Bachelor thesis

Department of Statistics

Nr 2021:13

Benford's law applied to sale prices on the Swedish housing market

Belit Ay
Nabiel Efrem

15 ECT credits in Statistics III, VT2021
Supervisor: Per Gösta Andersson

Abstract

Benford's law is based on an observation that certain digits occur more often than others in a set of numbers. This have provided researchers to apply the law in different areas including identifying digit patterns and manipulated data. To our knowledge, this have yet not been tested in the Swedish housing market. The purpose of this thesis is to examine whether the sale price for 171 643 tenant-owned apartments in Stockholm, Gothenburg and Malmö follow Benford's law. Numerous researchers have used this law for testing various types of data but based solely on the first digit distribution of their data. This study will furthermore test the second digit and the first two digits of our data. The tests used to evaluate our data's conformity to Benford's law include Kolmogorov-Smirnov test and Mean absolute deviation (MAD) test. We found that the second digit of sale prices did follow Benford's law, the first digit and the first two digits did not follow the law. The results show that Benford's law is a good method for identify certain digit patterns and further research is needed to draw the conclusion that sale price does not follow Benford's law as certain limitations on our data was identified.

Keywords: Benford's law, Sale price, Swedish housing market, Kolmogorov-Smirnov, Mean absolute deviation (MAD),

Table of contents

1 Introduction

The distribution of the first digits from a set of random numbers is not always uniformly distributed as one may think, instead it often follows a particular distribution which can be explained by Benford's law. This law was first discovered after an observation made by (Newcomb, 1881), but have in recent decade been used as a tool to identify certain digit patterns and possible manipulation in data. Some refer a set of numbers that obeys the law as natural occurring, and numbers that are selected from different distributions tend to show a higher degree of conformity to the law (Benford, 1938). The study of Benford's law is extensive and can be found in different areas such as election fraud (Mebane Jr, 2006) or mathematics (Hill, 1995) but the housing market is not as well studied.

The Swedish housing market is a recurring topic in the media where the discussions vary from the rising house prices to the bidding process when buying an apartment. The latter discussion stretches to the occurrence of false bidding (Aschberg & Mohlin , 2015), and a bid is not legally binding for any participants (Fastighetsmäklarinspektionen, 2020). This sparked an interest to study if Benford's law appears in the sale price for tenant-owned apartments. With data obtained from Svensk Mäklarstatistik will this thesis examine whether the sale price for 171 643 tenant-owned apartments in Stockholm municipality, Gothenburg municipality and Malmö municipality during 2013-2018 follow Benford's law. To clarify, from now on we will refer Stockholm municipality as only Stockholm, the same implies for Gothenburg and Malmö. We further look at each year individually and perform the same tests as for the entire period 2013-2018. Further analysis will be performed in order to possibly identify any digit patterns in data. The digit tests are based on the first digits, second digits and the first two digits with The Kolmogorov-Smirnov test and the Mean absolute Deviation (MAD) test. As the first digit test can give misleading confirmative results is it important to extent the tests to other digits.

1.1 Objective

The purpose of this study is to examine whether Benford's law appears in the Swedish housing market and if there are any digit patterns that can be identified. We will use sale price for tenant-owned apartments during 2013-2018 in Stockholm, Gothenburg and Malmö as our data to analyze. As no previous study has investigated this to our knowledge, we found this to be an interesting topic to study.

The main question this thesis aims to answer:

- Does sale price for tenant-owned apartments during 2013-2018 in Stockholm, Gothenburg and Malmö follow Benford's Law?

As Benford’s law allows for further digit analysis it is of huge interest to answer:

- Can we identify any digit patterns in our data?

- Does sale price for tenant-owned apartments for each year between 2013 to 2018 in Stockholm, Gothenburg and Malmö follow Benford's Law?

The main focus will be on the whole period since Benford’s law prefers data with a larger sample size, which we further discuss in coming sections. Another reason for examining individual years is to mainly identify any digit patterns for each year. The difference in sample size for the whole period and individual years supports the decision to split up the study in this way in order to obtain comparable results from our tests.

1.2 Benford’s Law

Let’s play a game, choose two random numbers from any random set of numerical data, it can be the surface area of lakes, the population of cities or the numbers found in a newspaper. You chose the numbers 635 and 1309. Let us first define a digit and a number: digit is a natural number taking on values between 1 and 9, were 0 can be a digit after the first position and a number as one or more digits combined (Benford, 1938, p. 552). From this definition we know that the first digit of the numbers 635 and 1309 is 6 and 1, respectively. If we continued to collect the first digit from these random sets of numerical data, can we expect certain probabilities for each number to be the first digit?

For those who are not familiar with Benford’s law might intuitively expect that each number from 1 to 9 have the same probability to occur as the first digit. Meaning that the first digit is expected to be uniformly distributed with equal probability of occurring as one of the numbers 1 to 9. This is however not invariably the case where previous research has observed that digits are distributed in a way that do not resemble the uniform distribution. It has been shown that the first digit follows a pattern where we can expect 1 to occur as the first digit approximately 30.10 % of the times, 2 approximately 17.61 % of the times etc. This phenomenon is called Benford’s law and was first discovered in the 19^{th} century by the astronomer and mathematician Simon Newcomb (1881) who noted that the pages in the logarithmic table with number 1 as the first digit were more used than pages with higher number as the first digit. Newcomb published a two page long article about his observation and further presented the logarithmic probabilities of numbers occurring as the first and second digit as shown in table 1.1 (Newcomb, 1881).

Table 1.1: Probabilities of occurrence for the first two digits

Dig.	First Digit.	Second Digit.
0	…	0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

Newcomb furthermore came to the conclusion that “*The law of probability of the occurrence of numbers is such that all mantissae of their logarithms are equally probable”*. (Newcomb, 1881, p. 40).

This is the mathematical basis of Benford's law and implies that the mantissa of the logarithm of a set of numbers should be uniformly distributed. We will further explain mantissa in later section of this thesis. The second digit, just as the first digit, are not uniformly distributed but their values are closer to the uniform distribution than the first digit, which can be seen from table 1.1 (Berger & Hill, 2015, pp. 1-3). The probabilities as shown in table 1.1 can be calculated as follows:

$$P(D_1 = d_1) = log_{10}\left(1 + \frac{1}{d_1}\right), \qquad d_1 = 1,2,\dots,9 \tag{1.1}$$

$$P(D_2 = d_2) = \sum_{k=1}^{9} log_{10}\left(1 + (10k + d_2)^{-1}\right), \qquad d_2 = 0,1,2,\dots,9 \tag{1.2}$$

Equation (1.1) is the formula for calculating the probability that a number has a certain first digit, d_1 and equation (1.2) is the formula for calculating that a number has a certain second digit, d_2 where k represents the first digits. Both equation (1.1) and (1.2) are derived from (Hill, 1995, p. 354). Newcomb did not present further theoretical explanation on this observation, and it was later rediscovered by the physicist Frank Benford (Benford, 1938) who corroborated Newcomb's observation by studying 20 datasets with unrelated subjects taken from different sources. The datasets had a total of 20 229 observations, that varied from mathematical tabulations to the surface area of 335 rivers and the numbers from the front page of newspapers. Benford's findings gave support to Newton's observation, but he further applied to the law. One further discovery was that a sample of individual numbers without relationship tend to obey to the Benford distribution better than individual numbers that do have relationship to each other, such found in mathematical tabulations (Benford, 1938, pp. 556-557). This was the reason Benford called this phenomenon "The law of anomalous numbers" and it was after he published his results that this law became known as Benford's law.

Benford further explained the possible cause of this law with a geometric series. An outgiving that (Miller, 2015, pp. 12-13) takes on with the explanation that a process with a constant growth, lower digits tend to have a longer duration compared to higher digits. Meaning that more time is spent on the first digit 1 to become 2, than 8 to become 9 with a constant growth rate. Miller presents an example of a yearly constant growth rate of 4 % and how it takes over 17 years for 1$ to be worth 2$, and less than 3 years for 9$ to be worth 10$.

After Benford published "The law of anomalous numbers" in 1938, many researchers aimed to prove the law and one important finding was made by (Hill, 1995) who noted that if random samples taken from randomly selected distributions, the first digit of the combined sample will converge to the Benford's distribution. This finding has helped to explain the occurrence of Benford's law from different sets of numbers since the combination of numbers from different sources tent to obey the law.

1.3 Previous studies

There are numerous of previous studies regarding Benford's Law in varied fields. However, few studies have been made in the area of house prices. A study showed that sale prices for multifamily apartments in New Jersey and office buildings of class A and B in New York obey Benford's law (Pomykacz, et al., 2017). The study also describes which sorts of data in terms of real estate is naturally consistent with Benford's law, these could for instance be real estate prices, values and rents. The importance of this work is that it provides research on sale prices, as this thesis aims to study.

In a study investigating whether Benford's law holds for the winning bids in selected eBay auctions, (Giles, 2007) found that the price for professional football tickets do obey Benford's law. In other words, the aim was to test if the numbers were naturally occurring. The study was made based on the winning bids in 1 161 auctions for professional football tickets and only based on prices that ended from a bidding process. The importance of (Giles, 2007) for our study is that sale and auction prices are both results of a bidding process which is influenced by the human action. There are certain distributions that do not obey Benford's law such as prices found in superstores and ATM withdrawals because these numbers are influenced by the human mind (Mittermaier & Nigrini, 1997). For example, supermarket prices often end with 9 such as 4.99$ or 49.99$.

Another research regarding auction prices and Benford's law was made by (Endress, 2014), which showed that the "buy now" prices of in-game items sold at auction houses in the virtual gaming world obey the law. Players in these virtual games can buy and sell in-game items thru auction houses where items can be sold either by bidding or for a fixed price called "buy now" price. Together, these studies provide some evidence that sale price and numbers that are influenced by human mind can show conformity to Benford's law.

