Lisbon School of Economics & Management
Universidade de Lisboa

LISBOA
UNIVERSIDADE DE LISBOA

MASTER
APPLIED ECONOMETRICS & FORECASTING

MASTER´S FINAL WORK
INTERNSHIP REPORT

DOCUMENT SPECIFICALLY MADE FOR OBTAINING THE MASTER'S DEGREE

PROVISIONAL DOCUMENT

COMMERCIAL PROPERTY RENT INDEX FOR PORTUGAL: EXPLORING A FISCAL ADMINISTRATIVE MECHANISM - E-FATURA

ANA RAQUEL PINA MAURÍCIO

SUPERVISION:
PROF. DR.ª ESMERALDA RAMALHO
DR. ÂNGELO TEIXEIRA

2022

Glossary:

CAE – Portuguese Classification of Economic Activities

CCI- International Trade Operator type

CPRI – Commercial Property Rent Index

FJR – Legal Form

FUE- Statistic Unit File

INE- Statistics Portugal

JEL – Journal of Economic Literature

MFW – Master's Final Work

OLS – Ordinary Least Squares

SIN – Institutional Sectors Nomenclature

VIF- Variation Inflation Factors

ABSTRACT

This internship provides an insight on the creation attempt of a new rent index for commercial properties in Portugal, using monthly data, from an existing and functioning administrative mechanism - *E-fatura*, over the period 2016-2021.
The mechanism *E-fatura* contains information about the Portuguese companies and individuals' monthly billing. The purpose is to analyse if this information could be enough to translate into a new rent index or a proxy indicator without having to recur to lengthy classic methods, such as adjusting inquiries, or having to do a new type of specific census. In other words, the purpose is to try to extend the utility of an established receipt management mechanism in order to find if it could provide a detailed portrait of the commercial properties' rent universe by itself. The information selected from *E-fatura*, although restricted to companies with economic activity related to the rental of real estate, is not exclusive to the renting universe of commercial properties and in that sense, it had to be limited and adapted to be in this line of work. Hedonic models were used for the index compilation and the method chosen was the time-dummy approach, first, by the adjacent time-dummy approach and second, by the (all-periods) pooled time-dummy method. After testing the different methods along with theoretical basis, a comparison between them is done, followed by an evaluation about the future viability of the index or proxy indicator. It was concluded that by using this approach, with all limitations considered, the computed index could capture macroeconomic dynamics and represent an overall correct picture of the market.

KEYWORDS: Commercial Properties; Price Indices; Hedonic Models; Time Dummy Method; Stratified Mean

JEL CODES: B23, C43, E31.

RESUMO

Este relatório de estágio descreve a tentativa de criação de um novo índice de rendas para propriedades comerciais em Portugal, com dados mensais, através de um mecanismo administrativo, *E-fatura*, durante o período de 2016 a 2021.
O mecanismo *E-fatura* contém informação sobre a faturação mensal de empresas e pessoas singulares portuguesas. O objetivo é analisar se a informação proveniente do mecanismo seria suficiente para calcular um novo índice de rendas sem ter de recorrer a métodos clássicos mais demorados tais como censos ou um novo questionário específico para esta situação. Em outras palavras, o objetivo é tentar alargar a utilidade de um mecanismo já estabelecido de gestão de faturas de forma a analisar se o universo de propriedades comerciais poderia ser retratado por este. A informação derivada do *E-fatura* apesar de ter sido restrita apenas a empresas com atividade económica relacionada com o arrendamento de imobiliário, não é exclusiva a este universo e, como tal, necessitou ser limitada e ajustada para enquadrar o propósito deste relatório. Para a agregação do índice foram utilizados modelos hedónicos pela abordagem *time-dummy*: primeiramente pela *time-dummy* adjacente e consequentemente pela *time-dummy* englobando todos os períodos (pooled time-dummy). Após testar diferentes métodos, tendo por base a teoria apresentada, estes métodos foram comparados e avaliados quanto à sua futura viabilidade como índice. Com a abordagem de *time-dummy* adjacente, foi concluído que o índice final calculado tinha a capacidade de capturar dinâmicas macroeconómicas e representar o mercado de propriedades comerciais de uma forma geral.

PALAVRAS-CHAVE: Propriedades Comerciais; Índices de Preços; Modelos Hedónicos; Método *Time Dummy*; Média Estratificada

CÓDIGOS JEL: B23, C43, E31.

Index

Table Index

Figure Index

ACKNOWLEDGMENTS

Firstly, I want to thank Dr. Ângelo Teixeira for all the support, patience, and shared knowledge, which were indispensable for the realisation of this report and for my future professional life.

To Prof. Drª Esmeralda Ramalho for being available and supportive along every step of the report. Also, a thank you to all professors of the Econometrics & Forecasting Master's for all their teachings, which without them this report would not be possible.

To INE, for providing me with this opportunity, in particular to Dr. Vitor Mendonça and all other members of the National Accounts Department (DCN) for making me feel welcome and providing me with good conditions throughout the internship.

Least but not lastly, to my family for all the support, patience and love they have given me throughout my life. Thank you.

1. INTRODUCTION

National Statistics play a major role in providing an unbiased picture of a country's economy, society, and environmental state. The information provided by National Statistics is crucial for a country's development. Patterns can be found such as trends and cycles. Information provided by National Statistics unravels the strengths, weaknesses, and threats to the general wellbeing of a society. Since this information covers data geographically and over time, it can represent guidelines for government policies. One of the main issues with National Statistics comes from the lack of adequacy of available information about specific subjects inside different sectors, which makes the process of obtaining a rich data base with all the information needed almost unreachable. A consequence of such is that only the simplest estimation methods can be implemented to obtain quantities of interest. The main source of data comes either from census or specific inquiries, both lengthy and costly processes. These mechanisms are hard to adjust to the problems evolution' with a long interval time-window which also presents a problem. The information available is, for that reason, reduced with usually missing observations. More and more, considering the information needs, administrative data has started being used to try to compute new indicators for those sectors; Nordbotten (2010). This report was an exploratory analysis of an existing administrative database in Portugal to assess the possibility of using it to produce a statistical indicator for the commercial properties market. As stated above, there is a lack of adequate data for some economy sectors which does not seem to be countered by classic methods. In order to fill that shortage, the use of administrative mechanisms could present an addition to those classic methods.

The goal of this report is to use the information available in *E-fatura* from 2016-2021, to attempt to create an index or proxy indicator for the commercial property's market where there is room for innovation. More specifically, the objective is to study the use of the invoicing between companies whose main activity (identified by CAE) is the renting of real estate – these companies will be referred as the issuers - and other companies from several economic activities - the acquirers – which would rent these properties. These invoices will be used as an input to compile an index of the value of rents charged on commercial properties using a hedonic model approach. The information will be adjusted so that we are only working in the renting universe of enterprises as property owner´s, or

the closest possible to that scenario[1] and will be split into four different main groups in accordance with acquirers' CAEs, with the purpose of obtaining information on different sectors.

Administrative data has been increasingly more available to the construction of statistics since it has been tested to have a significant effect on countering the lack of data stated above without having to recur to the classical, costly methods. In recent years, there were already some approaches to construct price indexes using different types of administrative data. The following bibliographic references constitute a basis for this report and an example of the above. In Evangelista & Raposo (2016) a **transactions-based commercial property price index** was derived from administrative data on transfer and property tax records by a hedonic regression using stratification. The results have shown it was possible to compute a coherent index from that type of administrative data. Another attempt on a similar price index with administrative data is the **house price index** that uses data from two different taxes-Municipal Tax on Real Estate Transfer (IMT) and Local Property Tax (IMI), given by Evangelista & Teixeira (2014). This House Price Index aims at measuring price inflation of dwelling purchases for residential purposes. As method, a hedonic price index was implemented using the adjacent time dummy approach. Both the House Price Index and the previously mentioned index, the Commercial Property Price Index, are in current production by Statistics Portugal. One more paper useful for the development of this report is the **Rent Price Index;** Evangelista, Moreira & Teixeira (2019). The dataset used for this index was a combination of different administrative records from Portuguese Tax and Customs Authority (AT), the National Energy Agency (ADENE), Census and other information. Using hedonic methods, the final remarks underline that the results were coherent and therefore could be a beginning for future developments of a commercial property rent index[2].
The report is structured as follows. In Chapter 2, a theoretical concept scope along with a description of the methodology which will be applied further. In Chapter 3, there is a

[1] In Portugal, in the case of individuals as owners of commercial property, a specific system is defined for rental contracts. Most landlords, whenever they start a new rental contract, or issue a new receipt, have to communicate that information to the Portuguese Tax and Customs Authority. Regarding properties owned by collective entities there is no such system, therefore this work seeks to analyse that portion of the market.

[2] This index only covers properties owned by individuals who are obliged to issue electronic receipts.

description of the data along with the adaptations/restrictions applied. In Chapter 4, the results of the different strata are considered and discussed along with the overall combined index. Finally, in Chapter 5 concludes and underlines a few final remarks for future development.

2. Theoretical Concept Scope

2.1. Hedonic Price Models

"Etymologically, the term "hedonics" is derived from the Greek word hedonikos, which simply means pleasure. In the economic context, it refers to the utility or satisfaction one derives from the consumption of goods and services"; Chin & Chau (2003). Hedonic price models aim to explain the price of an asset according to the premise that price is determined by two forces: the asset's internal characteristics and external factors that may affect its price. Therefore, these models capture a consumer's willingness to pay for what they perceive are differences between the assets which may increase or decrease its intrinsic value. They are usually applied to the real estate market, since the price of a property is both defined by its internal characteristics – e.g., if the building has an elevator, how many rooms has the property - but also by the external factors- e.g., its closeness to local businesses and transportation services, or if the neighbourhood is deemed problematic. One of the first economists to present the usefulness of hedonic models to counter limitations of the classic consumer's theory was Lancaster (1966).

Lancaster (1966) extended the consumption theory analysis to its application into the production theory. A crucial assumption was that goods possess several characteristics on which consumers' preferences rely on instead of just on the good itself. The conventional consumer's theory given by early economists like Slustky & Hicks-Allen to Debrew(1960) and Uzawa (1960) went through "a long process of refinement" which stripped the theory of all irrelevant hypothesis, following a parsimony principle on how to extract the minimum of results from a minimum number of assumptions. In this framework any good is a good, no matter their characteristics.

Lancaster (1960) wanted to prove the intrinsic characteristics of a good were indeed relevant to the consumer's theory. Previous to the work by Lancaster some other economists took a similar approach like Morishima (1959), Quandt (1956) and Becker

(1965) but these economists presented approaches to specific problems. Lancaster instead of trying to find a solution to a specific problem, presented a general alternative of the traditional consumer's theory. He concluded that a model which took in consideration different characteristics in goods, was richer in heuristic and predictive explanatory power than the model estimated according to the conventional consumer's theory. Also deals easily with common sense characteristics of actual consumer's behaviour. Dynamics which were missing in traditional approaches.

Later another economist, Rosen (1974), presented an evolution from Lancaster's ideas by deriving a model of product differentiation based on the hedonic hypothesis that goods are valued according to their bearing attributes/characteristics as a package. Econometrically, implicit prices are estimated by regressing product prices on their intrinsic characteristics in the construction of hedonic price indexes. Rosen's ideas differ from Lancaster's manly in the fact that arbitrage is assumed impossible on the assumption of indivisibility. Rosen assumes that a package englobed by goods with similar characteristics give the same level of utility therefore cannot be untied and concludes about the consequence of the constructing implicit markets for characteristics embodied in differentiated products. When goods can be treated as a tied package of characteristics, observed market prices are comparable in those terms as well. The economic content of the relationship between observed prices and observed characteristics becomes evident once price differences among goods are recognized as equalizing differences for the alternative package they embody. Hence, estimated hedonic price-characteristics functions typically identify neither the demand nor supply functions. Throughout the years many other economists, either refuted their claims or tried to develop further these ideas since both present good basis for development. Although the model has its disadvantages such as the specification of each goods characteristics to explain the price hedonically, it also has significant advantages when it comes to other methods.

Freeman (1979) underlined the main advantage of using hedonic models for the housing market being the fact that the inputs needed to compute the models are very low compared to other methods. "(…) One only needs to have certain information, such as the property price, the composition of housing attributes, and a proper specification of the functional relationships. It is a straightforward approach because only the coefficients of the estimated hedonic regression are needed to indicate the preference structure. (…) The

implicit price of each housing attribute, ceteris paribus, can be derived from the regression coefficients. Thus, the hedonic price approach allows us to estimate the individual effects of each housing attribute on housing prices, holding all other factors constant."; Chin & Chau (2003). The hedonic price models applied to the properties market, specially to commercial properties, face a challenge of trying to find equal characteristics and grouping properties which are very heterogeneous. Due to this market heterogeneity, an index without any type of adjustment will be distorted by quality differences between properties in different periods. For quality adjustment purposes the sample will go through stratification to diminish this distortion on quality differences from one period to the other. Thus, several approaches were followed, with main focus on the adjacent time-dummy approach.

2.2 Methodology Framework

Hedonic Models are used in Portugal for Housing Statistics such as for compiling the House Price Index and are also used in several other countries for the same purpose such as in the United States of America (Sander & Polansky, 2009), France (Gouriéroux &Lafferrére, 2009), China (Zheng et al., 2010), Japan (Shimizu et al.,2010) and others. A hedonic price model will be used in this report to try to compile an index capable of capturing the rental effects through the monthly invoicing data gathered by *E-fatura*. Unlike other administrative data, like the Local Property Tax (IMI) or Municipal Transfer Tax (IMT), *E-fatura* does not present direct information about the property's so-called characteristics, such as the number of floors or the area, of the property. Instead, its focus is on the business characteristics (e.g., the business volume or the number of employees). The challenge is to check if it would be possible to obtain a viable index through a hedonic price model just based on this data[3].

The hedonic price model will be combined with stratification (detailed below in section *3.3*). This will allow a better understanding of the behaviour and characteristics of each sector. In terms of compilation methods, panel data model approaches could also be an

[3] Each record available in the database represents the total monthly amount invoiced by a issuer whose economic activity is the rental of real estate, and a acquirer. Both the issuer and the acquirer have been restricted to companies, with the intention of covering only the rental of commercial property.

option, still the hedonic model approach using stratification was preferred over this one, chosen accordingly to the methodology document for conducting commercial price indices by INE[4]. There will be four main strata- Retail Commerce, Wholesale Commerce, Services and Industry. Each stratum will represent a sub-index with a different specification regarding the hedonic price model.

