

Article

A Stochastic Approach for Product Costing in Manufacturing Processes

Paulo Afonso ^{1,*} , Vishad Vyas ¹, Ana Antunes ¹ , Sérgio Silva ² and Boris P. J. Bret ² 

¹ Centro Algoritmi, University of Minho, 4710-057 Braga, Portugal; b12192@algoritmi.uminho.pt (V.V.); ana.antunes@dps.uminho.pt (A.A.)

² Bosch Car Multimedia Portugal, S.A., 4705-285 Braga, Portugal; Sergio.Silva2@pt.bosch.com (S.S.); Boris.Bret@pt.bosch.com (B.P.J.B.)

* Correspondence: psafonso@dps.uminho.pt

Abstract: Nowadays, manufacturing companies are characterized by complex systems with multiple products being manufactured in multiple assembly lines. In such situations, traditional costing systems based on deterministic cost models cannot be used. This paper focuses on developing a stochastic approach to costing systems that considers the variability in the process cycle time of the different workstations in the assembly line. This approach provides a range of values for the product costs, allowing for a better perception of the risk associated to these costs instead of providing a single value of the cost. The confidence interval for the mean and the use of quartiles one and three as lower and upper estimates are proposed to include variability and risk in costing systems. The analysis of outliers and some statistical tests are included in the proposed approach, which was applied in a tier 1 company in the automotive industry. The probability distribution of the possible range of values for the bottleneck's cycle time showcase all the possible values of product cost considering the process variability and uncertainty. A stochastic cost model allows a better analysis of the margins and optimization opportunities as well as investment appraisal and quotation activities.

Keywords: stochastic models; costing systems; variability; uncertainty; risk



Citation: Afonso, P.; Vyas, V.; Antunes, A.; Silva, S.; Bret, B.P.J. A Stochastic Approach for Product Costing in Manufacturing Processes. *Mathematics* **2021**, *9*, 2238. <https://doi.org/10.3390/math9182238>

Academic Editors: Yi-Kuei Lin, Lance Fiondella, Cheng-Fu Huang and Ping-Chen Chang

Received: 7 August 2021
Accepted: 9 September 2021
Published: 12 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To control the costs of products and services, companies should have accurate information of the relevant cost objects [1,2]. Therefore, it is very important to have a proper costing system and controlling of the cost [3]. Furthermore, companies now are facing competition globally, pushing the development of new products, using different and more complex production processes, which enforce even more sophisticated costing systems [4]. Major technical decisions taken in the manufacturing industry related to new products and processes must be supported by complete, accurate, and timely information about costs and profitability [5].

Nevertheless, many companies continue to perform the costing of the product in a traditional way, e.g., allocating costs to products proportionally to the quantities produced. Nowadays, companies are characterized by complex systems with multiple products being manufactured in multiple assembly lines. In such situations, traditional costing systems cannot be used. It is very often that cost becomes distorted in a traditional costing system as accounting decisions are made years ago [1], when the range of products in the company was narrow.

A good costing system helps managers to understand the detailed cost of different short- and long-term activities and processes. The costs of products and services and other relevant cost objects include raw material costs, direct labor costs, indirect costs and other expenses from non-production departments. To allocate adequate costs to cost objects, appropriate and consistent information is needed to be successful [6].

In cost management, the main focus of researchers and practitioners is on solving problems related to the allocation of resource costs to cost objects (e.g., products), overhead cost analysis, etc. To solve these problems, they develop and apply deterministic cost models. Deterministic cost models play a pivotal role in the understanding of product costing and to perform profitability analysis and pricing strategies. On the other hand, cost estimation is important for quotation and budgeting exercises and to support the design of the business plans. Hence, cost estimation models are often used to reduce the cost of the product or optimize costs. The relevance of the selected method influences the quality of the estimations [7].

Actual manufacturing systems are characterized by a high level of variability and uncertainty, which can drastically affect the cost of the product. To control and monitor the process of manufacturing, it is important to consider the uncertainty to achieve the required levels of consistency, quality, and economy [8]. The results can depend on the uncertainty incorporated by the unknown variables that are not modelled. Uncertainty factors in manufacturing processes are demand, cycle time, resources, etc. The uncertain situations in the stochastic approach demand the use of probability distributions [9]. Such scenarios can be created for uncertainty based on optimistic, pessimist and neutral forecast values and their probability [10]. The mean values of the uncertain parameters can be used for the development of the stochastic model [11].

Thus, the motivation to develop this work is, as there is space for it, improvement in the costing of the product in practice (approaches and tools used by companies) and conceptually (concepts and models). It is indeed necessary to accommodate all the variabilities and uncertainty that prevail during the manufacturing process. This paper focuses on developing a stochastic approach to costing systems that considers the variability in the process cycle time of the different workstations in the assembly line. Such an approach will provide the range of values for the product costs, allowing a better perception of the risk associated to these costs, instead of providing a single value of the cost.

Stochastic analysis is associated with the analysis of events in which at least one of the components of the process is random. It is usually applied in situations that face a set of random variables over time. The literature presents several approaches to deal with the stochastic behavior, with Markov models and simulations being some of the most common. A Markov process is described as such when only the current value of the variable is taken into account to predict its future value. Process simulations are associated with the random generation of values, assuming that they are in a given interval or in accordance with, for example, the variable's mean value and respective standard deviation [12]. Usually, values are generated following a normal distribution of mean x and standard deviation s . The stochastic frontier analysis can be used to identify the efficiency and productivity of manufacturing processes. A model can be designed for each plant, process or production line, and these models can be combined to achieve the overall performance, bridging the gap between the current productivity and the objective to be achieved [13].

Thus, in order to achieve that, it is essential to use mathematical concepts to understand, in a first instance, the variability and then to identify a range of values for the product cost applying statistical methods. With the applied mathematics, a better perception of the product process can be achieved, and better decisions can be made. This paper extends and complements recent research on applied mathematics in the industrial engineering setting, where stochastic modelling and new mathematical models for the allocation of process costs to products were used to improve the performance evaluation, optimization of maintenance costs, supply chain costs, investment decisions and production inventory with flexible manufacturing [14–17]. Recent work by [9] shows the use of the stochastic approach in production planning and to optimize the profit as uncertainties are involved. By using the methodology proposed in this paper, it will showcase the historical statistical data of a defined period so that the variability in the process can be observed and the real-time cost of the product can be calculated. This can contribute significantly to support continuous improvement in production systems. Moreover, it can also bridge the gap

between financial and production departments, integrating production and accounting information. The comparison between the standard and the real cost can be done in a better way, which will facilitate both operational and strategic decision making.

In this research project, a six-step methodology was developed and applied. Firstly, a data analysis is performed to obtain relevant descriptive statistics and identify the outliers which must be removed from the product cost analysis. After removing the outliers, the descriptive analysis must be done again to understand the data. The third step is to perform the hypothesis test to compare cycle times across the different assembly lines. The next step is to obtain the confidence interval for the mean, which provides information about the potential risk and variability in the cost of the product. Additionally, quartiles Q1 and Q3 provide a range of potential values for the computation of product costs, highlighting the inherent risk of such cost.

The proposed model was applied in a tier 1 manufacturer of the automotive industry. Specifically, one product was taken into consideration on which analysis was performed. To manufacture this product, 18 workstations were needed. Only the bottleneck workstation of the assembly line was considered since it represents the cycle time of the assembly line. So, any variability and uncertainty in the bottleneck will affect the cycle time of the assembly line resulting in changes in the cost. The cycle time of the bottleneck workstation was gathered, and after analyzing and removing the outliers, a descriptive analysis was performed on it. After performing the statistical tests, it was evident the existence of variability and uncertainty. Following the stochastic approach, the range of cost was calculated, using the quartiles and confidence interval for the mean. This stochastic range of cost accommodates the variability and uncertainty that can prevail during the manufacturing process in a specific period for a certain type of product.

In the next section, a literature review on uncertainty and variability, and stochastic approaches in costing systems is presented. To counter the impact of uncertainty and variability on cost computation, the use of a stochastic approach is proposed. The methodology is explained in Section 3, and the results of its application in a case study developed in a tier 1 manufacturer of the automotive industry are presented in Section 4. The computation of the product costs and relevance of the proposed methodology are discussed in Section 5. Section 6 presents the main conclusions and opportunities for further research.

2. Literature Review

2.1. Uncertainty and Variability in Cost Models

Uncertainty, in the modelling process, is due to insufficient knowledge, which implies that predictions will differ from reality [18]. There are significant sources of uncertainty associated to the cost estimation process. During the estimation of the cost, the information of the cost available is usually the historic product costs; sometimes, this information has a high level of uncertainty, first in the design stage, and later, in the manufacturing process, e.g., the dimensions of the product, process cycle times [19]. The uncertainty in the cost estimation can be reduced by providing more complete information about the manufacturing process, product design, product support, reliability, and requirements of disposal [20].

When an event is modelled, a decision must be made by the decision maker whether it is necessary to include the uncertainty that prevails with the event. In some cases, it is enough to approximate the uncertain events by a deterministic model. However, if uncertainty happens to be significant and can impact the results, then it is important to include it in the model, and uncertain events should be studied. In this kind of situation, it is necessary to choose a proper method to model the uncertainty [21]. Uncertainty and variability are present in many real-life events, e.g., variation in the cycle time of assembly lines. It may occur due to several reasons, such as lack of information, ambiguity, complexity in the information, errors in measurements, beliefs instead of real information, etc. [21]. The magnitude of uncertainty depends on the information available and the

complexity of the event. Uncertainty is also related to the possibility of error from not having ample information about the event and the related aspects [22].

