PI Game

A comparison of the parametric and non-parametric model in the NVP with holding costs

Group 13

BDF *

June 2021

Abstract

The News Vendor Problem optimizes the order quantity using the price, cost and demand distribution, however the cost of discarding products was not yet included. We have changed the News Vendor Problem to include these holding costs. To estimate the optimal order quantity, a parametric or non-parametric model can be used. Our simulations suggest that the parametric models estimations are generally of a higher quality, however this may not be true when an incorrect demand distribution is assumed. When analyzing the data from a European bakery chain, we found that the data appears to be normally distributed in weekends and log-normally distributed during the workweek. This led us to estimate the optimal order quantity using the parametric and non-parametric model.

Keywords: News Vendor Problem; probability; parametric and non-parametric comparison; bakery optimization.

1 Introduction

In this project we are going to take a look at the News Vendor Problem (NVP) including holding cost, it is a typical problem in Operations Research. The vendor does not know how much he will sell throughout the day and determining the optimal amount of stock is difficult due to the fact that the demand is uncertain. By solving the NVP we can not only maximize profit, we can also reduce excess waste.

The NVP has two models, the parametric model and the non-parametric model. When using the parametric model you assume a demand distribution, where as when using the non-parametric model you do not have such a requirement. We are interested in which of the two models estimations are of higher quality. This is important, because it allows us to make the most accurate estimations possible, and thus maximize profit. We will compare these models using Monte Carlo simulations.

We have also received data from a bakery chain. In our quest to minimize excess waste, and primarily maximize profit, we are looking at the data and determining the optimal order quantity. Calculating all this data manually is almost impossible, hence we will use Python to easily find descriptive statistics, time series, and histograms. These descriptive statistics will be used to estimate the demand distribution. The estimation of the demand distribution is used to estimate the optimal order quantity and its confidence interval using the parametric model. Then using the non-parametric model the optimal order quantity and its confidence interval will also estimated.

*Micha den Heijer (2690717), Maurits Moers (2706968), Cemre Suler (2693488), Sean Lee (2713026), and Olaf Oostenbrug (2711573). BDF is an acronym for "Beter Dan Freek", which translates to Better Than Freek (Byrman), our classmate whose group we consider to be our biggest competitor.

2 Theoretical questions

2.1 Problem Description

In the theoretical part of this research project we are looking into the advantages and disadvantages of the parametric (Lin, 2021d) and non-parametric (Lin, 2021c) model of the NVP, and into holding costs.

We will first look at the theoretical side of the NVP. The classical NVP only depends on the price (p), cost (c) and true demand distribution ($F_Y(\cdot)$) to estimate optimal order quantity (Lin, 2021a). However in reality unsold goods can lead to a holding cost (c_h, i.e. cost to destroy unsold goods) per unit, which would affect the NVP model. Therefore the profit function will change to include the additional holding cost. We are interested in the question: "How does the optimal order quantity and its expected profit react when including holding costs?".

The parametric and non-parametric models for NVP would change accordingly. In a parametric model, we assume that demand is distributed by a known continuous distribution. When using the non-parametric model we do not assume any distribution. This may be advantageous, because then the demand distribution does not have to be estimated. We are comparing the accuracy and bias of the estimated optimal order quantity, and its impact on the expected profit, by simulating demand and comparing it for different target service levels (τ). Therefore we are interested in the research question: "How do the parametric and non-parametric models of the NVP compare at different target service levels?".

In the previous experiment we assumed that the distribution of the demand is known, however this may not be true. Previous literature (Levi, Perakis, and Uichanco, 2015) has suggested that when the true demand is non-normally distributed, but the data is fit to a normal distribution, the estimation could result in a sub-optimal order quantity. Therefore we are going to simulate non-normal distributions and try to estimate the optimal order quantity, to answer the research question: "How does the parametric model deal with assuming a normal distribution, when the true demand distribution is a non-normal distribution?".

2.2 Methodology

2.2.1 Holding cost

To add the holding cost into the News Vendor Problem, the cost (c) and price (p), can be rewritten to include the holding cost (c_h).

$$\tilde{c} = c + c_h$$
$$\tilde{p} = p - c$$

In these equations $\tilde{p}$ refers to the profit per product sold, and $\tilde{c}$ refers to the cost per unsold product. This results in the profit function:

$$\Pi\left(Q, Y; \tilde{c}, \tilde{p}\right) = \tilde{p}\, \min\{Q,\, Y\} - \tilde{c}\, \max\{Q - Y,\, 0\} \tag{1}$$

To find the optimal order quantity we need to solve (2). This function assumes that the optimal order quantity exists and is unique.