For additional quality adjustment purposes, the sub-indices will follow the adjacent time dummy approach, the same methodology used by INE to compute indices of this sort (e.g., Commercial Property Price Index or the House Price Index). In this method the monthly price change is estimated by using two adjacent months at a time. This approach follows the assumption the parameter's coefficients remain constant between those two adjacent months. It represents a particular case of the rolling time-dummy approach introduced by Shimizu et al. (2010). In this method the estimation is done over a fixed number of periods. It functions in a chained logic, i.e., when new information is available the model can be re-estimated by just rolling the time-dummy by one period. The minimum time-window is two and that is why the adjacent time dummy is a particular case of this one. It refers to minimum time-window of a rolling time dummy approach. A limitation of the general rolling time-dummy method is that if all periods are used for a time-window the method is reduced to the simple time-dummy method. The pooled time dummy method assumes the characteristics coefficients are constant through all periods (Triplett, 2006), meaning that they have the same impact on the dependent variable, which is a highly unrealistic approach specially in the commercial properties' universe, where there are multiple different factors affecting each period and it is important to distinguish those in order to better understand the behaviour of the market. Despite being unrealistic for this case, due to the commercial properties' universe itself and due to the limitations of the data, the simple time-dummy was also computed for this report for comparison purposes. Non adjusted values, or simple stratified means, were computed as well to compare the evolution of each method and get a clearer view of the overall picture. All these methods results are represented graphically in appendix 2.

[4] The addressed document only refers three strata- Commerce, Services and Industry/Warehouses. The division of the sector Commerce into Wholesale and Retail as different strata was not suggested by the document.

The adjacent time-dummy approach provides information for all pairs of adjacent months in the sample m= (M-1, M), for each invoice observation i=1,…, N and for every chosen characteristic k=1,…,K. Each stratum has its own regression, since each one has different characteristics.

Empirically, this approach can be applied as in the following hedonic regression:

$$Log(P_{i,m,j}) = \alpha + \sum_{k=1}^{K} \beta_k X_{i,k,j} + \theta D_{i,m,j} + \epsilon_{i,m,j} \quad (1)$$

Where:

- Log($P_{i,m,j}$) is the logarithm of the value of the ith invoice transaction in month m and strata j;
- $X_{i,k,j}$ is the value of the kth characteristic of the ith invoice transaction of the strata j;
- $D_{i,m,j}$ is the adjacent time-dummy indicator:

$$D_{i,m,j} = \begin{cases} 1, if\ m = M \\ 0, otherwise \end{cases} \quad (2)$$

- $\epsilon_{i,m,j}$ is the error term which is assumed to satisfy the conditional mean zero assumption:

$$E(\epsilon_{i,m,j} | X_{i,1,j}, \dots, X_{i,k,j}, D_{i,m,j}) = 0 \quad (3)$$

Being an adjacent time dummy method, θ and the rest of the parameters may vary between each pair of adjacent months unlike in the classic pooled time-dummy method, which implies the coefficients to be the same through all periods, as mentioned above.

To obtain the sub-indices values between two adjacent months, for each stratum j and each year y, one needs to compute the following simple transformation:

$$I_{y,m,j} = \exp(\hat{\theta}) \quad (4)$$

Let:

I: Sub-index values;

P: Total invoice values;

y: year with y =2016, 2017, …, 2021;

m: month with m =1,2,3, … ,12;

j: strata with j =1,2,3,4;

With the aim of compiling all sub-indices into a general Commercial Property Rent Index (CPRI), weights for each one of the strata had to be computed. These weights, in year y, are computed using the total annual invoice values ($\sum_{j=1}^{4} P_{y,j}$), of all transactions between the issuer and the acquirer, and the total annual invoice values for each stratum ($P_{y,j}$). The computation for the indices' weights ($v_{y,j}$) for each strata j in year y was done as follows:

$$v_{y,j} = \frac{P_{y,j}}{\sum_{j=1}^{4} P_{y,j}} \tag{5}$$

On the other hand, each sub-index is updated to the first month, *m*=1, of the first year of the sample, *y*=2016, (i.e., update each subindex to January 2016=100):

$$\frac{I_{y,m,j}}{I_{2016,1,j}} \tag{6}$$

The next step is to update the sub-index values to the base-year 2016 [5] for each strata j (i.e., update each subindex to 2016=100):

$$a_{y,m,j} = \frac{\left(\frac{I_{y,m,j}}{I_{2016,1,j}}\right)}{\left(\frac{I_{2016,m,j}}{\sum_{m=1}^{12} I_{2016,m,j}}\right)} \tag{7}$$

For the chained index, the sub-index values in equation (7), $a_{y,m,j}$, were weighted according to each strata j:

$$a_{y,m,j}.v_{y,j} \tag{8}$$

[5] For the year 2016, there are not values for January, due to the method applied. So, for means the year of 2016 was computed like having 11 months whilst the other years were computed using the 12 months.

And then combined into a single index:

$$q_{y,m} = \sum_{j=1}^{4} a_{y,m,j} \,.\, v_{y,j} \tag{9}$$

Subsequently, the unified indices values ($q_{y,m}$) were updated in two steps. The first was for the values to have as base the average of 2016 (y=2016) obtaining:

$$r_{y,m} = \frac{q_{y,m}}{\left(\frac{q_{2016,m}}{\sum_{m=1}^{12} q_{2016,m}}\right)} \tag{10}$$

The second step was to have these values updated to January 2016, (m=1, y=2016) in a chained method as:

$$d_{y,m} = \frac{r_{y,m}}{r_{2016,1}} \tag{11}$$

The final step in the creation of the final unified yearly index was to average the values according to each year:

$$f_{y} = \frac{d_{y,m}}{\sum_{m=1}^{12} d_{y,m}} \tag{12}$$

Although there is not a general agreement in literature about the correct functional form specification for regression hedonic models (Halvorsen & Pollakowski, 1981; Cropper et al., 1988). The functional formed adopted for this report is the log linear as described above in equation (1). This functional form has several advantages for this case:

1. As can be seen below in the descriptive statistics the distribution is skewed to the right. By logging the dependent variable, the distribution becomes more symmetric and therefore less prone to outliers.
2. All values predicted will be positive, which makes sense in this analysis. There is no negative price transaction or rent value.
3. Using time-dummies in the approach, the price index is parsimoniously obtained by exponentiating the time-dummies' coefficients.

3. DATA DESCRIPTION

3.1. Fiscal Administrative Data Description

E-fatura is a fiscal evasion control mechanism established in Portugal since 2013. It registers the monthly invoicing of enterprises and individuals. Whenever an acquisition of goods or services occurs, the *E-fatura* system transmits this transaction directly to the Tax and Customs Authority. The availability of the invoice does not depend on the buyer's request. The database for this report consists of the information obtained from *E-fatura* between 2016 and 2021 containing 2 981 105 observations[6]. Although it refers to the invoicing of issuers with main activity related to the rental of real estate (CAE's 682 or 68321), it is not exclusive to the rental commercial properties and, consequently, had to be adapted to a collective universe. As we have already mentioned, although these are invoices from companies with CAE's related to the rental of real estate, some invoices may correspond to secondary activities. In addition, it was necessary to guarantee that only commercial properties were covered. Thus, it was applied a B2B (business to business) filter to the data to ensure that only transactions between enterprises were considered, excluding individual agents. In this way it is more likely to be dealing with rental observations between agents with similar characteristics, and more likely to be excluding residential properties rentals. One advantage of this mechanism is that it is in current production, making the process of getting recent, new data easier. If this mechanism would be viable for the creation of an index or proxy indicator, it would help to create a reliable mean from which INE could keep up with the evolution of the commercial properties' market, since the information provided by the mechanism would be available as soon as two months from when it was generated. Some limitations are the quality of the data for the problem at hands. Since the available variables are not directly qualitative attributes of the properties- like their size, the surrounding areas, or the building's attributes. Moreover, the issuer (fiscal identity numbers) NIF's comes associated with the enterprise headquarters, which means one cannot do an unbiased analysis of geographical clusters and therefore obtain a clear spatial picture of the

[6] Some of the variables in the database come from INE's Statistical Units File, which has essential information for each of the fiscal identification numbers issued in Portugal.

invoicing values' behaviour. It should be noted that each record corresponds to the total amount invoiced, in the month, between the two agents. Furthermore, there is no indication of the number of items, or in this case, properties, to which the invoice value corresponds. There is no guarantee that the billing between the issuer and the acquirer corresponds exclusively to rents, our goal. Thus, all restrictions adopted were aimed at ensuring that each record that remained would reflect invoices relating to rental contracts.

3.1.1 Invoice Value - Dependent Variable

The **invoice values** (o_Valor_Tributavel) represent, as we said before, the amount of monthly invoicing registered between two agents: the issuer and the acquirer. These two agents are identified in the database by their NIF, being the issuer the one who rents and the acquirer the renter. They are united as a single individual by combining them into a single observation, i.e., each combination of NIF's – one NIF from the issuer and other NIF from the acquirer which registers an invoice transaction between them, counts as an individual for the analysis.

Invoice Values						
Mean	*Median*	*Min*	*Max*	*Std. Deviation*	*Kurtosis*	*Skewness*
3 095,32	91,22	-12 924 898,42	41 292 120,45	49 895,97	219 635,10	326,89

Table 1:Invoice Values' Descriptive Statistics, 2016-2021, in euros (Mean, Median, Min, Max)

This variable has a wide amplitude of observations indicated by its standard deviation being equal to 49 895,97, an extremely high value. By the descriptive statistics one can also conclude the variable is far from the normality assumptions and is distributed by a heavy tailed distribution, meaning most of its values are concentrated on the tails instead of in the middle of the sample. This can also be observed by the difference between the mean and median values.

The minimum invoice value is negative. This type of negative invoice relates to payment adjustments or breached contract clauses which may imply a cash reimbursement to one

of the parties. Due to these situations and consequent observations being outliers - representing only 0,5% of all observation in the sample-, not fitting with the general purpose of this report and contributing highly to the sample's high variance, they will be left out of the analysis from this point forward.

The above reflects the heterogeneity of the invoices obtained by the mechanism, some are very low, some are very high, this will be a difficulty for the delimitation of the data to the commercial property universe. Since the invoice observations may belong to completely different products transactions, between individuals with very different characteristics, is difficult to obtain just one general specification.

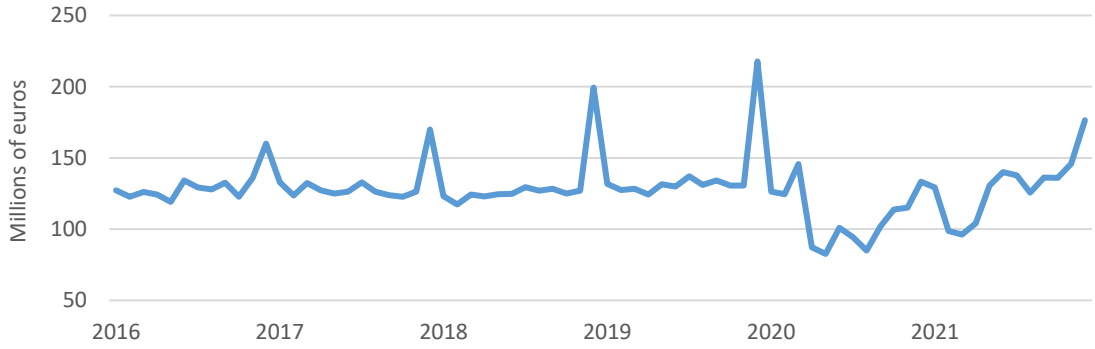

Figure 1:Overall Behaviour of Invoice Values, in millions of euros, 2016-2021

In order to reduce the sample variance and try to get as close as possible to a commercial property universe, some data restrictions are imposed. As can be seen above in Figure 1, the data until 2020 presents signals of seasonality. All spikes represented yearly until that year correspond to the month of December, the exact reasons for this seasonality are unknown but may be justified by payment adjustments before the end of the year, such as paying missing monthly invoices in December, increasing therefore the total invoice value. For certain CAEs like Commerce, which explain approximately 30% of the sample, these spikes could be explained by the increase in number of sales related to the holidays. This topic will not be further explored in this report due to lack of means so far but may present interesting results for future analysis with more data and covariates of interest. The covid-19 pandemic also makes this analysis by now more complicated since it is clear from Figure 1, the pandemic's affects on the structure of the invoices' values. The values appeared to have a similar structure until 2020 with the same peaks of seasonality and similar medians through the years, but after 2020 the overall behaviour of the variable shifts and appears to be returning to previous structure by the end of 2021.

3.1.2 Explanatory Variables

The covariates were extracted from the *E-fatura* and FUE data. Most explanatory variables represent business characteristics of the issuer and the acquirer. Some have suffered transformations during the production of this report to better describe the phenomenon at hands and produce better estimates. Each variable has its observations for both the issuer (*emitente*) and the acquirer (*adquirente*). The original variables were related to the company's state of activity (e.g., suspended, active); legal form (e.g., quota society, anonymous society); international trade operator type (e.g., national/international importer; national/international exporter); number of remunerated employees; institutional sector (e.g., non-financial, financial, public administration); social capital; business volume; total assets; total imports and exports; the company's establishment date; proportion of foreign/private/public social equity; also the fiscal numbers, economic activity code and each company headquarters' location information. Both original and created variables are presented in detail in Appendix 1.

3.2 Restraints, Adaptions and Limitations to the Data

As previously stated, the database is not exclusive to the commercial properties' renting universe, which means adaptations and limits need to be implemented for that effect. An objective is to reduce the data to only transactions between NIF's corresponding to enterprises/businesses and not singular NIF's. For that reason, only collective NIF's are used in the sample. Regarding the dependent variable- **Invoice values**- it was trimmed, negative values were excluded from analysis. Also, the sample presented a high concentration of neglectable values, values lower than 5€, which do not present an economic sense for the purpose of this report. Therefore, the minimum limit applied was the 20^{th} percentile of the variable, or if the value for the 20^{th} percentile was smaller than 100€, it would apply the minimum limit of 100€. The 99^{th} percentile was chosen as the maximum limit.[7] By applying these limits, the descriptive statistics of the variable changed considerably as can be seen in the table below:

[7] The percentiles were computed for all period, 2016-2021, for each acquirer's two-digit CAE.

