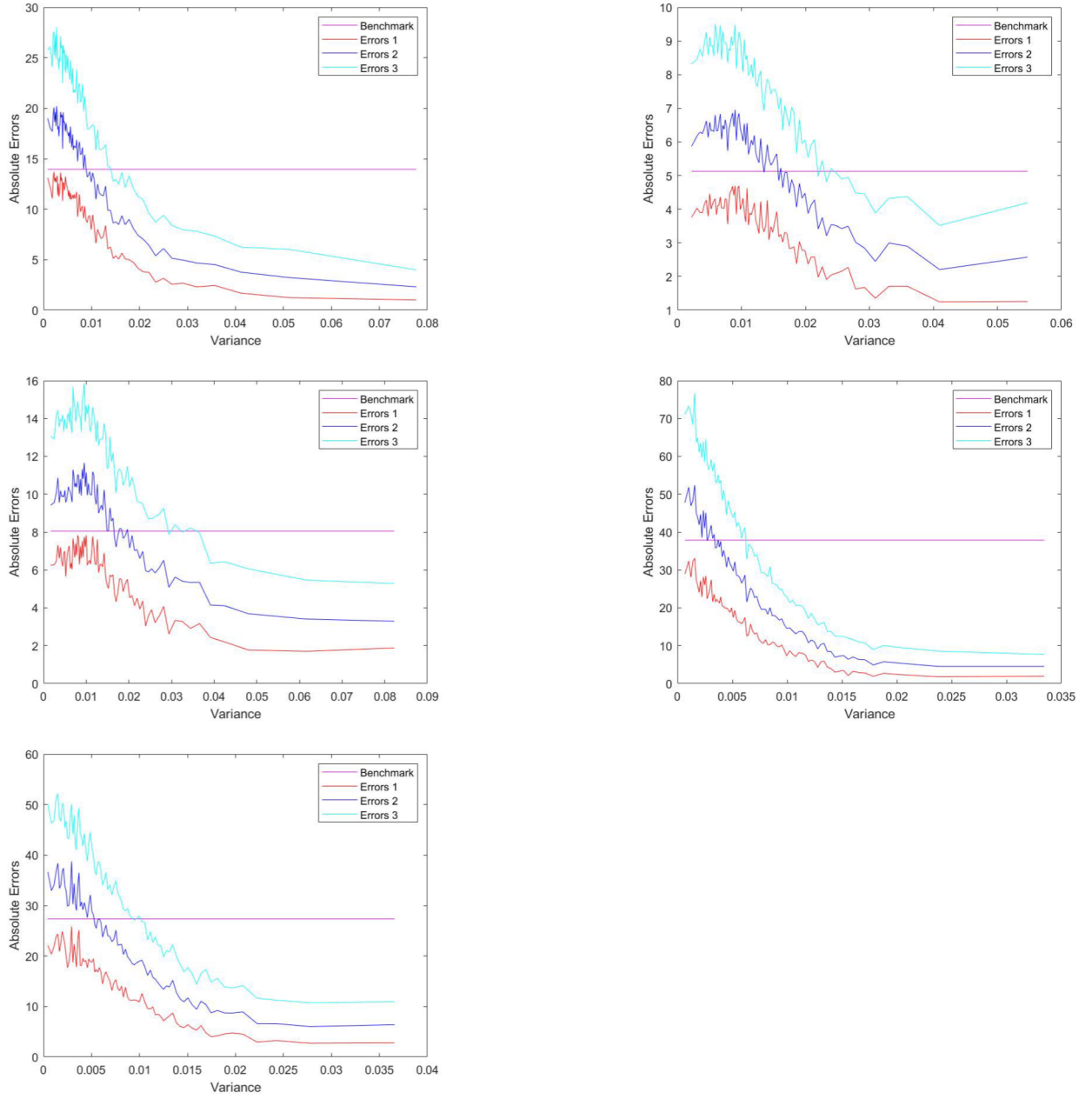


Figure 1: Prediction errors (in absolute value) for the Errors 1, the Errors 2 and the Errors 3 of the variance considering all the initial weight distributions. [Top] (Left) Uniform distribution; (Right) Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . [Center] (Left) Beta distribution with  $\gamma = \beta = 0.5$ ; (Right) Beta distribution with  $\gamma = \beta = 2$ . [Bottom] (Left) Beta distribution with  $\gamma = 1$  and  $\beta = 0.5$ .



As the subgraphs of Fig. 1 show, we can note a decreasing trend of errors towards zero in all the cases taken into consideration. High volatility of the system components is linked to the reduction of errors made in the prediction, and it is due to the decrease of the uncertainty about the systems failure time. Data are very spread out from the mean, and we can capture a wider range of values.

Variance is very performing, and the errors stabilize around zero for all initial weights distributions except for the 50<sup>th</sup> and the 90<sup>th</sup> percentile of the Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . In this case, from a certain threshold of variance onwards, we observe a small increase in errors, which emphasizes that excessive fluctuations affect the predictability of the variance for medium or very large errors. When we detect a high variance, we pay close attention to data without generalizing the values never seen before. So, for the negative exponential distribution, too high values and excessive turbulence within the observations, however, are overfitting, which can highlight distortions that lead the model to become less good at predicting new data.

With low variance values, the model is underfitting the residual failure times, and it cannot identify the real moment of system failure.

From Fig. 1 in the initial parts of the various curves, we observe a very scattered trend characterized by many spikes with errors that stabilize at a certain level of variance. The prediction of the systems' failure times from these levels of variance is almost perfect. In these paths windows, the components are very dependent on each other; this allows us to predict how long it will take for the system to fail.

Moreover, we can note a different trend in Errors 3 compared to Errors 1. The line of the biggest errors decreases faster toward zero, while in the line of the smallest errors, there is a slight flat behaviour. This flat behaviour depends on the fact that when we have a very low variance (close to zero), in a big error, the gain in the prediction is more evident than what we will see in a small error.

The different subgraphs also display the change in the initial distribution of components.

According to the existing literature, the ideal shape of the distribution of the weights that minimizes prediction errors is the negative exponential one (see [48], [77], [4] and [25]). In our study it corresponds to the Beta distribution with  $\gamma = 1$  and  $\beta = 3$ .

In fact, the best predictive efficacy belongs to the asymmetric case with  $\gamma = 1$  and  $\beta = 3$  and to the symmetric one with  $\gamma = \beta = 0.5$ . In the first case, the predictive gain is very high, although there is a worsening in the final trend, which corresponds to very high levels of variance.

However, in other cases, we always obtain excellent results through rational expectations with increasing levels of variance. In fact, we appreciate a constant decrease to near zero. The symmetric Beta with  $\gamma = \beta = 2$  together with the asymmetric Beta with  $\gamma = 1$  and  $\beta = 0.5$  are the cases that perform worse as regards the errors in absolute value, but in correspondence with appropriate variance values, they reach errors close to zero with trends flattening out due to ever-increasing

values of the indicator.

Therefore, we can deduce from this analysis the levels of variance that provide an optimal balance without overfitting or underfitting the residual failure times. This specific application suggests that the study of the variance can be a powerful tool to predict the residual failure times of a stochastic system in case of high values of variance, which are crucial to understanding the behaviour of prediction models.

[40] dealt with software fault prediction models by establishing that the lower the variance, the more reliable the system is. Also [83], who dealt with error estimator methods for evaluating prediction models, confirmed with their studies that good performance is linked to low variance.

On the contrary, in our study, high variance values are more informative and allow a better prediction, performing better than lower values and with fewer fluctuations.

Switch now to the skewness.

In Fig. 2 we can see the different trends of skewness errors.

Through the skewness of the components' weights, we investigate the nature and dynamics of the shape of the distribution. Skewness measures how close the data are to the symmetry or the lack of symmetry in data distribution.

The skewness has been used in predicting future market returns with excellent results (see [41]). Also in our study, it is a good predictor depending on the values assumed by the distribution.

Compared to the previous indicators, skewness fluctuates more in the central part of the curves. It exhibits large spikes, especially at values close to zero when the data are fairly symmetric. The highest errors are achieved in conditions of symmetry, negative asymmetry or slight positive asymmetry of the distribution. Regardless of the initial distribution considered, a forecast bias is evident around a skewness range from -1.5 to +2.5.

Overall, the trend of average errors and percentiles is very scattered, and we do not notice big differences between the various curves. The best accuracy is obtained by applying our predictive model in cases of Beta distribution with  $\gamma = \beta = 0.5$  and Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . The least predictive gain and very high maximum error levels are reached in the case of Beta distribution with  $\gamma = \beta = 2$  (as in the study of the two previous indicators).

While generally, highly skewed data affect the accuracy of the predictive model (see [47]), the obtained results provide evidence that extremely positively skewed weights pick up the best information with good predictive gain.

As part of the use of statistical moments in forecasting models, [71], [1], [70] and [44], state that the lowest moments of the distributions (second moment in our case, i.e. the variance) are more efficient than the highest moments which are attested to be less stable and reliable (third, skewness and fourth, kurtosis). With our analysis, we confirm what these researchers concluded.