The study of Benford's law can be found in other fields such as mathematics (Hill, 1995; Pinkham, 1961), election forensics (Mebane Jr, 2006) and tax auditing (Nigrini, 1996). The latter research showed that low income taxpayers were more likely to manipulate numbers on their tax returns than high income payers. Benford's law is also efficient in detecting fraud in accounting data (Mittermaier & Nigrini, 1997). Nigrini appears in many citations by other researchers in the field of Benford's law and have published books (Nigrini, 2020; Nigrini, 2012) with practical applications and methods for fraud detection, digit analysis and auditing with Benford's law. The latter books have been extensively used in this thesis in order to both understand and use Benford's law as a method to answer our hypothesis.

The interest in studying sale prices came on its own initiative and also through inspiration from (Giles, 2007; Pomykacz, et al., 2017) and mainly (Nigrini, 2012; Nigrini, 2020) contribution to the research area. As earlier mentioned, there are few studies in the area of house prices and up to now, there are no study in sale price in the Swedish housing market to our knowledge. In recent years, news articles regarding the Swedish housing market have appeared including the occurrence of false bidding in order to initiate sales (Aschberg & Mohlin , 2015). Thus far, the bidding process is not regulated by law and a bid is not legally binding for any of the participants (Fastighetsmäklarinspektionen, 2020). The number of complaints regarding the bidding process have been increasing the past years (Fastighetsmäklarinspektionen, 2021) and this have led to real estate firms implementing identification controls via bank-ID in order to participate in bidding (Hellekant, 2019).

We want to emphasize that this thesis does not aim to detect fraudulent activity in the bidding process. It is important to understand how to interpret the results from a digit analysis with Benford's law and it starts with what type of dataset is used. We will further explain which datasets are suitable to use in coming section. But even if data do not follow Benford's law it does not directly imply that data is a result of fraudulent activity, it can be used for identifying suspect numbers for further analysis, it can point you to the right direction if used properly (Durtschi, et al., 2004). Which also was suggested by (Nigrini, 2020, p. 79) that a set of data not conforming to Benford do not always indicate fraud or error, it is important to always make further analyze of the data.

2 Method

This section presents each method used to answer the hypothesis of the study that was earlier presented. The chosen methods consist of both hypothesis testing and analysis from visual interpretation and are primarily influenced by Professor Mark Nigrinis work on Benford's law (Nigrini, 2012; Nigrini, 2020). A vast majority of previous research made on Benford's law refers to Mark Nigrini and his work on Benford's law and research either applies Nigrinis methods or refers to his work. The first subsection "First step" will cover some essential subjects in Benford's law such as the mathematical basis and data's fit for an Benford analysis. The second subsection "Digit tests" will cover diverse digit tests in order to conform data to Benford's law and analyzing digit patterns which is our second hypothesis to answer. Third and lastly, "Test statistics" will cover our chosen tests to validate conformity.

2.1 First step

2.1.1 Mantissa

As first discovered by (Newcomb, 1881), a collection of numbers should conform to Benford's law if the mantissas of the logarithm are uniformly distributed. This is the mathematical condition for Benford's law to hold.

The logarithm of a number will result in two parts: the first part is the integer of the number which is called the *characteristic* part and the second part is the decimal value of the number that is called the *mantissa*. In other words, the mantissa is the decimal part of the logarithm. For example, $log_{10}(20) = 1.3010$ the characteristic equals 1 and the mantissa equals 0.3010. The mantissa can only take on values between 0 and 0.9999, not 1 since this would then be an integer (Weisstein, u.d.).

Let's say that we have a set of numbers that are near perfect fit of Benford's law. If we further took the logarithm these numbers, extracted the mantissas and sorted these in ascending order and plotted this with the ordered values on x-axis would we get:

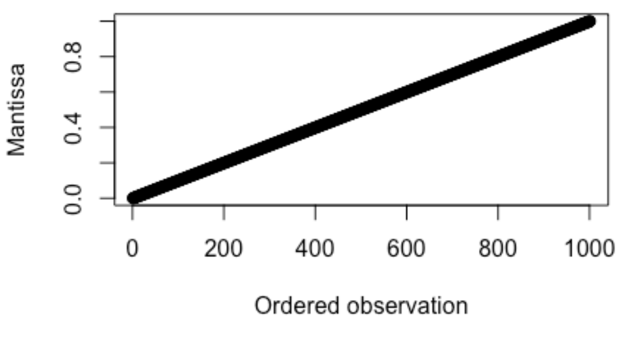

Figure 2.1: Uniformly distributed mantissas [0,1) with n=1000

Table 2.1: Properties for $Mantissa \sim U[0,1)$

Mantissa	
Average	0.50
Variance	1/12
Skewness	0
Kurtosis	-6/5

One important factor is that as the number of observations tends to infinity, the mean of the mantissas will get closer to the value 0.5, which will provide a set of numbers that is closer to the Benford distribution (Nigrini, 2012, p. 11).

The visual inspection of the mantissas can be essential, aside from being the mathematical basis for the law it is possible that the first digit test can give misleading results on whether data obey Benford's law, as we will further explain in coming section. Table 2.1 present the properties that should hold for the mantissas of a set of numbers in order to exhibit a near perfect fit of Benford's law (Nigrini, 2012, p. 11). It is still possible for a set of numbers to show conformity to Benford's law if the properties do not hold perfectly. No test will be done based on the mantissas; it is only for visual interpretation as we will perform other methods for testing if data follows Benford's law.

We will use the software R with the package *benford.analysis* in order to obtain the values and graph for the mantissas for the whole dataset, sale price 2013-2018.

2.1.2 Conformity

It's of huge importance that the data used to perform the analysis follows the requirements of the Benford's law. This should be validated in prior to the analysis with Benford's law. The methods to evaluate the fit of data consists of both knowledge from previous research and guidelines provided from Mark Nigrini. If data do not meet any or a only few requirements, is it considered a weak fit and there a possibility that data contains unusual duplicates and anomalies (Nigrini, 2012, p. 22).

First and foremost, data that define the magnitude of facts or events are preferred, for example the population of countries or cities, and for financial data can this be market value (Nigrini, 2012, p. 20). Sale price for apartments that are sold after a negotiation should correspond to the market value (Fastighetsmäklarförbundet, u.d.). Second of all, data that are restricted to arbitrary minimum or maximum limits are not preferred such as a city that are defined for a population count of a minimum 20 000, or a commission charge taken from a stockbroker of the amount of 75$. The latter data will most likely produce an excess of the number 7 as the first digit and 0 as the second digit (Nigrini, 2012, p. 22). Third, data that contains telephone numbers, bank account numbers or personal identification numbers such as social security number do not obey Benford's law. Fourth, data that are influenced by the human mind such as supermarket prices and ATM withdrawals are not preferred and will not obey the law (Mittermaier & Nigrini, 1997). The fifth and last guideline is that the observations should not be too centered around the mean, the optimal data have a high variation and data should contain more small values than large. The latter holds true for real circumstances since there is more small villages, corporations and lakes than big cities, huge corporations like Apple and big lakes (Nigrini, 2012, p. 22).

It has previously been observed that positively skewed data are more likely to obey Benford's law, (Wallace, 2002) suggests that a larger ratio of mean to median will show higher conformation. To further understand why Benford's law applies to financial data, such used in this thesis, (Nigrini, 2020, p. 94) explains that data that are based on single events such as a single purchase obey Benford's law because of what he refers to the *mixture of distribution* argument. This argument is based on findings from previous research such as in (Hill, 1995) as earlier mentioned, a set of numbers containing random samples from random distributions will converge to the Benford distribution.

Another important factor for data to show conformity is the sample size. There is no correct answer to this but as stated in the previous section, as sample size increase to infinity the mean of the mantissas will converge to 0.5. (Nigrini, 2012, p. 20) Also noted that data with at least 4 digits will show a closer fit for Benford's law, this was first observed by (Benford, 1938). It is preferable to have at least 1000 observations until we can expect data to conform to Benford's law (Nigrini, 2012, p. 20).

2.2 Digit tests

2.2.1 The first digit test

Benford's law, also called the first digit phenomenon, explains the distribution of the first digits which (Newcomb, 1881) first observed. Many researchers have tested the hypothesis if their data follows Benford's law with different methods but based solely on testing the first digit distribution of their data, such as (Giles, 2007; Benford, 1938; Endress, 2014; Pomykacz, et al., 2017). The problem with only testing the first digit is that it can visually resemble a Benford distribution like figure 2.1 but the mathematical basis about the mantissas being uniformly distributed are violated (Nigrini, 2012, p. 15).

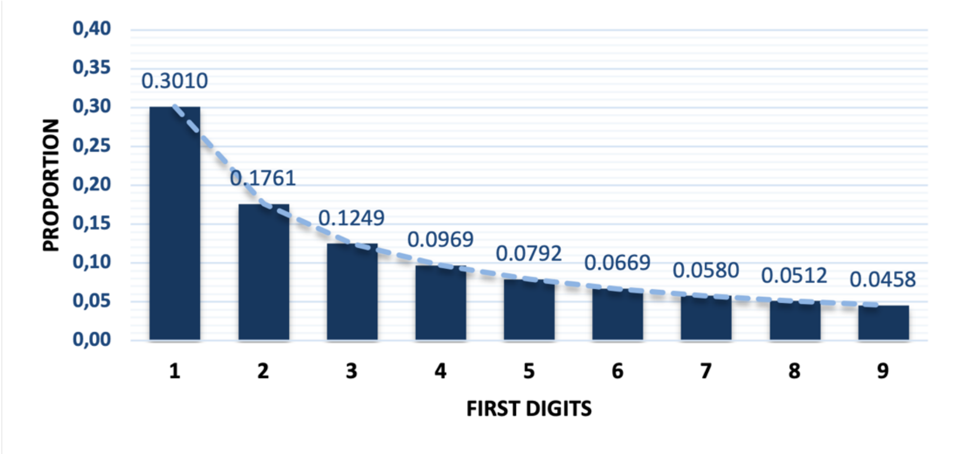

Figure 2.2: The first digit distribution

The first digit distribution will present the proportions of first digit in the data(figure 2.2), this is the same first digit proportions as (Newcomb, 1881) discovered. Even though the first digit test by itself is not enough to confirm that data follows Benford's law because of the chance of violating the mathematical basis, will the first digit test be performed because it is relevant to the subject and to possibly identify Benford patterns in the data.