Invoice Values							
N	***Mean***	***Median***	***Min***	***Max***	***Std. Deviation***	***Kurtosis***	***Skewness***
647 546	6 531,95	2 235,89	100	857 190,37	21 768,82	458,53	17,82

Table 2: Invoice Descriptive Statistics after Invoice Limits, 2016-2021, in euros (Mean, Median, Min, Max)

Comparing Table 1 and Table 2, there are significant differences after limiting the sample. There is now a much minor discrepancy between the mean and the median, this effect is also captured by the abrupt drop in the values registered for kurtosis and skewness, which allowed the analysis to continue with a more homogeneous and less asymmetric sample. A disadvantage from limiting the invoice values is the significant drop in the number of observations. Before there were more than 2 million observations and after limits were applied, just around 600 thousand observations remained. The aim was essentially, obtaining a representative sample of the population of interest. If an invoice transaction between a combination of NIFs is less than 100€ it does not seem plausible that it represents a renting transaction, therefore it would be acceptable to lose this observation. In order to try to get even closer to the commercial properties' renting universe, additional measures were taken:

1. For every NIF combination there was a minimum invoices transaction. Yearly at least one invoice per combination should be present in the sample.
2. For each combination of NIF's there must be more than one contiguous invoice for that observation to be included in the sample.

A dummy variable (*contiguo*), with the purpose of capturing the contiguity of invoice transactions - i.e., if the invoicing for the same NIF combination is present in the sample in two consequent months - was created for this effect.

$$Contiguity = \begin{cases} 1, & if\ invoice\ is\ contiguous \\ 0, & otherwise \end{cases} \quad (14)$$

The idea for this variable is to analyse if the invoices of a NIF combination are not scattered along the sample but, instead, if they are, in a chained fashion, contiguous to one another.

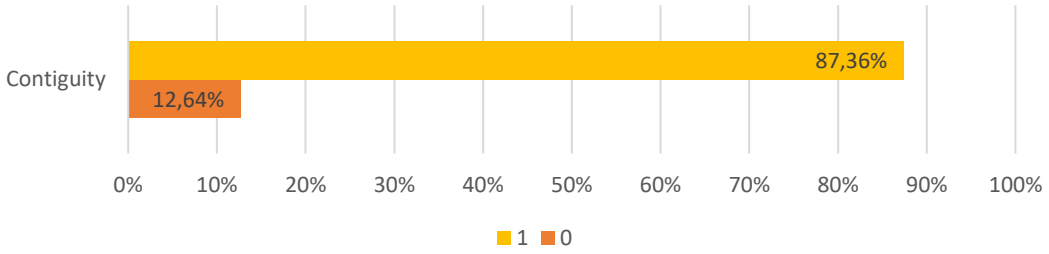

Figure 2:Sample Contiguity Proportion, in percentage, 2016-2021

Figure 2 above presents the invoice values after the limits were applied, the contiguity of the sample proved very high. At least 80% of the sample invoices had contiguity to some degree. Solo transactions and/or transactions that were only present in 2 out of the 72 periods are left out of the sample. These observations are unlikely to belong to the renting universe.

In respect to repetitiveness, a criterion for repeatability was created. Repeatability mentioned onwards represents the invoicing between a combination of NIF's which repeats for at least x% of the whole sample.

Two different types of repeatability were assessed:

a. Repeatability according to Contiguity.
b. Repeatability according to the general proportion of Invoices between a NIF combination.

Both present the same idea: The ratio goes from 0 to 1 and the closer it is to 1-a. the higher percentage of contiguity or -b. the largest percentage of invoicing transactions. The larger this ratio the better, since it presents a continuous or long trade between a combination of NIFS offering an approximation to a regular renting contract, which in this report is expected to maintain year from year for a NIF combination[8]. The two types of repeatability also produced similar results, with more than half the sample with a repeatability rate higher than 0.5.

The focus will be on Repeatability by contiguity. The reasoning for this choice is that the proportion of invoices between a combination of NIF's does not matter if they don't

[8] The following warning must however be given. The guarantee of contiguity of invoicing between two agents tells us nothing about what is being invoiced. In other words, we know that there is billing between both, but we do not know if it corresponds to the same properties, for example. As we will see later, this situation would be easier to analyse if the rent were always a constant value.

approximate the renting universe. In standard renting context, invoices occur usually in a monthly basis (still considering that there are other transaction modalities, such as trimestral) and since the data is monthly, it makes sense to focus on this type of repeatability in detriment of the other[9].

A minimum level of 0.6 repeatability of contiguous invoices was applied. Although the level appears to be high, below this level there are just 9% of the sample. By applying this limit, approximately 91% of the data was still considered. Again, this comes as a trade-off between the number of observations and unbiasedness. As stated above, the focus will be on the unbiasedness of the results, therefore those 9% of observations were dropped from the model.

A more radical limit of 0.9 along with the data without any restrictions according to repeatability will be compared to the results obtained when using the chosen 0.6 level further in the report. The aim will be to show if there are differences in the results just by restraining the number of contiguous invoices per combination of NIFs and if there are differences significant to explain the phenomenon.

In order to obtain a more homogeneous sample, only individuals with observations in all variables were considered. Only using individuals with information on all covariates significantly reduced the problem of missing information in the sample.

Although multicollinearity does not affect the consistency of the model, it does affect its estimates accuracy and consequently their statistical significance. Multicollinearity was analysed according with Variance Inflation Factors (VIF), $VIF_k = \frac{1}{1-R_k^2}$.

VIF is calculated for each covariate. It measures, as a ratio, the variance of a model containing just that covariate with a model with all the covariates. If VIF>10 the covariate indicated a high collinearity problem with some other variable and the specification was adjusted according to this premise. To control for heteroskedasticity, in the presence of which OLS becomes inefficient despite still being consistent, heteroskedasticity-corrected standard errors were used in the analysis for robust results.

[9] In the information received, the period to which the invoice corresponds is not indicated. The contiguity aims to ensure some security in terms of the period to which the invoice corresponds and to be able to compare invoices from identical periods.

A limitation of this analysis is the lack of geographical information in the database. Even though there are variables regarding the NIF'S territorial location (NUTS III), these variables give the information of the enterprise's headquarters and not on the actual property being rented, making a geographical analysis recurring to regional clusters[10] biased and not conclusive. Therefore, it is not included in the model

3.3 Stratification

The stratification considers four strata: Industry, Retail Commerce, Wholesale Commerce and Services. In the Table 3 below are the descriptive statistics of each stratum[11] from 2016 to 2021.

	Retail Commerce	***Wholesale Commerce***	***Industry***	***Services***	***Total Invoice Values***
N	202 468	86 380	49 612	236 128	574 588
Mean	9 710,38	3 235,25	2 309,25	6 073,76	6 603,43
Median	5 335,26	1 459,38	700	1 699,52	2 455,32
Std. Deviation	14 753,22	5 367,34	6569,91	29 010,68	20 914,88

Table 3: Descriptive Statistics by Strata, Mean and Median in euros, 2016-2021

Retail Commerce and Services sectors represent most of the total Invoice's Values. This was expected from theory since these are the two sectors more prone to rental. In fact, their focus is more on the products/services provided to the client than on buying the property where they operate. Enterprises of these sectors are often changing location or expanding their businesses. Also, there are much more new companies entering these sectors' markets unlike in the case of Industry or Wholesales. In these last strata, the

[10] In Evangelista, R., Moreira, H., & Teixeira, Â. (2019) a regional cluster analysis was conducted, and it would be interesting for this universe as well.

[11] The descriptive statistics are given after all previous referred restrictions were applied. For each sector the CAE's were chosen, according to their relevance and respective weight in the sample. A description of the chosen CAE's by sector is available in the appendix.

owners of the enterprises usually find more advantage in acquiring the warehouses/factories where they operate instead of renting.

The high standard deviations through all sectors reflect the heterogeneity of the commercial properties market, especially in the Retail and Services sectors. The difference between the means and the medians, indicate the distribution is skewed. The fact that the means are always higher than the medians gives that the skewness is positive which is a common characteristic of price distributions; Evangelista & Raposo (2016).

Retail Commerce, as stated above, is one of the largest sectors. Shopping centres[12] usually located in the principal cities, rent many properties to mostly salesman but also to services and constitute 67% of all issuers in the sample. These properties often have different characteristics, and with the variables limitation from this database, is hard to distinguish them. Also, if a company is renting a property supposedly on the 1st floor but then decides to stay in the same shopping centre but changes location either to a different floor or different location on that same floor, the rent changes. Another factor which is hard to control for in shopping centre rents are their complexity since they are usually a function of fixed and variable components. The rent in a shopping centre may depend, among other aspects, on the sales value of the establishment in question. Shopping centres constitute a big part of Retail Commerce[13], which is expected specially in a time where outdoors commerce is in decay.

[12] The classification as "shopping centre" was obtained from the list of members of the Portuguese Association of Shopping Centres in the 2017 yearbook. All issuer NIFs in the list used were classified as shopping centre. It is, therefore, not an official classification. The results obtained or disaggregated by shopping centre should be interpreted as a sample of the total universe, since some are not (or were in 2017) members of the association.

[13] Shopping Centres after applying the 0.6 level of repeatability represent around 40% of all Retail observations

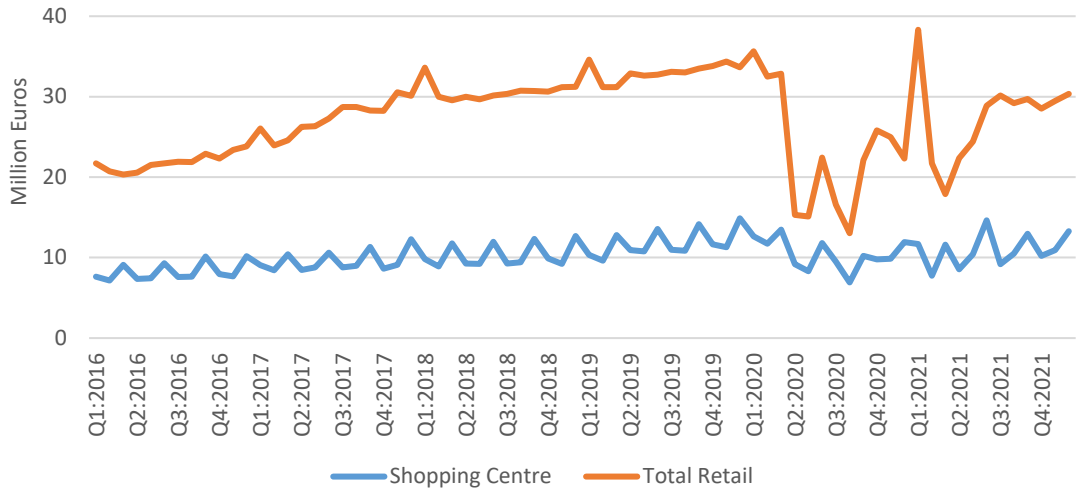

Figure 3:Retail Commerce vs Shopping Centres, Q1:2016-Q4:2021, in millions of euros

The data presented above in Figure 3 presents signs of seasonality through 2016 to 2021, which influences the Retail Commerce strata data (Figure 6). An interesting case is the one of 2020: after the 1st trimester of 2020 the total retail strata registered a significant fall in the volume of invoice transactions whilst the same did not seem to happen in the shopping centres. Although registering a small fall, the volume of invoice transactions remains constant which could be due to the components of the complex rent they apply.

An interesting characteristic of the shopping centres invoice structure is that around 60% of the shopping centres in the sample registered a plus 0,9 repeatability rate for both Retail Commerce and the Services sectors. Services are also highly affected by shopping centres. The overall behaviour of this sector in the sample (see Figure 4) is approximately the same as the behaviour of Services within Shopping Centres unlike in the example above regarding Retail, which proved to be much more seasonal and not so reflective of the overall behaviour of Retail as a whole.

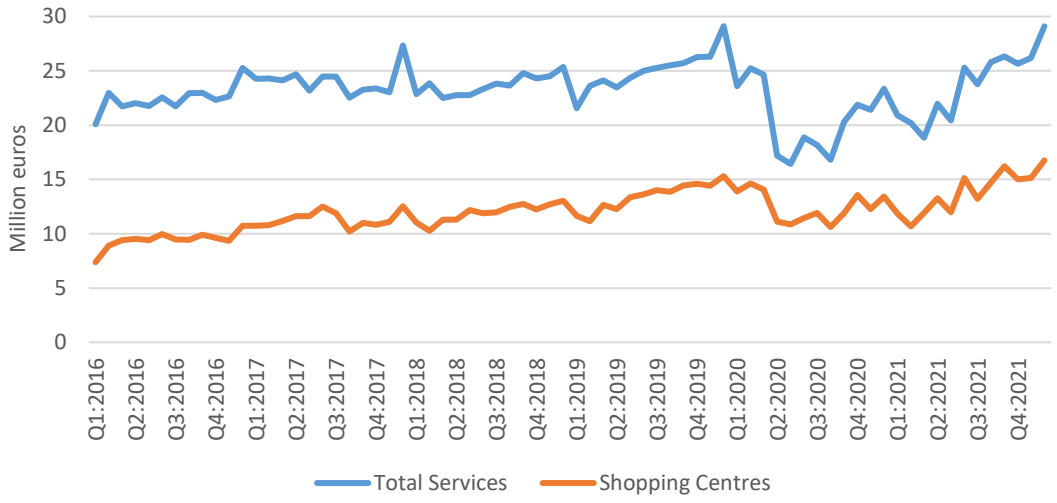

Figure 4:Services vs Shopping Centres, Q1:2016-Q4:2021, in millions of euros

The relevance of shopping centres in these two sectors mentioned was somewhat expected and it is a fact to take into consideration when specifying the sectors and restraining them. Since the specification when there are many observations within the shopping centre sphere may differ from other commercial properties outside it.

The stratification was done according to the acquirer's CAE. In this fashion one is able divide the several observations according to their main economic activity and guarantee the stratification only captures the information of the aimed sector. This is only possible due to the issuer's CAE have been restricted in the sample to only CAE 68[14]. This CAE which represents just real estate activities, serves as a guarantee the transaction between issuer and acquirer rests in the non-individual renting universe.

In Figure A1 and Figure A2 (in Appendix 2) one can see the evolution of invoice transactions value through 2016 to 2021 pre-stratification and post-stratification respectively. Both graphs reflect an overall upwards trend in the sample until the 1st quarter of 2020, point where the pandemic hit and therefore the impacts are there reflected with a very significant drop in the invoice values followed by a recuperation by the 3rd quarter of

[14] CAE3 68 is composed by only two sub-CAE3: CAE3 682 and CAE3 683, properties' rental and non-individual real estate activities respectively.