There are different definitions for uncertainty [7]. For example, in the financial domain, there is a fine distinction between uncertainty and risk. Indeed, it is normal to speak of risk when the future is unknown but there is known a probability distribution of such a probable future. On the other hand, uncertainty occurs when the probability distribution is itself unknown [23]. For Zimmermann [21], “uncertainty implies that in a certain situation a person does not dispose about information which quantitatively and qualitatively is appropriate to describe, prescribe or predict deterministically and numerically a system, its behavior or other characteristic”.

It is evident from different sources that uncertainties and variabilities can impact cost estimation. If more information is available, it is possible to reduce the uncertainty in the estimation of the cost, specifically, details about manufacturing and design of the product [7]. In the field of cost calculation and costing systems, many methods can be used. In the past, a sensitivity analysis was used to estimate the cost and uncertainty by measuring the level of risk. It can provide brief information to the decision maker about what can happen if the input changes, considering the best-case and worst-case situation. However, this method has its limitations, namely, when there is a considerable number of variables, we have difficulties in estimating what will occur and understanding how the relationship among the variables can explain the phenomenon under study. To lessen the gravity of this limitation, various probabilistic methods are developed. In these methods, the uncertainty related to cost information is expressed through probability functions. Probabilistic methods help decision makers by providing more information, but they require a larger amount of data and more sophisticated statistics than the sensitivity analysis [24].

Managing the cost is important to understand the cost behavior, as the variability in the market, prices, and work methodology affect the cost. This variability can affect the cost of the product; hence, the efficiency of the company can be affected. A better understanding of the product variability is necessary for a better control over the cost. A proper risk assessment demands a good understanding of the impact of uncertainty and variability on costs [25].

Several methods, including activity-based costing, have been proposed but they have limitations to incorporate adequately uncertainty and variability. It has been proposed to use the fuzzy approach, which is based on the rationality of uncertainty because of factors such as vagueness and inaccuracy [26]. In the cost–volume–profit analysis, various factors, such as risk and uncertainty, are ignored; thus, it severely limits the usefulness of this methodology. By using a fuzzy set, it is possible to manage inaccuracy, giving access to evaluate the cost–volume–profit decision-making process [27]. Monte Carlo simulation and fuzzy set theory are better methods to handle the uncertainty and inaccuracy in the data for cost models [22]. When uncertainties are partially or totally random, they are usually represented by probability density functions, which can be used also in costing systems [28]. When the decision maker faces an uncertain problem, it can be expressed as uncertain ratios also known as fuzzy numbers [29]. Techniques based on the fuzzy numbers have been proposed to manage uncertainty in cost models [22]. There is research based on linear planning models using fuzzy numbers to deal with the uncertainty of the parameters [9]. For example, fuzzy-based-activity-based costing can help to compute the cost [30]. Additionally, it can allow managers to control the cost and define key performance indicators, supporting decision making [30].

2.2. Stochastic Approaches in Cost Models

A stochastic model can be defined as the collection of random variables that are arranged in a specific mathematical set, which is associated with an element of the set [31]. A stochastic model is a tool to estimate the probability distribution of the possible results by allowing the random variable in one or more inputs over time. The random variables are

based on the variability and uncertainty found in the historical data for the selected time, using standard time series methods. The distribution of the probable results is derived from the number of stochastic projections which reflect the random variable in the inputs [32].

A stochastic cost model will set up a projection model, which allows the analysis of single products, several products in the same assembly line and all products in the entire company. It can use random variation to understand what conditions (risk, variability, and uncertainty) can affect the cost of the product. In the end, the distribution of the outcome can portray not only the most likely cost of the product, but the whole possible range of product costs from a reasonable set of assumptions and constraints. The most likely estimate is given by the distribution curve also known as the probability density function, which is typically also the mode of the curve [32].

The advantage of a stochastic approach is that, unlike the deterministic approach which only gives a value to the product cost, this method of cost modelling gives the entire range of possible cost values of the product, considering the related variability, risk, and uncertainty. It can reflect better real-world situations, providing different ranges of possible results. Additionally, by running several rounds of calculations, and using many different estimates of future economic situations, the model will show the range of costs, showing the potential upside and downside of each.

Stochastic frontier analysis models are used as a statistical benchmark to provide an overview of the industrial sector. They can also be adjusted to understand plant productivity. However, the general form of the stochastic frontier analysis is difficult to implement in a complex manufacturing industry because of the problems of multicollinearity [13]. It is important to have accurate information; with modern complex systems, deterministic models are impractical, as they cannot represent the disturbance and uncertainty adequately. So, it is necessary to apply stochastic models [8].

The deterministic cost model was extended by Ioannou et al. [32] to systematically account for the stochastic input. They performed a Monte Carlo simulation to obtain the probability distributions that allow estimating the probabilities of exceeding a set of thresholds and determining the confidence interval. This stressed the importance of appropriate statistical modelling of the stochastic variables to reduce modelling uncertainties and contribute to better-informed decision making for making investments. An advanced stochastic model was applied to identify the most relevant parameters influencing cost and uncertainty. The stochastic model was implemented by using a @risk software extension. They also investigated the variables that behave as the most important cost drivers, which are behind the effective reduction of the cost, thus providing information on where additional efforts are required, effectively reducing the costs [33].

There are stochastic cost models addressing the influence of uncertainties in wind generation on the optimal operation of power systems [34]. In this particular case of wind generation, there are a lot of possible scenarios, so a scenario reduction algorithm was applied. The stochastic model gives a near possible optimal solution considering all possible scenarios. This particular solution cannot be optimal for one particular scenario, but it is robust over all possible realization of the uncertainties. An adaptive particle swarm optimization algorithm was proposed to solve the stochastic cost model, which overcomes traditional drawbacks, such as penalty coefficients, and parameter tuning. Sobu et al. [35] developed stochastic scenarios by using observed mean-values and standard deviation from data, and then stochastic operation cost optimization models for minimizing operation cost were formulated. Then, they used a particle swarm optimization approach.

Stochastic approaches were also applied in the context of life cycle costing to study the economic value of energy-efficient building retrofitting investments [36]. The authors also investigated the effect of interdependent stochastic variables, such as explicit evaluation. The economic evaluation is itself stochastic, so it can express both the expected value and the value inherent to uncertainty and risks. The actual validity of stochastic costing depends on the reliability of the conclusions that can be drawn. This reliability consists of the robustness of the results and accuracy with respect to the real system, which the

model aims to represent. The first aspect, in particular, not only generates the stochastic life cycle cost outcomes, but also allows us to assess the variability of the outcome given by the dataset.

Furthermore, Baldoni et al. [37] developed a new software tool for the evaluation of life cycle impacts and costs assessments, which can be used to support decision making. This tool allows evaluating the long-term trade-off between the economic and environmental performance of investment projects, while accounting for uncertainties in the input parameters. The software also offers several tools to perform sensitivity analysis.

Thus, by using a stochastic approach, the range of cost showing the potential upside and downside can be obtained (see Figure 1).

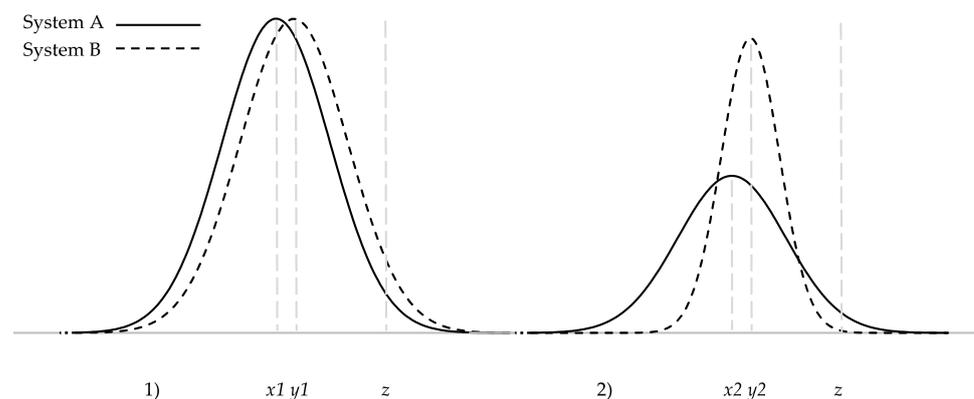


Figure 1. Comparison between deterministic and stochastic values in different situations.

In the figure above, based on [38], it can be observed that, from a deterministic perspective, were only x_1, y_1 and x_2, y_2 are considered, Systems A and B are far below risk z but the analysis is insufficient, and a probabilistic approach should be taken into consideration. Situations 1 and 2 show the range of possible values for each system under different conditions. In situation 2, System A, despite presenting a lower average (x_2), has a higher risk of reach z . Nevertheless, the risk of System B in situation 1 is higher. Thus, we should go beyond the simple deterministic analysis and understand the behavior of the phenomenon under study. This reasoning can be applied to products, activities and processes; a stochastic cost model makes it possible to measure such variability and its impact on the cost of the relevant cost objects. The range of costs, along with the measurement of predictions, should be provided. Such predictions should be based on the calculated risk for the relevant variables.

For example, in companies, the standard cost of the product can be compared with the estimations made for the different weeks of production, and the range of possible costs and all estimations can be compared with the real cost, which can be updated weekly.