$$Q^*(F_Y; \tilde{c}, \tilde{p}) = \underset{Q \geq 0}{\arg\max}\ \mathbb{E}_Y\left[\Pi(Q,\, Y;\, \tilde{c},\, \tilde{p})\right] \tag{2}$$

The expected value of the profit function (1) can be obtained:

$$\begin{aligned}
\mathbb{E}_Y\left[\Pi(Q,\,Y;\,\tilde{c},\,\tilde{p})\right] &= \int_{\mathbb{R}} \tilde{p}\,\min\{Q,\,Y\} - \tilde{c}\,\max\{Q-Y,\,0\}\,dF_Y(y) \\
&= \tilde{p}\int_{\mathbb{R}} \min\{Q,\,Y\}\,dF_Y(y) - \tilde{c}\int_{\mathbb{R}} \max\{Q-Y,\,0\}\,dF_Y(y) \\
&= \tilde{p}\left[\int_{-\infty}^{Q} y\,dF_Y(y) + Q\int_{Q}^{\infty} 1\,dF_Y(y)\right] \\
&\quad - \tilde{c}\left[Q\int_{-\infty}^{Q} 1\,dF_Y(y) + \int_{Q}^{\infty} y\,dF_Y(y) - \int_{-\infty}^{\infty} y\,dF_Y(y)\right] \\
&= \tilde{p}\left[Q - \int_{-\infty}^{Q} F_Y(y)\,dy\right] - \tilde{c}\left[\int_{-\infty}^{Q} F_Y(y)\,dy\right] \\
&= \tilde{p}\,Q - (\tilde{p}+\tilde{c})\int_{-\infty}^{Q} F_Y(y)\,dy
\end{aligned} \tag{3}$$

Because F_Y is invertible, and by solving the first-order condition,

$$\tilde{p} - (\tilde{p}+\tilde{c})\,F_Y(Q^*) = 0$$

we obtain

$$Q^*(F_Y;\tilde{c},\tilde{p}) = F_Y^{\leftarrow}\left(\frac{\tilde{p}}{\tilde{p}+\tilde{c}}\right) = F_Y^{\leftarrow}\left(\frac{p-c}{p+c_h}\right). \tag{4}$$

We assume demand from a normal distribution with $\mu = 100$ and $\sigma = 10$, cost $c = 1$, and price $p = 1.5$. We calculate the optimal order quantity (Q^*) for $c_h \in [-1, 5]$ and graph it into a sensitivity plot. We also calculate the expected profit (3) for $c_h \in [-1, 5]$ and graph it into a sensitivity graph. For both graphs we set $c \in \{0.9, 1.0, 1.1\}$, and fix $p = 1.5$.

2.2.2 Parametric Model vs. Non-Parametric Model

To test the accuracy of the parametric and non-parametric estimations at different target service levels (τ), we create a Monte Carlo simulation. Because in this simulation we are interested in the accuracy at a certain target service level, we are not interested in the specific price, cost or holding cost. Therefore we set $c = 1 - \tau$, $c_h = 0$, and $p = 1$. This results in $\frac{p-c}{p+c_h} = \tau$. Thus we can define $Q^*(\tau) := Q^*(F_Y; c, p) = F_Y^{\leftarrow}(\tau)$.

For the Monte Carlo simulation we generate $n \in \{10, 50, 100, 200\}$ (random) demand from normal and log-normal distributions, with $\mu \in \{50, 100, 200, 500\}$, and $\sigma \in \{1, 5, 10, 20\}$ for the normal distribution and $\mu \in \{4, 5, 6, 7, 8\}$ and $\sigma \in \{0.2, 0.5, 0.8, 1\}$ for the log-normal distribution.

Then we try to estimate the optimal quantity ($\widehat{Q}_n^k(\tau)$, with $k \in \{\mathrm{P}, \mathrm{NP}\}$) using the parametric (Lin, 2021d) and non-parametric model (Lin, 2021c) at target service levels $\tau \in \{0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99\}$. This is repeated $M = 1000$ times. The quality of the estimations is compared by the Root Mean Squared Error (RMSE, 5), the Profit Loss Ratio (PLR, 6), and the Mean Percentage Error (MPE, 7).

$$\mathrm{RMSE}_n^k(\tau) = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left[Q_n^{k,j}(\tau) - Q^*(\tau)\right]^2} \tag{5}$$

$$R(\widehat{Q}_n^{k,j};\ \tau) := \mathbb{E}_Y\left[\Pi(Q,\,Y;\,c = 1-\tau,\,p = 1)\right]$$

$$\mathrm{PLR}_n^k(\tau) = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{R(Q^*;\tau) - R(\hat{Q}_n^{k,j};\tau)}{R(Q^*;\tau)}\right| \tag{6}$$

$$\text{MPE}_n^k(\tau) = \frac{100\%}{M} \sum_{j=1}^{M} \frac{Q^*(\tau) - \widehat{Q}_n^{k,j}(\tau)}{Q^*(\tau)} \tag{7}$$

The RMSE (5) clearly describes the error of the estimated optimal quantity, the PLR (6) describes the decrease in expected profit because of the error in the estimation and the MPE (7) describes the bias in the estimation of the optimal quantity ($\widehat{Q}^*(\tau)$).