2020. By the end of the sample the invoice transaction values were once again at pre-pandemic values.

Figure 5:Proportion of Retail & Wholesale in the total of Commerce (on the left); weight of each strata (on the right)

The total values are highly influenced by the Commerce sector, in particular by the Retail sub-sector, this is clear since the impacts shown in the behaviour of Commerce are impulses registered in the behaviour of Retail and consequently reflected in the total behaviour. This can also be explained by the weight of Retail in the overall Commerce, and its overall weight; see Figure 5.

Even though Services also represent a large portion, its behaviour is a lot smoother and constant throughout the years in analysis. Although smoother it is still affected by the pandemic effects but not in the same magnitude.

Wholesale Commerce and Industry are the less representative of the four sectors due to the lack of observations, both do not seem to present any trend, seasonality, or structural break, which appears unrealistic because supposedly the effects of the pandemic were felt by all sectors. Since the sample for these sectors is so small, they may not present an accurate representation.

Both graphs also present seasonality. This seasonality is more obvious in pre-stratification (Figure A1) and seems to be smoothed after stratification (Figure A2). Still after stratification some peaks of seasonality maintain, mostly derived from the Commerce sector which is expected since it is the sector most prone to this problem (e.g., holidays, sales, among others). The peaks are always in the 1st quarter of 2017, 2018 and 2019. In 2020 the structure is understandably different still after the decay in values there was a peak around

the 3rd quarter of the year. In 2021 the values peak again in the 1st quarter and with this peak it was also registered a return to pre pandemic values.

With both the stratification and the restrictions applied, the amplitude of values was trimmed, diminishing the variance of the sample, and consequently providing more reliable estimates.

4.Regression Results

After applying the restrictions and the stratification, several specifications were tested for each sub-index before obtaining the final specifications presented in this report. These tested specifications included not only original variables but also different combinations between the variables presented in Appendix 1 (the combined variables which were not selected for any index are not presented in this report for the sake of space and relevance). The sub-indices were computed as previously stated, by two different methods: pooled time-dummy and adjacent time-dummy. The two methods led generally to similar results. The Adjacent Time Dummy will be the one by which the aggregated index and overall analysis will rely on by the reasons stated in *2.2.* Nonetheless, for comparison purposes, year-on-year rates for each sector were computed by each method along with unadjusted data and the results are included in the Appendix 3.

4.1.Sub-Indices

Each sub-index corresponds to a strata, so there will be four sub-indices. These provide us with a more accurate scenario for each sector, because in this fashion each sub-index has its own specification which allows for the sub-index to be better specified than if they were from the start computed as a whole and the specification would have to be more general and not so fitting. In this way a better interpretation is taken regarding the overall behaviour and characteristics of the sector.

Each specification will be tested through the Ramsey's RESET specification test[15]:

$$\begin{cases} H_0: The\ model\ is\ well\ specified \\ H_1: \qquad Not\ H_0 \end{cases} \tag{15}$$

[15] A 5% level was used for the results.

The specifications are tested for its general form - year by year for each of the 12 months- and by pairs of adjacent months for the time dummy method. There are specifications for some periods which were rejected. There are many reasons for this to happen. A main problem and reason for this is the omitted variable problem, which is also a problem with this database. The omitted variable problem causes the estimators to be biased. Other reason may be problems which cause macroeconomic unbalance, like the Covid-19 pandemic, which created structural breaks, causing the specification that worked before to no longer fit the data after the unbalance. The lack of data for a specific year whereas in the next or previous year there is more information, may also cause the specification to vary in those years. So, some degree of flexibility was applied in this report when it comes to the specifications. The models for periods which indeed were not rejected by the specification test represent the majority of the sample and the acceptance of the specification for periods in which the rejection was suggested is motivated by the forecasting approach followed in the index construction. Results for the RESET test and Estimates for each specification are presented in Appendix 3.

Below follows a description of each sub-index's overall results.

4.1.1. Retail

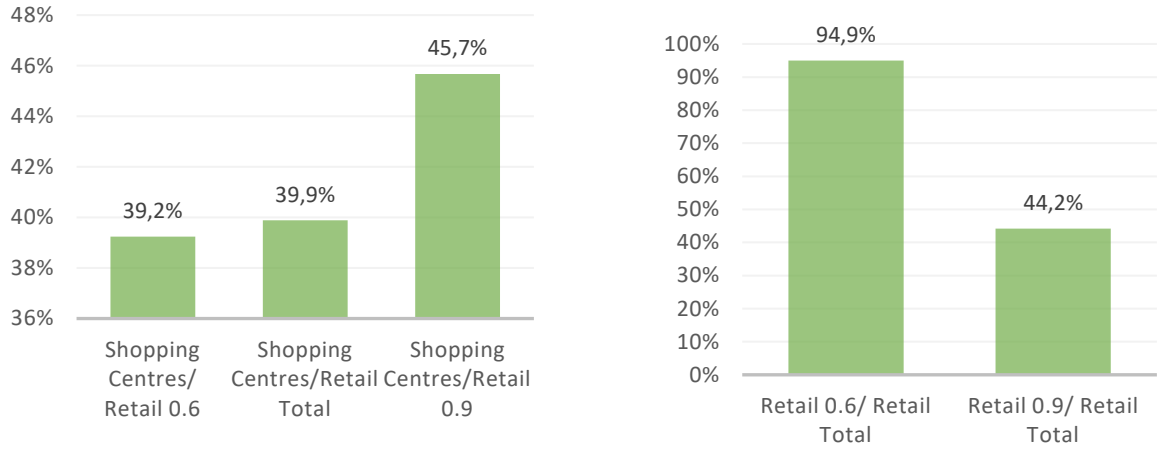

Figure 6:Retail/Shopping Centre Ratios, in % (on the left); Retail Sub-index Proportions, in % (on the right)

Retail is one of the largest and most representative sub-sectors in the analysis. For this sub-index, two levels of repeatability were assessed: the 0.6 level and the 0.9 level. The 0.9 level was also chosen to check differences in the specification and over-all behaviour

of the sub-sector. Figure 6 displays Figure 6:Retail/Shopping Centre Ratios, in % (on the left); Retail Sub-index Proportions, in % (on the right)the magnitude of the repeatability proportion according to the number of invoice transactions along with the shopping centre's proportion in each level of repeatability. When the limit of 0.9 is applied there is a significant drop on the number of observations whereas in the 0.6 level one loses just 6% of those. This same drop of observations does not apply to the proportion of shopping centres. As stated above, around 60% of shopping centres register invoice transactions with over 0.9 repeatability rate and this fact is implicit in the figure above. When it comes to the overall behaviour of the Retail sub-index year-on-year rate (Figure 7), it is practically the same for the 0.6 level of repeatability and the 0.9 level (in the Appendix 2) from 2017 to the 1st trimester of 2020, keeping a constant level of variation. From the 1st trimester of 2020 onwards, coinciding with the start of the Covid-19 pandemic, this pattern changed and in different magnitudes for the 0.6 and 0.9 levels. The 0.6 level presents a drop in values which kept negative variations comparing to 2019 until the beginning of 2021, when an upwards trend starts to appear reflecting the recovery of invoice values after the pandemic, this upwards trend peaks (104%) in August of 2021.

Still in Figure 7, whereas in the 0.9 level the variations in 2020, the year most affected by the pandemic, are only negative for some periods, for the other periods the variations are positive compared to the value of 2019. This may be partly due to the Retail Commerce that was considered as essential (e.g., food retail) and therefore did not fully close like other sector's activities which were forced to close due to the pandemic. The major peak is in April 2021 (140%), the homologous period to when the pandemic had started and the lowest value was registered. After this period the variations were lower and by the last trimester of 2021 there also appears to be an upwards trend.
It is also interesting to notice that the establishments which register a higher level of repeatability show a slightly different reaction to a heterogeneous shock, the pandemic, than those with a lower level of repeatability.

As for the hedonic model results, most variables are significant at the 1% level for both the issuer and the acquirer. For the issuers side the main conclusions follow: For each additional invoice, the total invoice value is expected to increase 0,7%. If the invoice between the two parties has a repeatability rate over 90% the total invoice value is estimated to increase by 11,3%. If a shopping centre is involved, the total invoice value

is expected to decrease by 36,4% (economically expected for the reasons stated above in 3.3 Stratification). If the issuer's company was established before year 2000 then the total invoice value is expected to be 59% lower than a company established after year 2000. If the issuer has a high number of establishments (more than five) the total invoice value is expected to be 41% lower than those with a small number of establishments. If the issuer's equity if fully private or foreign the total invoice value is expected to be higher by 33,4% and 31,9% (respectively) than if the issuer's equity were fully public. If the issuer has international presence- abroad establishments, imports or exports- the total invoice value is expected to be 28,9% higher than if the issuer did not have international presence. If the issuer has a positive ratio between the volume of business and total assets, the total invoice value is expected to be 45,6% higher than those who do not. For the acquirer side: If the acquirer has international presence, the total invoice value is expected to be 33,7% higher than those who do not. If the equity if fully private or foreign the total invoice value is expected to be 20,2% and 23,2% higher than acquirer with full public equity. If the company was established prior to 2000, the total invoice value is expected to be 15,8% lower than companies created after 2000. If the acquirer is a non-financial foreign society the total invoice value is expected to be 31,1% higher than for other forms.

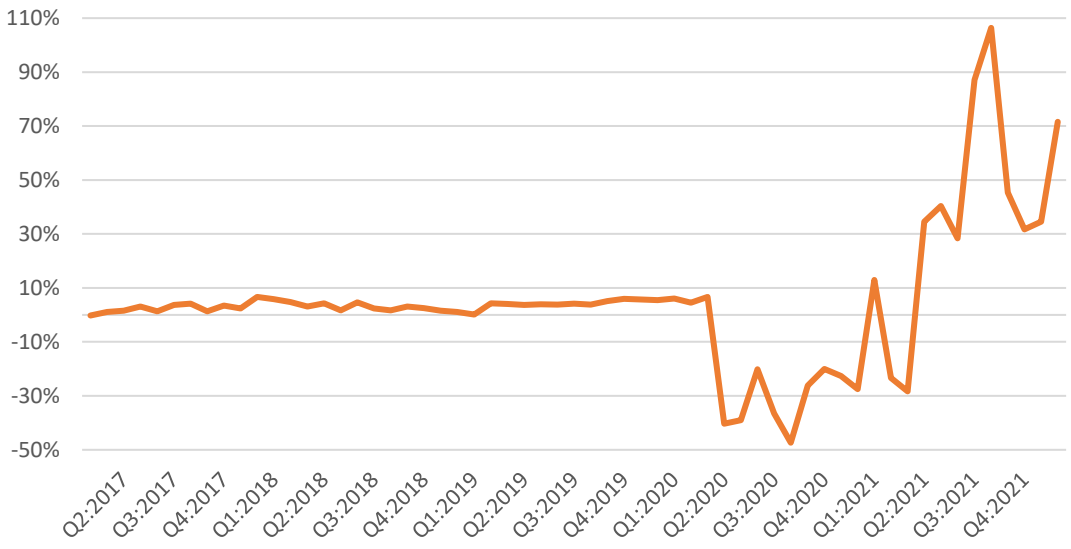

Figure 7:Year-on-year rate of change for Retail, repeatability 0.6, in percentage, Q2:2017-Q4:2021

4.1.2. Services

Services is one of the largest sectors in the sample along with Retail. For the Services sector, the proportion of shopping centres to the sample is very significant.

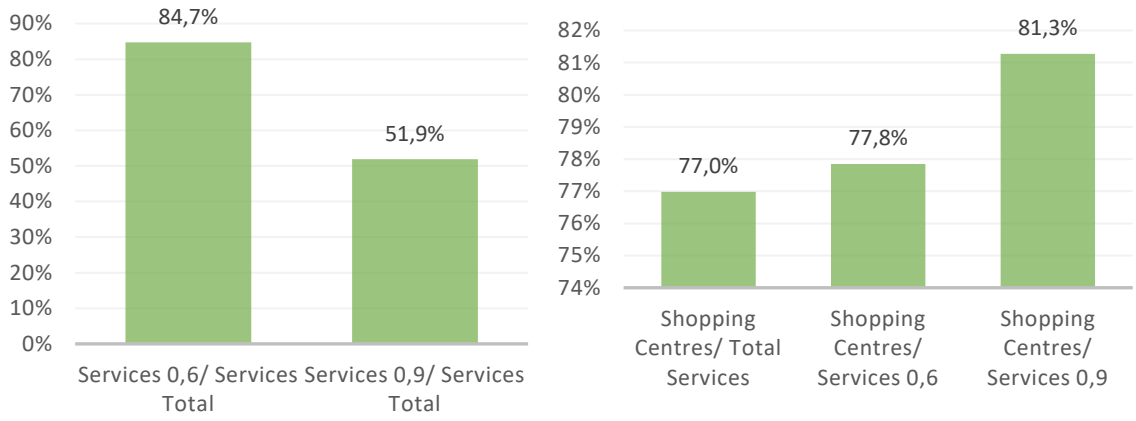

Figure 8:Services Sub-index Proportions (on the right) and Service/Shopping Centres Ratio (on the left)

From Figure 8 one can see the overall proportion of shopping centres regarding Services and their proportion after the restraints were applied to the sample. Over 75% of all Services sub-sample invoice value transactions include a shopping centre. The more the sub-sample was restrained, according to its level of repeatability, the bigger the proportion of shopping centres. The latter means invoice value transactions involving shopping centres are expected to repeat more, i.e., having longer periods of invoicing, than when considering other commercial property for service not involving a shopping centre in the issuer.