3. Materials and Methods

Nowadays, manufacturing analytics are important to derive insights about the impacts on the organization of internal and external changes and variability [39]. Thus, statistical methods are an important tool since they can be used to deal with the variability in observed data. Moreover, data can be organized and summarized to understand the information available. Descriptive statistics are widely used to identify the important features of the data, for example, the mean, standard deviation, quartiles, minimum and maximum values, range and coefficient of variation.

This analysis is important for decision making, in general, and engineering and manufacturing, in particular [40]. For example, the detection of deviant behavior (large or small variations) in costs can be detected with measures, such as standard deviation and coefficient of variation.

Descriptive statistics, such as the minimum, maximum, mean, and standard deviation, were used to analyze product cost management data [41]. Another study was made to analyze the impact of strategic costing techniques, where the descriptive statistics mean and standard deviation were also applied to verify whether the new strategy achieved successful performance when compared with prior years [42]. Furthermore, the first quartile was also used for measuring machinery usage [43]. Chen et al. [44] defined the cost of care for congestive heart failure using the quartiles, where the lowest, middle, and highest costs are associated to the first quartile, second to third quartile, and more than the third quartile, respectively. The coefficient of variation is another metric used in previous studies; for example, operations' cycle times were used to analyze the optimal allocation of storage space in production lines [45]. According to [40], estimations using the mean could be close or far from the true mean; in order to avoid this, it can be used instead a range of the potential values, such as a confidence interval.

Therefore, in this paper, to analyze the product development process variability, we propose a methodology based on six main steps:

1. Firstly, data analysis must be performed to identify the descriptive statistics and outliers by activity and process, namely, mean, standard deviation, quartiles, minimum and maximum values, range and coefficient of variation. In the case study, the analysis was made by the workstation and line. If the outliers are caused due to external causes, then they must be removed;
2. After this removal, a new descriptive analysis must be performed to conduct a critical assessment. In this step, it is intended to identify what is happening and whether it is possible to find differences between lines for a specific workstation and product. The analysis of the outliers is important to analyze the efficiency of the process and to identify opportunities to improve the process;
3. The third step is related to performing hypothesis tests, where it is intended to compare the workstation in different lines and identify in each one whether there are significant differences;
4. The next step is to perform the confidence interval for the mean, considering that the level of confidence is 95%. Note that instead of considering the mean value to compute the product cost, the confidence interval provides information on the variability and the potential risk of the cost. In this case, the variability is analyzed since it is a range of values;
5. Furthermore, the quartiles values can be another possibility to take into consideration, as it also gives an idea of the risk associated to the product cost. The interval for the values can be computed, using the first and the third quartile thus, focusing on the 50% of the values around the median. The value associated to Q3 is a measure for the product cost risk because 25% of the produced units will have a higher cost than such a value. This is a conservative approach for cost risk analysis, and the 90 percentile can be used to signalize the risk of a too-high cost. On the other hand, Q1 represents a reference value for quotations because prices lower than this value will push the margin to negative values. Again, alternatively, a less conservative approach can be used considering, in this case, the 10 percentile. Thus, both the lower and upper limits can be used as risk measures;
6. Finally, the values achieved in the confidence interval and the quartiles (first and third) can be used to compute the product cost, considering each workstation or aggregated byline (usually, considering the bottleneck of the line).

Note that the proposed approach can be applied differently, depending on what is intended to be achieved. The six steps presented before can be simplified or developed, if necessary. For example, it may be only necessary to identify the confidence interval for the mean instead of the quartiles, or vice versa.

The implementation of the proposed methodology was made, using the 3.8 version of the Python software. The pandas' library was used to import the data to be analyzed and to produce the descriptive statistics, using the `read_csv` and `data.describe` functions,

respectively. Furthermore, the graphs were displayed using the matplotlib library with the plot function. Thereafter, the scipy.stats library was also used to obtain the confidence interval for the mean (norm.interval function) to verify that the data follow a normal distribution (kstest function) and to perform the nonparametric tests (mannwhitneyu function) [46–48]. Note that when the sample is large and the data do not follow a normal distribution, the mean and standard deviation are unknown, and the confidence interval can be performed using Equation (1) [40]. The normal.interval function considers the expression defined, where \bar{x} and s are the sample mean and standard deviation, n the sample number of observations, α is the significance level that it is pretended to be used. Thus, $z_{\frac{\alpha}{2}}$ is the chosen z -value, also known as the critical value [40].

$$\bar{x} - z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}} \quad (1)$$

To compare two independent samples, one can use parametric and non-parametric tests. A parametric test must be used when the population follows a normal distribution, has equal variance and is continuous. However, non-parametric tests are applied when at least one of the parametric assumptions is not validated. Note that parametric tests are more robust than non-parametric ones and consider less information to make stronger conclusions. Therefore, Student's t -test is a parametric test commonly used to identify if the mean of one sample is different from a known mean or to identify if there are differences between the mean from two identical samples. Furthermore, the Mann–Whitney test is a non-parametric alternative to evaluate if two samples are from the same population [49]. In order to check if the sample follows a normal distribution, Shapiro–Wilk and Kolmogorov–Smirnov are well-known tests for the task. Where the Shapiro–Wilk test is commonly used in small samples ($n \leq 50$), the Kolmogorov–Smirnov test is applied in other cases [50].

Moreover, the Wilcoxon Sign Rank Test was used for the validation and analysis of the computed costs. It is a t -test alternative since it is a non-parametric test. Thus, this test intends to evaluate whether the median of one sample is different from a known value instead of using the mean value [40]. In this case, the test was used to assess whether the calculated values of the costs present significant variations over the weeks in relation to the planned/standard cost. The Wilcoxon test was performed and paired, with the assumption of two-tailed distributions and for a significance level of 5%. This test allows us to evaluate the differences or disparities of the median values of the data, being useful to understand whether observations or values of the same variable, recorded at different times, present significant variations or not. This way, it allows us to evaluate the adequacy of the cost model and evidence the existence of cost variability, justifying the stochastic analysis of costs in detriment of the traditional deterministic approach.

4. Analysis of Results

The proposed methodology to include variability in costing systems was applied following the six steps explained before. A data sample was considered, corresponding to a weekly period (seven consecutive days), considered normal (they were not considered holiday periods, breaks or other), referring to all production lines (A, B, C and D) where the product is produced. These are data obtained from the production information system, where the cycle time values for each day, per line and workstation, are recorded. These, in turn, were processed, removing the outliers from the cycle times per line and workstation. Next, the mean, quartile, extreme values (maximum and minimum), standard deviation and coefficient of variation were computed.

In each line, composed of several workstations, the bottleneck (workstation with the highest cycle time) was identified, and its frequency (count) was identified, corresponding to the number of units produced. In total, it is a sufficiently large sample, with about 38,000 observations, which are distributed by all the lines in a variable way, but also with considerable frequency values, that is, we have frequencies much higher than 50 (even the minimum value exceeds 3000).

The empirical data were obtained in a Tier 1 manufacturer of the automotive industry that produces instrumentation systems, navigation systems, and steering sensors, among others, partners with most car brands, is a worldwide leader in the areas of automotive and industrial technology, and provides products and services for professional and private use, making it an interesting case study.

Nowadays, for a company to be able to respond to customer demand and bring value through its products, it must be able to produce with great flexibility and diversity. To do so, an enormous complexity in the production process is necessary. Having complex processes in the assembly lines causes variation in the cycle time of the workstations, which will consequently affect the cost of the product. Thus, if a company wants to be competitive, it must understand and control the variation of several activities that compose the production process. This demands a stochastic approach in controlling the activities of production processes.

The company under study is characterized by the development and production of navigation systems for the automotive industry, mainly car displays. The development of these products starts from prototype construction to series production.

A product was selected, produced in 4 different production lines. These lines are considered semi-automatic lines since they require manual assembling (performed by operators) and automatic assembling. B, C and D lines are composed of 17 and line A by 18 workstations—one workstation can have one or more machines. All products pass through different tests, most of them automatic, but also tests with human intervention. In the last workstation, the product is labelled and then the process is finished. Before arriving at these 4 lines, the product already undergoes through other processes in the factory, with the studied process being the final one, before shipping the product to the client. Small lines, A and B, produce fewer quantities and therefore, have fewer operators allocated. Lines A and B have a different number of machines per workstation. Big lines C and D are considered large lines because their production volume is much higher than that of the small lines.

In order to analyze the variability between lines (A, B, C, and D), the bottleneck's cycle time (workstation 17) was analyzed. Figure 2 presents the cycle time of each piece produced in workstation 17 in the 4 production lines in one week. The data (i.e., cycle times and daily produced quantities) were collected for the period between the 18th and the 24th of December 2020. Both the quantities and process times were different in each production line, so these data clearly highlight the variability that exists in the production process. So, the cycle time at each workstation was recorded for each product unit manufactured during that week, highlighting the correct bottleneck of the production line. Once confirmed that the 17th workstation represented the bottleneck, the tests were made on that workstation, as it would define the production line cycle time. All the recorded cycle times from each assembly line were extracted from the company's management information systems to the statistics software, where various tests were made on the data.

As mentioned earlier, there are four different assembly lines involved in producing the product under scrutiny. Lines A and B are considered small lines, and lines C and D are considered the big lines. The difference between the small and big lines is the amount of equipment at each workstation. Big lines have more equipment, compared to the small lines. As they have more equipment in the workstations, big lines can process more parts in parallel. Hence, big lines produce faster and in greater quantity. Big lines produce around 15,000 parts per week, whereas small lines produce around 3000 parts. Production planning and scheduling prioritizes big lines, and small lines complement the big ones.

