Tests considering literature benchmarks

Before generating our benchmark, we tried to use or adapt instances from literature and previous works. However, as will be explained in detail below, we failed, as most instances do not consider infeasibilities related to the production capacity, i.e., situations when the makespan found by the scheduling problem is greater than the capacity. Literature instances were feasible for most cases, contrary to our article's objective.

In addition to the lack of infeasible instances, the literature benchmarks generally consider only one type of distribution for data generation. For example, James & Almada-Lobo (2011) generate data based on uniform distributions. In our paper, the benchmark also considers Normal and Poisson distributions aiming to simulate several possible cases. Further, we create three different scenarios. In the first one, the setup time is higher than the production time. In the second, the processing time is higher than the setup time, while in the third scenario, the processing and setup times are generated with the same distribution.

To indicate that the literature benchmarks are incompatible to our work, we adapt the benchmark of James & Almada-Lobo (2011) to the three scenarios considered in our paper. The authors study single and parallel machine capacitated lot-sizing and scheduling, considering instances with and without capacity variations. They use this capacity variation to deal with cases where the capacity may not meet the demand, thus replacing backorder strategies. In our paper, we already consider backorder as an alternative. Then, we only test instances with no capacity variation. We emphasize that, in the work of James & Almada-Lobo (2011), the objective function aims to minimize the sum of setup and inventory costs. Then, demand unmet may occur for different reasons that in our work.

In their work, the scheduling deals with sequence-dependent setup times and costs. To adapt their instances to our study, we exclude the setup costs generated by the authors and add backorder costs following the distribution $B_j \sim U(10, 13)$. We define this interval considering the distribution used by James & Almada-Lobo (2011) to generate inventory costs, $H_j \sim U(2, 9)$. The authors consider 15 and 25 production quantities and planning horizons of 5, 10, and 15 time periods. Instances vary from 0 to 9. They define the capacity based on a parameter named Cut, which indicates the capacity utilization. Parameter Cut is set to 0.6 and 0.8. Then, we test 120 instances.

In https://github.com/fernandafalves/Paper_Lotsizing_Scheduling, we provide the files related to the new tests. The code was named "Test_SingleMachineNoCapVar.txt". The modified data is in the folder "SingleMachineNoCapVar". Table 1 shows a summary of the results obtained with all the tests. For each experiment, we present the number of infeasible instances.

When solving the instances of James & Almada-Lobo (2011) with the modification of setup costs, we obtained the results presented in the file "ResultIU_TestI.txt". As we can observe in Column 1 of Table 1 (Test I), no instance was infeasible related to production capacity.

In the first scenario tested, the setup times were greater than the production times, see the work of James & Almada-Lobo (2011). We perform a second test considering a scenario in which the processing times are greater than the setup times. First, we change the processing time of the data of James & Almada-Lobo (2011) to a value of 12-time units for all the products. We show the results in "ResultIU_TestII.txt" (Column 2 of Table 1, Test II). As we can see, 44 out of 120 instances resulted in infeasibility. However, one of our work objectives is to generate the parameters considering its dependency with the product quantities to enable a fair comparison between several cases, e.g., to compare the strategies fairly even when considering 4 or 100 products. In this case, the objective functions for four products are of the same order of magnitude than the instances with 100 products. When we observe the results using the benchmark of James & Almada-Lobo (2011), the objective function goes from a minimum of 108054.00 to a maximum of 1639411.00.

Then, the benchmark of James & Almada-Lobo (2011) does not suit the objective of comparing several instances sets with the same order of magnitude for the objective function.

Lastly, we consider the third scenario when processing and setup times present the same source distribution. In this case, we set these parameters to one unit. We did not find any infeasible solution ("ResultIU_TestIII.txt", Column 3 of Table 1, Test III).

We conclude that we can't use the benchmark of James & Almada-Lobo (2011) to achieve our paper's objectives, i.e., to consider instances with infeasibilities and presenting objective functions with the same order of magnitude when considering different product quantities. Thus we have created our instances that are made available for future comparisons.

Table 1: Number of infeasible instances.

Test I ($setup > p$)	Test II ($p > setup$)	Test III ($setup = p$)
0	44	0

References

James, R. J., & Almada-Lobo, B. (2011). Single and parallel machine capacitated lotsizing and scheduling: New iterative mip-based neighborhood search heuristics. *Computers & Operations Research*, *38*(12), 1816–1825.