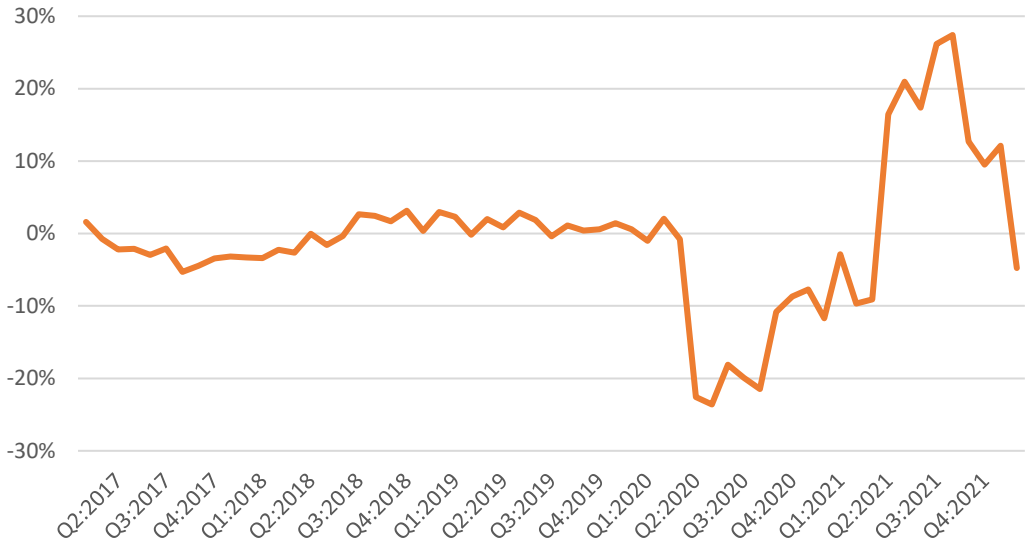

Figure 9:Year-on-year Rate of change for Services, in percentage, Q2:2017-Q4:2021

Regarding the year-on-year rate (Figure 9), where the effects of phenomena like seasonality are smoothed allowing for a more accurate analysis, the invoice transaction values were on an upward trend until the first trimester of 2020 after which the pandemic hit and with it the values dropped very significantly when compared to the homologous values of 2019, since many services like restaurants and accommodations were shut down for most part of the year, these values only recovered after the 1st trimester of 2021, where one can see the positive effects on the Services' invoice transactions values after the reopening of the economy. There is a visible recovery and even growth compared to pre-pandemic levels.

As for the hedonic regression the variables in the issuers' side, they were more relevant to explain the model than the acquirer's variables. The following conclusions derived from variables with at least 5% level of significance (in this case all variables are included).

For the issuer's side, each additional invoice is expected to increase the total invoice value by 0,3%. If there is a shopping centre in the invoicing, the total invoice value is expected to be 50,4% lower than for an invoicing without a shopping centre party. This is economically expected for the reasons stated above in 3.3 Stratification. An 1% increase in the issuer's total assets is expected to increase the total invoice value by 14,4%, this may be caused by the bigger an issuer's company is the more properties it may have or properties with higher commercial value. If the issuer's company was established prior to

year 2000, the total invoice value is expected to be 22,2% lower than if the company was established after year 2000. If the issuer has less than 5 establishments, the total invoice is expected to be 62,5% higher than if the issuer has more than 5 establishments. If the issuer is classified as an anonymous society the total invoice value is expected to be 9,5% higher than if the issuer as a different juridical classification. If the issuer has full private equity, the total invoice value is expected to be 17,6% higher than other types of equity participation (public or foreign). If the ratio between the business volume and the total assets of an issuer is positive, the total invoice value is expected to be 46% higher than if the company has a negative ratio. If the issuer has more than 100 yearly acquirers the total invoice is expected to be 62,9% lower than if the issuer has less than 100 acquirers per year. The previous may be related to the quality of the properties even though one would expect that more acquirers would reflect into a higher total invoice value, the results show the opposite. For the acquirer's side, if the total assets of an acquirer increases 1%, the total invoice value is expected to increase by 15,9%. If the acquirer's company was established between year 2000 and 2012, the total invoice value is expected to be 7,9% lower than if it was established in any other temporal interval. If the company has a small total assets value, i.e., the total assets being lower than the mean of that sector, the total invoice value is expected to be 64,3% lower than for companies with the total assets larger than the sector' mean. If the acquirer has a business volume between 150 thousand € and 2 million €, the total invoice value is expected to be 22% higher than other levels of business volume.

4.1.3 Wholesale

The proportion of observations retained after applying the 0.6 repeatability rate is over 90% like in the Retail sub-index. Such high rates are only represented in the commerce sector even though the number of observations in Wholesale are significantly lower. This suggests that in the sample used, most commerce related invoice value transactions repeat in contiguous periods for more than 60 % of all invoicing between a pair of NIF's.

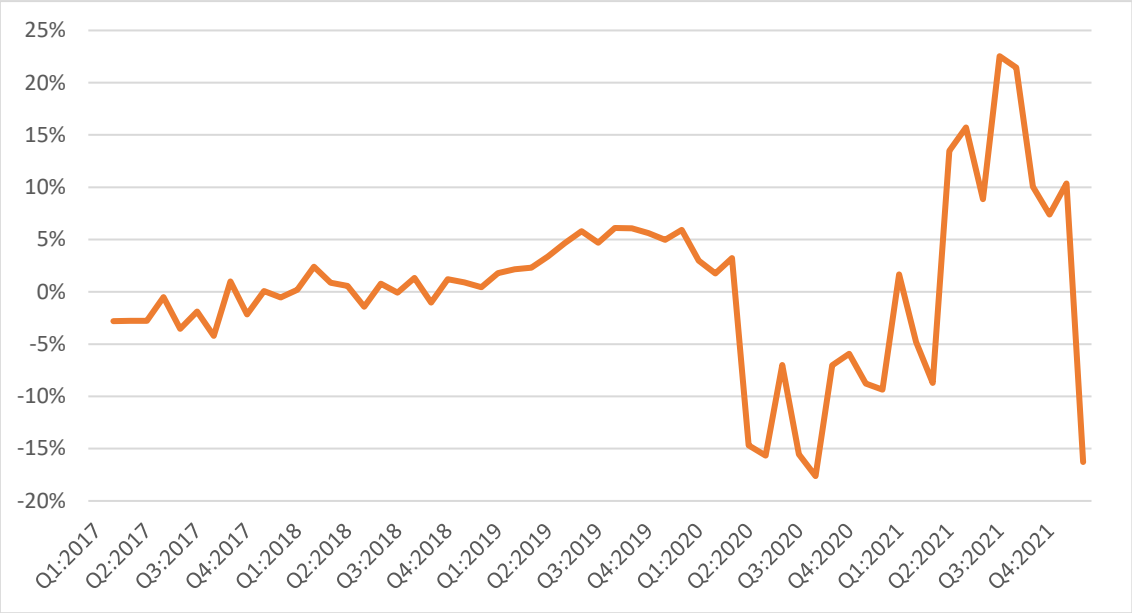

Figure 10:Year-on-year Variation Rate for Wholesale, in percentage, Q2:2017-Q4:2021

The year-on-year rate of change, Figure 10, on Wholesale Commerce seems to be on a slight positive trend until the pandemic, showing a growth in 2019 comparing to 2018. After the pandemic hit in April 2020, the sector reflects these impacts through negative rates of change compared to 2019 until April 2021. In this period the rate of change increased significantly showing signs of recovery in the sector comparing to the homologous period. Although the year of 2021 ends with a significant drop in the invoice values transacted in the sector comparing to December of 2020. This drop in values is a different result for that period in homologous terms than the other sectors in analysis. Every other sector presents an increase in homologous values in December 2021. The reason for this drop in unknown, may be due to this sector having less observations in the sample than other sectors like Retail and Services and the observations in the sample may not be enough to present the true impact of the economy recovery or may indeed be caused by some sector differences when it comes to renting commercial properties. On the hedonic regression results, variables in the issuers' side were more significant to explain the model than the acquirer's variables. The following conclusions derived from variables with at least 5% level of significance. If there is more one invoice the total invoice value is expected to increase by 0,2%. If the total assets of a company increase by 1% the total invoice value is expected to increase by 7,2%. If the company was established prior to 2000 or between 2000 and 2012, the total invoice value is expected

to be 40,8% and 33,9% lower than companies established after 2012. If the issuer has more than 5 establishments the total invoice values is expected to be 50,6% lower than issuers with less than 5 establishments. If the issuer is a commercial society, the total invoice values are expected to be 10% higher than otherwise. If the company has full private equity, the total invoice values are expected to be 27,5% higher than if the equity were public or foreign. If the issuer has an international part, the total invoice value is expected to be 23,2% higher than for those who do not have that international part. If the issuer has between 15 and 100 acquirers per year, the total invoice value is expected to be 43,7% lower than other levels of acquirers per year. For the acquirer's side, an increase of 1% of the total assets is expected to translate in the total invoice value being higher by 13,2% and if the acquirer has a positive business volume to total assets, the total invoice value is expected to be 17,7% higher than those who do not have a positive ratio.

4.1.4 Industry

Along with the Wholesale sector, this is the least representative sector. The reduced number of observations in this sector may be related to the fact that most enterprises invest in buying the properties where they operate instead of renting commercial properties. This may occur because their focus in on production instead of the client like in commerce/services. The types of properties they seek have also quite different characteristics relative to the ones searched for in the Commerce or Services sectors.

Regarding the number of observations after the sub-sample was limited to the 0.6 repeatability rate, above 80% of observations are kept. The year-on-year rate, Figure 11, presents negative rates of change for 2017 comparing to 2016, after which presents an increase through all 2018 year in homologous terms, in the 2nd and 3rd trimester of 2019 it also presented a rise in the rate.

This is the only sector to present such high negative values on year-on-year rates of change during 2020. Also presents the same upwards trend in 2021 of economy recovery comparing to the previous year. (see Figure 11)

The fact that this sector seems to be the less affected by the pandemic effects was sort of expected. For the reasons mentioned previously that enterprises in this activity do not rent most properties they have but could also be related to other aspects like enterprises not

needing to shut down due to being essential during the pandemic, as mentioned above in retail, and being less sensitive to the variation of sales.

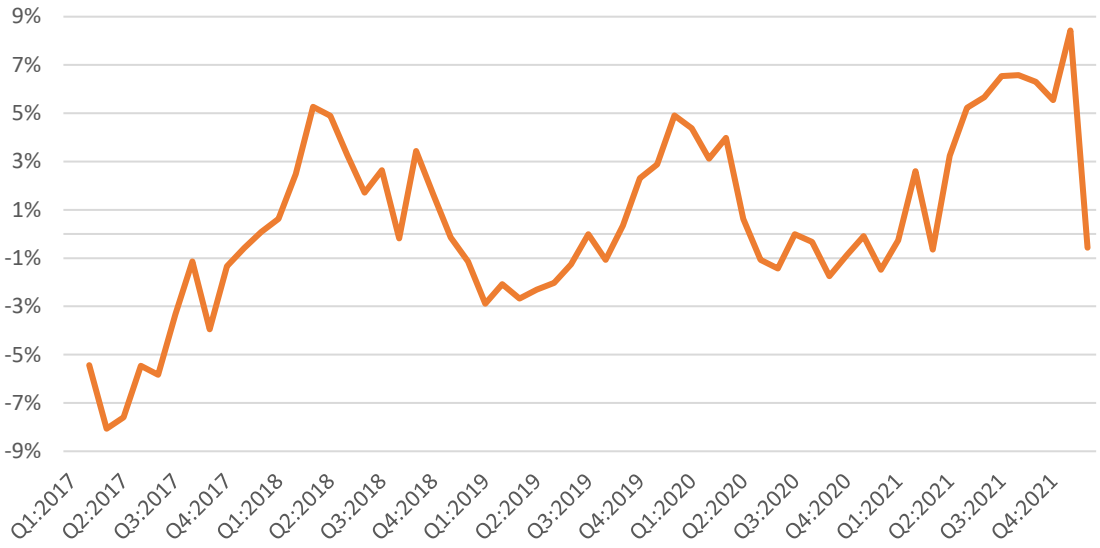

Figure 11:Year-on-year Variation Rate for Industry, in percentage, Q2:2017-Q4:2021

In the hedonic regression the variables in the issuers' side were more relevant to explain the model than the acquirer's variables. The following conclusions derived from variables with at least 5% level of significance. For the issuer's side, for each additional invoice the total invoice value is expected to increase by 0,5%. If the issuer's total asset increases 1%, the total invoice value is expected to increase 4,6%. If the company was established prior to year 2000, the total invoice value is expected to be 43,9% lower. If the company has full private equity, the total invoice value is expected to be 59,4% higher than if the company has a different equity participation. If the issuer has more than 100 yearly acquirers the total invoice is expected to be 36% lower than if the issuer has less than 100 acquirers per year (same reasoning than for the Services sub-sector). For the acquirer's side if the total asset if a company increases 1%, the total invoice value is expected to increase 15%. If the acquirer has a business volume between 150 thousand € and 2 million €, the total invoice value is expected to be 5,5% higher than other levels of business volume. If the company has a social equity between 5 thousand € and 100 thousand €, the total invoice value is expected to be 8,9% higher than companies with different levels of social equity.

4.2 Aggregated Index

After analysing and specifying each sub-index, all of them were combined according to the methodology described above in 2.2.1 into a single index. In a monthly perspective presented below in Figure 12, until the Covid-19 pandemic, which started along March/April 2020, the index presents a stable behaviour with an apparent seasonal pattern. A value increase on the invoice transactions is registered in January and this increase tends to be followed by a decrease in February approximately in the same amount. This pattern in the overall index is most likely influenced by the structure of the sub-indices of Retail Commerce and Services.

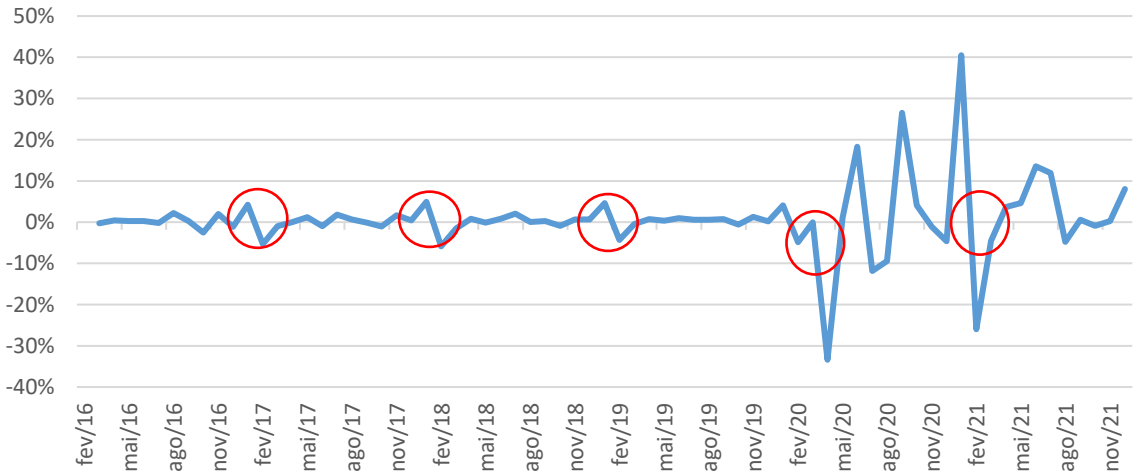

Figure 12:Monthly Rate of change, in percentage, fev-16 to dez-21

After the pandemic hit, its effects on the overall behaviour of the monthly index are clear.