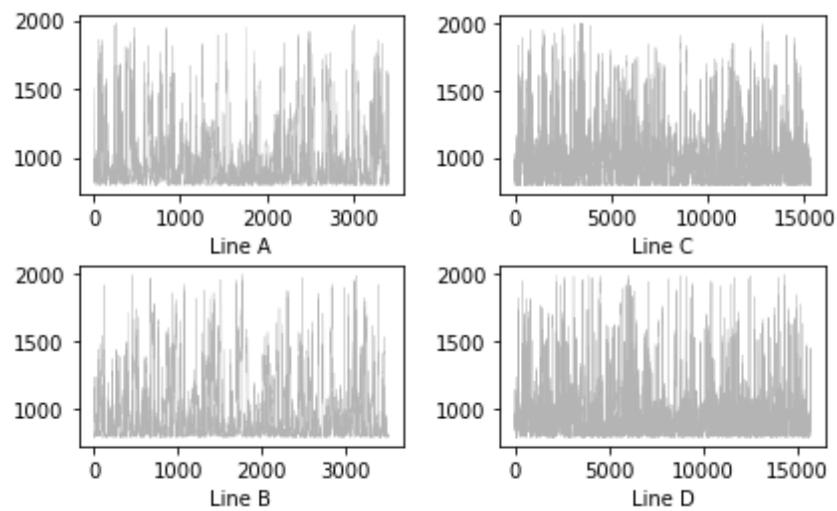


Figure 2. Cycle time (in seconds), per line.

A descriptive analysis was conducted. Table 1 presents the number of observations, mean, standard deviation, minimum, first, second and third quartile (Q1, Q2 and Q3), maximum, range and the coefficient of variation, for each line. The coefficient of variation is commonly used to identify whether the mean is representative. When this metric is less than 50%, the mean is representative. Otherwise, it is preferable to use a median instead.

Table 1. Descriptive statistics (in seconds), per line.

	Line A	Line B	Line C	Line D
Count	3407.00	3510.00	15,314.00	15,664.00
Mean	984.28	992.15	993.30	952.42
Standard Deviation	240.33	257.90	217.50	210.91
Minimum	789.00	786.00	786.00	784.00
Q1	823.00	807.00	843.00	815.00
Q2	886.00	868.00	938.00	888.00
Q3	1036.00	1089.75	1049.00	988.00
Maximum	1981.00	1997.00	1999.00	1998.00
Range	1192.00	1211.00	1213.00	1214.00
Coefficient of Variation	24.42	26.00	21.90	22.14

According to the results obtained, lines A and B have almost the same number of observations. The same conclusion can be drawn for lines C and D. Furthermore, according to the mean, line B has a longer cycle time when compared to line A. Besides that, line C is the one with the longer cycle time, when compared to line D. The minimum and maximum cycle times are nearly the same for all the lines. Another conclusion is that the mean is representative in all the lines, although there is variability since the range of the values (the difference between the maximum and minimum values) is high. Thus, it is important to understand the cause of these high values to avoid wrong conclusions.

According to what was observed, lines A and B present very close mean values, the difference being 7.87 s, while in lines C and D, the difference between the mean values is 40.88 s. In terms of standard deviation values, the difference is greater between lines A and B than between lines C and D, being, respectively, 17.57 and 6.59 s; the small lines show a tendency for greater variations in cycle times, around the mean. The interquartile

range is 213 and 282 s for lines A and B, respectively, and 206 and 173 s for lines C and D, respectively. There is a greater difference in the small lines compared to the big lines.

For the coefficient of variation, the values on the small lines are close (differential of 1.58 s) but on lines C and D, they are even more similar (differential of only 0.24 s). All lines show variation, although the highest values are observed in the small lines.

In general, the pairs of lines ((A, B); (C, D)) have characteristics that resemble each other, namely, count, mean, standard deviation and coefficient of variation, and, at the same time, allow the distinction between the two types of lines (small and big lines).

The available capacity and cycle times of the machines is fundamental to allocate the cost of resources used to the cost objects. The variability in cycle times gives us also information on the variability of the cost. Therefore, it is necessary to study the variability of the cycle time, and the average confidence interval can be a way to do it. Hence, in Table 2, it is shown the confidence interval for the mean cycle time in each line (given by the cycle time of the line's bottleneck, which is workstation 17).

Table 2. Confidence interval for the mean cycle time (in seconds), per line.

	Line A	Line B	Line C	Line D
Lower Bound	976.21	983.62	989.86	949.11
Upper Bound	992.35	1000.68	996.75	955.72

We can see that line D has the smallest values, and lines B and C have the higher ones. With these results, there is a suspicion that there are differences between lines A and B and between lines C and D. Differences between lines should be identified and analyzed because they can result from different and not optimized planning, efficiency, demand requirements, etc. Considering the high variability in internal processes and external demand, these differences must be monitored on a weekly or monthly basis to support effective and timely action plans from a continuous improvement philosophy.

To analyze these differences and trigger eventual action plans, non-parametric tests were performed since the lines do not follow a normal distribution. To evaluate the differences between lines, the analysis was conducted, considering line pairs A and B, C and D. Thus, the Mann–Whitney test was performed to assess differences between the lines. The hypotheses to take into consideration were as follows:

Hypothesis 1 (H1). *There are no significant differences between lines in terms of the execution (cycle) time.*

Hypothesis 2 (H2). *There are significant differences between lines in terms of the execution (cycle) time.*

Table 3 presents the p -value for the Mann–Whitney test (Mann–Whitney) and the mean value for each line pairs. According to these results, Hypotheses H1 and H2 are rejected since the p -value is less than the level of significance ($\alpha = 0.05$). Therefore, there are significant differences between the cycle time in lines A and B. The same conclusion can be drawn for lines C and D. According to the mean, lines B and C have higher cycle times than lines A and D, respectively. This variation between the small and big lines can influence the product cost and represent opportunities for improvement in process costs. In other words, if computed by the line, it is expected that the product cost will be higher in line B than in line A.

After this analysis, it is important to verify if there are outliers. Thus, Figure 3 presents the boxplot to visualize the cycle time variation in each line. With this visualization, it is possible to identify outliers and, since there are too many, they contribute to a very high variability. Hence, it is essential to understand why these values are happening to reduce such variability.

Table 3. Results of Mann–Whitney test.

Lines	Mann–Whitney	Mean (seconds)
A and B	<0.001	984.28 and 992.15
C and D	<0.001	993.3 and 952.42

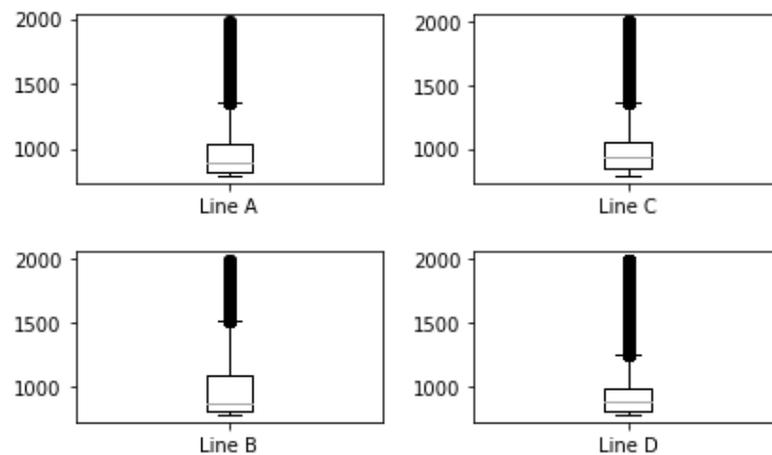


Figure 3. Boxplot for the cycle time (in seconds), per line.

The mean cycle time is higher for the small lines, compared to the big lines. When the demand is lower than the total capacity given by the four lines, the company chooses to produce in the big lines at full capacity, complemented by the small lines. This causes those small lines to produce below their capacity, reducing the performance of the small lines, and causing higher cycle times compared to big lines.

In terms of product cost, if the cycle time has a higher variability, then the variability of cost will be higher. Minimizing the final cost is important to increase the margin; minimizing variability contributes to decreasing the cost risk. Outliers are caused by internal and external factors to the process, which should be managed differently, namely in the context of continuous improvement or within the costing system. A new analysis was conducted without the outliers to reduce the variability, which can be managed within the costing system. Figure 4 presents the cycle time, per line, for workstation 17 without the outliers. In a first analysis, it can be observed that the maximum value decreased in all lines.

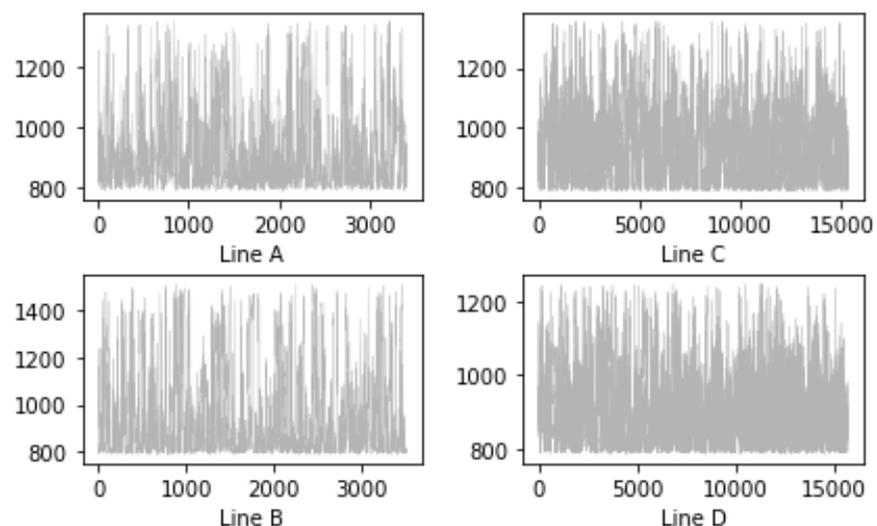


Figure 4. Cycle time per line (in seconds), without outliers.