In fact, the variance is the one that comes closest to zero errors, with some features of uncertainty

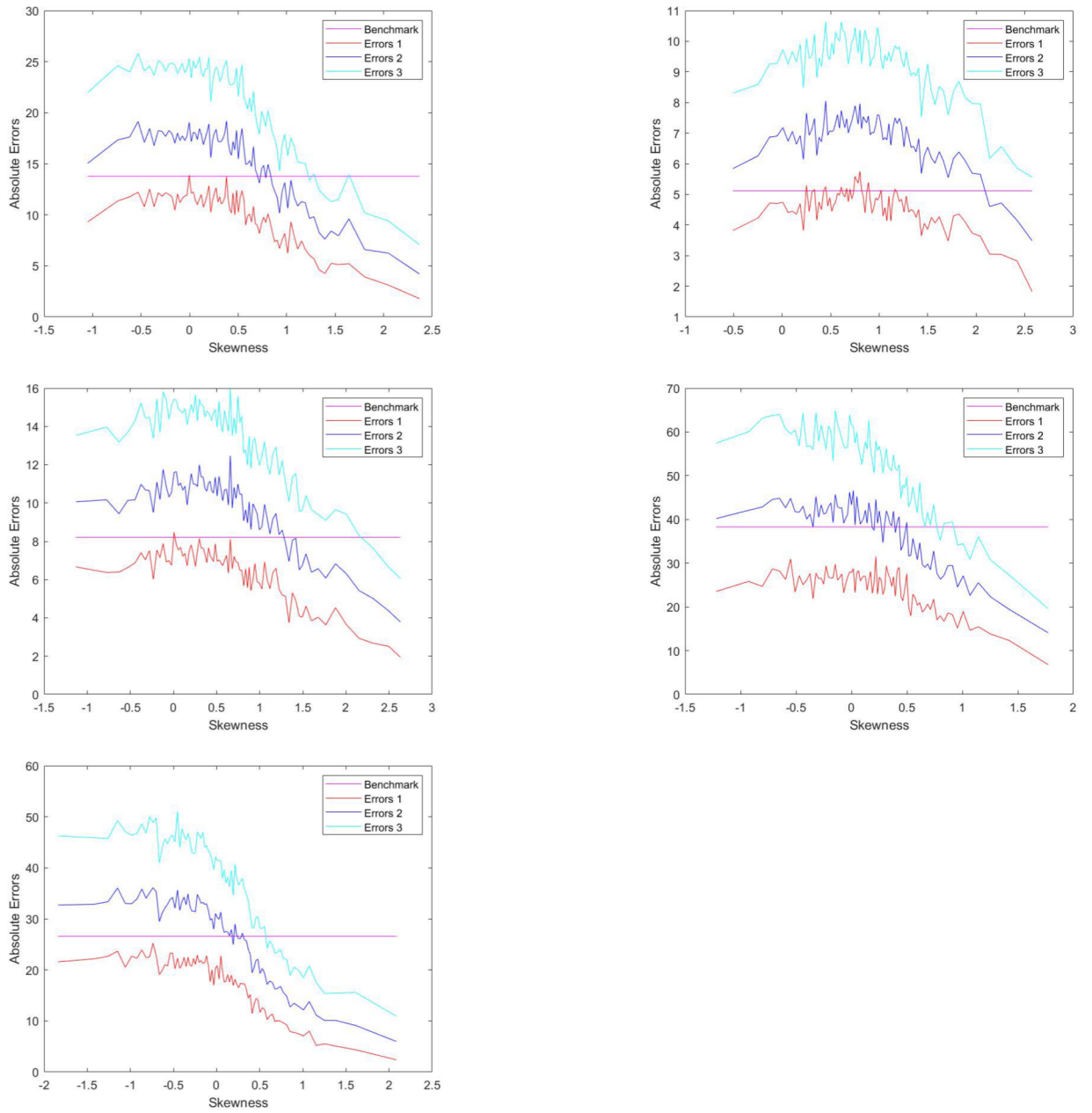


Figure 2: Prediction errors for the skewness. The five sub-graphs contain the five initial weight distributions. Refer to the caption of Fig. 1.

in the case of very low variability (components very similar to each other). The situation worsens for skewness and kurtosis. The third statistical moment is characterized by less regularity and mixed trends but with the achievement of very low error values. The fourth statistical moment has a more flattened shape, but it is less efficient overall. We now analyze the kurtosis. Kurtosis is a measure of the "tailedness" of the distribution through the description of the tails and the identification of the outliers.

In Fig. 3 we see the prediction errors concerning kurtosis.

The situation is scattered.

As in the study of variance, the two distributions that perform better and provide lower and decreasing errors are Beta distribution with  $\gamma = \beta = 0.5$  and Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . At a glance, the latter analysis is the best performing.

In the first three sub-charts, we can observe a different error trend between Errors 3 line and Errors 1 line. The trend of the smallest errors is almost linear, with constant errors for all the kurtosis values recorded in the simulations. Instead, the biggest errors follow a path that reaches maximum errors in correspondence of low kurtosis values and, therefore, in the presence of a platykurtic distribution of the weights. Kurtosis loses predictive power when the distribution has a lower likelihood of extreme weights than a normal distribution. When the weights are distributed around their average and we observe thinner tails, the use of rational expectations proves to be ineffective.

Our model, instead, becomes successful when there is a positive excess of kurtosis and the data are fat-tailed distributed. If we consider even the most extreme values, we obtain an increasingly inclusive range of cataloged weights. The application of rational expectations attests to be an excellent method of forecasting the residual failure times considering even the values with a greater distance from the mean.

The situation is completely reversed if we start from a symmetric weight distribution centred on 0.5 (Beta distribution with  $\gamma = \beta = 2$ ). In correspondence with distributions with non-fat tails, kurtosis does not work well. This case represents an exception showing anomalous behaviour. It provides very high error levels. It does not work because the initial values assigned to the weights extracted from this distribution are concentrated in 0.5. It underestimates the extreme values close to 0 and 1, which are considered in the other distributions. The minimum errors committed to predicting failure times through rational expectations are those at kurtosis levels around 8. It can only be used if we have distributions with extreme values.

On the other hand, if the weights are distributed with an asymmetry to the right (Beta distribution with  $\gamma = 1$  and  $\beta = 0.5$ ), kurtosis is a good predictor of thin tail distributions with continuously rising curves as the values increase.

Fig. 4 displays the prediction errors (in absolute value) for the Gini coefficient considering all the different five types of initial weights distribution.