The index reached new lows followed by new highs, reflecting the high volatility present in this period. One can also note that by the end of 2021 the results seem to tend to the previous, more stable values. This becomes clearer when observing the monthly rate of change, where the pattern of an increase in January is followed by a decrease in value by February (again signalled in red in the figure below) is corroborated. The return to previous pre-pandemic values around the end of 2021 is corroborated as well, since the rate of change seems to stabilise after highly fluctuating during the lockdown phase of the Covid-19 pandemic.

In the year-on-year analysis (Figure 13), the evolution of values is better comparable since effects of seasonality and trends are smoothed. Having 2016 as a base-year, the annual

changes revolved around mean 0%, i.e., the variations between years were small enough for the values to remain almost constant. Until March 2020, for the same reasons mentioned above, the year-on-year rate fluctuated significantly.

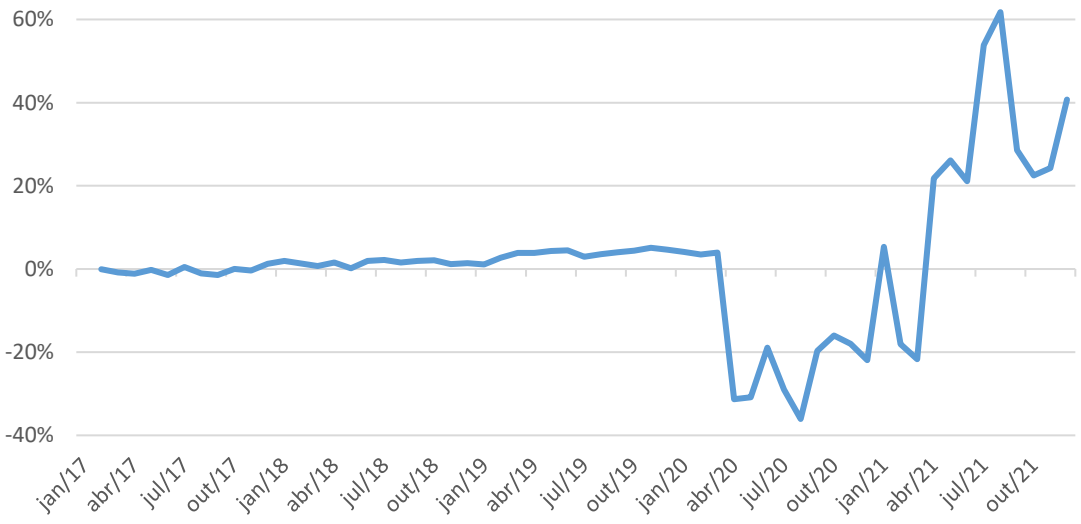

Figure 13:Year-on-Year Rates of Change, in percentage, jan-17 to dec-21

The variations of April 2020 reached a minimum as low as -31% when comparing to April 2019. This very significant drop seems to correspond to the lockdown period, period when every sub-sector drastically reduced its activity specially comparing to the previous year. And therefore, renting properties was also on hold, explaining the drop. The comparison between April 2020 and April 2021 has the exact opposite conclusion. April 2021 shows the positive impacts of having reopened the economy, in this period the invoice values in general registered an increase of 55% comparing to the April of the previous year. This high year-on-year rates of change for 2021 and 2020, express the economic conjecture Portugal and the world were facing at the moment due to the Covid-19 pandemic, and should not be seem like regular behaviour of the commercial properties market invoice transactions.

The final index, presented below in Figure 14 summarizing all of the above, is presented in an annual optic for a clearer visualization of the invoice transactions value behaviour.

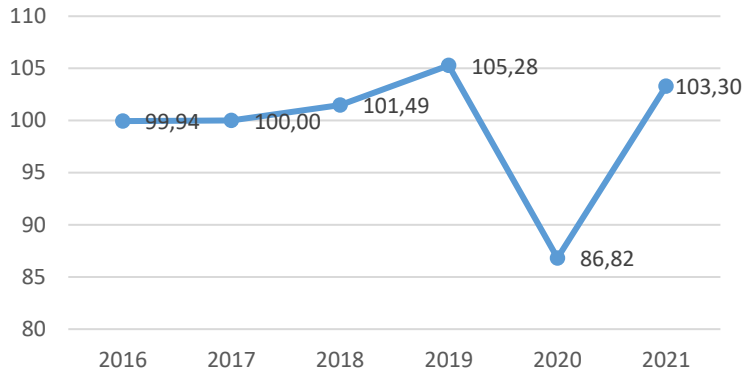

Figure 14:Annual Aggregated Index, 2016-2021

This annual index allows to take some conclusions. Up until 2020, the aggregated index presents a growth dynamic, in a stable outlook, as mentioned above by analysing the monthly and year-on-year rates. The decrease caused by the pandemic is presented by a drop in 2020 but in this annual perspective is clearer to see that in 2021 the index recovered from the fall caused by the pandemic, and not only recovered but now presents an upward trend. The invoice transaction values in 2021 seem to be increasing when comparing to the previous years. The reasons for these are unclear but this effect may cause a change in the structure of the commercial properties market from this point on. This effect was also not so exposed by just regarding the monthly analysis, which shows that presenting the data in various formats allows for a better understanding of its overall behaviour.

The annual index englobing all sub-indices was computed using data according to the 0.6 level of repeatability, the 0.9 level and using data without any adjustments regarding the level of repeatability. In Table 4 there are the annual rates of changes according to each specification. Although each one present rates with slightly different magnitudes, all of them share the same pattern when it comes to the overall behaviour of total invoice values.

Beyond that, these results, are an indicator that the data extracted from *E-fatura*, even though not circumscribed to the universe of commercial properties, seems to present a good overall picture of the macroeconomic conjecture of all sectors combined.

Bearing in mind the figures presented above, and our main objective, we would say that commercial property rents remained unchanged in 2017 compared to 2016, with annual

growth of 1-2% in 2018 and 3-4% in 2019. In 2020, with the COVID-19 crisis, rents saw a 15-17% reduction, only to grow again by 17%-21% in 2021.

Year	***0.9 Repeatability***	***0.6 Repeatability***	***No Repeatability***
2017	0%	0,1%	0%
2018	1,4%	1,5%	1,5%
2019	3,1%	3,7%	2,7%
2020	-17,3%	-17,5%	-15,8%
2021	21,2%	19%	17,4%

Table 4:Yearly Rates of Change, 2017-2021, in percentage

5. Conclusion and Final Remarks

Despite all limitations mentioned above in this report about *E-fatura*, in particular the fact that the mechanism is not exclusive to rental of real estate (invoicing), or even commercial properties and the fact that explanatory variables may not be directly related to properties, the compiled commercial property's index seems to be capable of capturing the evolution of each subsector and all sub-sectors combined according to the macroeconomic conjecture along the temporal period in analysis. It is to expect that with more and recent data, to be possible for this mechanism to be used for future analysis and to keep pace with the behaviour of commercial properties through enterprises' invoicing.

It was a good indicator that the pandemic effects were captured by the single sub-indices and by the combined, but it made it harder to understand the real specificity of each sector, since the next year 2021 is still influenced by 2020, which might have changed the specification form of each sector, in particular Retail and Services, since this phenomenon changed in general people's lives, more in specific the utilities' level one gets from specific products/services which indirectly affect each economy sector in different ways. So, with more recent information one could evaluate the specification chosen in this report and balance if it still makes sense or adjustments need to be done.

Regarding the method by which the subindices were computed, the choice was the Adjacent Time-Dummy, a particular case of the Rolling Time Dummy Method with the time window of two periods. An indication for future endeavours using this mechanism for this purpose would be to try different time windows, which could allow for better estimates and a better overall picture of the commercial properties' rental market behaviour. A suggestion would be for instance to try a time window of 12 months to diminish seasonal effects. In future works, it is also advisable to compare the results obtained with other methods alternative to the Time-Dummy, namely the imputation method or repricing.

It was mentioned the weight and specificities of shopping centres in the sample and how englobing transactions between an acquirer and a shopping centre. Due to their own values and characteristics being significantly different than when dealing with an issuer which is not a shopping centre, could bias the results. Computing specific indices for shopping centres could prove of interest and help to create less heterogenous indices by

keeping observations with characteristics closer to one another. Instead of trying to specify a more general form, a more informative and detailed index could be obtained.

Another aspect which could improve the analysis would be to being able to obtain territory variables related to each property to better understand the spatial pattern associated to each transaction and factors involved like repeatability, characteristics, and invoice values. Also, more observations on the Industry and Wholesale sectors, since these sectors had the least number of observations and therefore it is harder to be sure the specification and results can represent the population they try to sample.

Repeatability by contiguity was the one focus on this report for believing it better grasped the concept of commercial properties' rent. Restraining the data according to this criterion allowed to keep only the most relevant transactions in the sample, trying different levels than the 0.6 and 0.9 presented here may also demonstrate to be helpful. The closer one could get to restraining the overall data to the universe in question the better the mechanism will work for constructing the index.

Still considering there is a lot of room for improvement in terms of available data, relevant variables, and the factor of time, *E-fatura* as an administrative mechanism for computing a possible commercial properties' rent index or proxy indicator may be a viable option for the future and this report presents an initial steppingstone for the construction of such.

REFERENCES

Chin, T.L. & Chau, K.W. (2003). *A Critical Review of Literature on the Hedonic Price Model*. International Journal for Housing and Its Applications, 27(2), 145-165.

Cropper, M.L., Deck, L.B., & McConnell, K.E. (1988). *On the choice of functional form for hedonic price functions*. The Review of Economics and Statistics, 70(4), 668-675.

Evangelista, R., Moreira, H., & Teixeira, Â. (2019) *On the derivation of hedonic rental price indexes for commercial properties: Rent price determinants, data quality problems and evidence from the use of administrative data for the Portuguese market*. Consulted in 2022 at https://www.readkong.com/

Evangelista, R. & Raposo, I.G. (2016) *A transactions-based commercial property price index for Portugal*. Paper presented at the 15th Ottawa Group Meeting.

Evangelista, R. & Teixeira, A. (2014) *Using Different Administrative Data Sources to Develop House Price Indexes for Portugal*. INE Working Paper, 2014.

Gouriéroux, C. & Laferrère, A. (2009). *Managing hedonic housing price indexes: The French experience*. Journal of Housing Economics, 18(3), 206-213

Halvorsen, R. & Pollakowski, H. O. (1981). *Choice of functional form for hedonic price equations*. Journal of Urban Economics, 10(1), 37-49.

Instituto Nacional de Estatística (INE) (2017b). Índice de Preços das Propriedades Comerciais. Documento Metodológico. Versão 1.0. Consulted in 2022 at https://cse.ine.pt

Lancaster, J.K. (1966) *A New Approach to Consumer Theory*. The Journal of Political Economy, 74(2), 132-157.

Nordbotten, S. (2010). *The Use of Administrative data in Official Statistics- Past, Present and Future-With Special Reference to Nordic Countries*. Journal of Official Statistics, 205-223

Rosen, S. (1974) *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*. Journal of Political Economy, 82, 34-55.

Sander, H.A. & Polasky, S. (2009). *The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA*. Land Use Policy, 26(3), 837-845.

Shimizu, C., Takatsuji, H., Ono, H. & Nishimura, K.G. (2010). *Structural and temporal changes in the housing market and hedonic housing price indices*. International Journal of Housing Markets and Analysis, 3(4), 351-368.

Triplett, J. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*. OECD. Consulted in 2022 at https://www.oecd-ilibrary.org/

Wooldridge, J.M. (2003). *Introductory Econometrics: A Modern Approach* (2nd Ed.). Cengage Learning

Zheng, S., Kahn, M.E. & Liu, H. (2010). *Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities*. Regional Science and Urban Economics, 40, 1-10.

APPENDIX 1: EXPLANATORY VARIABLES

Original Variables	***Variables' Description***
NIF	Fiscal Number
CAE3	Economic Sector Activity Code
STA	State of Activity
FJR	Legal Form
NUTSIII	Territory Nomenclature Code
CCI	International Trade Operator type
NPS_REM	Number of People at Service with Remuneration
SIN	Institutional Sector
CPS	Social Equity
VVN	Volume of business
TOTAL_ACTIVO	Total Assets
IMPORTACOES	Total Imports
EXPORTACOES	Total Exports
DCO	Enterprise' date of establishment
PTE	Proportion of foreign social equity
PNV	Proportion of private social equity
PNP	Proportion of public social equity

Table 5:Original Variables' Description

Covariates' Names	***Covariates' Description***
Ln_Total_Activo [16]	Logarithm of the total assets of NIF i.
Ln_VVN [16]	Logarithm of the total Business Volume of NIF i.
N_faturas	Number of invoices traded between a combination of an issuer and acquirer's NIF.
Repeticao_90	If the contiguous invoices between a combination of NIF's has a repeatability rate higher than 0.9.
Pre_2000 [16]	If the enterprise represented by NIF i was founded prior to year 2000.
00_12 [16]	If the enterprise represented by NIF i was founded between year 2000 and year 2012.
Pos_2012 [16]	If the enterprise represented by NIF i was founded after year 2012.
Centro_comercial	A dummy variable =1 if NIF i corresponds to a Shopping Centre; =0 otherwise.
N_peq_estab_pt [16]	Dummy Variable =1if the number of establishments belonging to NIF i is smaller than 5.
N_gran_estab_pt [16]	Dummy Variable =1if the number of establishments belonging to NIF i is larger than 5.
Norte [16]	Dummy Variable =1 if NIF i is located in NUTSIII North; =0 otherwise.
Centro [16]	Dummy Variable =1 if NIF i is located in NUTSIII Center; =0 otherwise.
Lisboa_AM [16]	Dummy Variable =1 if NIF i is located in NUTSIII Lisbon Metropolitan Area; =0 otherwise.