The next step of the proposed methodology is to perform the descriptive statistics to identify which metrics change when the outliers are removed. Therefore, Table 4 presents the descriptive statistics, and we can see that most statistics have decreased, except for the minimum, which remained the same. Besides that, the range decreased considerably, as was expected, and, according to the coefficient of variation, the mean is still representative. Moreover, there is more evidence that the cycle time is different in the small and big lines since the means are slightly different.

Table 4. Descriptive statistics of cycle time (in seconds) without outliers, per line.

	Line A	Line B	Line C	Line D
Count	3070.00	3305.00	14,096.00	14,286.00
Mean	917.85	948.39	940.65	896.19
Standard Deviation	129.38	191.53	121.74	96.41
Minimum	789.00	786.00	786.00	784.00
Q1	820.00	805.00	835.00	811.00
Q2	869.00	854.00	923.00	873.00
Q3	979.75	1040.00	1015.00	956.00
Maximum	1355.00	1513.00	1358.00	1247.00
Range	566.00	727.00	572.00	463.00
Coefficient of Variation	14.10	20.19	12.94	10.76

Furthermore, the confidence interval for the mean cycle time is presented in Table 5, considering the confidence level to be 95%. These values also decreased, and the amplitude is, also, smaller. With these results, it is expected that there are differences between the cycle times per line.

Table 5. Confidence interval for the mean per line, without outliers.

	Line A	Line B	Line C	Line D
Lower Bound	913.27	941.86	938.64	894.61
Upper Bound	922.43	954.92	942.66	897.77

Regarding the analysis of the measures without the presence of outliers, the count values are very similar when analyzing the pairs of lines, A and B, and C and D. There is a greater difference in the means between these two pairs of lines and the respective standard deviation values. The values are lower compared to those obtained with the presence of outliers but more differentiated between lines of the same type. The coefficients of variation are also lower for all lines, but there is a greater difference between them when analyzing pairs of lines, A and B, and C and D.

To verify if there are differences in the cycle times per line, the Mann–Whitney test was performed, and Table 6 presents the results achieved. According to the p -value, in the Mann–Whitney test, there are significant differences between lines A and B. The same conclusion can be drawn for lines C and D. Lines B and C have a higher cycle time when compared with lines A and D, respectively. Thus, the conclusions are the same when all the available information is used. However, it is important to remember that we intend to analyze the variability within product cost, where extreme values can lead to wrong conclusions.

Table 6. Mann–Whitney results without outliers.

Lines	Mann–Whitney	Mean
A and B	<0.001	917.85 and 948.39
C and D	<0.001	940.65 and 896.19

After these analyses, the last step of the proposed methodology is to identify how many values are in each quartile to provide optimistic and pessimistic estimations for the product cost instead of a deterministic cost. Therefore, Table 7 presents the number of observations in each quartile, where the first count is the first 25% of the data, the second for 25 to 50% of the data, and the last one for 50 to 75% of data. For example, in line A, there are 795 observations with the cycle time being less than or equal to 820. The product cost for these cycle times will be the lowest when compared with the other quartiles because there is a reduced consumption of the resources. Thus, using these values, it is possible to propose a range for the product cost and measure cost risk, particularly, using the cycle time achieved in Q1 and Q3, respectively.

Table 7. Quartiles and number of observations for each line.

Line	Q1	Count	Q2	Count	Q3	Count
A	820	795	869	743	979.75	764
B	805	838	854	819	1040	823
C	835	3536	923	3527	1015	3515
D	811	3588	873	3572	956	3572

Taking into consideration the results achieved and presented in Table 7, the boxplot (Figure 5) was performed to visualize the variability of the data. Thus, there are new outliers, which are included in the variability of the process that is intended to be allocated to product cost. Initial outliers are supposed to be removed or, if not, allocated to the product as general costs not specific to the process/line. Identifying the different levels of cost and understanding their behavior is so important for allocating them to products. Costs can be specific to each produced unit, to the batch, the process, general costs of the product or general cost of the company/business. High variability in cycle times can be explained by reasons related to all these different levels.

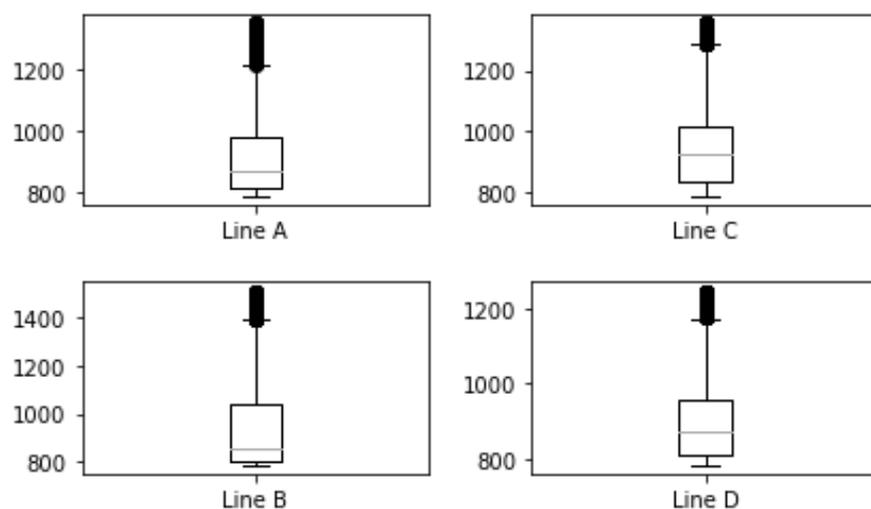


Figure 5. Boxplot for the cycle time per line, without outliers.

Thus, for the inclusion of variability in the computation of product costs, the cycle times associated to Q1 and Q3 and the confidence interval for the mean are used. Both can be calculated or estimated for each workstation or just considering the bottleneck of the line (in this case, workstation 17). The analysis made was used to compare production lines; thus, it was centered on the bottleneck which defines the production speed of the line. After this high-level approach to optimize production lines, a detailed analysis within each line should be made to analyze and optimize workstations.

5. Discussion

The stochastic analysis of production cycle times is fundamental to include variability and risk within costing systems. Besides the variability in the production processes, we can have also variability caused by changes in the demand and variability in the value of the resources used. Process variability is particularly relevant in costing systems and for optimization purposes, and it is the focus of this research work.

Manufacturing product costs can be explained through the typical three components: direct materials, direct labor and indirect costs (such as energy, amortization, area, etc.). Direct labor plus indirect costs represent the conversion costs. Costs can vary with the production, which is called variable costs, or not vary, which is called fixed costs. The unitary product cost can also include non-manufacturing costs (e.g., logistics costs, and sales and administrative costs), typically allocated on a volume basis. Such a complete cost can be compared to the price in order to evaluate the profitability of the product. However, a first analysis of the margins must be based on the manufacturing cost from which several actions can be made in the shop floor, e.g., optimization of processes, and waste reduction, among others.

5.1. Main Assumptions

The cost analysis made in this case is focused on the manufacturing cost and on process variability. Further work can be done to extend it to the other dimensions of variability and the non-manufacturing costs. Thus, to calculate the cost of the product, these are the key inputs of the cost model, namely, the following:

- Quantities demanded by the client;
- Available time to produce the product in the line: (n° of days \times shifts per day \times minutes per shift \times 60);
- Workstations—the stations where the work associated with each process is carried out;
- Number of equipment per workstation and respective investment costs (i.e., depreciation);
- Cycle times per unit produced;
- Tariffs for the different resources used (e.g., area, energy, maintenance).

A general expectation is to have an overall equipment effectiveness (OEE) of 90% having in consideration the possible losses while producing. This efficiency of 90% multiplied by the time will lead the real time expected by the production line. The main resources are related to labor, depreciation, maintenance, auxiliary material, energy, area, other internal costs, tooling, etc. Considering the planned quantities and the budgeted costs, a specific tariff for each category of resources can be calculated. Summing all those tariffs, we can obtain the general tariff for the line. Table 8 below shows the general tariff for each line for the year 2021.

Table 8. General tariff (in EUR) per line.

Line	General Tariff
A	0.1178
B	0.1243
C	0.1470
D	0.1470

Tariffs are different if the resources used and/or available capacities are different. Lines C and D have used similar resources and offer identical capacity levels. Having the values of the tariffs, we can calculate the cost of the product in the different lines. To calculate the cost of the product, one must multiply the general tariff by the cycle time.

According to the statistical analysis performed and presented in the previous section, we can obtain the range for product cost considering process variability in each line. The first quartile of the cycle time is considered the lower range, and the third quartile is considered the upper range value, giving us an interval of the expected variation and allowing to estimate the risk of cost. The lower range helps with budgeting exercises, quotations, and the development of new products because it represents the potential lower costs of the product. The upper range gives an alert that margins can be compromised if the efficiency of the line is not improved. In this case, a conservative approach was followed, taking the values for Q1 and Q3; however, these limits could be calculated using the 10th and the 90th percentile.