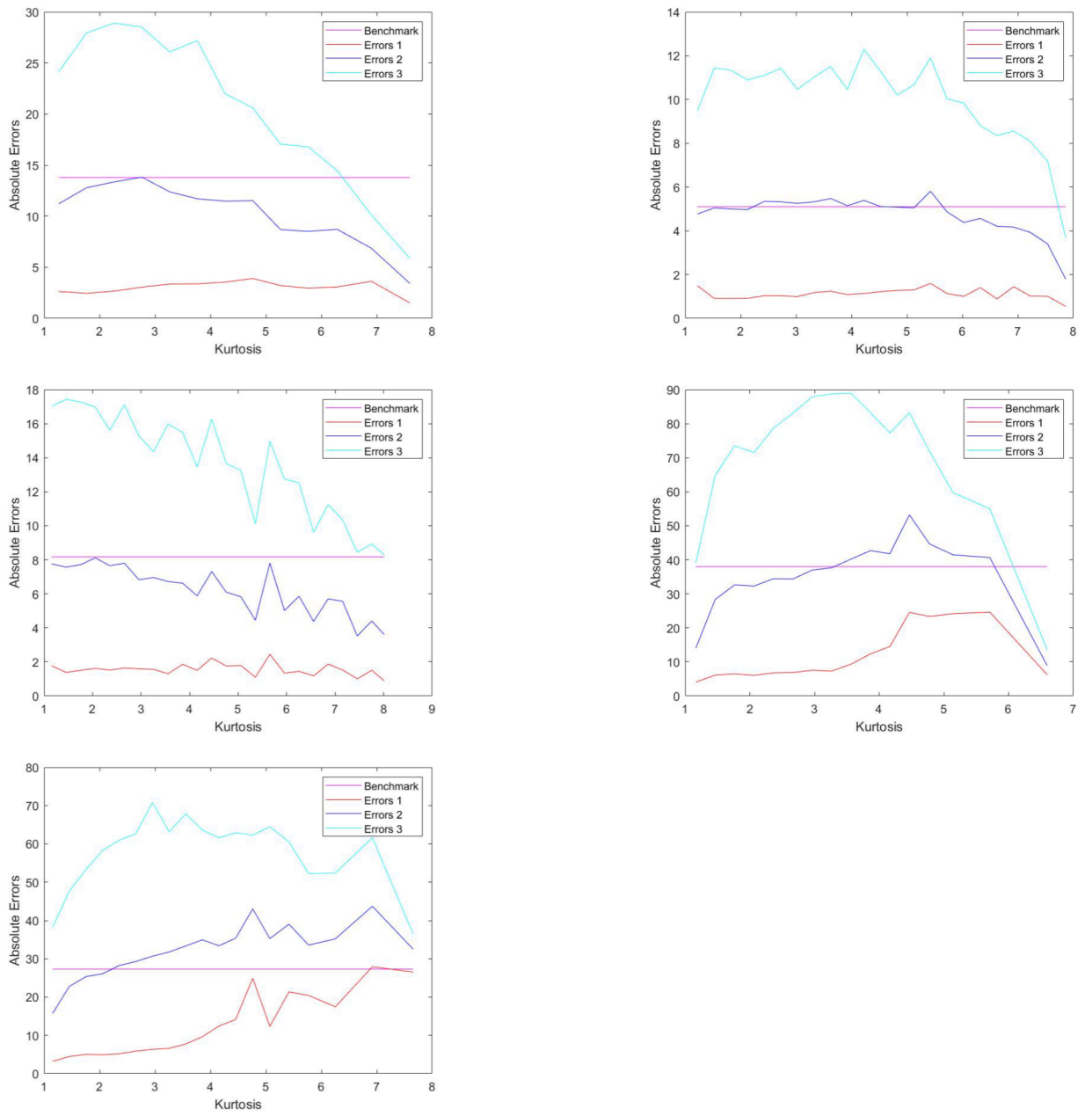


Figure 3: Here we present the study of kurtosis. See for the details Fig 1.

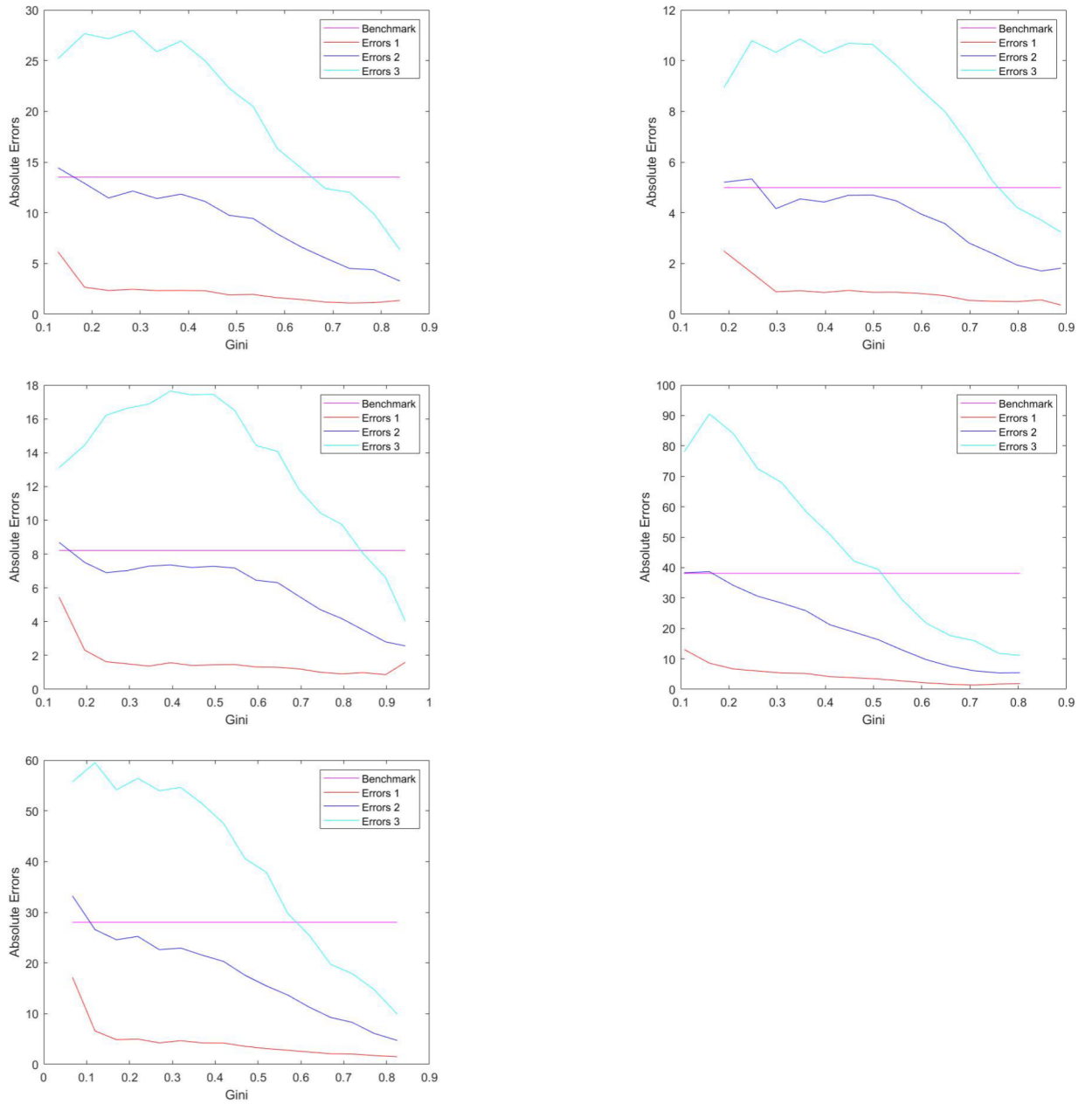


Figure 4: Case of the Gini coefficient. Magenta line of errors is the Benchmark. Red line, blue line and light blue line are Errors 1, Errors 2 and Errors 3, respectively (see 1).

We can consider the Gini coefficient as a measure of accuracy to assess the performance of our prediction model of residual failure times. In fact, the increase in the coefficient corresponds to an increase in inequality among the values taken into consideration in the analysis. Values close to 1 indicate that the range of possible weights cataloged and on which the study is based is very wide.

Not many researchers have used the Gini coefficient as a prediction tool. [66] have argued that it is very useful to use the Gini when dealing with failure prediction models.

The behaviour of errors is similar to that of kurtosis in some scenarios, but there are differences to note. However, the comments previously left for kurtosis about the different trends of the three curves can also be extended for the study of this indicator.

First of all, the 10<sup>th</sup> percentile has a linear trend, but for values of the Gini coefficient near zero (the initial trend), we observe a peak that decreases rapidly, settling at error levels close to zero.

Secondly, the symmetric Beta distribution with  $\gamma = \beta = 2$  and the asymmetric Beta with  $\gamma = 1$  and  $\beta = 0.5$  behave in the opposite way to that of kurtosis. In fact, although the level of initial errors is very high, we can visualize a predictive gain that grows exponentially as the values of the Gini coefficient increase.

We can see visually through error analysis the improvement in the accuracy of the predictions for increasing coefficient values up to the attainment of prediction errors tending to zero. Near the maximum polarization of the components, we reach excellent levels of prediction.

For very small Gini values, we can deduce that we are in the initial time of the simulations when no node or few nodes have failed. In fact, the values of the weights will be more equidistributed, and the equidistribution will decrease as the components fail and the weights of the failed nodes are proportionally redistributed among the components still active. Low coefficient values correspond to less cataloged information and greater uncertainty associated with forecasts through rational expectations.

We end this first part of the analysis by focusing on the Shannon entropy illustrated in Fig. 5.

There are no contributions in literature on the use of entropy as a prediction tool. Thus our analysis fills this gap.

The results show a very irregular error trend characterized by many spikes that alternate touching a wide range of values. Any improvement is very random, and using rational expectations is ineffective as a prediction tool.

For each type of initial distribution – by comparing the values with the previous analysis – the errors are the worst ever obtained, and the possibility of distortions occurring by applying our model on entropy is evident.

The Shannon entropy is the least informative and least useful statistical indicator in this forecasting model. There is no regularity in the trends, and we fail to capture information on which entropy values are best suited for better predicting failure times.