[16] Each variable, with exception of some, has correspondence for an issuer's NIF side and an acquirer's NIF side. E.g., In the data base there is a LN_Total_Activo_emitente (issuers') and LN_Total_Activo_Adquirente (Acquirers')

Alentejo [16]	Dummy Variable =1 if NIF i is located in NUTSIII Alentejo; =0 otherwise.
Algarve [16]	Dummy Variable =1 if NIF i is located in NUTSIII Algarve; =0 otherwise.
Ilhas [16]	Dummy Variable =1 if NIF i is located in NUTSIII Madeira or NUTSIII Açores; =0 otherwise.
Soc_Comercial [16]	Dummy Variable =1 if NIF i corresponds to a Commercial Society.
Soc_Nfin_npt [16]	Dummy Variable =1 if NIF i corresponds to a non-financial society under foreign control.
Admin_Central [16]	Dummy Variable =1 if NIF i corresponds to Portugal's Central Administration.
Sin_11 [16]	Dummy Variable =1 if NIF i is registered as a Non-Financial Society.
Sin_12 [16]	Dummy Variable =1 if NIF i is registered as a Financial Society.
Sta_20 [16]	Dummy Variable =1 if NIF i is currently in activity.
Sta_27 [16]	Dummy Variable =1 if NIF i is an already existing enterprise currently in restructuring.
Sta_40 [16]	Dummy Variable =1 if NIF i has closed activity.
Sta_41 [16]	Dummy Variable =1 if NIF i has closed activity by bankruptcy.
Sta_47 [16]**Error! Bookmark not defined.**	Dummy Variable =1 if NIF i has closed activity due to fusion/merger.
Fjr_32 [16]	Dummy Variable =1 if NIF is legally declared as an anonymous society.

Fjr_34 [16]	Dummy Variable =1 if NIF is legally declared as a limited liability company.
Ln_ativo_ratio	Logarithm of the ratio between the acquirer's total assets and the issuer's total assets.
Comercio	Dummy Variable=1 if the acquirers' Nif i belongs to a CAE3 corresponding to the Commerce Sector.
Industria	Dummy Variable=1 if the acquirers' Nif i belongs to a CAE3 corresponding to the Industry Sector.
Servicos	Dummy Variable=1 if the acquirers' Nif i belongs to a CAE3 corresponding to the Services Sector.
Atv_Com_peq [16]	Dummy Variable=1 if Nif i belongs to the Commerce Sector and has a total asset smaller than the mean of that sector.
Atv_Serv_peq [16]	Dummy Variable=1 if Nif i belongs to the Services Sector and has a total asset smaller than the mean of that sector.
Atv_Ind_peq [16]	Dummy Variable=1 if Nif i belongs to the Industry Sector and has a total asset smaller than the mean of the sector.
PNV_1 [16]	Dummy Variable=1 if NIF i has a fully private equity.
PNP_1 [16]	Dummy Variable=1 if NIF i has a fully public equity.
PTE_1 [16]	Dummy Variable=1 if NIF i has a fully foreign equity.
Part_Int [16]	Dummy Variable=1 if NIF i has international presence.
VVN_Atv_Pos [16]	Dummy Variable=1 if NIF i has a positive ratio of volume of business to total assets.

VVN_1 [16]	Dummy Variable=1 if NIF i has a volume of business larger than 5 million €.
VVN_2 [16]	Dummy Variable=1 if NIF i has a volume of business between 150 thousand € and 5 million €.
VVN_3 [16]**Error! Bookmark not defined.**	Dummy Variable=1 if NIF i has a volume of business smaller than 150 thousand €.
N_ADQ_Ano_1	Dummy Variable=1 an issuer's NIF has 1 acquirer per year.
N_ADQ_Ano_2	Dummy Variable=1 an issuer's NIF has more than 1 acquirer per year and less than 15 acquirers.
N_ADQ_Ano_3	Dummy Variable=1 an issuer's NIF has more than 15 acquirers per year and less than 100 acquirers.
N_ADQ_Ano_4	Dummy Variable=1 an issuer's NIF has more than 100 acquirers per year.
Sqrt_NPS [16]	Square Root of the number of people at service with remuneration.
NPS_Rem_1 [16]	Dummy Variable=1 if NIF i has less or exactly 15 remunerated employees.
NPS_Rem_2 [16]	Dummy Variable=1 if NIF i has more than 15 remunerated employees.
Ln_CPS [16]	Logarithm of the Social Equity Capital of NIF i.
CPS_1 [16]	Dummy Variable=1 if NIF i has a Social Equity less than 5 thousand €.
CPS_2 [16]	Dummy Variable=1 if NIF i has a Social Equity between 5 thousand € and 100 thousand €.
CPS_3 [16]	Dummy Variable=1 if NIF i has a Social Equity larger than 100 thousand €.

Table 6:Created Covariates Description

APPENDIX 2: ADDITIONAL FIGURES

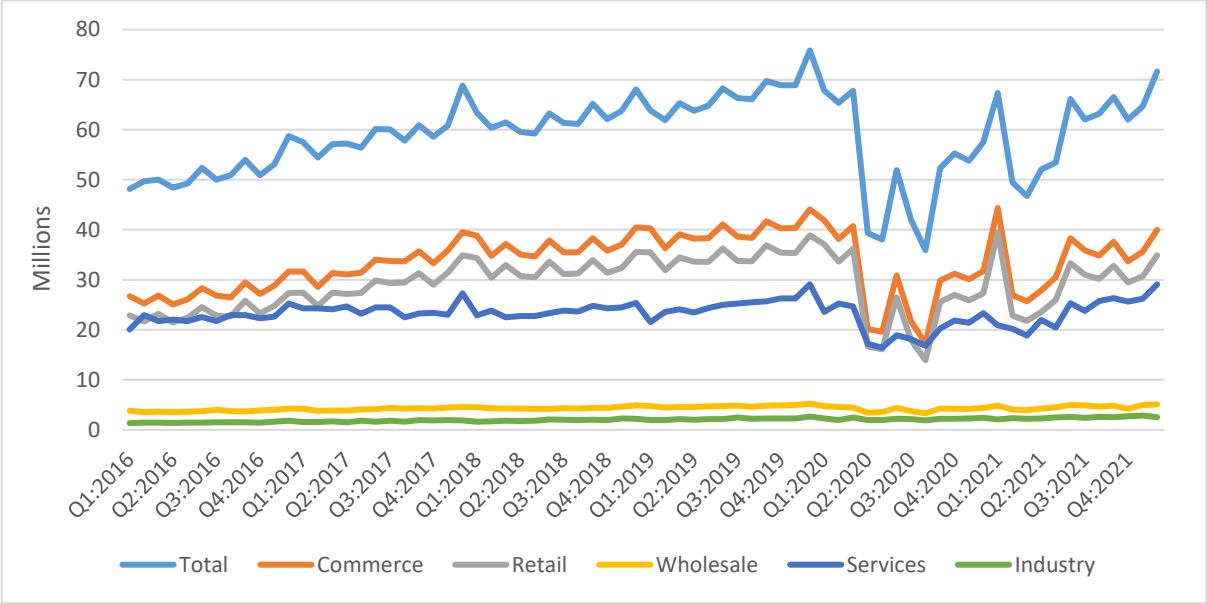

Figure A.1: Overall Invoice Values Behaviour prior to Stratification, in millions of euros, Q1:2016-Q4:2021

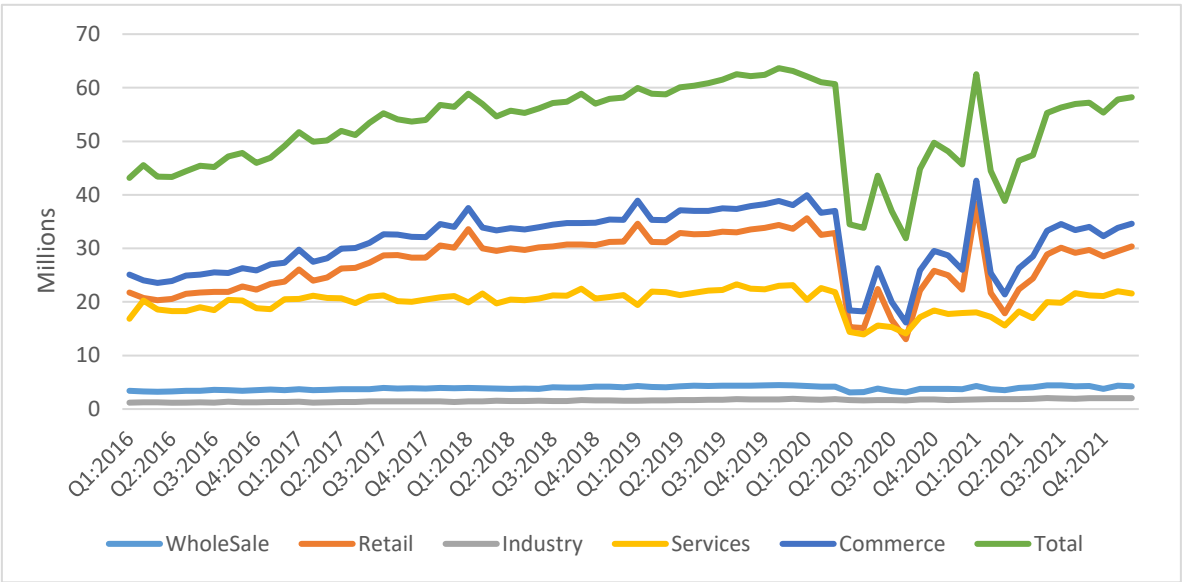

Figure A.2:Overall Invoice Values Behaviour after Stratification in millions of euros, Q1:2016-Q4:2021

YEAR-ON-YEAR RATES[17]

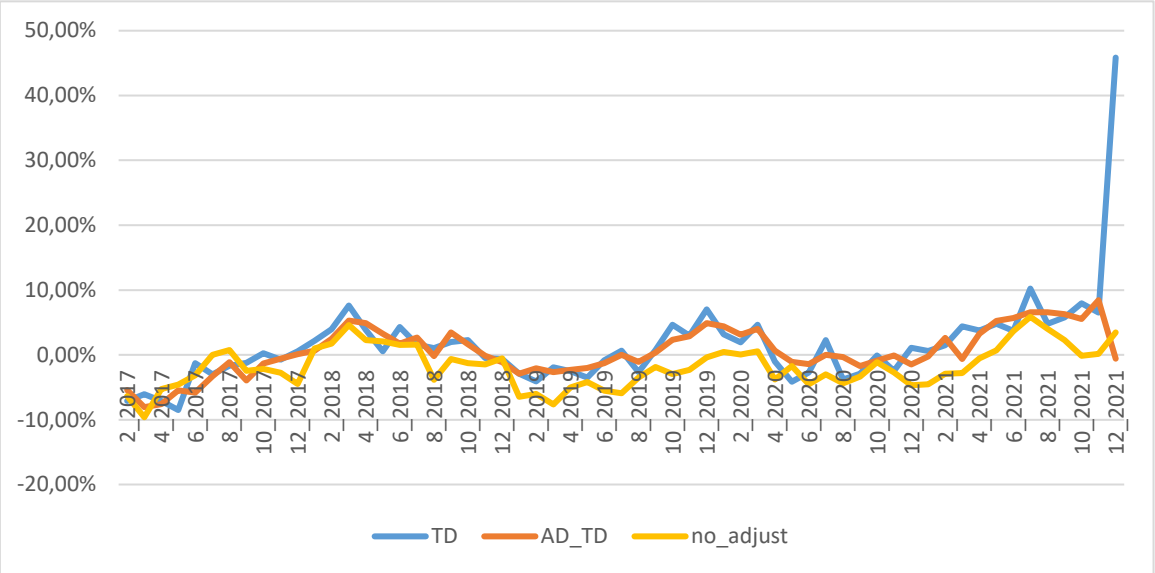

Figure A.3:Industry year-on-year rate

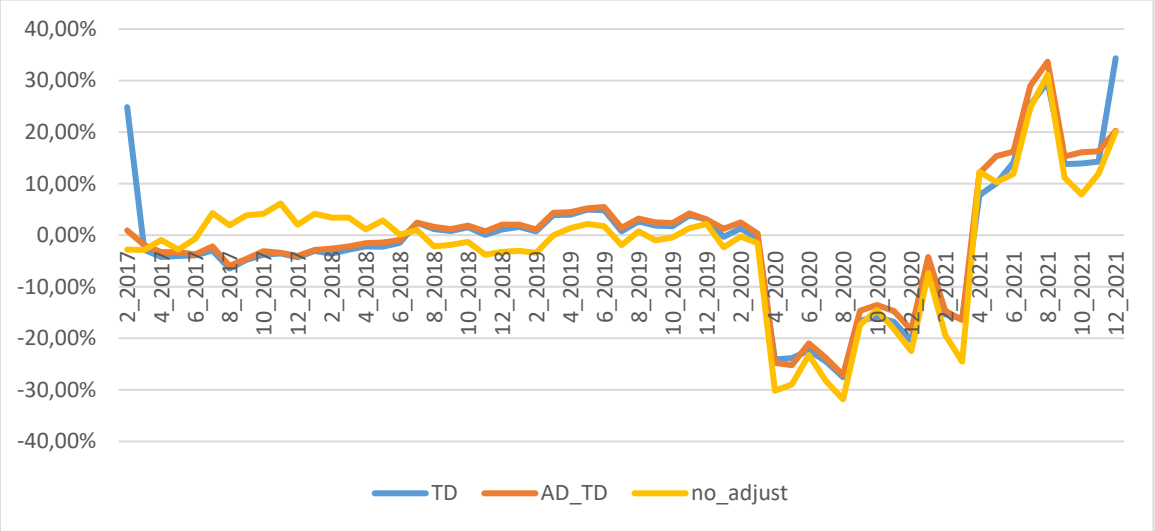

Figure A.4:Services year-on-year rate

[17] With TD: Pooled Time Dummy Method; ADJ_TD: Adjacent Time Dummy Method; no_adjust: simple stratified means