5.2. Computation of the Costs

Table 9 presents the range for product costs considering the values for the first and third quartiles, which cover 50% of the values around the median, considering one week of analysis. It is important to note that in lines A and B, there are 18 parts that are produced in parallel, whereas in lines C and D, there are 36 parts.

Table 9. Range of product cost considering Q1 and Q3, per line.

Line	Tariff	Process Time (Seconds)		Parts in Parallel	Cost (EUR)	
		Lower Range	Upper Range		Lower Range	Upper Range
A	0.1178	820	979.75	18	5.366	6.411
B	0.1243	805	1040	18	5.558	7.181
C	0.1470	835	1015	36	3.409	4.144
D	0.1470	811	956	36	3.311	3.903

In Table 10, the range of the cost can be observed based on the mean value of process time with a confidence interval of 95%. By using this smaller amplitude, cost variability is reduced, and the results are more related to the standard efficiency of the process.

Table 10. Range of product cost considering the confidence interval for the mean, per line.

Line	Tariff	Process Time (Seconds)		Parts in Parallel	Cost (EUR)	
		Lower Range	Upper Range		Lower Range	Upper Range
A	0.1178	913.22	922.43	18	5.976	6.036
B	0.1243	941.86	954.92	18	6.504	6.594
C	0.1470	938.64	942.66	36	3.832	3.849
D	0.1470	894.61	897.77	36	3.652	3.665

It can be observed that the tariff for each line is different because the number of equipment is different, except for the big lines (C and D). In line B, there is more equipment than in line A but the planned quantities are almost the same. Thus, the amortization cost per product unit in line B increases. Hence, the parts produced in line B are costlier. Lines C and D are a replication of each other, so they have similar tariffs. The amount invested in the equipment in the big lines is bigger than in the small lines, but at the same time, the quantities produced in these lines are significantly higher. Therefore, the product produced in the big lines has a smaller cost despite having more equipment in each workstation. The

cycle time of the assembly line decreases with the increase in the amount of equipment in the workstations.

The methodology proposed here is related to some work done in the recent past. For example, Zanjani et al. [11] developed a stochastic model using the mean values of the uncertain parameters for the probability distribution. The developed model was applied in a milling industry with the purpose of supporting the production planning. Additionally, Sobu and Wu [35] developed stochastic scenarios by using observed mean-values and standard deviation from data, and then based on these stochastic scenario data, stochastic operation cost optimization models for minimizing operation cost were formulated. These models were used to measure uncertainty in power generation and renewable energy.

By using this methodology firstly, it is easy to understand how the assembly line is performing. It can be identified that the number of parts produced falls under the standard cycle time allocated for the production. With a stochastic approach, the range of the cost is available, which can help the manager to make decisions about the planning of the production, as this approach can facilitate the understanding about the real-time cost of the product along with the allocation of production quantities for each assembly line. It is important to note that, even though the assembly lines are replicates of each other, there may be some variability in them. Lines C and D, despite being exactly alike, have a significant difference between them.

5.3. Cost Analysis per Line and Product

For a better understanding of the variation in the data and analysis of their unpredictability, the mean values of cycle times corresponding to 12 weeks of three consecutive months were extracted and analyzed. Each of these weeks corresponds equally to a period of seven days.

For each of the lines, a confidence interval for the mean of 95% was calculated, as well as the first and third quartiles. To better understand the variation in the final costs per line, these were calculated according to the values presented in Table 8, that is, multiplying the cycle times obtained by the respective tariff.

Thus, the cost variation intervals were found when considering the confidence interval for the mean and the interquartile range. The values are shown in Figure 6.

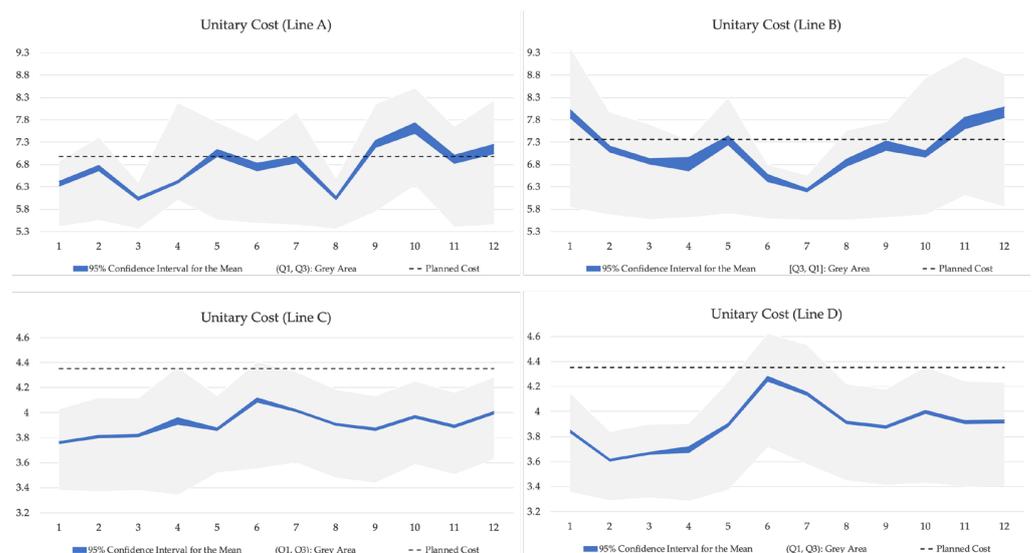


Figure 6. Stochastic cost analysis, per line. Values in euros for each of the 12 weeks.

Starting with lines A and B, we see that costs vary over time, above or below what was planned. Lines C and D tend to present their mean costs lower than planned. The upper limits (i.e., the 3rd quartile values) for the small lines are almost twice as high as for

the big lines. In the big lines, values vary between EUR 3.3 and 4.7, while in the small lines, they can reach maximum values of around EUR 9.4.

Furthermore, in most cases, the planned and expected cost values are above the value of Q3, that is, the planning presupposed obtaining a higher cost than the reality. This is not necessarily positive because it could represent an excessive pressure in the product development and quotation stages.

Considering all lines combined, the cost of the product does not exceed the planned cost, globally. The risk of higher costs given by the values related to the third quartile is not significant, and it is higher in the last weeks. Nevertheless, the average cost increases consistently over alternating weeks of increases and decreases in cost, as we can see in Figure 7.

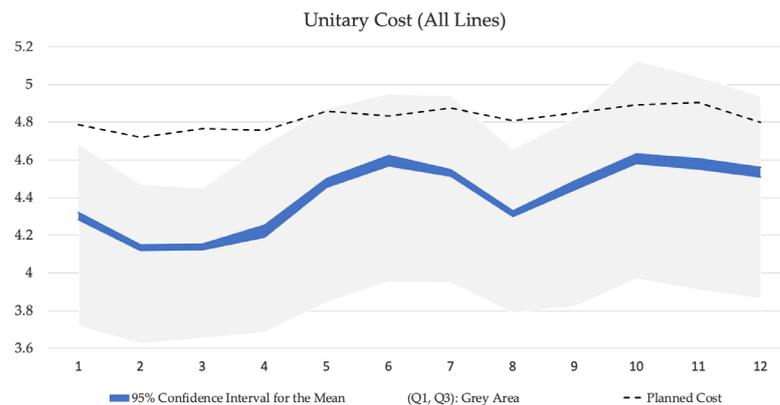


Figure 7. Stochastic product cost analysis. Values in euros for each of the 12 weeks.

The Wilcoxon Sign Rank Test was used for the validation and analysis of the proposed methodology and the computed costs. The test was used to assess whether the calculated values of the costs present significant variations over the weeks in relation to the planned/standard cost (Table 11).

Table 11. Results of the *t*-test for the comparison of planned and real costs.

	T-Value	Real versus Planned Cost
A	0.1671409760	3.9%
B	0.1298144462	−3.7%
C	0.0000023061	−10.0%
D	0.0000383380	−15.4%
All Lines	0.0000087905	−10.1%

Note: Significance level of 5%.

As we can see from the results obtained and presented in Table 11, lines A and B present test values of 0.167 and 0.130, respectively. We are led to conclude that these do not present significant differences between the planned value and the observed values. This denotes a greater tendency for the cost values, actual and planned, to come closer together. Notably, in absolute terms, the median values of the observed costs present a variation around 4% compared to the planned values.

As regards lines C and D, the test values obtained are 2.30613×10^{-6} and 3.8338×10^{-5} , respectively. In other words, there are significant differences between the values planned and those obtained. In real terms, this means that these lines are more sensitive to having cost values that are significantly different from the planned ones, due to different levels of productivity, planning efficiency and process variability. The median values are indeed different, with the actual values being 10 and 15% lower than planned.

The results obtained are consistent with the company’s situation that allowed the product cost to be lower than planned, namely, because big lines work very efficiently and significantly below the target cycle time used by the finance department to produce the

annual budget. Most of the production is scheduled for the big lines (around 15,000 parts per week, compared to 3000 parts per week in small lines). In addition, the number of equipment is higher in big lines, so more parts are produced in parallel. Thus, the cost of the product tends to be much lower than the standard cost defined by the finance department.

In general, the Wilcoxon test shows that the average real costs, considering all lines, tend to be lower than the planned ones. On the other hand, this difference is significant for big lines (lines C and D). Small lines (lines A and B) do not present significant differences between planned and observed values. However, given the influence of the big lines on total production, performing the same test and under the same conditions, we observe that there are significant differences between the planned product costs and actual observed costs. In this case, we obtained a test value of 8.79051×10^{-6} with actual median values 10% lower than planned, considering all production lines.