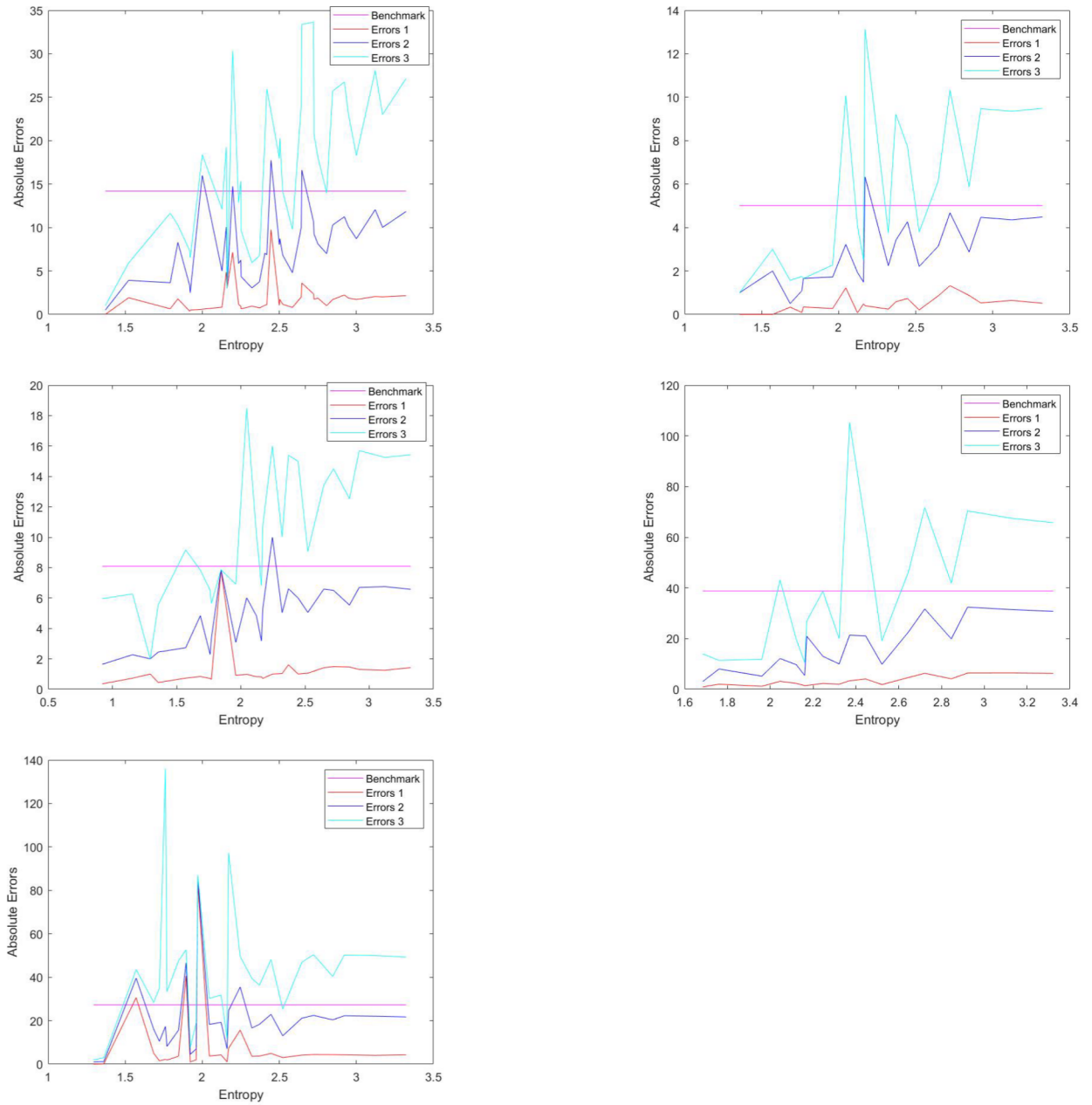


Figure 5: Shannon entropy analysis. As in the previous charts: [Top] (Left) Uniform distribution; (Right) Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . [Center] (Left) Beta distribution with  $\gamma = \beta = 0.5$ ; (Right) Beta distribution with  $\gamma = \beta = 2$ . [Bottom] (Left) Beta distribution with  $\gamma = 1$  and  $\beta = 0.5$ .



Comments relating to previous charts need to be corroborated by further analysis.

We proceed with a second part in which we can add investigations performed by illustrating a series of histograms containing the absolute frequencies of the prediction errors.

The aim is to understand the extent of the errors according to the statistical indicator and the weights' initial distribution.

In each graph, there are two sub-graphs; one in which the error values are divided into classes of width 10, with the residual values included in the last class whose size is variable; the other where we eliminate the zero value (no errors made in the prediction), so that we can provide a representation of the frequencies for each error value. The analysis of the variance is reported in Fig. 6.

In Figs. 7 and 8 we show the error frequencies for skewness and kurtosis, respectively.

We can observe the analysis of the distribution concentration in Fig. 9 where we find the histograms for the five different initial distributions of the Gini coefficient.

Ultimately, we find Shannon entropy (see Fig. 10).

The comments for this second analysis can be quite generalized for all statistical indicators. In fact, the trend of the error frequencies is very similar in the various cases, with variance, kurtosis and Gini coefficient, which are the best prediction tools, with more linear and regular trends, while skewness and Shannon entropy show deviations that invalidate the study.

And this confirms the first part of the analysis.

In general, observing the frequency of the prediction errors, these reach a very high level for errors located in the lower classes and then decrease in a strictly monotonous manner as the error values increase (higher classes).

The predictive ability is not trivial because the error values with higher frequency are found when the error is close to zero, and the curve's trend falls very quickly.

The errors corresponding to 0 and 1 (zero or almost zero errors) are the most frequent. This confirms our analysis's effectiveness, which provides a consistent reliability rate with many correct predictions. The frequencies decrease dramatically in correspondence with high prediction errors.

These graphs give us an overview of the order of magnitude of these curves. In fact, there is an inversely proportional relationship between the frequency of errors (ordinate axis) and the error classes (abscissa axis).

The most dense and excessive errors are those concerning the symmetric Beta distributions with  $\gamma = \beta = 2$  and the asymmetric one with  $\gamma = 1$  and  $\beta = 0.5$ . In these cases, the frequency of the lowest errors is much higher than in other scenarios. Very often, our predictions are correct. However, the high number of error classes gives us the perception of the high probability of making very large errors when the predictions are wrong.

Best predictive gain ever for Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . This confirms that the negative exponential distribution is the one that is most suitable in the context of forecasting failure

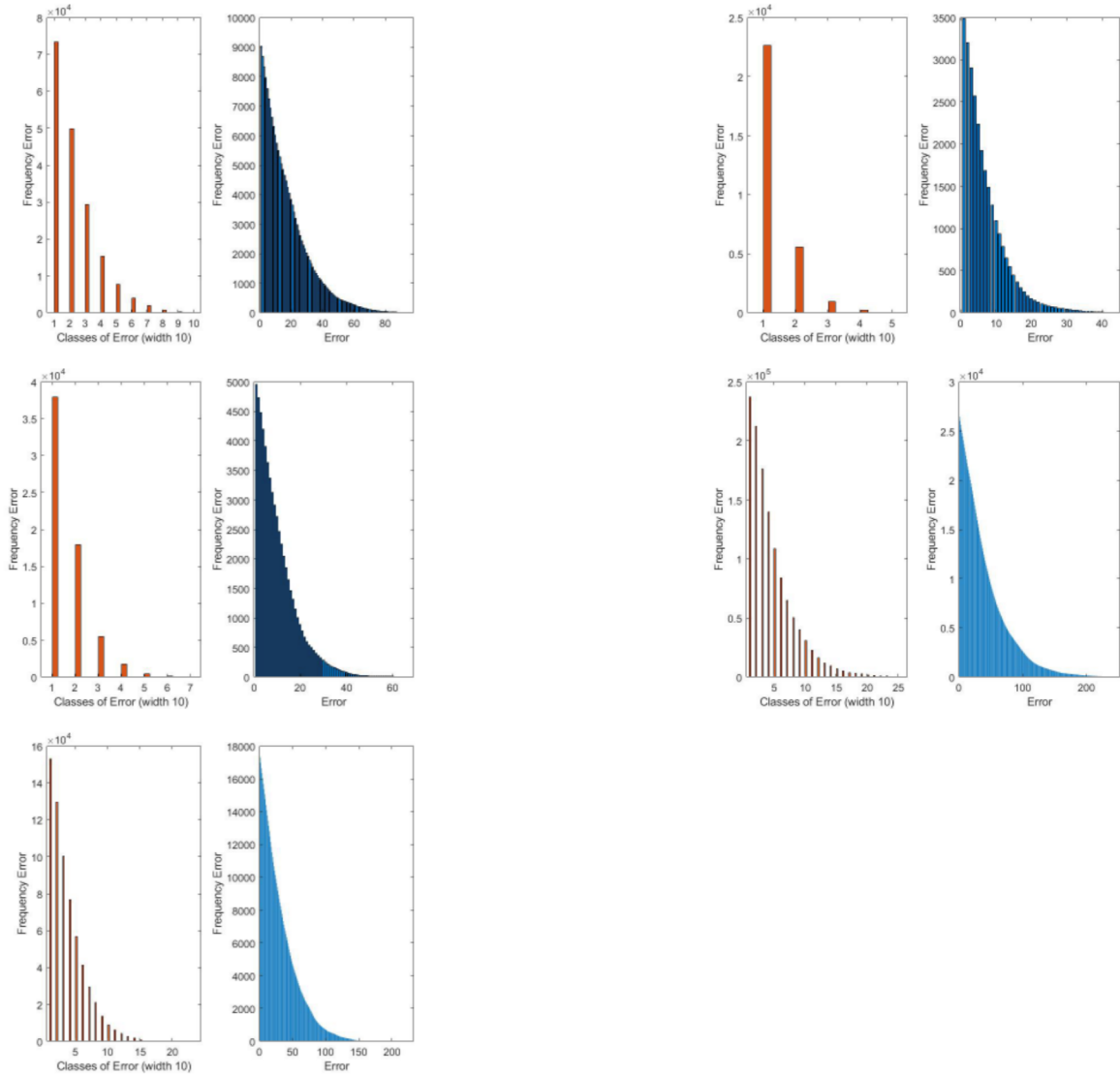


Figure 6: Histograms of the frequency of the variance prediction errors. In the first sub-graph we show the frequency for error classes of width 10 (each bar contains 10 error values in ascending order, except the last class where the final error values are inserted). In the second sub-graph, instead, we have the error frequency of each single value excluding zero from the representation. These are the errors related to the analysis of variance considering all the initial weight distributions. [Top] (Left) Uniform distribution; (Right) Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . [Center] (Left) Beta distribution with  $\gamma = \beta = 0.5$ ; (Right) Beta distribution with  $\gamma = \beta = 2$ . [Bottom] (Left) Beta distribution with  $\gamma = 1$  and  $\beta = 0.5$ .