This experiment will generate a lot of data, by dividing the RMSE or the PLR of the the non-parametric model, by the same value of the parametric model, you can more clearly see which model estimates at a higher quality.

2.2.3 Wrong Distribution Parametric Model

To test how the parametric model reacts when choosing the wrong distribution we simulate a log-normal distribution with $\mu \in \{5, 10\}$ and $\sigma \in \{0.5, 1\}$ and make an estimation of the optimal order quantity ($\widehat{Q}^*$) using the assumption that the data is normally distributed. We simulated this for $\tau \in \{0.01, 0.5, 0.99\}$. This is done using the parametric model as defined in Lin, 2021d. To assess the quality of the estimation we use the RMSE (5), PLR (6) and MPE (7).

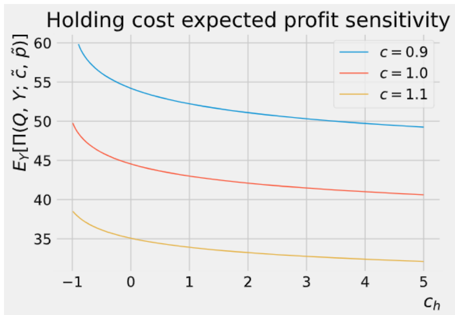
2.3 Results

2.3.1 Holding cost sensitivity

The sensitivity of the holding cost on the optimal quantity is displayed in Figure 1a. In this plot on the x-axis the holding cost is displayed, and on the y-axis the optimal order quantity is displayed. There are 3 lines, which each display their own cost. Note that $c = 0.9$ is undefined from $c_h \leq -0.9$.

The expected profit of the optimal quantity is displayed in Figure 1b. In this plot on the x-axis the expected profit is displayed, and on the y-axis the holding cost is displayed.

(a) Figure 1a: Sensitivity graph of Q^* by changing c_h.

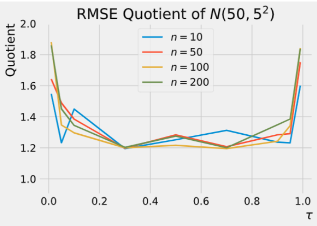


(b) Figure 1b: Sensitivity graph of expected profit by changing c_h.

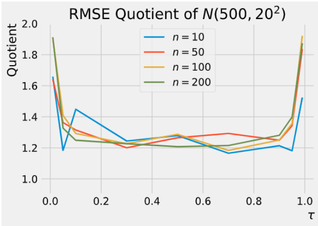
2.3.2 Parametric vs. Non-Parametric Quality

From our simulation we got too much data to include in this document. Therefore we decided to include all relevant data, as most of the data tells a similar story. All the other graphs can still be viewed in the Jupyter Notebook.

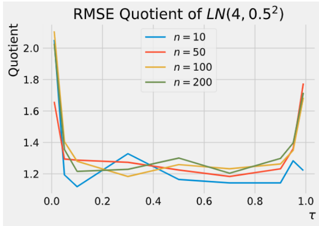
Root Mean Squared Error Quotient



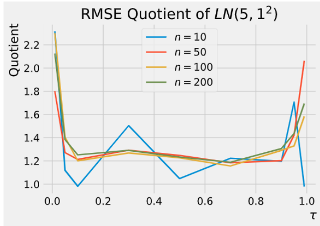
(a) Figure 2a: Normally distributed with parameters: $\mu = 50$; $\sigma = 5$



(b) Figure 2b: Normally distributed with parameters: $\mu = 500$; $\sigma = 20$

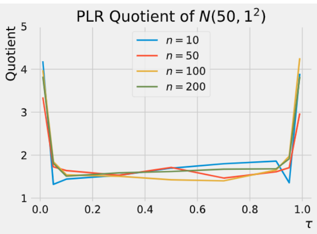


(c) Figure 2c: Log-normally distributed with parameters: $\mu = 4$; $\sigma = 0.5$