We are led to infer that big lines considerably influence the variation of the final cost of the product and that the explanation for real costs being lower than planned lies in the operating conditions of these lines. The Wilcoxon test allows us to deduce that the median values, in total, are very different (47,341.42341 and 42,580.85791, comparing planned and observed values, respectively), and that the actual average values are about 10% lower than the planned value.

The Wilcoxon test reinforces the scientific validity of the significance of the cost variation, if any, and furthermore, shows us that the proposed methodology is able to present and describe that same variation and its effects on the product cost.

In this analysis, the workstation that represents the bottleneck, per line, was considered for the computation of the cycle time, allowing a view of the minimum time required to produce an article in the production line. Further work can be developed to support a much more detailed analysis, considering all the stations that compose the line and, consequently, the respective specific costs associated with each workstation. With the methodology adopted here, it will be interesting to notice the variation, along the same lines, of the different costs by workstation and by line and its consequent variations over time. Moreover, an intensive outliers' analysis must be performed since it increases the variability in the process, and it is important to understand these occurrences to reduce them.

By following this methodology, it is possible to know the real-time cost of the product, which can facilitate controlling the cost of the product in a timely manner. The manager can better decide the allocation of quantities to each line, as each line can provide different margins of profit based on the quantities produced and the variability of the process time in each assembly line. Investment in new equipment can be also verified by this approach, regarding whether it will be profitable or not. Thus, investment appraisal exercises will also benefit from the use of a stochastic approach in product cost calculations.

5.4. Final Remarks

The proposed methodology was applied in a real context, and the main remark to take into consideration is that the presence of outliers can lead to wrong misperceptions, that is, in Table 1, it was not clear that there are differences between lines A and B, considering the mean values. Therefore, when the outliers were removed, this was more perceptible (Table 4) and, in terms of variability, the range was halved. Note that the removal of the outliers was done for those caused by external factors. Despite the hypothesis test having the same conclusions with (Table 2) and without outliers (Table 5), the range between the lower and upper bound also decreased considerably. This means that part of the variability was removed.

The lines use different resources and have also different capacity levels. Small line B has the highest costs and, incidentally, the highest fare. The big lines, C and D, have very similar costs and the same fare. Furthermore, it is also expected that costs have their own variability, which was not studied here. The combination of cycle times variability,

costs variability and demand or planning variability will make the model too complex. However, all of these variabilities should be taken into consideration.

Statistics, namely, the study of averages and respective variations in values by lines, allows us to have a broad view of the production time and, consequently, respective costs. It is possible to verify that, depending on the type of line, these values differ, allowing inferences about different trends and variation intervals for the average cycle times and cost.

With the presentation of the confidence intervals for the mean, it is possible to obtain a notion of the expected variation, in terms of production times, helping to forecast costs. Comparing the actual and planned results, we confirm the existence of cost variability, which, in some cases, may be above the expected value, and in others, below it. In other words, the clarity of the uncertainty in cost forecasting is expressed. With a confidence interval of 95%, it is possible to predict that the big lines have a greater tendency to have lower than expected mean cost values, compared to small lines.

The study presented also allowed us to determine the importance of studying outlier elements, as when these are extracted, we are led to a clearer analysis, closer to what we may consider common (without major variations and differences in values). In other words, it became easier to see that the non-consideration of aberrant elements helps in predicting results within something that can be considered as expected.

With the approach presented here, an important contribution is made to estimate something that is uncertain and that can vary greatly over time. Even though the cycle time considered only refers to the bottleneck station, the results show the existence of divergence in costs in the two types of lines.

Despite the results obtained, it should be noted that the cycle time considered was related only to the bottleneck station. In other words, although other workstations operate simultaneously, the specific production time of those workstations was not considered. For further work, it may be considered the times of all workstations that make up the line to understand how this influences the final cost. In addition, it will be important to extend the study to the relationship between cycle times, if any can be found. It opens doors to the analysis of the average cycle times and respective variations of each workstation in each of the lines to estimate and predict the respective costs with high confidence.

In addition, outliers can be analyzed carefully to understand which and what types of lines are more sensitive to large variations in time cycles, that is, if lines with differing cycle times lead to different costs and/or large cost variations.

The main obstacles and difficulties faced in the implementation of this methodology were related to the access and integration of financial and production information. This process takes some time, as data must be extensively collected, and various tests must be performed on it. The presentation and visualization of the results in an automated and simplified manner must also be improved, which will contribute to the routinization and institutionalization of the entire process. Business intelligence and analytics tools are particularly useful in this context. Company's managers are experienced with such tools (e.g., Tableau software) but there is still needed a better integration among databases, reporting models, routines, and procedures.

6. Conclusions

Nowadays, traditional costing systems based on deterministic cost models are not enough to adequately support decision making. As it was identified in previous studies, the mean and standard deviation are widely used to analyze the product cost. However, making estimations using the mean, in some cases, can lead to wrong conclusions [11] since the estimation can be far from the true value. Thus, the confidence interval can be used instead of the mean. Other alternatives are to consider the quartiles or the coefficient of variance.

This paper proposes a stochastic approach to costing systems, which considers the variability in the process cycle time. This approach provides a better perception of the risk associated with product costs. The confidence interval for the mean and the use of quartiles

1 and 3 as lower and upper estimates are proposed to include variability and risk in the costing systems.

The developed six-step methodology was applied in a tier 1 manufacturer of the automotive industry. Only the bottleneck workstation of the assembly line was considered since it represents the cycle time of the line, but the analysis can be extended to all workstations in the production line. The allocation of the cost to line can be enough for product costing, to compute margins, and compare costs and revenues supporting the profitability analysis. Nevertheless, the optimization of production processes demands a deeper analysis, where costs should be highlighted by activities, workstations or individual machines.

With the work carried out here, doors are opened for future analysis of cost variations over time with an awareness of the associated risks. The use of descriptive statistics gives the ability to understand and evaluate the behavior of cycle times and their influence on costs. In particular, the study of a confidence interval for the mean and the interquartile range gives us insights into what we can expect, and the risk of obtaining values that are much higher or lower than what was predicted. Variability can be also studied in the demand and production quantities and in the tariffs that allocate resource costs to the activities and processes.

The proposed methodology can be extended to developing stochastic cost models based on activities or other more sophisticated costing systems. Indeed, by using an activity-based costing approach, for example, every activity can be observed, and thus, the variability for each activity can be recorded, and more accurate product costs can be obtained. A dynamic model for cost analysis, cost estimations and controlling dashboards based on stochastic assumptions can also be developed to understand and observe the cost from a more complete perspective.

From a managerial perspective, it is very advantageous, as these methods provide information in real time of the cost of the product, including the inherent variability. The pace of production can be understood, and the possible bottleneck of the assembly line can be taken into analysis, as it may vary from week to week. Additionally, this will help to reduce the gap between financial and production departments and bring them on common grounds. One of the main difficulties faced by managers is related to the collection of the data and performing the steps mentioned, as these steps may take some time.

Author Contributions: Conceptualization, P.A.; methodology, A.A.; validation, S.S.; formal analysis, V.V., A.A., S.S.; investigation, V.V., A.A., S.S.; data curation, V.V., A.A., S.S.; writing—original draft preparation, V.V., A.A.; writing—review and editing, P.A., B.P.J.B.; project administration, B.P.J.B.; funding acquisition, P.A., B.P.J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by: European Structural and Investment Funds in the FEDER component, through the Operational Competitiveness and Internationalization Programme (COMPETE 2020) [Project n° 39479; Funding Reference: POCI-01-0247-FEDER-39479].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaplan, R.S.; Robin, C. How Cost Accounting Systematically Distorts Product Costs. In *Accounting and Management: Field Study Perspectives*; Bruns, W.J., Jr., Kaplan, R.S., Eds.; Harvard Business School Press: Boston, MA, USA, 1987; Chapter 8, pp. 204–228.
2. Cooper, R.; Kaplan, R. Measure costs right: Make the right decisions. *Harv. Bus. Rev.* **1988**, *66*, 96–103.
3. Pember, A.; Lemon, M. *Measuring and Managing Environmental Sustainability: Using Activity-Based Costing/Management (ABC/M)*; The Consortium for Advanced Management–International: Austin, TX, USA, 2012.
4. Drury, C.; Tayles, M. Product costing in uk manufacturing organizations. *Eur. Acc. Rev.* **1994**, *3*, 443–470. [[CrossRef](#)]
5. Kaplan, R.S. The Evolution of Management Accounting. *Acc. Rev.* **1984**, *59*, 390–418. Available online: www.jstor.org/stable/246701 (accessed on 28 May 2021).