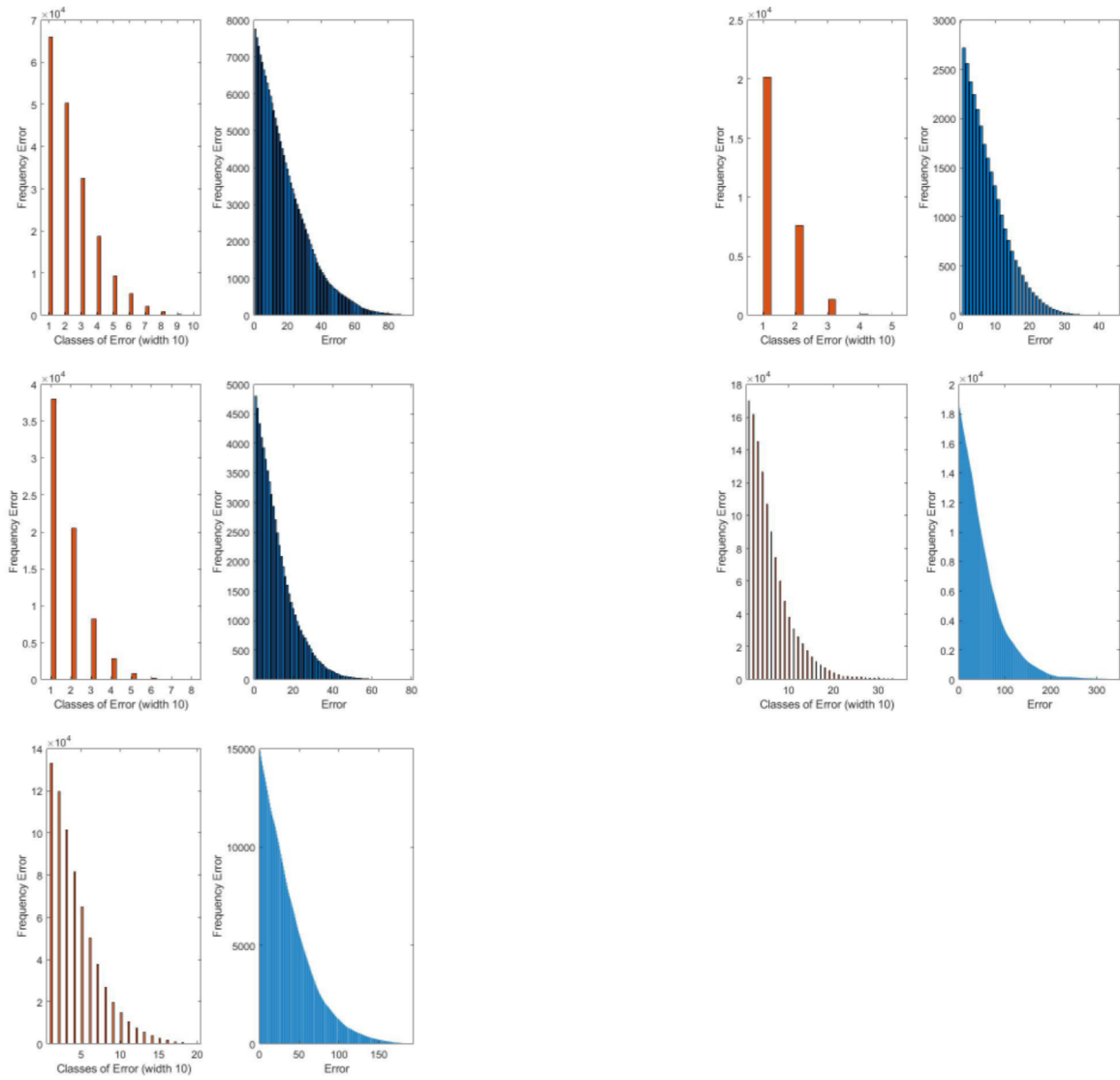


Figure 7: Here the analysis of the frequency error of skewness. More details in Fig. 6.

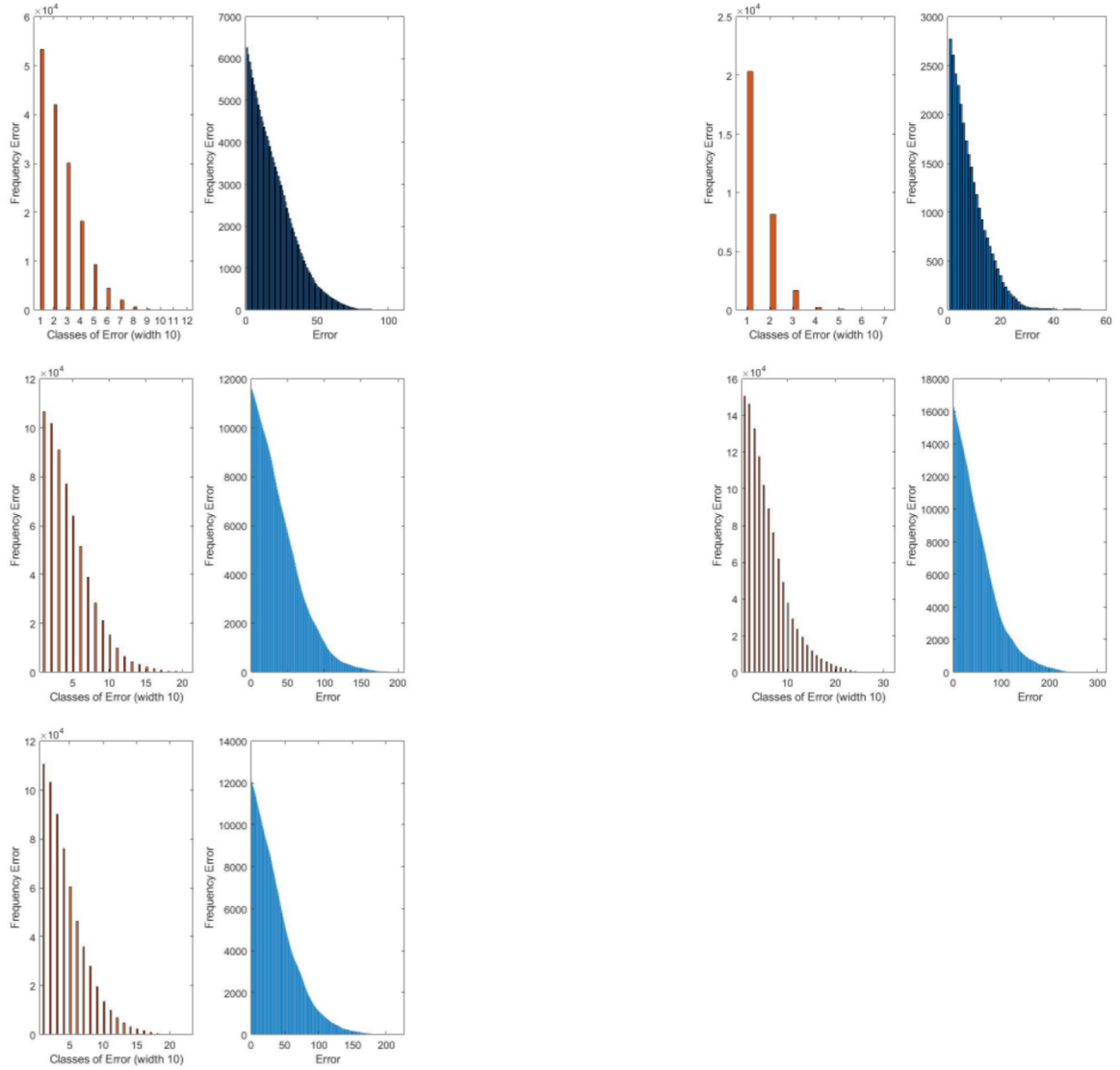


Figure 8: Histograms related to the frequency of the kurtosis prediction errors. See caption of Fig. 6 for a complete description of the sub-graphs.