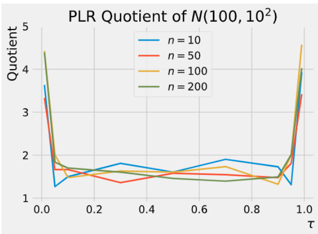


(d) Figure 2d: Log-normally distributed with parameters: $\mu = 5$; $\sigma = 1$

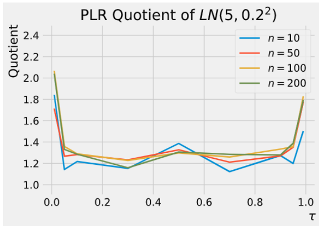
Probability Loss Ratio Quotient



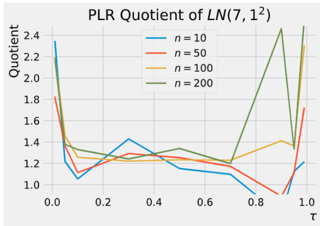
(a) Figure 2e: Normally distributed with parameters: $\mu = 50$; $\sigma = 1$



(b) Figure 2f: Normally distributed with parameters: $\mu = 100$; $\sigma = 10$

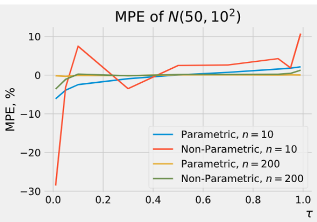


(c) Figure 2g: Log-normally distributed with parameters: $\mu = 5$; $\sigma = 0.2$

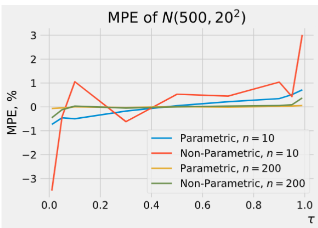


(d) Figure 2h: Log-normally distributed with parameters: $\mu = 7$; $\sigma = 1$

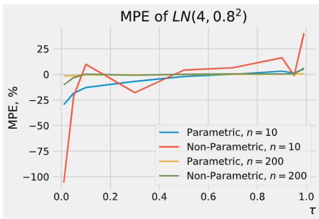
Mean Percentage Error



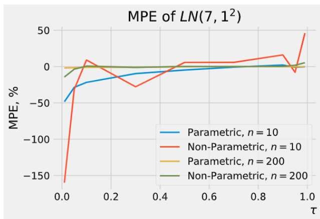
(e) Figure 2i: Normally distributed with parameters: $\mu = 50$; $\sigma = 10$



(f) Figure 2j: Normally distributed with parameters: $\mu = 500$; $\sigma = 20$



(g) Figure 2k: Log-normally distributed with parameters: $\mu = 4$; $\sigma = 0.8$



(h) Figure 2l: Log-normally distributed with parameters: $\mu = 7$; $\sigma = 1$

2.3.3 Incorrect Distribution Parametric Model

At $\tau \in \{0.01, 0.5, 0.99\}$, $c = 1 - \tau$, $p = 1$ we get the following error data. Not all data is shown, but it is included in the source file.

τ	0.01				0.5				0.99	
n	10	50	100	200	10	50	100	200	10	200
RMSE	152.1	114.5	107.8	102.2	35.60	24.09	21.80	20.73	87.09	87.43
PLR	0.0093	0.0073	0.0071	0.0069	0.1739	0.1269	0.1181	0.1181	1.452	1.846
MPE	25.80%	21.97%	21.53%	20.87%	-13.78%	-13.77%	-13.25%	-13.34%	140.5%	184.6%

Table 1: Error metrics when estimating $LN(5, 0.5^2)$ distribution as normal distribution

τ	0.01				0.5				0.99	
n	10	50	100	200	10	50	100	200	10	200
RMSE	173.3	120.0	80.6	59.0	28.96	13.39	9.17	6.71	31.73	6.47
PLR	0.0107	0.0062	0.0046	0.0033	0.1534	0.0674	0.0442	0.0307	0.5638	0.1105
MPE	29.92%	-0.14%	7.38%	4.31%	4.20%	0.82%	0.51%	0.32%	-53.15%	-6.01%

Table 2: Error metrics when estimating $LN(5, 0.5^2)$ distribution with the non-parametric model

τ	0.01				0.5				0.99	
n	10	50	100	200	10	50	100	200	10	200
RMSE	128700	100700	95200	88000	20790	15920	15220	14680	64430	76670
PLR	0.5255	0.4173	0.3990	0.3733	0.6915	0.6593	0.6562	0.6488	31.84	33.65
MPE	46.85%	39.36%	38.30%	36.99%	-64.91%	-65.87%	-65.62%	-64.87%	3184%	3366%

Table 3: Error metrics when estimating $LN(10, 1^2)$ distribution as normal distribution