There will be no conclusion about conformity to the law based solely on this test. The first digit test will provide us information about the first digit distribution of our data and if it conforms to Benford's law with the help of Mean Average Deviation (MAD) measure and the Kolmogorov-Smirnov test which will be discussed in further section. If the first digit distribution conforms to Benford's law, is it also important to observe the mantissas as mentioned in previous section. It is worth mentioning that even if data conforms to Benford's law does it not assure absence of anomalies and other errors in data (Nigrini, 2020, pp. 103-105). If there are four datasets that are analyzed and three of them do obey Benford's law would the focus be on the dataset that showed the opposite, it indicates higher risk of containing error or fraud (Nigrini, 2012, p. 74).

The first digit test is performed in the software R with the packages *benford.analysis* and *BenfordTests*. The data used in this thesis is sale price in SEK for tenant-owned apartments sold during 2013 to 2018 in Stockholm, Gothenburg and Malmö, with 171 643 observations. The first digit of all observations is extracted from data to a frequency table. From 171 643 bins do we now have 9 bins with the first digits ranging from the numbers 1 to 9 with the count of each of the first digits from the initial data. From this table will further graphs and tests be performed, and the packages provide the test results for Mean Average Deviation (MAD). The Kolmonogrov-smirnov test are performed with the same software but with the package "*BeyondBenford*".

2.2.2 The first two digit test

The first two digit graph will display both the first digit and the second digit from data which can provide more information about both their distributions at once. We will be able to identify various types of traits such as unusual duplicates, anomalies and Benford tendencies (Nigrini, 2012, p. 78). The second digit graph can be used to detect certain behaviors such as avoiding controlled threshold, while the first two digit can provide us further information about which first two digits are of importance (Nigrini, 2012, p. 76). For example, if there is a controlled threshold for deposit machines at 2 500$, the second digit graph can show spikes at number 4 indicating excess deposits just under the threshold of 2 500$, and the first two digit graph can further confirm if this behavior hold true by showing spikes at 24. Examining the first two digits can also display psychological pricing where spikes occur just under psychological barriers for example 48, 49, 98 and 99 indicates excessive amounts made just such as 500, 5000, 100 or 10000 dollars (Nigrini, 2012, pp. 78-80). The first two test will not only further point us to the right direction, but it will also narrow down the sample size if suspicion is raised at the first digit test (Nigrini, 2020, p. 109).

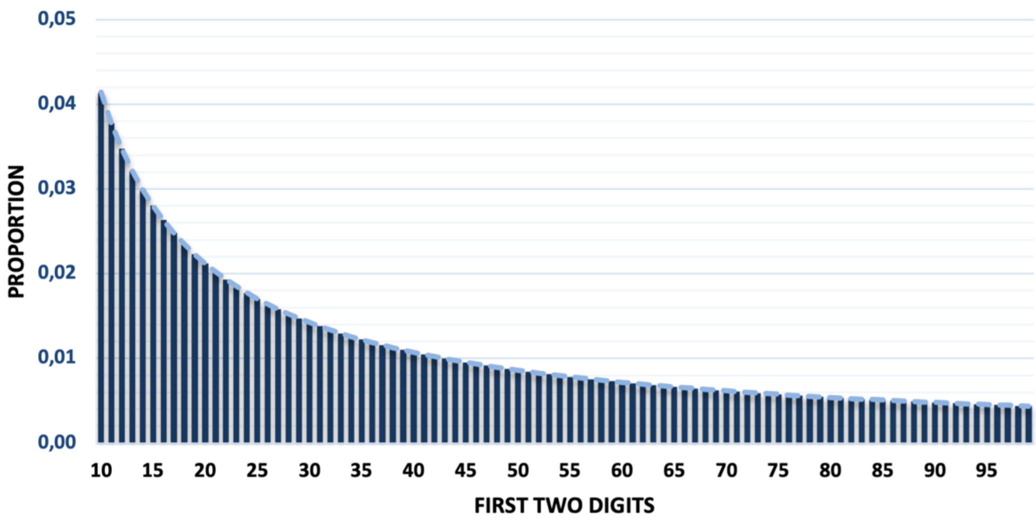

Figure 2.3: The first two digit distribution

Since the first digit can show misleading conformity to Benford's law is it valuable to complement with the first two digit test to get further evident results. (Nigrini, 2012, pp. 15-18) demonstrates how a confirmative first digit test gave misleading results by proving that the mantissas were not uniform distributed, and this was further supported by the first two digit test which failed to show conformity to Benford's law (Nigrini, 2012, pp. 15-18).

The first two digit test is performed in the software R with the packages *benford.analysis* and *BenfordTests*. The first two digit of all observations is extracted from data to a frequency table. From 171 643 bins do we now have 90 bins with the first two digits ranging from the numbers 10 to 99 with the count of each of the first two digits from the initial data. From this table will further graphs and tests be performed, and the packages provide the test results for Mean Average Deviation (MAD). The Kolmonogrov-Smirnov test are performed with the same software but with the package "*BeyondBenford*".

2.2.3 The second digit test

The second digit test alone is not usually a method to test if data conforms to Benford's law, however even if the first digits does not obey Benford's law is it still possible that the second digit does.

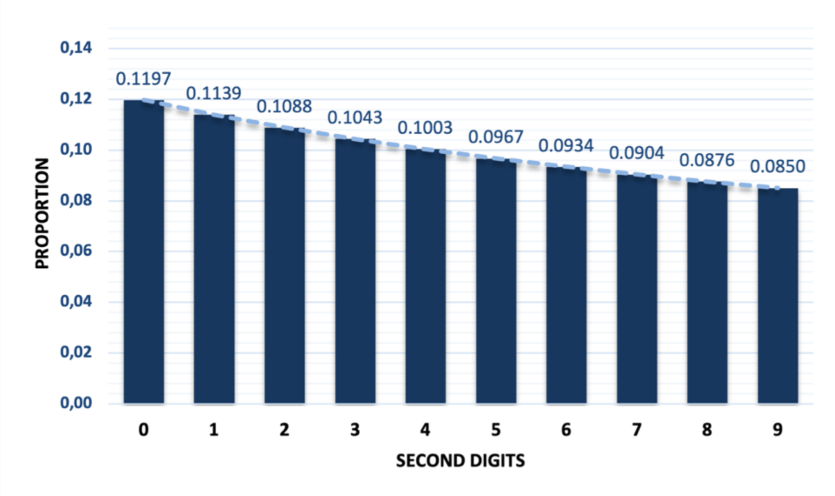

Figure 2.4: The second digit distribution

Figure 2.4 shows the second digit distribution with Benford proportions that was earlier presented in table 1.1. The second digit distribution can also show other digit patterns that can be useful in certain situations. One type of digit pattern that the second digit is suitable to detect are round numbers which is usually found in payments or price data where the second digit more often are 0 and 5 (Nigrini, 2012, pp. 75-80). Other effects that give rise to round numbers is the effect of negotiation, which can be seen in numbers such as sale price (Nigrini, 2020). Studies where second digit test have come to use is for example in election forensic (Mebane Jr, 2006) and with financial data where (Nigrini, 2005) was able to identify rounding up behavior in earning reports. The second digit can also find patterns as a result from targeting specific numbers and as mentioned in previous section, is useful to detect that avoiding controlled threshold behavior (Nigrini, 2012, pp. 75-80).

The second digit test is preformed using the software Excel. The data in question, sale price in SEK, are in tabulated form with observations of 171 643. The first two digit of all observations are then extracted from data to a frequency table. From 171 643 observations do we now have 10 digits ranging from 0 to 9 with the count of each of the second digits from the initial data. From this table will further graphs and we only perform the Mean absolute deviation (MAD) test since the second digit is not used for conforming whole data to Benford's law.

2.3 Test statistics

2.3.1 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (K-S) test (Kolmogorov, 1933) is a nonparametric test often used for assessing data's conformity to Benford's law by comparing the expected cumulative distribution function $F_b(x)$ and the actual cumulative distribution function $F_0(x)$. The Kolmogorov-Smirnov statistic use the greatest distance between the expected cumulative distribution function and actual cumulative distribution function, in a vertical direction. For example, the first digit distribution has nine numbers ranging from 1 to 9, the test will use the largest difference in absolute value between the digit's cumulative distribution functions for both our data's first digits and Benford's first digits, the supremum (sup_x) means the largest value (Kotz & Johnson, 1992, pp. 93-103). We define $x = 1,2,...,9$ For the first digit test and $x = 10,11,...,99$ for the first two digit test.

The Kolmogorov-Smirnov statistic:

$$D = sup_x|F_0(x) - F_b(x)| * \sqrt{N} \tag{2.1}$$

We test the null hypothesis against the alternative hypothesis:

$$H_0: The\ sale\ price\ data\ follows\ Benford's\ distribution$$

$$H_A: The\ sale\ price\ data\ do\ not\ follow\ Benford's\ distribution$$

The Kolmogorov-Smirnov statistic (2.1) is obtained from (Joenssen & Muellerleile, 2015). The K-S test takes sample size into account, which makes it sensitive for small deviations the larger data we have. This is not always ideal when having big sample sizes, but the test is still widely used by researchers with big sample data. Other test statistics that are commonly used for assessing conformity to Benford's law is the Pearson's chi-square test, but as the K-S test it takes the number of observations into account which makes it sensitive to small deviations.