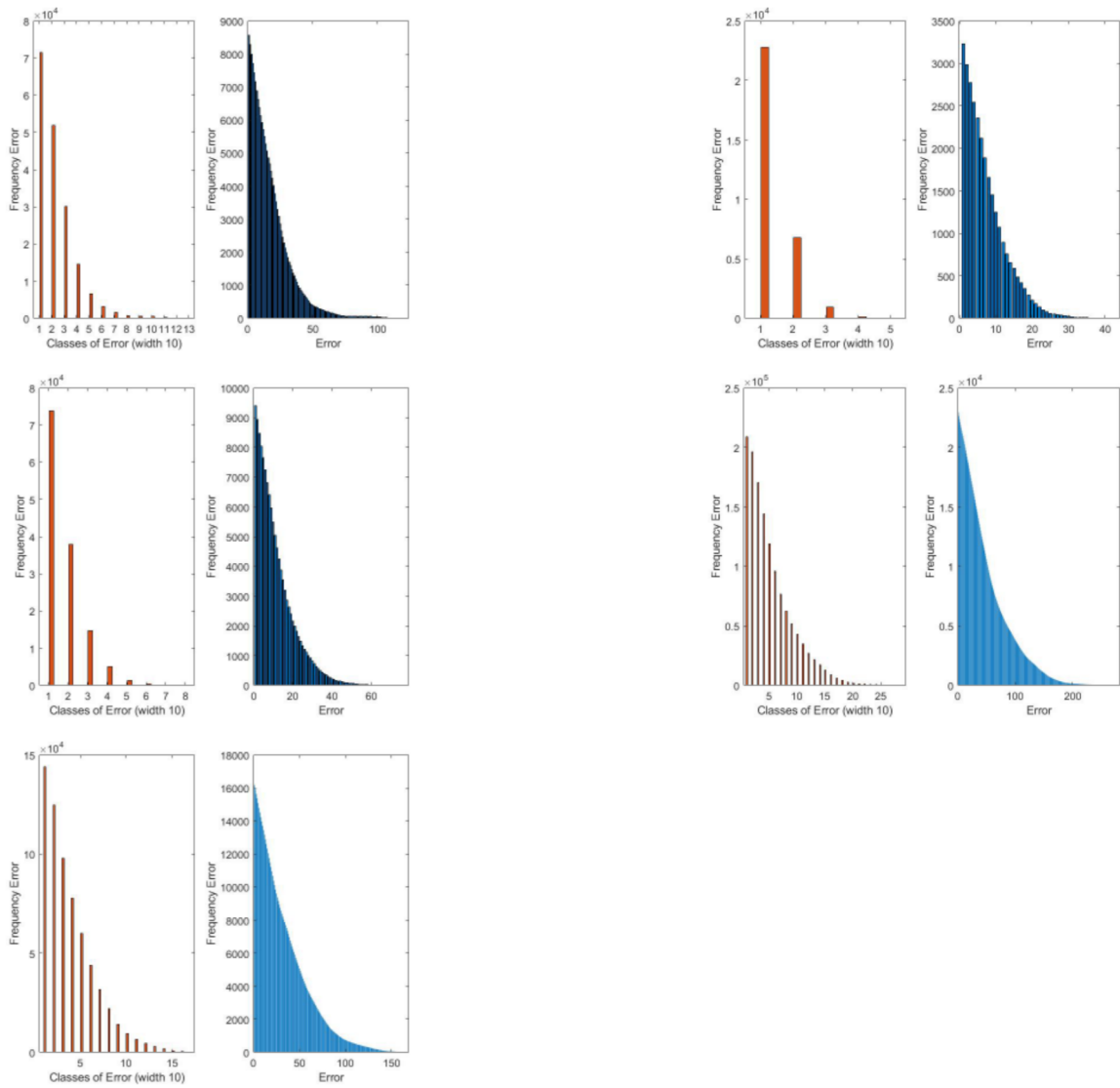


Figure 9: Frequency errors related to the study of Gini coefficient considering five initial weight distributions. [Top] (Left) Uniform distribution; (Right) Beta distribution with  $\gamma = 1$  and  $\beta = 3$ . [Center] (Left) Beta distribution with  $\gamma = \beta = 0.5$ ; (Right) Beta distribution with  $\gamma = \beta = 2$ . [Bottom] (Left) Beta distribution with  $\gamma = 1$  and  $\beta = 0.5$ .

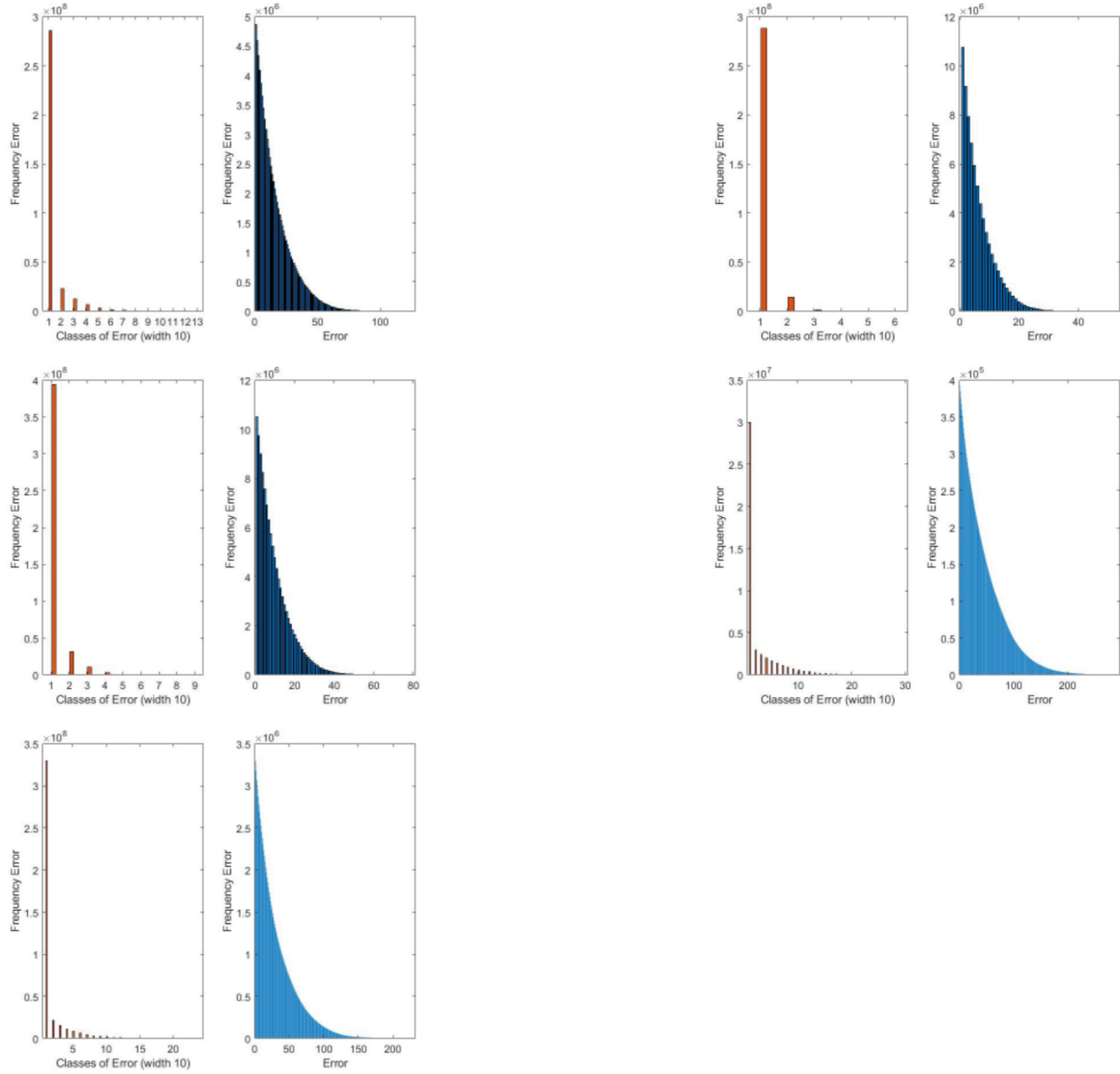


Figure 10: Histograms of the frequency of the prediction errors. Refer to the caption of Fig. 6 for a comprehensive explanation of Shannon entropy graphs.

times. And also, with this aspect of the analysis, we are consistent with the study of the first part, which deals with each individual indicator in relation to errors in absolute value.

## 6 Conclusions

In the context of reliability systems, our goal is to improve prediction performance using rational expectations by conditioning the residual failure times of simulated stochastic systems to the memory of past events. The key to our approach is the information stored over time through rational expectations with a particular focus on the effect of different levels of statistical indicators on prediction results. We make a transversal analysis over time: we emphasize the contribution of the statistical indicator, showing how the error varies as a function of the measure considered aggregated based on time.

We examine how the paths of the errors are influenced by several indicators very important for the informative content they provide, the different levels of these indicators and the initial weight distribution used in the study.

To deeply investigate our predictions, we take into account three different quantiles of the error distribution: 10%, 50% and 90% of the distribution. We thus have a comparison between small, medium and large errors.