τ	0.01				0.5				0.99	
n	10	50	100	200	10	50	100	200	10	200
RMSE	132500	147200	71500	56070	8665	3744	2732	2043	4578	598.7
PLR	0.5366	0.3953	0.2582	0.2021	0.3109	0.1349	0.1000	0.0736	1.661	0.2211
MPE	45.02%	-5.54%	11.38%	6.54%	4.85%	1.04%	0.39%	0.55%	-158.0%	-10.93%

Table 4: Error metrics when estimating $LN(10, 1^2)$ distribution using the non-parametric model

2.4 Conclusion & Discussion

2.4.1 Holding Cost Sensitivity

To test how the optimal quantity changes when introducing the holding cost we have calculated the optimal order quantity for values for $c_h \in [-1, 5]$. We assumed normally distributed demand with $\mu = 100$ and $\sigma = 10$. We tested for $c \in \{0.9, 1.0, 1.1\}$.

In Figure 1a we clearly see that a lower holding cost c_h relates to a higher optimal quantity. This is logical, as when holding costs are high, a low chance of left over stock would be preferred. However when holding costs are low, or maybe even negative (where a negative holding cost could represent the left over value of the stock, e.g selling old paper to a recycling company), you would be more willing to take the risk of overbuying.

A similar relationship to the holding cost (c_h) can be seen when looking at the expected profit in Figure 1b. The decrease in expected profit could easily be explained by not having the risk of spending a lot of money of possibly paying to destroy unsold goods, or having the opportunity of buying more, without having the risk of paying much to dispose unsold goods.

Although we believe that this result can be generally applied to all demand distributions and other prices, there is a possibility that a demand distributions for which this result might not hold. It could also be interesting to look into the possible cost of not having a product in stock, as some costumers might choose to shop elsewhere if not enough stock is kept.

2.4.2 Parametric vs. Non-Parametric Quality

To test the the quality of both the parametric and non-parametric models we created a Monte Carlo simulation. To estimate the accuracy of the optimal order quantity we defined the RMSE (5), to assess the bias of the optimal order quantity we defined the MPE (7), and to test the impact of the error in the estimation we defined the PLR (6).

Our results are generally quite clear and consistent. When dividing the RMSE of the non-parametric model by the RMSE of the parametric model we consistently get values between 1.2 and 1.4. This suggests that the parametric models estimations are of a higher quality than the non-parametric model. Generally we can also see that when we set $\tau \in \{0.01, 0.1, 0.9, 0.99\}$ the quotient increases quickly. This suggests that the estimations non-parametric model are of a much lower quality at target service levels close to the boundaries.

When looking at the PLR, we can clearly see that generally the parametric model is more profitable than the non-parametric model. This follows logically, as the RMSE already suggested that the quality of the optimal order quantity estimations was of a higher quality. We can also similarly see that at the boundaries of the target service levels the parametric model estimations are much more profitable than the non-parametric model.

Looking at the MPE, we can clearly see that non-parametric estimations close to the boundaries of the target service levels are relatively biased. Low target service levels often result in overestimations (as a negative MPE signifies an overestimation), and high target service levels result in underestimations. This is especially true when we have a low sample size. Logically this makes sense, as $\lceil 0.01 \cdot 10 \rceil = 1$. However with just ten data points often the lowest data point is not in the 1st percentile.

Notably there are some exceptions. When estimating for a log-normal distribution with a high σ (i.e. a σ above 0.7) there are some cases where it seems that the non-parametric model outperforms the parametric model. This is only for low n and it does not appear consistently, and thus we believe it is because of the low M.

2.4.3 Incorrect Distribution Parametric Model

To test how the parametric model reacts when assuming the normal distribution, while the true demand distribution is non-normal, we generated random demand from a log-normal distribution with $\mu \in \{5, 10\}$ and $\sigma \in \{0.5, 1.0\}$. We tried estimating the optimal order quantity using the parametric model with the assumption that the data was normally distributed, and compared it to the non-parametric model.

The results are quite clear, when using a normal distribution to approximate the log-normal distribution the results are sub-optimal. Looking at the MPE (where a negative value means an overestimation), the normal estimation consistently overestimates demand at $\tau = 0.5$, and thus overestimates the optimal order quantity. At $\tau \in \{0.01, 0.99\}$ demand is underestimated, and slightly outperforms the non-parametric model at $n \in \{10, 50\}$ according to the RMSE and PLR, although at higher n, the non-parametric model is much more accurate.

This is particularly the case when $\sigma = 1$. Where the parametric model reaches a PLR of 33.65. Although a $LN(5, 1)$ or $LN(10, 1)$ distribution does not look anything like a normal distribution, and thus it would be unlikely that this distribution would be confused for a normal distribution.