6. Alessandrini, S.; Dell'Isola, F.; Porfiri, M. A revival of electric analogs for vibrating mechanical systems aimed to their efficient control by PZT actuators. *Int. J. Solids Struct.* **2002**, *39*, 5295–5324. [[CrossRef](#)]
7. Goh, Y.M.; Newnes, L.B.; Mileham, A.R.; McMahon, C.A.; Saravi, M.E. Uncertainty in through-life costing-review and perspectives. *IEEE Trans. Eng. Manag.* **2010**, *57*, 689–701. [[CrossRef](#)]
8. Orellana, R.; Carvajal, R.; Escárate, P.; Agüero, J.C. On the uncertainty identification for linear dynamic systems using stochastic embedding approach with gaussian mixture models. *Sensors* **2021**, *21*, 3837. [[CrossRef](#)]
9. Wang, C.N.; Nhieu, N.L.; Tran, T.T.T. Stochastic chebyshev goal programming mixed integer linear model for sustainable global production planning. *Mathematics* **2021**, *9*, 483. [[CrossRef](#)]
10. Zhang, X.; Prajapati, M.; Peden, E. A stochastic production planning model under uncertain seasonal demand and market growth. *Int. J. Prod. Res.* **2011**, *49*, 1957–1975. [[CrossRef](#)]
11. Zanjani, M.K.; Ait-Kadi, D.; Nourelfath, M. A stochastic programming approach for sawmill production planning. *Int. J. Math. Oper. Res.* **2013**, *5*. [[CrossRef](#)]
12. Abounadi, J.; Bertsekas, D.; Borkar, V.S. Learning algorithms for Markov decision processes with average cost. *SIAM J. Control Optim.* **2002**, *40*, 681–698. [[CrossRef](#)]
13. Oh, S.C.; Shin, J. The assessment of car making plants with an integrated stochastic frontier analysis model. *Mathematics* **2021**, *9*, 1296. [[CrossRef](#)]
14. Singh, S.R.; Prasher, L. A production inventory model with flexible manufacturing, random machine breakdown and stochastic repair time. *Int. J. Ind. Eng. Comput.* **2014**, *5*, 575–588. [[CrossRef](#)]
15. Noh, J.S.; Kim, J.S.; Sarkar, B. Stochastic joint replenishment problem with quantity discounts and minimum order constraints. *Oper. Res.* **2019**, *19*, 151–178. [[CrossRef](#)]
16. Singh, S.R.; Singh, C. Supply chain model with stochastic lead time under imprecise partially backlogging and fuzzy ramp-type demand for expiring items. *Int. J. Oper. Res.* **2010**, *8*, 511–522. [[CrossRef](#)]
17. Singh, C.; Singh, S.R. Supply chain model for expiring items following ramp-type demand with stochastic lead time under crisp and fuzzy environment. *Int. J. Fuzzy Syst. Appl.* **2020**, *9*, 64–91. [[CrossRef](#)]
18. Oberkampf, W.L.; DeLand, S.M.; Rutherford, B.M.; Diegert, K.V.; Alvin, K.F. New methodology for the estimation of total uncertainty in computational simulation. In Proceedings of the 40th Structures, Structural Dynamics, and Materials Conference and Exhibit, St. Louis, MO, USA, 15 April 1999. [[CrossRef](#)]
19. Scanlan, J.; Rao, A.; Bru, C.; Hale, P.; Marsh, R. DATUM project: Cost estimating environment for support of aerospace design decision making. *J. Aircr.* **2006**, *43*, 1022–1028. [[CrossRef](#)]
20. Corbett, J.; Crookall, P.R. Design for Economic Manufacture. *CIRP Ann. Manuf. Technol.* **1986**, *35*, 93–97. [[CrossRef](#)]
21. Zimmermann, H.J. Application-oriented view of modeling uncertainty. *Eur. J. Oper. Res.* **2000**, *122*, 190–198. [[CrossRef](#)]
22. Nachtmann, H.; Needy, K.L. Methods for handling uncertainty in activity based costing systems. *Eng. Econ.* **2003**, *48*, 259–282. [[CrossRef](#)]
23. Miller, E.M. Risk, uncertainty, and divergence of opinion. *J. Financ.* **1977**. [[CrossRef](#)]
24. Datta, P.P.; Roy, R. Cost modelling techniques for availability type service support contracts: A literature review and empirical study. *CIRP J. Manuf. Sci. Technol.* **2010**, *3*, 142–157. [[CrossRef](#)]
25. Afonso, P.; Jiménez, V.J. Costing Systems for Decision Making under Uncertainty Using Probabilistic Models. In *Analyzing Risk through Probabilistic Modeling in Operations Research*; IGI Global: Pennsylvania, PA, USA, 2016. [[CrossRef](#)]
26. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**. [[CrossRef](#)]
27. Yuan, F.C. The use of a fuzzy logic-based system in cost-volume-profit analysis under uncertainty. *Expert Syst. Appl.* **2009**, *36*, 1155–1163. [[CrossRef](#)]
28. Byrne, P. Fuzzy DCF: A contradiction in terms, or a way to better investment appraisal. In *Cutting Edge '97, RICS*; RICS: London, UK, 1997.
29. Kishk, M.; Al-Hajj, A. A Fuzzy Model And Algorithm to Handle Subjectivity In Life Cycle Costing Based Decision-Making. *Archit. Built Environ. J. Financ. Manag. Prop. Constr.* **2000**, *5*, 93–104.
30. Durán, O.; Durán, P.A. Activity Based Costing for wastewater treatment and reuse under uncertainty: A fuzzy approach. *Sustainability* **2018**, *10*, 2260. [[CrossRef](#)]
31. Makridakis, S.G.; Fildes, R.; Hibon, M.; Parzen, E. The Forecasting Accuracy of Major Time Series Methods. *Statistician* **1985**, *34*, 261–262. [[CrossRef](#)]
32. Ioannou, A.; Angus, A.; Brennan, F. Stochastic Prediction of Offshore Wind Farm LCOE through an Integrated Cost Model. *Energy Procedia* **2017**, *107*, 383–389. [[CrossRef](#)]
33. Mora, E.B.; Spelling, J.; van der Weijde, A.H. Global sensitivity analysis for offshore wind cost modelling. *Wind Energy* **2021**. [[CrossRef](#)]
34. Swaroop, P.V.; Erlich, I.; Rohrig, K.; Dobschinski, J. A stochastic model for the optimal operation of a wind-thermal power system. *IEEE Trans. Power Syst.* **2009**, *24*, 940–950. [[CrossRef](#)]
35. Sobu, A.; Wu, G. Optimal operation planning method for isolated micro grid considering uncertainties of renewable power generations and load demand. In Proceedings of the IEEE PES Innovative Smart Grid Technologies, Tianjin, China, 21–24 May 2012. [[CrossRef](#)]

36. Baldoni, E.; Coderoni, S.; D'Orazio, M.; di Giuseppe, E.; Esposti, R. The role of economic and policy variables in energy-efficient retrofitting assessment. A stochastic Life Cycle Costing methodology. *Energy Policy* **2019**, *129*, 1207–1219. [[CrossRef](#)]
37. Baldoni, M.; Bergenti, F.; Seghrouchni, A.e.; Winikoff, M. Special issue on Current trends in research on software agents and agent-based software systems. *Auton. Agents Multi Agent Syst.* **2021**, *35*. [[CrossRef](#)]
38. Park, C.; Park, P.K.; Mane, P.P.; Hyung, H.; Gandhi, V.; Kim, S.H.; Kim, J.H. Stochastic cost estimation approach for full-scale reverse osmosis desalination plants. *J. Memb. Sci.* **2010**, *364*, 52–64. [[CrossRef](#)]
39. Meister, M.; Beßle, J.; Cvikov, A.; Böing, T.; Metternich, J. Manufacturing analytics for problem-solving processes in production. *Procedia CIRP* **2019**, *81*, 1–6. [[CrossRef](#)]
40. Montgomery, D.C.; Runger, G.C. *Applied Statistics and Probability for Engineers*; Wiley: Hoboken, NJ, USA, 2018.
41. Okoye, P.V.C.; Egbunike, F.C.; Meduoye, O.M. Product Cost Management via the Kaizen Costing System: Perception of Accountants. *J. Manag. Sustain.* **2013**, *3*, 114. [[CrossRef](#)]
42. Alsoboa, S.; Al-Ghazzawi, A.; Joudeh, A. The Impact of Strategic Costing Techniques on the Performance of Jordanian Listed Manufacturing Companies. *Res. J. Financ. Acc.* **2015**, *6*, 116–127.
43. Sobczyk, T.; Koch, T. A method for measuring operational and financial performance of a production value stream. *IFIP Int. Fed. Inf. Process.* **2008**, *257*, 151–163. [[CrossRef](#)]
44. Chen, Y.H.; Ho, Y.L.; Huang, H.C.; Wu, H.W.; Lee, C.Y.; Hsu, T.P.; Cheng, C.L.; Chen, M.F. Assessment of the clinical outcomes and cost-effectiveness of the management of systolic heart failure in Chinese patients using a home-based intervention. *J. Int. Med. Res.* **2010**, *38*, 242–252. [[CrossRef](#)] [[PubMed](#)]
45. Hillier, F.S.; So, K.C. The effect of the coefficient of variation of operation times on the allocation of storage space in production line systems. *IIE Trans. Inst. Ind. Eng.* **1991**, *23*, 198–206. [[CrossRef](#)]
46. McKinney, W. *Pandas: Powerful Python Data Analysis Toolkit—Pandas Documentation*. 2020. Available online: <https://pandas.pydata.org/pandas-docs/version/0.25.3/> (accessed on 4 March 2021).
47. Hunter, J.; Dale, D.; Firing, E.; Droettboom, M. Matplotlib: Visualization with Python. *Abgerufen*. 2020. Available online: <https://matplotlib.org> (accessed on 4 March 2021).
48. Bonald, T.; de Lara, N.; Lutz, Q.; Charpentier, B. Scikit-network: Graph analysis in python. *J. Mach. Learn. Res.* **2020**, *21*, 185:1–185:6.
49. Kaur, A.; Kumar, R. Comparative Analysis of Parametric and Non-Parametric Tests. *J. Comput. Math. Sci.* **2015**, *6*, 336–342.
50. Yazici, B.; Yolacan, S. A comparison of various tests of normality. *J. Stat. Comput. Simul.* **2007**, *77*. [[CrossRef](#)]