2.3.2 Mean Absolute Deviation (MAD) test

The Mean Absolute Deviation (MAD) is used for measuring the extent of data's conformity to Benford's law (Nigrini, 2012, pp. 158-160). MAD measures the difference of each digit proportion in the dataset and the digit proportion of Benford in absolute terms and then averages the sum of this difference. The difference measures the deviation between the digit proportion in the dataset and the digit proportion of Benford.

$$Mean\ Absolute\ Deviation\ (MAD) = \frac{1}{k}\sum_{i=1}^{k}|AP - EP| \tag{2.2}$$

Where AP represents the actual proportion, which is the digit proportion in the dataset used in this thesis, EP represents the expected proportion which is the digit proportions of Benford and k represents the number of bins. The number of bins depends on which digit test are preformed, the first digit test contains 9 bins and the first two digit test 90 bins. The actual proportion are calculated by dividing the count of respective digit with the total number of bins. Low value of the deviation indicates that the digit proportions of the data is close to the digit proportion of Benford and vice versa. The critical values used are obtained from (Nigrini, 2012, p. 160) and presented in table 2.2. The difference between the MAD test and other test statistics is the absence of N in the equation (2.2) which makes MAD a more preferable measure since it does not take sample size into account.

Table 2.2: Conclusion range for the Mean absolute deviation test

Conclusion range	**First digits**	**Second digits**	**First two digits**
Close conformity	0.000 to 0.006	0.000 to 0.008	0.0000 to 0.0012
Acceptable conformity	0.006 to 0.012	0.008 to 0.010	0.0012 to 0.0018
Marginally acceptable conformity	0.012 to 0.015	0.010 to 0.012	0.0018 to 0.0022
Nonconformity	Above 0.015	Above 0.012	Above 0.0022

3 Data

The research data used in this thesis were obtained from (Svensk Mäklarstatistik, 2021) who accounts for over 95% of all estate sales through a realtor. The data contains 171 643 observations on sale prices in SEK for tenant-owned apartments in Stockholm municipality, Gothenburg municipality and Malmö municipality during the period from January 2013 to December 2018 with the date of transaction contract date. We obtain the observations of all districts that together is defined as a municipality for each area. Each district is further separated to individual data frames using the software R and then combine all observations in one data frame with *rbind* in order to study all municipality simultaneously. This thesis will both study the entire time period and for each year meaning we will have total seven outcomes for all municipalities for each test we perform.

Table 3.1: Number of observations in each municipality

Stockholm	Gothenburg	Malmö	Total
93 229	46 439	31 975	171 643

4 Results

This section presents the distribution of the mantissas and the dataset over the whole study period for assessing the fit of data for a Benford analyze. Thereafter, the distributions for the first, first two and second digits and the tests performed to assess data's conformity to Benford's law. All data presented in this section is observing the whole study period, 2013-2018, and for the individual years are presented in Appendix A, B, C and D.

4.1 Mantissa

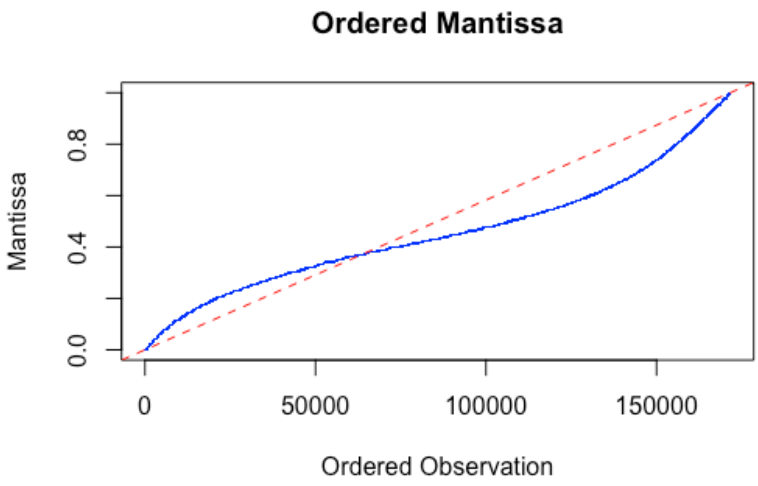

Figure 4.1: The ordered mantissas for sale price 2013-2018

Table 4.1: Mantissa values for sale price 2013-2018

Mantissa	
Average	0.45
Variance	0.05
Skewness	0.37
Kurtosis	-0.33

Figure 4.1 displays the ordered values of the mantissas for the sale price data, which is represented as the blue line, and a fitted line for the uniformly distributed mantissas ranging between [0,1). The values in table 4.1 corresponds to the blue line in figure 4.1 where we can observe that the mantissas of sale price are not perfectly uniformly distributed. The mantissas of uniformly distributed mantissas have a skewness of 0 (table 2.1), which gives the symmetrical

line whereas sale price data have a skewness of 0.37 which can be observed in figure 4.1 with. Table 4.1 shows that the average and the variance of the mantissas for sale price are lower than the values for the uniformly distributed mantissas (table 2.1).

4.2 Conformity

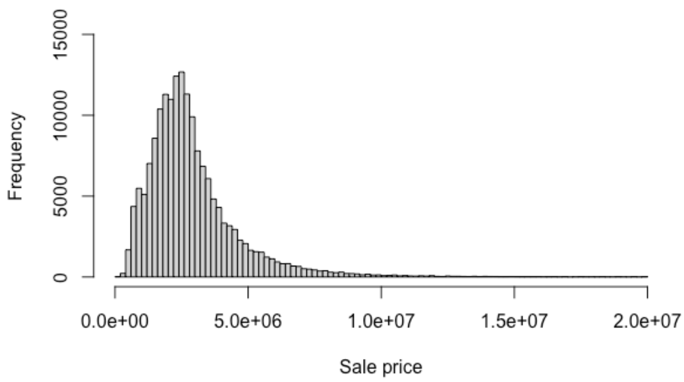

Figure 4.2: The distribution of sale price 2013-2018

Table 4.2: Descriptive statistics for sale price 2013-2018

Obs.	Mean	Median	Mode	Min	Max
171 643	2 915 731	2 550 000	2 500 000	15000	20 000 000

The distribution of data used in this thesis is presented in figure 4.2 which contains 171 643 observations corresponding the sale price of sold tenant-owned apartments during the period of 2013-2018 in Stockholm, Gothenburg and Malmö. A majority of the observations lie between the range of 0 and 5 000 000 SEK and the maximum value stretches to 20 000 000 SEK as can be seen from figure 4.2. This shows that our data contains more smaller values which is aligned with Benford's law as earlier stated. The median is smaller than the mean which is the characteristics of a positively skewed data. It was noted by (Benford, 1938) to use numbers with at least 4 digits and table 4.2 shows that the minimum value of sale price is 15 000 which contains 5 digits.

The distribution for each year individually has positively skewed data with majority of the observations lie between the sale price range of 0 and 5 000 000 SEK (Appendix A), which is similar to the whole period 2013-2018 figure 4.2.

4.3 The first digit test

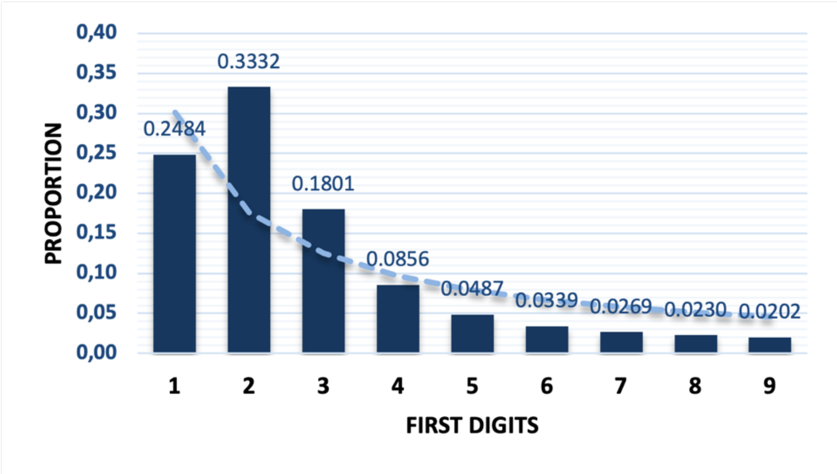

Figure 4.3: The first digit distribution for 2013-2018

Table 4.3: Results for first digit test

Period	Mean Absolute Deviation (MAD)	Kolmogorov-Smirnov test
2013-2018	0.04717	D = 66.145, p-value < 2.2e-16

Figure 4.3 displays the first digits proportion of our dataset over the whole period where the blue dotted line represents the probability of the first digits for Benford's law, as first presented in table 1.1. The probability for number 1 to occur as the first digit is according to Benford's law 30.1 % and the probability drops as the numbers gets higher. This is however not always true for our dataset were number 2 have the highest probability to occur as the first digit as shown in figure 4.3. we also observe that the two numbers that have higher proportions than Benford's is number 2 and 3, this means that sale price data contains more numbers starting with 2 and 3 than what is expected from a perfect Benford set. Number 4 is closest to the Benford proportion. Since we have a low p-value we reject the null hypothesis as can be seen from table 4.3 for the Kolmogorov-Smirnov test. The Mean absolute deviation gives too high value for sale price to conform to Benford's law as the nonconformity range for the first digits is 0.015 and over (table 2.2).

The first digit distribution for individual years do not have the same proportions over the whole period. The first year in our period, 2013, shows the most Benford tendencies, meaning the occurrence of number 1 as being the first digit have the highest probability and the probability is descending for higher numbers. But this is not true for the years 2015-2018. Comparing 2014 with 2013, there is a change in the proportions for the number 2, and from 2015 to 2018 is the number 2 the most occurring first digit (Appendix B). Since we have a low p-value we reject the null hypothesis for the Kolmogorov-Smirnov test. The Mean absolute deviation gives too high value for sale price to conform to Benford's law as the nonconformity range for the first digits is 0.015 and over (table 2.2), (Appendix B).