3 Empirical questions

3.1 Problem Description

After having taken a look at the theoretical section, we will now address what all this means in practice by devising an advice for a bakery on how much bread to bake. In order to do this, we have received data over a period of three years, from four different stores of one European bakery chain. We assume that all the data is correct, and that all bread needs to be sold the same day it is baked. Earlier in the theoretical section, we found the formulas needed to calculate the optimal order quantity when holding costs are involved (4). Now, armed with everything we found in that section, and all the data of the bakeries, we finally have everything we need to answer the question posed at the start of this paper: 'How much bread should the individual bakeries bake, in order to maximize profit?'.

To get the best estimation of the optimal quantity we have to solve some problems and we have some choices to make. The first choice is between a parametric (Lin, 2021d) and non-parametric (Lin, 2021c) approach, and in the case of a parametric approach we need to find the best fitting distribution while also remembering the possible patterns in our data. So we want to know: "What are patterns that can be seen in the data?".

Using a Jarque-Bera test (Lin, 2021d) we can confirm or disprove our assumptions about the normality of the demand distribution. When we know what the demand distribution is, we can calculate the optimal order quantity using the parametric model. Thus we are interested in the research question: "What is the optimal order quantity and its confidence interval estimated using the parametric model for Friday, Saturday and Sunday?".

However we might estimate the demand distribution wrong. Therefore we also want to calculate the optimal order quantity using the non-parametric model. So we are interested in the research question: "What is the optimal order quantity and its confidence interval estimated using the non-parametric model for Friday, Saturday and Sunday?".

3.2 Methodology

3.2.1 Describing the Data

To answer the main question it will be of importance to dissect the data that we have been given. Firstly, we calculate important descriptive statistics that summarizes the collection of data: the sample mean, sample variance, skewness, and (Fishers) kurtosis. Secondly, we make histograms to graphically represent the data.

Furthermore, to describe the data more accurately, we dissect our data by splitting it in multiple intervals. In our case, splitting by days gives us a more accurate representation of the correct distribution (Lin, 2021c). Besides the increase in accuracy, we are now able to compute new important descriptive statistics and plot histograms. We can use all this to estimate the demand distribution.

We will also use the Jaque-Bera (JB) test to test for normality (8). The JB test requires a computed critical value and we reject normality for higher JB values.

$$\begin{aligned} \text{JB} &= n\left(\frac{\hat{s}_n^2}{6} - \frac{(\hat{\kappa}_n - 3)^2}{24}\right) \xrightarrow{\mathcal{D}} \chi_2^2 \\ \text{critical value} &= F_{\chi_2^2}^{-1}(1-\alpha) \end{aligned} \tag{8}$$

3.2.2 Parametric Model

For our parametric framework we use the Maximum Likelihood Estimator (MLE) to find the mean and the variance. With these values we can calculate the estimated optimal order quantity $\widehat{Q}^*$ using the inverted distribution, price (p), cost (c) and holding cost (c_h) using (4). After this we calculate our $1 - \alpha = 0.95$ confidence interval with the direct method (Lin, 2021b). We will use this method to calculate the optimal order quantity and its confidence interval for Friday, Saturday and Sunday.

3.2.3 Non-Parametric Model

The big advantage of a non-parametric framework (Lin, 2021c) is not having to choose the right distribution. Using the non-parametric framework we calculate both our estimation of the estimated optimal quantity order ($\widehat{Q}$) using (9) and the confidence intervals using order statistics (10) for Friday, Saturday and Sunday.

$$\begin{aligned}\widehat{Q}_n^{NP}(c, p, c_h) &= \inf\left\{y \in \mathbb{R} : \widehat{F}_n(y) \geq \frac{p-c}{p+c_h}\right\} \\ &= \inf\left\{y \in \mathbb{R} : \sum_{i=1}^{n} \mathbb{1}\{Y_i \leq y\} \geq \frac{p-c}{p+c_h} n\right\} \\ &= Y_{\left(\left\lceil \frac{p-c}{p+c_h} n \right\rceil\right)}\end{aligned} \tag{9}$$

$$\begin{aligned}\tau_{c,p} &= \frac{p-c}{p+c_h}, \quad z_{1-\frac{\alpha}{2}} = \phi^{-1}\left(1-\frac{\alpha}{2}\right) \\ U_n &= \left\lceil n\tau_{c,p} + z_{1-\frac{\alpha}{2}} \sqrt{n\tau_{c,p}(1-\tau_{c,p})} \right\rceil \\ L_n &= \left\lceil n\tau_{c,p} - z_{1-\frac{\alpha}{2}} \sqrt{n\tau_{c,p}(1-\tau_{c,p})} \right\rceil\end{aligned} \tag{10}$$