It was interesting to understand how results change depending on the characteristic considered for the analysis and the initial distribution of the configurations' weights.

To our knowledge, this has not been done before.

Some results are in agreement with what is stated in the literature, such as, for example, ideal initial distributions in the predictor field (the negative exponential distribution is the best performing). In contrast, other results offer new and original perspectives.

## References

- [1] Amari, S. V., Pham, H., & Misra, R. B. (2012). Reliability characteristics of  $k$ -out-of- $n$  warm standby systems. *IEEE Transactions on Reliability*, 61(4), 1007-1018.
- [2] Andersen, J. V., & Sornette, D. (2005). Predicting failure using conditioning on damage history: demonstration on percolation and hierarchical fiber bundles. *Physical Review E*, 72(5), 056124.
- [3] <https://halshs.archives-ouvertes.fr/halshs-01673338v2>
- [4] Asadi, M., & Bayramoglu, I. (2006). The mean residual life function of a  $k$ -out-of- $n$  structure at the system level. *IEEE Transactions on Reliability*, 55(2), 314-318.

- [5] Atici, F. M., Ekiz, F., & Lebedinsky, A. (2014). Cagan type rational expectation model on complex discrete time domains. *European Journal of Operational Research*, 237(1), 148-151.
- [6] Azaron, A., Katagiri, H., Sakawa, M., & Modarres, M. (2005). Reliability function of a class of time-dependent systems with standby redundancy. *European Journal of Operational Research*, 164(2), 378-386.
- [7] Barlow, R. E. (1985). Combining component and system information in system reliability calculation. In *Probabilistic Methods in the Mechanics of Solids and Structures* (pp. 375-383). Springer, Berlin, Heidelberg.
- [8] Barlow and Heidtmann: Barlow, R. E., & Heidtmann, K. D. (1984). Computing k-out-of-n system reliability. *IEEE Transactions on Reliability*, 33(4), 322-323.
- [9] Barro, R. J. (1976). Rational expectations and the role of monetary policy. *Journal of Monetary Economics*, 2(1), 1-32.
- [10] Basu, A. P., & Ebrahimi, N. (1991). Bayesian approach to life testing and reliability estimation using asymmetric loss function. *Journal of Statistical Planning and Inference*, 29(1-2), 21-31.
- [11] Bhattacharya, S. K. (1967). Bayesian approach to life testing and reliability estimation. *Journal of the American Statistical Association*, 62(317), 48-62.
- [12] Blanchard, O. J., & Watson, M. W. (1982). Bubbles, rational expectations and financial markets (No. w0945). National Bureau of Economic Research.
- [13] Boland, P. J., & El-Newehi, E. (1998). Statistical and information based (physical) minimal repair for k out of n systems. *Journal of Applied Probability*, 35(3), 731-740.
- [14] Boland, P. J., & Proschan, F. (1983). The reliability of k out of n systems. *The Annals of Probability*, 11(3), 760-764.
- [15] Canfield, R. V. (1970). A Bayesian approach to reliability estimation using a loss function. *IEEE Transactions on Reliability*, 19(1), 13-16.
- [16] Cerqueti, R. (2021). A new concept of reliability system and applications in finance. *Annals of Operations Research*, doi:10.1007/s10479-021-04150-9.
- [17] Chiang, D. T., & Niu, S. C. (1981). Reliability of consecutive-k-out-of-n: F system. *IEEE Transactions on Reliability*, 30(1), 87-89.
- [18] Chow, G. C. (1991). Rational versus adaptive expectations in present value models. *Econometric Decision Models* (pp. 269-284). Springer, Berlin, Heidelberg.



- [19] Coit, D. W., & Liu, J. C. (2000). System reliability optimization with k-out-of-n subsystems. *International Journal of Reliability, Quality and Safety Engineering*, 7(02), 129-142.
- [20] Da Costa Bueno, V., & do Carmo, I. M. (2007). Active redundancy allocation for a k-out-of-n: F system of dependent components. *European Journal of Operational Research*, 176(2), 1041-1051.
- [21] Davies, K., & Dembińska, A. (2019). On the number of failed components in a k-out-of-n system upon system failure when the lifetimes are discretely distributed. *Reliability Engineering & System Safety*, 188, 47-61.
- [22] Delcey, T., & Sergi, F. (2019). The Efficient Market Hypothesis and Rational Expectations. How Did They Meet and Live (Happily?) Ever After.
- [23] Di Crescenzo, A., & Pellerey, F. (2011). Improving series and parallel systems through mixtures of duplicated dependent components. *Naval Research Logistics (NRL)*, 58(5), 411-418.
- [24] El-Sayyad, G. M. (1967). Estimation of the parameter of an exponential distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 29(3), 525-532.
- [25] Eryilmaz, S. (2011). Dynamic behavior of k-out-of-n: G systems. *Operations Research Letters*, 39(2), 155-159.
- [26] Eryilmaz, S. (2014). Multivariate copula based dynamic reliability modeling with application to weighted-k-out-of-n systems of dependent components. *Structural Safety*, 51, 23-28.
- [27] Eryilmaz, S. (2019).  $(k_1, k_2, \dots, k_m)$ -out-of-n system and its reliability. *Journal of Computational and Applied Mathematics*, 346, 591-598.
- [28] Eryilmaz, S., & Bozbulut, A. R. (2014). Computing marginal and joint Birnbaum, and Barlow–Proschan importances in weighted-k-out-of-n: G systems. *Computers & Industrial Engineering*, 72, 255-260.
- [29] Friedman, M. (1957). Introduction to “A theory of the consumption function”. A theory of the consumption function (pp. 1-6). Princeton University Press.
- [30] Ge, G., & Wang, L. (1990). Exact reliability formula for consecutive-k-out-of-n: F systems with homogeneous Markov dependence. *IEEE Transactions on Reliability*, 39(5), 600-602.
- [31] Gokhale, S. S., Lyu, M. R., & Trivedi, K. S. (1998). Reliability simulation of component-based software systems. In *Proceedings Ninth International Symposium on Software Reliability Engineering*, 192-201.
- [32] Gunawan, S., & Papalambros, P. Y. (2006). A Bayesian approach to reliability-based optimization with incomplete information. *Journal of Mechanical Design*, 128(4), 909-918.

- [33] Gupta, N., Misra, N., & Kumar, S. (2015). Stochastic comparisons of residual lifetimes and inactivity times of coherent systems with dependent identically distributed components. *European Journal of Operational Research*, 240(2), 425-430.
- [34] Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, 50(5), 1269-1286.
- [35] Heidtmann, K. D. (1982). Improved method of inclusion-exclusion applied to k-out-of-n systems. *IEEE Transactions on Reliability*, 31(1), 36-40.
- [36] Huang, J., Zuo, M. J., & Fang, Z. (2003). Multi-state consecutive-k-out-of-n systems. *IIE Transactions*, 35(6), 527-534.
- [37] Huang, J., Zuo, M. J., & Wu, Y. (2000). Generalized multi-state k-out-of-n: G systems. *IEEE Transactions on reliability*, 49(1), 105-111.
- [38] Hwang, F. K. (1986). Simplified reliabilities for consecutive-k-out-of-n systems. *SIAM Journal on Algebraic Discrete Methods*, 7(2), 258-264.
- [39] Jain and Gopal: Jain, S. P., & Gopal, K. (1985). Recursive algorithm for reliability evaluation of k-out-of-n: G system. *IEEE Transactions on Reliability*, 34(2), 144-150.
- [40] Jiang, Y., Lin, J., Cukic, B., & Menzies, T. (2009, November). Variance analysis in software fault prediction models. In *2009 20th International Symposium on Software Reliability Engineering*, 99-108.
- [41] Jondeau, E., Zhang, Q., & Zhu, X. (2019). Average skewness matters. *Journal of Financial Economics*, 134(1), 29-47.
- [42] Khaledi, B. E., & Shaked, M. (2007). Ordering conditional lifetimes of coherent systems. *Journal of Statistical Planning and Inference*, 137(4), 1173-1184.
- [43] Keilis-Borok, V., Shebalin, P., Gabrielov, A., & Turcotte, D. (2004). Reverse tracing of short-term earthquake precursors. *Physics of the Earth and Planetary Interiors*, 145(1-4), 75-85.
- [44] Kinatader, H., & Papavassiliou, V. G. (2019). Sovereign bond return prediction with realized higher moments. *Journal of International Financial Markets, Institutions and Money*, 62, 53-73.
- [45] Krishnamurthy, S., & Mathur, A. P. (1997). On the estimation of reliability of a software system using reliabilities of its components. In *Proceedings The Eighth International Symposium on Software Reliability Engineering*, 146-155.