4.4 The first two digit test

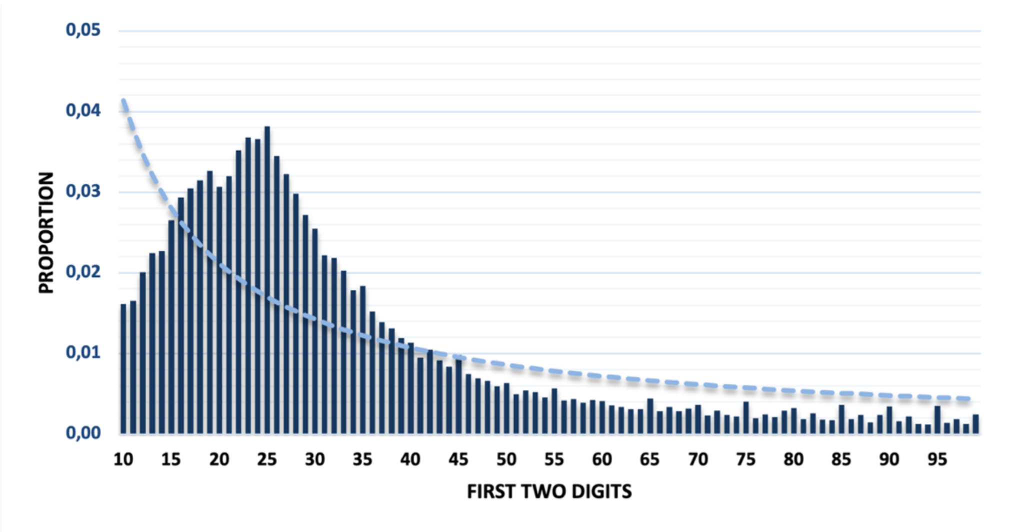

Figure 4.4: The first two digit distribution for 2013-2018

Table 4.4: Results for the first two digit test

Period	Mean Absolute Deviation (MAD)	Kolmogorov-Smirnov test
2013-2018	0.00534	D = 66.389, p-value < 2.2e-16

Figure 4.4 displays the first two digits proportion of our dataset over the whole period where the blue dotted line represents the probability of the first two digits for Benford's law. The two first digits that are above the Benford proportions is between 16 and 40. The two first digits that have the highest probability to occur is 25, meaning that sale price that started with 25 was the most occurring under the period of 2013-2018, sale prices that includes in this rage is 25 000, 250 000 or 2 500 000. It is also possible to observe small spikes for 35, 45, 50, 55, 65, 70, 75, 80, 85, 90, 95 and 99. And these spikes are more noticeable for higher numbers of the first two digits. Since we have a low p-value we reject the null hypothesis as can be seen from table 4.4 for the Kolmogorov-Smirnov test. The Mean absolute deviation gives too high value for sale price to conform to Benford's law as the nonconformity range for the first digits is 0.0022 and over (table 2.2).

4.5 The second digit test

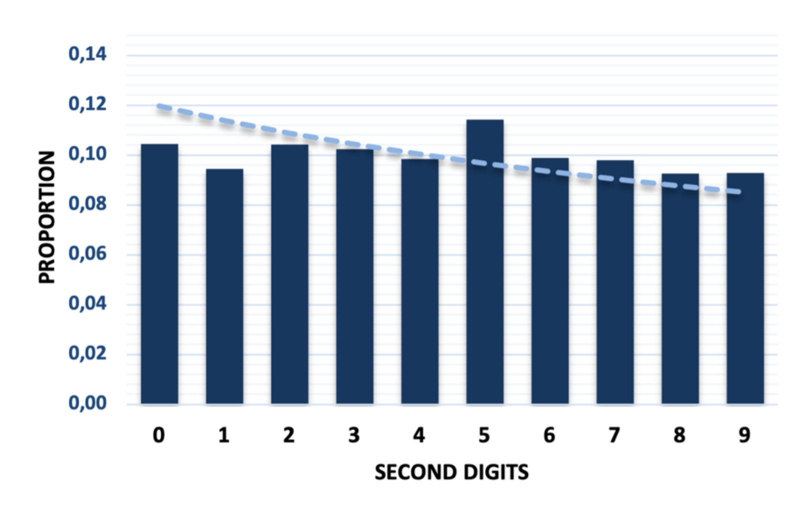

Figure 4.5: The second digit distribution for 2013-2018

Table 4.5: Results for the second digit test

Period	Mean Absolute Deviation (MAD)
2013-2018	0.00866

The second digits as earlier mentioned, are more uniformly distributed than the first digit which can be seen in figure 4.5. A visual inspection shows that number 3 and 4 is closest to Benford, followed by 2, 8 and 6. Number 1 and 5 deviates from Benford the most at opposite directions. The Mean absolute deviation gives a value 0.00866 (table 4.5) in the acceptable conformity range for the second digits is which is from 0.008 to 0.010 (table 2.2).

5 Discussion and conclusion

From the visual inspection of the mantissas from figure 4.1 shows that sale price data deviates from the fitted line which corresponds to the uniformly distributed mantissas [1,0). Although data can still show conformity, it is not expected for sale price data to obtain a perfect fit of Benford's law for both the first and first two digits due to this deviation. We will not interpret how the magnitude of the deviation of our results for the mantissas will affect the conformation for Benford's law.

The most important step before using Benford's law as a tool for analysis is to expect that the data used should conform to the Benford distribution. These steps were presented in the method section "First step". Our data hold some of the criterions for conformation such as having a large sample size and more small values which can be seen in table 4.2, this make data have a positive skewed distribution. And as earlier stated that financial data, such as sale price is often confirmative to Benford's law. Another important factor for data to show conformation is that the observations should not be too centered around the mean. In the earlier stage of this thesis did we only have data covering the central parts of the three municipalities, but as we learned more about Benford's law and the importance of variation in data due to (Hill, 1995) discovery of random sampling from random distributions which was also confirmed by (Nigrini, 2020, p. 94). This additional knowledge made us extend the geographical area we study from the central parts to all districts within each municipality in order to obtain more variation, which further increased our sample size. There are three options in order to increase the variation in sale price, first is to extend the geographical area, second is to increase the observed time period and third is doing both. The other concern we had in the early stage of the study was how long time period we should include and the answer to this was not clear because one of the main focus for choosing right data is to have a big sample size so there were studies made based on daily data and others for longer periods. But the big sample size argument is because it often gives larger variation in the data and with the extension of the geographical area which gave us 171 643 observations. The reason for expanding the geographical area is to include a larger area in Stockholm, Gothenburg and Malmö with both low and high sale prices.

The Mean absolute deviation test for the first digits gave a value of 0.04717 (table 4.3) which lies in the nonconformity range (table 2.2). The Kolmogorov-Smirnov test gave us a low p-value which rejects the null hypothesis that the first digit distribution of sale price for tenant-owned apartments for the whole period 2013-2018 do not obey Benford's law (table 4.3). Suspicion was raised when the first digit test for each individual year also did not obey Benford's law (Appendix B). The reason to be suspicious arise when one expects to obtain confirmative results and receive the opposite. The first action to take is to understand why this occurred. Observing figure 4.3 shows that number 2 is more often occurring than number 1 as the first digit, one possible reason for this can be the amount of sale price in relation to the municipality in question together with the constant growth rate argument as mentioned in the "Benford's law" section. Observing table 4.2 we see that the first digit can be 1 between the sale price ranges of 150 000-199 999, 1 000 000-1 999 999 and 10 000 000-19 999 999, and further observing figure 4.2 will show us that the most observations occur in the 1 000 000-1 999 999-sale price interval. It is however not so convincing to expect that the majority of tenant-owned apartments to be sold for any of the price intervals, especially in Stockholm during the studied time period. The first sale price interval is too low, the second one is the most occurring, but we surely think this applies in areas outside the central part of the cities and the third sale price interval is little too high but most likely to occur central areas. The gap between the second and the third sale price interval is large, meaning the time it takes to go from each sale price interval takes time. Recalling that we are taking the time interval for the studied period in consideration, other unobserved factors lie behind the reason for this pricing such as the inflation rate and if the study had a longer time interval or more cities would it be possible to observe other patterns.

The Mean absolute deviation test for the second digits gave a value of 0.00866 (table 4.5) which lies in the acceptable conformity range (table 2.2), meaning the second digits distribution of sale price for tenant-owned apartments for the whole period 2013-2018 obey Benford's law. We did expect that the numbers 0 and 5 would occur significantly more times than other numbers as the second digit due to round numbers, which can arise as a result of a negotiation. We can observe that the number 5 is occurring most of the times as the second digit, as expected, but the number 0 is not significantly higher than the rest of the numbers as seen in figure 4.5. Only having number 5 as a spike shows the effect of round numbers but it may also indicate that the average sale price has so many digits that there is more room for negotiation rather than settling for a value with 0 as the second digit. For example, sale price of 2 000 100 SEK, 2 050 000 SEK or 2 098 900 SEK, the difference in value when the second digit is 0 may not be so rewarding for the individual selling their apartment. The second digit showing conformity to Benford's law do not indicate that the whole data follows Benford's law, but it was a surprising result.

The Mean absolute deviation test for the first two digits gave a value of 0.00534 (table 4.4) which lies in the nonconformity range (table 2.2). The Kolmogorov-Smirnov test gave us a low p-value which rejects the null hypothesis that the first two digit distribution of sale price for tenant-owned apartments for the whole period 2013-2018 do not obey Benford's law (table 4.4). The results for the first two digit test did not come as a surprise since the first digit did not obey Benford's law. We observe the frequent occurrence of number 2 that was first seen in figure 4.3 is also observable in figure 4.4 where the first two digits between the range of 20 to 29 is significantly over the Benford proportions. The spike seen in the second digit figure 4.5 for number 5 is also observable in the first two digits figure 4.4.