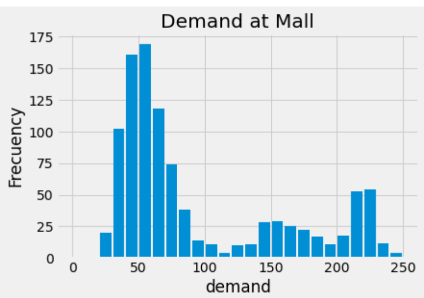
3.3 Results

3.3.1 Describing the Data

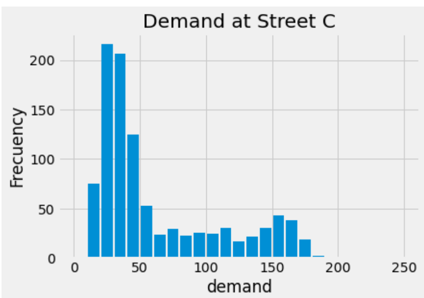
We have found the following descriptive statistics, histograms and time series. Note that we report the Fishers kurtosis.

	Mall	Street A	Street B	Street C
Mean	95.10	78.10	63.10	62.61
Variance	4177.96	2951.03	2334.67	2262.74
Standard deviation	64.64	54.32	48.32	47.57
Skewness	0.99	1.23	1.09	1.11
Kurtosis	-0.53	0.14	-0.25	-0.16

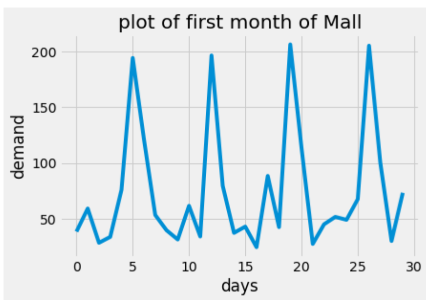
Table 5: Descriptive statistics



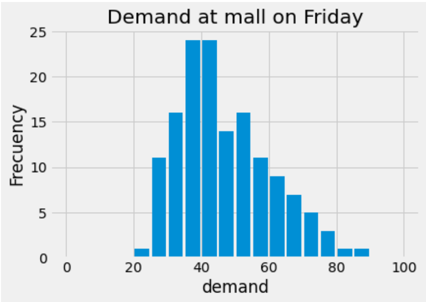
(a) Figure 3a: Histogram of the data of mall.



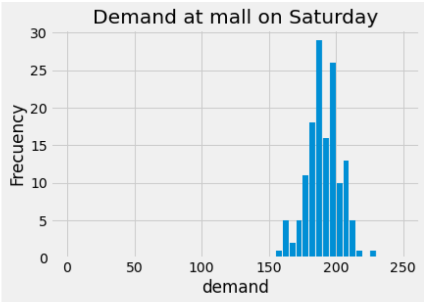
(b) Figure 3b: Histogram of the data of Street C.



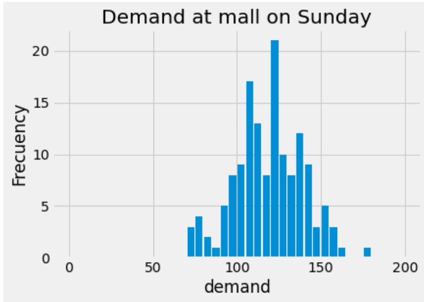
(c) Figure 3c: Timeseries of the first month of mall.



(a) Figure 3d: Histogram of the data of mall on Friday



(b) Figure 3e: Histogram of the data of mall on Saturday.



(c) Figure 3f: Histogram of the data of mall on Sunday.

Mall	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Mean	49.07	47.83	46.69	47.20	46.78	190.98	119.07
Variance	217.56	162.42	255.69	207.54	184.65	154.55	408.00
Standard deviation	14.75	12.74	15.99	14.41	13.59	12.43	20.20
Skewness	0.71	0.58	0.60	0.63	0.68	-0.10	0.00
Kurtosis	0.26	0.01	0.25	0.53	-0.19	0.16	-0.09

Table 6: Descriptive statistics of mall separated by days

Using the Jarque-Bera test we are able to reject normality for Monday to Friday, however we are unable to reject normality for Saturday and Sunday.

3.3.2 Parametric Estimation

The estimated optimal order quantity and its confidence intervals for the Mall and Street C, given by the parametric model.

		Lower bound	optimal order quantity	Upper bound
Mall	Friday	32.32	34.89	37.46
	Saturday	205.77	207.32	208.88
	Sunday	132.99	136.08	139.16
Street C	Friday	27.05	28.97	30.89
	Saturday	148.85	150.98	153.11
	Sunday	91.74	95.08	98.45

Table 7: Mall optimal quantity and CI

3.3.3 Non-Parametric Estimation

The estimated optimal order quantity and its confidence intervals for the Mall and Street C, given by the non-parametric model.