- [46] Kuo, W., Zhang, W., & Zuo, M. (1990). A consecutive-k-out-of-n: G system: the mirror image of a consecutive-k-out-of-n: F system. *IEEE Transactions on Reliability*, 39(2), 244-253.
- [47] Larasati, A., Hajji, A. M., & Dwiastuti, A. (2019). The relationship between data skewness and accuracy of Artificial Neural Network predictive model. *IOP Conference Series: Materials Science and Engineering*, 523(1), 012070.
- [48] Li, X., & Zuo, M. J. (2002). On the behaviour of some new ageing properties based upon the residual life of k-out-of-n systems. *Journal of Applied Probability*, 426-433.
- [49] Li, W., & Zuo, M. J. (2008). Reliability evaluation of multi-state weighted k-out-of-n systems. *Reliability Engineering & System Safety*, 93(1), 160-167.
- [50] Lucas Jr, R. E. (1972). Expectations and the Neutrality of Money. *Journal of economic theory*, 4(2), 103-124.
- [51] Martz, H. F., & Wailer, R. A. (1990). Bayesian reliability analysis of complex series/parallel systems of binomial subsystems and components. *Technometrics*, 32(4), 407-416.
- [52] Martz, H. F., Wailer, R. A., & Fickas, E. T. (1988). Bayesian reliability analysis of series systems of binomial subsystems and components. *Technometrics*, 30(2), 143-154.
- [53] Mastran, D. V. (1976). Incorporating component and system test data into the same assessment: a Bayesian approach. *Operations Research*, 24(3), 491-499.
- [54] Mastran, D. V., & Singpurwalla, N. D. (1978). A Bayesian estimation of the reliability of coherent structures. *Operations Research*, 26(4), 663-672.
- [55] McGrady, P. W. (1985). The availability of a k-out-of-n: G network. *IEEE transactions on reliability*, 34(5), 451-452.
- [56] Milczek, B. (2003). On the class of limit reliability functions of homogeneous series “k-out-of-n” systems. *Applied Mathematics and Computation*, 137(1), 161-176.
- [57] Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, 29(3), 315-335.
- [58] Nakagawa, T. (1985). Optimization problems in k-out-of-n systems. *IEEE transactions on reliability*, 34(3), 248-250.
- [59] Navarro, J., del Águila, Y., Sordo, M. A., & Suárez-Llorens, A. (2013). Stochastic ordering properties for systems with dependent identically distributed components. *Applied Stochastic Models in Business and Industry*, 29(3), 264-278.

- [60] Navarro, J., & Eryilmaz, S. (2007). Mean residual lifetimes of consecutive-k-out-of-n systems. *Journal of Applied Probability*, 44(1), 82-98.
- [61] Navarro, J., Pellerey, F., & Di Crescenzo, A. (2015). Orderings of coherent systems with randomized dependent components. *European Journal of Operational Research*, 240(1), 127-139.
- [62] Navarro, J., Ruiz, J. M., & Sandoval, C. J. (2005). A note on comparisons among coherent systems with dependent components using signatures. *Statistics and Probability Letters*, 72(2), 179-185.
- [63] Navarro, J., Samaniego, F. J., Balakrishnan, N., & Bhattacharya, D. (2008). On the application and extension of system signatures in engineering reliability. *Naval Research Logistics (NRL)*, 55(4), 313-327.
- [64] Navarro, J., & Spizzichino, F. (2010). Comparisons of series and parallel systems with components sharing the same copula. *Applied Stochastic Models in Business and Industry*, 26(6), 775-791.
- [65] Oe, S., Soeda, T., & Nakamizo, T. (1980). A method of predicting failure or life for stochastic systems by using autoregressive models. *International Journal of Systems Science*, 11(10), 1177-1188.
- [66] Ooghe, H., & Spaenjers, C. (2010). A note on performance measures for business failure prediction models. *Applied Economics Letters*, 17(1), 67-70.
- [67] Pham, H., & Upadhyaya, S. J. (1988). The efficiency of computing the reliability of k-out-of-n systems. *IEEE Transactions on Reliability*, 37(5), 521-523.
- [68] Phillips, M. J. (1980). k-out-of-n: G systems are preferable. *IEEE Transactions on Reliability*, 29(2), 166-169.
- [69] Rai et al: Rai, S., Sarje, A. K., Prasad, E. V., & Kumar, A. (1987). Two recursive algorithms for computing the reliability of k-out-of-n systems. *IEEE Transactions on Reliability*, 36(2), 261-265.
- [70] Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21(2), 201-214.
- [71] Reijns, G. L., & Van Gemund, A. J. (2007). Reliability analysis of hierarchical systems using statistical moments. *IEEE Transactions on Reliability*, 56(3), 525-533.
- [72] Risse, T. (1987). On the evaluation of the reliability of k-out-of-n systems. *IEEE transactions on reliability*, 36(4), 433-435.

- [73] Sanyal, S., Shah, V., & Bhattacharya, S. (1997). Framework of a software reliability engineering tool. In Proceedings 1997 High-Assurance Engineering Workshop, 114-119.
- [74] Sargent, T. J., Fand, D., & Goldfeld, S. (1973). Rational expectations, the real rate of interest, and the natural rate of unemployment. *Brookings Papers on Economic Activity*, 1973(2), 429-480.
- [75] Sargent, T. J., & Wallace, N. (1975). "Rational" expectations, the optimal monetary instrument, and the optimal money supply rule. *Journal of Political Economy*, 83(2), 241-254.
- [76] Sarje, A. K., & Prasad, E. V. (1989). An efficient non-recursive algorithm for computing the reliability of k-out-of-n systems. *IEEE Transactions on Reliability*, 38(2), 234-235.
- [77] Sarhan, A. M. (2005). Reliability equivalence factors of a parallel system. *Reliability Engineering & System Safety*, 87(3), 405-411.
- [78] Shebalin, P., Keilis-Borok, V., Zaliapin, I., Uyeda, S., Nagao, T., & Tsybin, N. (2004). Advance short-term prediction of the large Tokachi-oki earthquake, September 25, 2003, M= 8.1 A case history. *Earth, Planets and Space*, 56(8), 715-724.
- [79] Sheu, S. H., Liu, T. H., Tsai, H. N., & Zhang, Z. G. (2019). Optimization issues in k-out-of-n systems. *Applied Mathematical Modelling*, 73, 563-580.
- [80] Sornette, D., & Andersen, J. V. (2006). Optimal prediction of time-to-failure from information revealed by damage. *Europhysics Letters*, 74(5), 778.
- [81] Taghipour, S., & Kassaei, M. L. (2015). Periodic inspection optimization of a k-out-of-n load-sharing system. *IEEE Transactions on Reliability*, 64(3), 1116-1127.
- [82] Tian, Z., Zuo, M. J., & Yam, R. C. (2008). Multi-state k-out-of-n systems and their performance evaluation. *IEEE Transactions*, 41(1), 32-44.
- [83] Twomey, J. M., & Smith, A. E. (1998). Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 28(3), 417-430.
- [84] Van Gemund, A. J., & Reijns, G. L. (2012). Reliability analysis of  $k$ -out-of- $n$  systems with single cold standby using Pearson distributions. *IEEE Transactions on Reliability*, 61(2), 526-532.
- [85] Varian, H. R. (1975). A Bayesian approach to real estate assessment. *Studies in Bayesian econometric and statistics in Honor of Leonard J. Savage*, 195-208.
- [86] Wang, Y., Li, L., Huang, S., & Chang, Q. (2012). Reliability and covariance estimation of weighted k-out-of-n multi-state systems. *European Journal of Operational Research*, 221(1), 138-147.

- [87] Wu, J. S., & Chen, R. J. (1994). An algorithm for computing the reliability of weighted k-out-of-n system. *IEEE Transactions on Reliability*, 43, 327-328.
- [88] Yacoub, S., Cukic, B., & Ammar, H. H. (2004). A scenario-based reliability analysis approach for component-based software. *IEEE Transactions on Reliability*, 53(4), 465-480.
- [89] Zarezadeh, S., Mohammadi, L., & Balakrishnan, N. (2018). On the joint signature of several coherent systems with some shared components. *European Journal of Operational Research*, 264(3), 1092-1100.
- [90] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394), 446-451.
- [91] Zhang, Y. (2018). Optimal allocation of active redundancies in weighted k-out-of-n systems. *Statistics & Probability Letters*, 135, 110-117.
- [92] Zhang, Y. (2020). Reliability analysis of randomly weighted k-out-of-n systems with heterogeneous components. *Reliability Engineering & System Safety*, 205, 107184.
- [93] Zhang, N., Fouladirad, M., & Barros, A. (2019). Reliability-based measures and prognostic analysis of a K-out-of-N system in a random environment. *European Journal of Operational Research*, 272(3), 1120-1131.
- [94] Zhang, M. H., Shen, X. H., He, L., & Zhang, K. S. (2018). Application of differential entropy in characterizing the deformation inhomogeneity and life prediction of low-cycle fatigue of metals. *Materials*, 11(10), 1917.
- [95] Zuo, M., & Kuo, W. (1990). Design and performance analysis of consecutive-k-out-of-n structure. *Naval Research Logistics*, 37(2), 203-230.
- [96] Zuo, M. J., & Tian, Z. (2006). Performance evaluation of generalized multi-state k-out-of-n systems. *IEEE Transactions on Reliability*, 55(2), 319-327.