No further investigation will take place for this data as earlier mentioned, it is always important to analyze a dataset that do not obey Benford's law as it not always indicates fraud or some error. But let's assume that all criteria for our data to conform to Benford's law holds and we now proceed to test if data really do obey Benford's law. Firstly, the data in question is the sale price of tenant-owned apartments, a non-confirmative result could indicate manipulation of some sort, but what can these manipulations be? It can be for example an overvalued housing market or fake bidding. But one must also take the geographical area and timeline for the study into consideration. How convincing is it to assume that the sale price in Stockholm, Gothenburg and Malmö during 2013-2018 is a result of a fake bidding process? But if sale price for the years 2013, 2014, 2015, 2016 and 2017 conforms to Benford and not 2018, could this indicate some overvaluation in the housing market for the year 2018? perhaps. The point is that the use of Benford's law differs, and some may just want to test if a set of data conforms to this phenomenon, and some have the intention to further analyze digits or to find fraudulent activity, but it is always important to make further analyze of the results.

To understand the non-confirmative results, we consider all the arguments made for each digit test until now, and it may seem like the data need more observation in order to increase the variation in sale price even more. One factor that possibly can affect our results is the proportion of tenant-owned apartments in each district that are included in each municipality. Though each municipality consists of numerous districts with high variation of rent and sale prices, is the proportions of tenant-owned apartments and tenancy in each district in our data not known. With no reference on how high variation data should have, after expanding the geographical area with more observations did (Nigrini, 2021) believe the reason for the non-confirmative results may be because our sample is just covering the years 2013-2018. Future work is needed to determine the conformity of sale price to Benford's law, which possibly can be solved by either expanding the geographical area or the time period due to the low variation in our data.

The participation in a bidding process, as earlier mentioned, are not regulated by law in Sweden which still gives the incitements for possible fake bids. To draw better conclusion about data would one possible solution be to register every bid for each participant in the bidding process when buying an apartment. This would not be completely impossible since real estate firms have started with identification controls via bank-ID which is digital when individuals participate in bidding. If there was data on each bid from each participant, could it be possible to detect abnormal patterns. Such patterns could be for example constant amounts or certain percentage in relation to the starting price. A Benford analysis applied to this could be made to detect outliers with a summation test (Nigrini, 2020, p. 155), which sums the number of amounts for each first two digits.

As earlier mentioned, financial data such as market values are consistent with Benford's law (Nigrini, 2012, p. 20) and if we observe the first digits for individual years in Appendix B is it possible to notice that the first digit distribution for 2013 may not obey Benford's law but it does show Benford tendencies where number 1 occurs as the first digit most of the times and number 2 occurs less times then 1 and so on, but things change at 2014 and 2015. For years 2015 to 2018, number 2 is the first digit most of the times and observing the price development for Stockholm, Gothenburg and Malmö it appears like the price is increasing during the same years, and significantly increasing in Stockholm (Svensk Mäklarstatistik, 2021). So, is it possible that the change in the digit patterns is due to overvalued house market and can Benford's law help to predict housing bubbles? It is little uncertain to draw such conclusions based on our data, but future research should expand the area and possibly find more years before 2013 with Benford tendencies or conformation since we do not know if the patterns found in 2013 is random or a long-term pattern.

As the main goal of this study was to answer if sale price for tenant-owned apartments during 2013-2018 in Stockholm, Gothenburg and Malmö follow Benford's Law? we found from the first digit test (table 4.3) that our data do not follow Benford's law, which was further supported by the first two digit test (table 4.4).

As for the second question this thesis aimed to answer was is we can identify any digit patterns in our data? Which could be found in all digit figures presented, the first digit (figure 4.3, the second digit (figure 4.5) and the first two digits (figure 4.4). Patterns such as round numbers, the most occurring first or second digit and possible explanation to this. One surprising finding was that the second digit of sale price for the whole period (table 4.5) and each year (Appendix D) showed conformity to Benford’s law.

As for the third question this thesis aimed to answer was if sale price for tenant-owned apartments for each year between 2013 to 2018 in Stockholm, Gothenburg and Malmö follow Benford's Law? we found from the first digit test (Appendix B) that our data do not follow Benford’s law, which was further supported by the first two digit test (Appendix C).

References

Aschberg, R. & Mohlin , L. C., 2015. *Aftonbladet avslöjade mäklarfusk - nu ses lagen över.* [Online]
Available at: https://www.aftonbladet.se/nyheter/a/J1LrJ4/aftonbladet-avslojade-maklarfusk-nu-ses-lagen-over
[Accessed 14 05 2021].

Benford, F., 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 31 March, 78(4), pp. 551-572.

Berger, A. & Hill, T. P., 2015. *An Introduction to Benford's Law.* Princeton(New Jersey): Princeton University Press.

Durtschi, C., Hillison, W. & Pacini, C., 2004. The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Account*, 01 01, pp. 17-34.

Endress, M. B., 2014. *Benford's law and humanly generated prices in auction houses and buyout systems of virtual worlds,* Texas: Texas Woman's University Denton.
Available at: https://twu-ir.tdl.org/handle/11274/3642
[Accessed 6 May 2021].

Fastighetsmäklarförbundet, n.d. *Har jag betalat rätt pris?.* [Online]
Available at: https://fmf.se/kopprocessen/har-jag-betalat-ratt-pris/
[Accessed 06 05 2021].

Fastighetsmäklarinspektionen, 2020. *Budgivning.* [Online]
Available at: https://fmi.se/vad-galler-vid-formedling/budgivning/
[Accessed 06 5 2021].

Fastighetsmäklarinspektionen, 2021. *Främsta orsaker för anmälan 2014–2020.* [Online]
Available at: https://fmi.se/media/5xyba4op/tio-främsta-orsaker-för-anmälan-2014-2020.pdf
[Accessed 06 05 2021].

Giles, D. E., 2007. Benford's law and naturally occurring prices in certain ebaY auctions. *Applied Economics Letters*, 14(3), p. 157–161.

Hellekant, J., 2019. *Mäklare vill stävja fusk – inför krav på bank-id.* [Online]
Available at: https://www.svd.se/maklare-vill-stavja-fusk--infor-krav-pa-bank-id
[Accessed 06 05 2021].

Hill, T. P., 1995. A Statistical Derivation of the Significant-Digit Law. *Statistical Science*, November, 10(4), pp. 354-363.
Available at: https://doi.org/10.1214/ss/1177009869

Joenssen, D. W. & Muellerleile, T., 2015. *BenfordTests: Statistical Tests for Evaluating Conformity to Benford's Law.* [Online]
Available at: https://cran.r-project.org/web/packages/BenfordTests/index.html
[Accessed 13 05 2021].

Kolmogorov, A., 1933. Sulla determinazione empirica di una lgge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, Volume 4, pp. 83-91.

Kotz, S. & Johnson, N. L., 1992. *Breakthroughs in statistics.* 2 ed. New York: Springer-Verlag New York, Inc..
Available at: https://doi.org/10.1007/978-1-4612-4380-9_6

Mebane Jr, W. R., 2006. Election Forensics: Vote Counts and Benford's Law. *Summer Meeting of the Political Methodology Society, UC-Davis*, 18 07.

Miller, S. J., 2015. *Benford's Law.* New Jersey: Princeton University Press.

Mittermaier, L. J. & Nigrini, M. J., 1997. The use of Benford's law as an aid in analytical procedures. *Auditing*, 16(2), pp. 52-67.

Newcomb, S., 1881. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4(1/4), pp. 39-40.
Available at: http://links.jstor.org/sici?sici=0002-9327(1881)4:1/4%3C39:NOTFOU%3E2.0.CO;2-K

Nigrini, M. J., 1996. A Taxpayer Compliance Application of Benford's Law. 18(1), pp. 72-91.

Nigrini, M. J., 2005. An assessment of the change in the incidence of earnings management around the Enron-Andersen episode. *Review of Accounting and Finance*, 4(1), p. 92–110.

Nigrini, M. J., 2012. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection.* Hoboken(New Jersey): John Wiley & Sons, Inc..

Nigrini, M. J., 2020. *Forensic analytics: methods and techniques for forensic accounting investigations.* 2 red. Hoboken(New Jersey): John Wiley & Sons, Incorporated.

Nigrini, M. J., 2021. *Advice* [Personal communication, 20 April 2021].

Pinkham, R. S., 1961. On the Distribution of First Significant Digits. *The Annals of Mathematical Statistics*, 12, 32(4), pp. 1223-1230.

Pomykacz, M., Olmsted, C. & Tantinan, K., 2017. Benford's Law in Appraisal. *The Appraisal Journal*, 85(4), pp. 274-284.

Propertini, Inc., 2014. *Maths phenomenon explains why one-third of house prices begin with 1.* [Online]
Available at: https://propertini.com/press/propertini-press-release-benfords-law.pdf
[Accessed 14 05 2021].

Svensk Mäklarstatistik, 2021. *Prisutveckling.* [Online]
Available at: https://www.maklarstatistik.se/omrade/riket/stockholms-lan/stockholm/#/bostadsratter/arshistorik-prisutveckling
[Accessed 11 05 2021].

Svensk Mäklarstatistik, 2021. *Sale price data.* [Personal communication, 12 April 2021].

Wallace, W. A., 2002. Assessing the quality of data used for benchmarking and decision making. *Journal of Government Financial Management*, 51(3), p. 16–22.

Weisstein, E. W., n.d. *Logarithm.* [Online]
Available at: https://mathworld.wolfram.com/Logarithm.html
[Accessed 08 05 2021].