		Lower bound	optimal order quantity	Upper bound
Mall	Friday	33.43	36.32	37.30
	Saturday	204.68	206.73	209.42
	Sunday	129.97	137.57	140.71
Street C	Friday	28.04	29.43	32.70
	Saturday	147.89	151.01	154.82
	Sunday	90.49	94.62	98.38

Table 8: Mall optimal quantity and CI

3.4 Conclusion & Discussion

3.4.1 Describing the data

As seen in the previous results section, we have summarized the data in multiple histograms. The statistics and histograms of the Mall and Street C were similar to those of street A and B, thus we have not included those in this document, however they can be found in our source files.

When looking at the time series, it appears that every day of the week has its own demand distribution. Looking at the histograms, statistics and the JB test, it appears that demand in the weekend is normally distributed, as the skewness and Fishers kurtosis is close to zero. For workdays demand appears to be lognormally distributed as skewness is above zero, and the JB tests were rejected.

3.4.2 Parametric Estimation

When looking at the data we estimated the distributions of the demand. With this knowledge we made two parametric models when estimating the optimal order quantity. From our results the optimal order quantities are in Table 7.

The possibility exists that we have not chosen the correct demand distribution, and thus wrongly estimated the optimal order quantity. We have also assumed that all bread needs to be sold the same day it is produced, as is a requirement of the NVP. However bread may still be sold one day after is baked, this could change the optimal order quantity.

3.4.3 Non-Parametric Estimation

The estimation of the optimal order quantity and its confidence interval using the non-parametric model is in Table 8. Theoretically these values are the optimal order quantity, however it may be the case that when costumers want to buy a product and its not available they may switch to a competitor. Therefore it could be interesting to look at the cost of not having a product in stock, as this could further improve the estimations of the NVP.

4 Conclusion & Discussion

In conclusion, we have rewritten the NVP to include holding cost (2). We tested the sensitivity of this function (1a). We have also compared the parametric and non-parametric models, there we concluded that generally the quality of the estimations of the parametric model is of a higher quality. Although there are some notable exceptions where it seems that the the parametric models is outperformed at a high target service levels and low n by the non-parametric model. We believe that this is just coincidental, because of the low number of estimations ($M = 1000$), though more research could resolve this question.

When testing the parametric model with the wrong demand distribution assumption, we did find that the results were suboptimal. This shows the advantages and disadvantages of the parametric and non-parametric models. As the parametric model is more accurate, but only when you assume the correct demand distribution. The non-parametric model is generally less accurate, especially at high target service levels, tough it does not require an estimation of the demand distribution.

In the empirical part of this paper we have looked at the data of a European bakery chain. When looking at the data we concluded that during the workweek demand is log-normally distributed, and during the weekend the data is normally distributed. We used this assumption when estimating the optimal order quantity using the parametric model. Though this assumption could be wrong, and thus give us suboptimal results.

When estimating demand using the parametric and non-parametric model we did assume that the demand distribution changes per day of the week. However it could be the case that demand changes from month to month, or more generally demand for bread increases or decreases. Thus our demand distributions remained static. This assumption was necessary, because currently we do not have any tools for dealing with general trends (e.g. a monthly increase of demand of 3%). However future research could provide tools for dealing with such trends.

We also did not consider the cost of not having goods available. We assume that costumers do not mind us not having their favourite product. Though it could be possible that they start shopping at competitors. This could create a negative feedback loop, where lower demand is noticed, we keep less in stock. Thus at peak demand do not have enough for everyone, and thereby lose costumers. We did not consider this possibility when estimating the optimal order quantity for the bakery.

References

Levi, R., G. Perakis, and J. Uichanco (2015). "The Data-Driven Newsvendor Problem: New Bounds and Insights". In: *Operations Research* 63.6, pp. 1294–1306. DOI: 10.1287/opre.2015.1422. eprint: https://doi.org/10.1287/opre.2015.1422 (cit. on p. 2).

Lin, Y. (May 2021a). *News Vendor Problem, Introduction.* School of Business and Economics VU University Amsterdam (cit. on p. 2).

— (June 2021b). *The Confidence-Based Newsvendor Problem, Quantifying Uncertainty in Estimations.* School of Business and Economics VU University Amsterdam (cit. on p. 8).

— (June 2021c). *The Data-Driven Newsvendor Problem, Nonparametric and Feature-Based Estimations.* School of Business and Economics VU University Amsterdam (cit. on pp. 2, 3, 8, 9).

— (June 2021d). *The Newsvendor Problem, Monte Carlo and Maximum Likelihood Point Estimations.* School of Business and Economics VU University Amsterdam (cit. on pp. 2–4, 8).