Appendix A: Conformity for each year

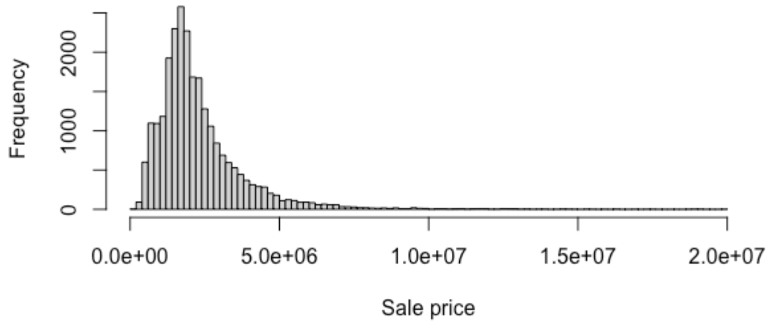

Figure A.1: The distribution of sale price 2013

Table A.1: Descriptive statistics for sale price 2013

Obs.	Mean	Median	Mode	Min	Max
24 801	2 319 140	1 950 000	1 600 000	117 200	20 000 000

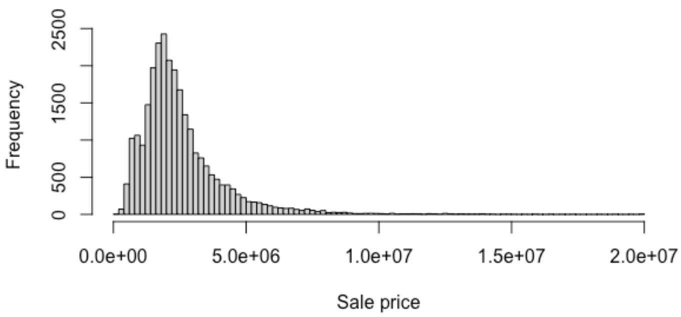

Figure A.2: The distribution of sale price 2014

Table A.2: Descriptive statistics for sale price 2014

Obs.	Mean	Median	Mode	Min	Max
26 470	2 536 371	2 160 000	2 100 000	167 000	20 000 000

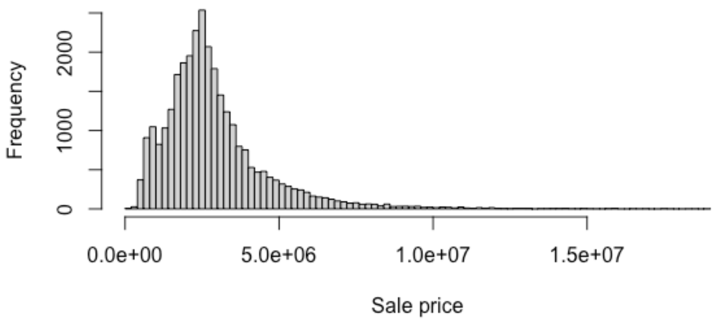

Figure A.3: The distribution of sale price 2015

Table A.3: Descriptive statistics for sale price 2015

Obs.	Mean	Median	Mode	Min	Max
30 209	2 910 966	2 550 000	2 500 000	15 000	19 000 000

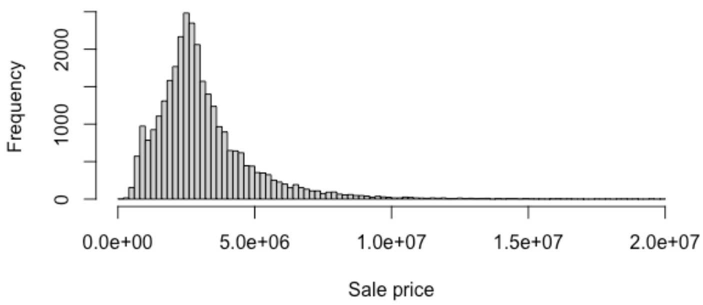

Figure A.4: The distribution of sale price 2016

Table A.4: Descriptive statistics for sale price 2016

Obs.	Mean	Median	Mode	Min	Max
30 655	3 133 574	2 750 000	2 500 000	100 000	20 000 000

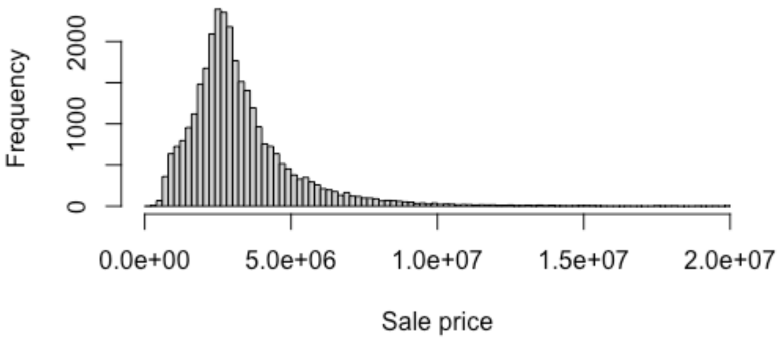

Figure A.5: The distribution of sale price 2017

Table A.5: Descriptive statistics for sale price 2017

Obs.	Mean	Median	Mode	Min	Max
30 622	3 287 384	2 885 000	2 500 000	90 000	20 000 000

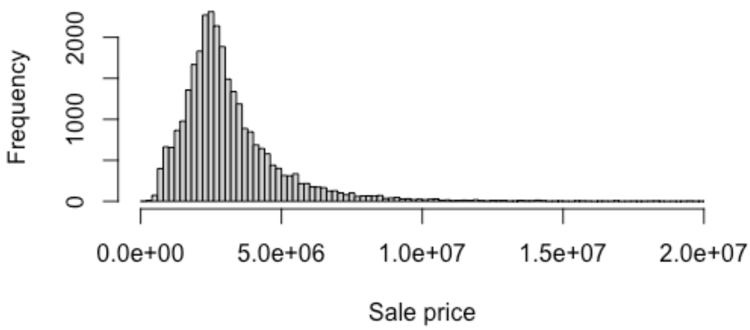

Figure A.6: The distribution of sale price 2018

Table A.6: Descriptive statistics for sale price 2018

Obs.	Mean	Median	Mode	Min	Max
28 886	3 155 393	2 750 000	2 500 000	110 000	19 995 000

Appendix B: First digit test for each year

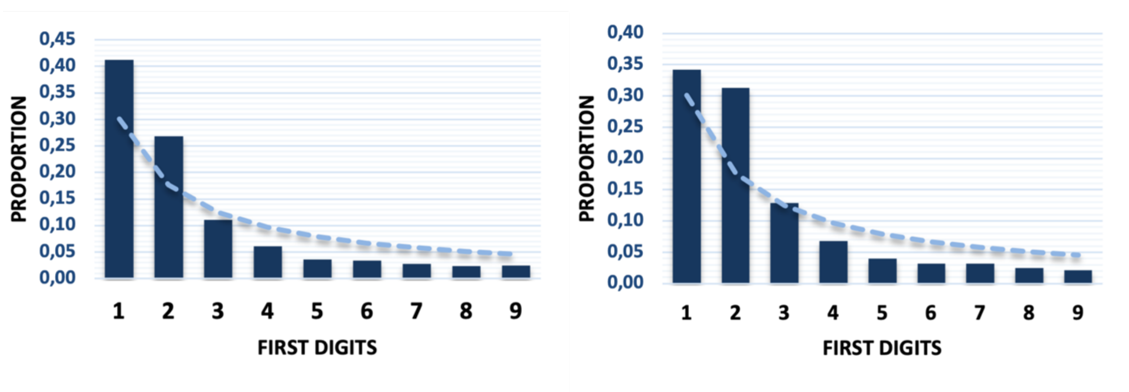

Figure B.1: The first digit distribution for 2013 and 2014

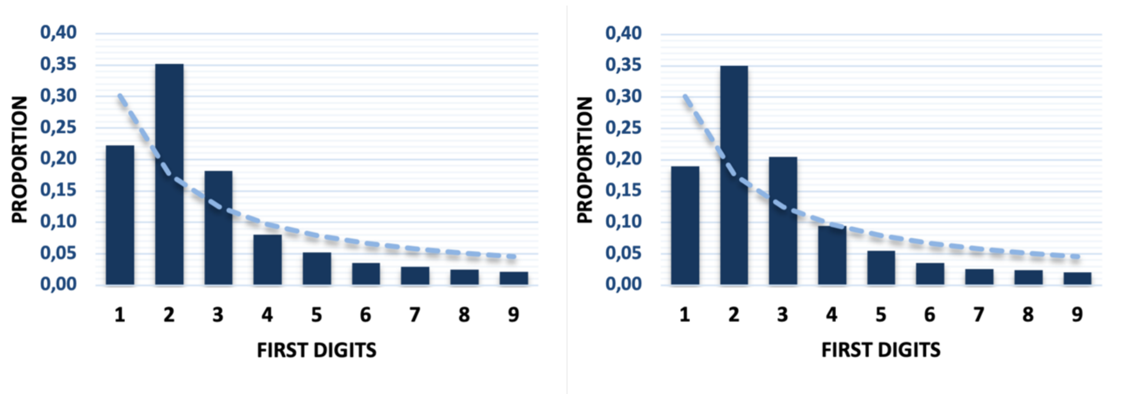

Figure B.2: The first digit distribution for 2015 and 2016

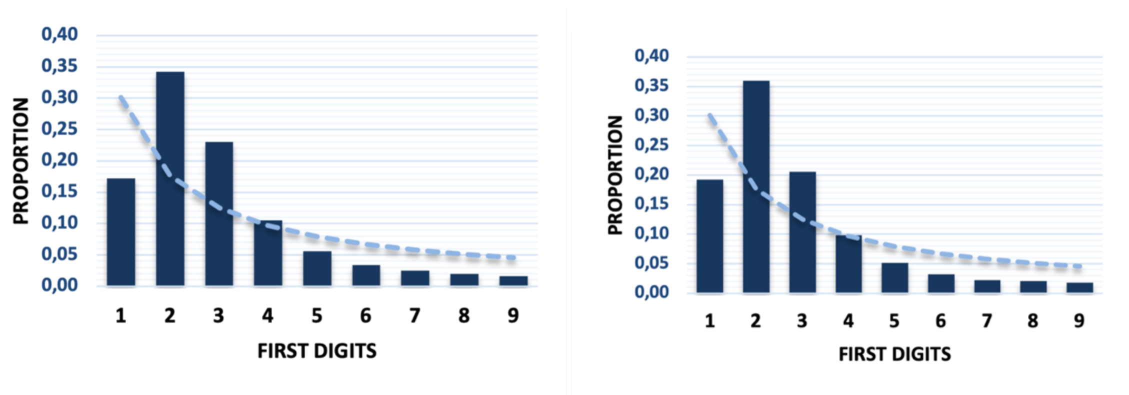

Figure B.3: The first digit distribution for 2017 and 2018

Table 4.1: Results for the first digit test

Period	Mean Absolute Deviation (MAD)	Kolmogorov-Smirnov test
2013	0.04522	D = 32.047, p-value < 2.2e-16
2014	0.04020	D = 29.432, p-value < 2.2e-16
2015	0.05159	D = 26.733, p-value < 2.2e-16
2016	0.05647	D = 25.01, p-value < 2.2e-16
2017	0.06225	D = 26.511, p-value < 2.2e-16
2018	0.05900	D = 26.633, p-value < 2.2e-16

Appendix C: First two digit test for each year

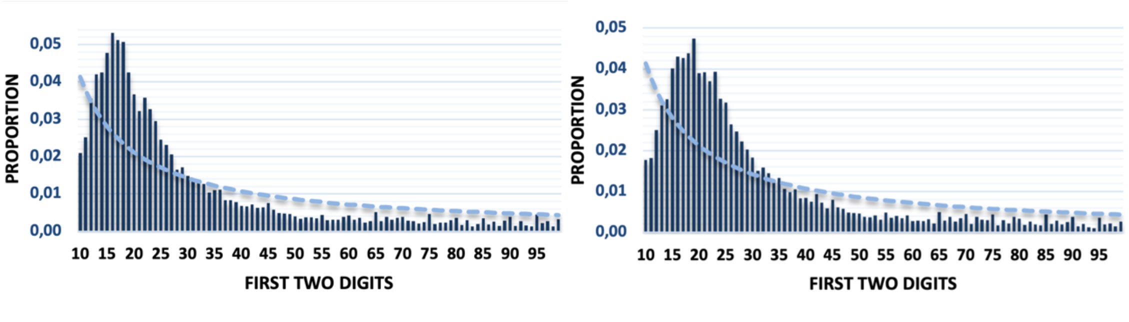

Figure C.1: The first two digit distribution for 2013 and 2014

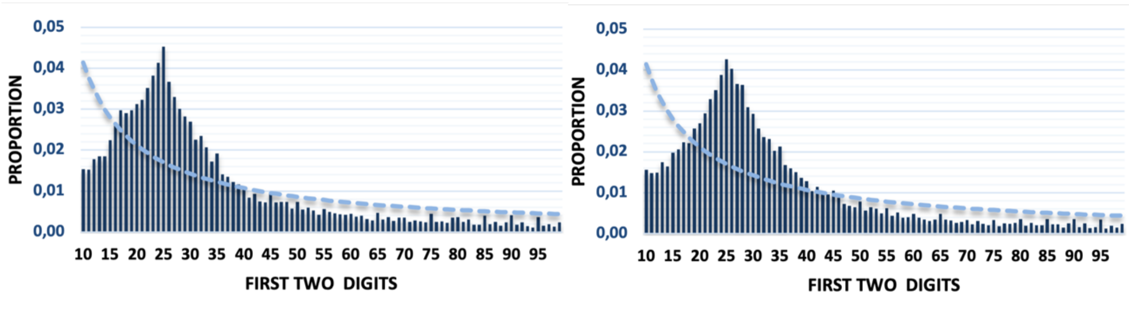

Figure C.2: The first two digit distribution for 2015 and 2016

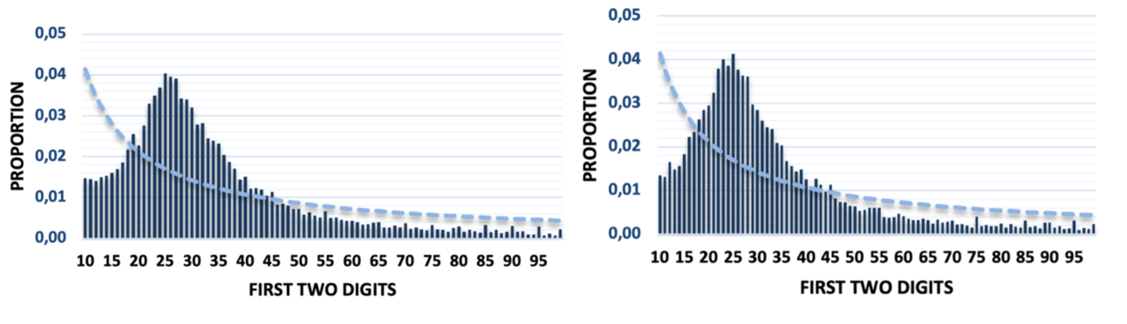

Figure C.3: The first two digit distribution for 2017 and 2018

Table 4.1: Results for the first two digit test

Period	Mean Absolute Deviation (MAD)	Kolmogorov-Smirnov test
2013	0.00527	D = 32.148, p-value < 2.2e-16
2014	0.00537	D = 30.478, p-value < 2.2e-16
2015	0.00556	D = 26.785, p-value < 2.2e-16
2016	0.00582	D = 25.754, p-value < 2.2e-16
2017	0.00639	D = 27.264, p-value < 2.2e-16
2018	0.00625	D = 27.798, p-value < 2.2e-16

Appendix D: Second digit test for each year

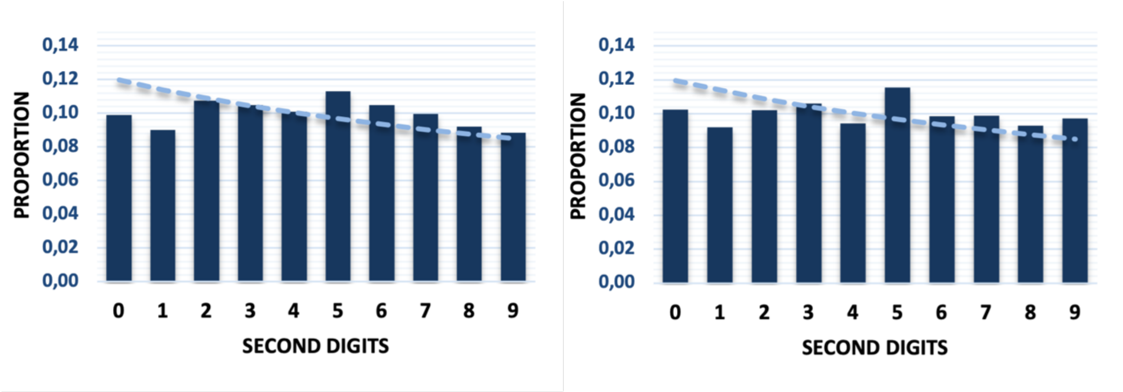

Figure D.1: The second digit distribution for 2013 and 2014

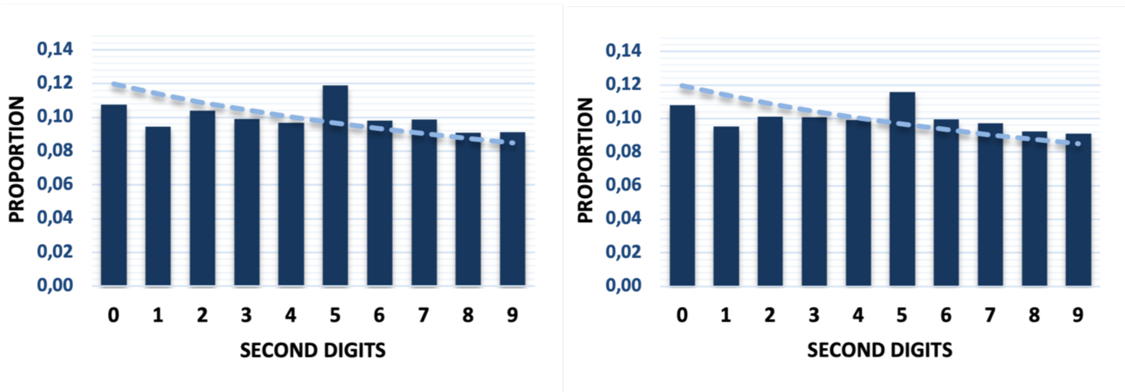

Figure D.2: The second digit distribution for 2015 and 2016

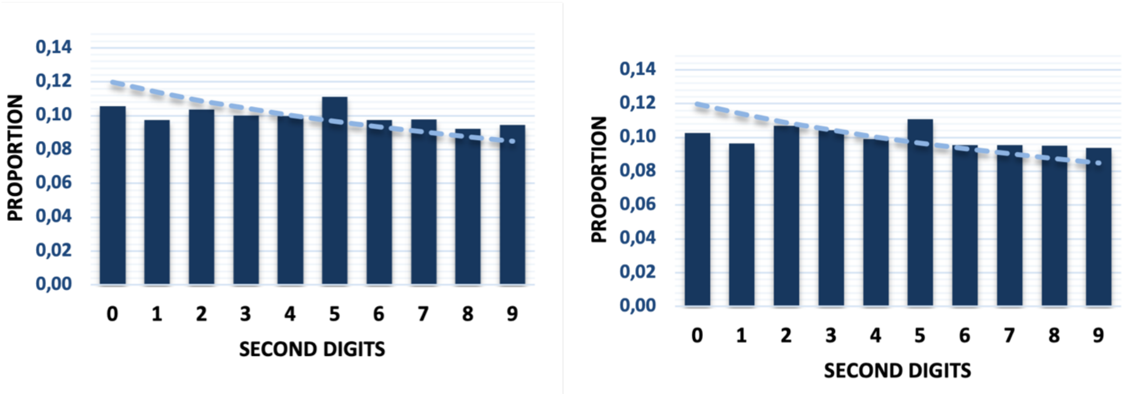

Figure D.3: The second digit distribution for 2017 and 2018

Table D.1: Results for the second digit test

Period	Mean Absolute Deviation (MAD)
2013	0.00919
2014	0.01036
2015	0.00904
2016	0.00852
2017	0.00802
2018	0.00756