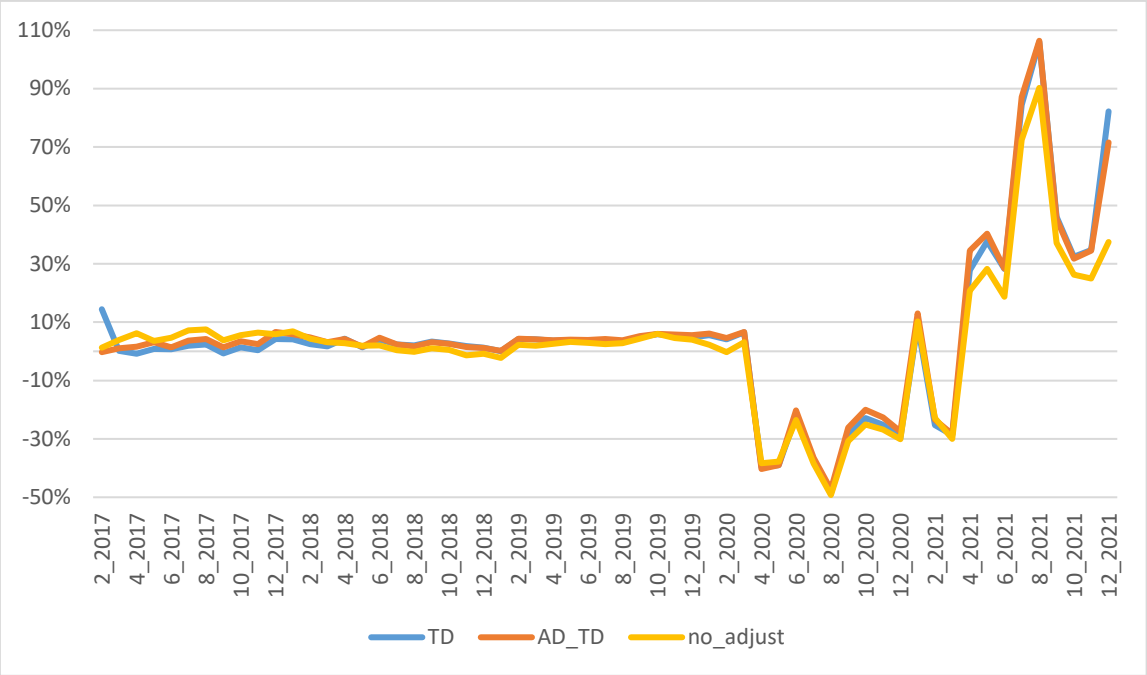

Figure A.5:Retail 0.6 year-on-year rate

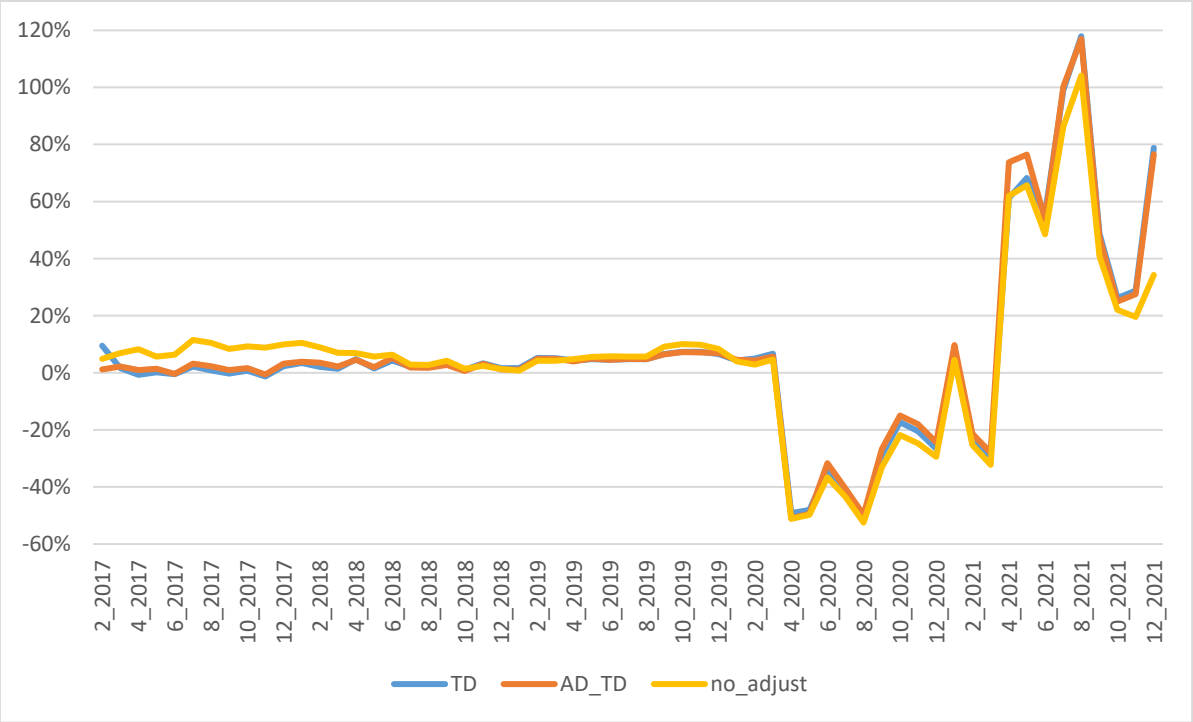

Figure A.6:Retail 0.9 year-on-year rate

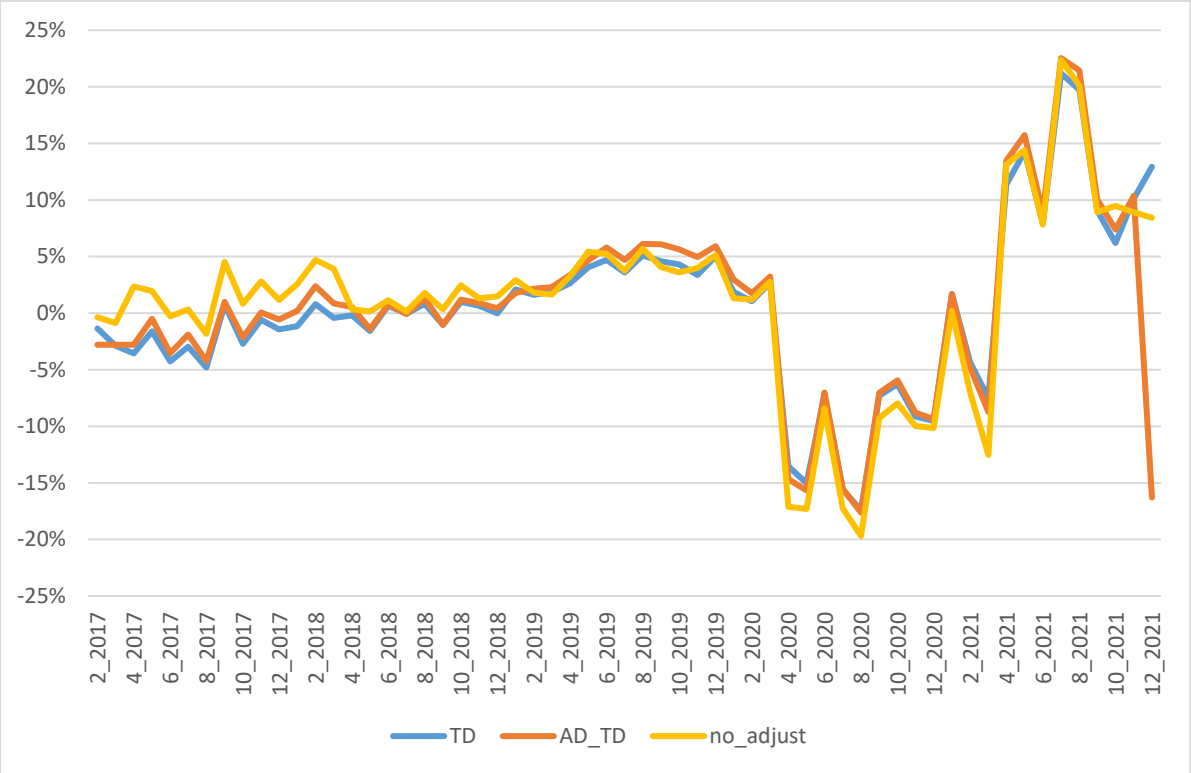

Figure A.7:Wholesale year-on-rate

APPENDIX 3: REGRESSION RESULTS

RESET Test				
Date	*Retail*	*Wholesale*	*Services*	*Industry*
1_2016	0,4336	0,5399	0,0018	0,1115
2_2016	0,6116	0,472	0,0013	0,1532
3_2016	0,5736	0,959	0,0211	0,179
4_2016	0,5342	0,7497	0,0538	0,1762
5_2016	0,2311	0,89	0,0396	0,5027
6_2016	0,1483	0,653	0,0526	0,8375
7_2016	0,2847	0,9561	0,0101	0,5005
8_2016	0,573	0,8692	0,0074	0,9847
9_2016	0,09	0,9947	0,0136	0,3288
10_2016	0,1555	0,7868	0,0535	0,5332
11_2016	0,3965	0,4477	0,3126	0,8528
12_2016	0,266	0,6988	0,193	0,6152
1_2017	0,4996	0,2981	0,1219	0,4999
2_2017	0,9089	0,3151	0,1314	0,7431
3_2017	0,9755	0,5049	0,2175	0,7906
4_2017	0,8641	0,767	0,4744	0,8385
5_2017	0,9741	0,8991	0,4478	0,8536
6_2017	0,8878	0,511	0,3231	0,3281
7_2017	0,8839	0,5832	0,2588	0,1019
8_2017	0,8112	0,8194	0,3546	0,4806
9_2017	0,6572	0,7476	0,7012	0,5182
10_2017	0,5317	0,5406	0,9769	0,4773
11_2017	0,9347	0,6092	0,944	0,6389
12_2017	0,702	0,9076	0,262	0,1454
1_2018	0,8879	0,8668	0,134	0,0926
2_2018	0,6631	0,9157	0,7879	0,1305
3_2018	0,2543	0,8468	0,9275	0,1253
4_2018	0,1423	0,6018	0,7789	0,1806
5_2018	0,3653	0,9063	0,7955	0,6319
6_2018	0,2738	0,9551	0,5189	0,4678
7_2018	0,4907	0,7979	0,3894	0,2509
8_2018	0,5767	0,9337	0,2761	0,5507
9_2018	0,2132	0,8781	0,7687	0,7027
10_2018	0,2403	0,9685	0,9491	0,8737
11_2018	0,4836	0,9159	0,734	0,8506
12_2018	0,901	0,5158	0,2244	0,61
1_2019	0,7523	0,5179	0,1557	0,5181
2_2019	0,4687	0,4989	0,3419	0,7756
3_2019	0,4722	0,7756	0,4024	0,3548
4_2019	0,1371	0,9668	0,2494	0,1789

5_2019	0,0995	0,9223	0,1961	0,4203
6_2019	0,1487	0,9428	0,4562	0,7893
7_2019	0,1521	0,9418	0,7643	0,4174
8_2019	0,1603	0,962	0,5282	0,1999
9_2019	0,1253	0,9035	0,3006	0,6832
10_2019	0,0931	0,7887	0,2034	0,6945
11_2019	0,0671	0,6829	0,0426	0,697
12_2019	0,2044	0,7713	0,0085	0,8551
1_2020	0,9504	0,9679	0,0124	0,7314
2_2020	0,5188	0,8737	0,0333	0,5687
3_2020	0,9033	0,0576	0,3849	0,3616
4_2020	0,0119	0	0,1563	0,2595
5_2020	0,0001	0,0013	0,0765	0,7199
6_2020	0	0,0011	0,0923	0,4643
7_2020	0	0	0,2544	0,2094
8_2020	0,2523	0,0134	0,4035	0,006
9_2020	0,3511	0,6252	0,6396	0,0135
10_2020	0,2807	0,4632	0,7485	0,1155
11_2020	0,0377	0,7272	0,6812	0,0757
12_2020	0,398	0,9361	0,7102	0,3287
1_2021	0,8296	0,5942	0,4689	0,6
2_2021	0	0,0754	0,7879	0,36
3_2021	0	0,2022	0,4143	0,0656
4_2021	0,012	0,2333	0,5251	0,0616
5_2021	0,2322	0,5808	0,9073	0,0726
6_2021	0,523	0,8216	0,6561	0,3085
7_2021	0,4033	0,5361	0,4193	0,6519
8_2021	0,0548	0,263	0,5705	0,9017
9_2021	0,1148	0,5666	0,9924	0,4717
10_2021	0,4785	0,647	0,7182	0,1678
11_2021	0,6482	0,6908	0,8096	0,5616
12_2021	0,7469	0,8052	0,5381	0,9706
# of failed monthly specifications	8	5	11	2
% of failed monthly specifications	11%	7%	15%	3%

Table 7: RESET specification test results

Specification By Strata with 0.6 repeatability [18] [19]				
	Retail	**Wholesale**	**Services**	**Industry**
EMITENTE				
N_faturas	0,007***	0,002***	0,003***	0,005***
Repeticao_90	0,113***	-0,140*	-	-
Centro_comercial	-0,364***	-	-0,504***	-
Ln_ativo_ratio	0,028	-	-	-
Ln_total_activo_emit	-	0,072***	0,144***	0,046***
Ln_VVN_emit	-	-	-	-0,025
Emit_pre2000	-0,59***	-0,408***	-0,222***	-0,439***
Emit_00_12	-	-0,339***	-	-
Emit_cae3_682	-	-0,292***	-0,049***	-0,665
Npeq_estb_emit_pt	-	-	0,625***	0,241*
Ngran_estb_emit_pt	-0,411***	-0,506***	-	-
Emit_norte	0,459***	-	-	0,186
Emit_algarve	0,108**	-	-	-
Emit_centro	0,011	-	-	-0,134
Emit_lisboa_am	0,346***	-	0,189***	0,31
Emit_alentejo	-	-	-	-
Emit_ilhas	-	-	-	-
Emit_soc_comercial	-	0,101**	-	-
Emit_soc_nfin_npt	-	-	-	-
Emit_fjr_32	-	-	0,095**	-
Emit_atv_ind_peq	-	-	-	-0,39
Emit_pnv_1	0,334***	0,275***	0,176***	0,594***
Emit_pte_1	0,319***	0,167	-	-

[18] *** p<0,01; ** p<0.05; * p<0.1. Let p represent p-value

[19] : This stratification refers to the Adjacent Time dummy method of March 2016, illustrative for the other periods.

Emit_part_int	0,289***	0,232***	-	-
LN_VVN_ativo_emit	-	-	0,047***	-
Emit_VVN_atv_pos	0,456***	0,854*	0,460**	-
Emit_VVN_3	-	-0,196*	-	-
Emit_VVN_2	-	-	0,003***	-
N_adq_ano_3	-	0,437***	-	-
N_adq_ano_4	-	-	-0,629***	-0,36***
Sqrt_NPS	-0,033***	-	-	-
ADQUIRENTE				
Ln_total_activo_adq	-	0,132***	0,159***	0,150**
Adq_Part_Int	0,337***	-	-	0,142*
Adq_lisboa_am	0,173***	-	-	-
Adq_Centro	-	-0,173**	-	-
Adq_alentejo	-	-	-	-
Adq_ilhas	-	-0,42	-	-
Adq_PNV_1	0,202***	-	-	0,183
Adq_PTE_1	0,232***	-	-	-
Adq_pre2000	-0,158***	-	-	-
Adq_00_12	-	-	-0,0769**	-
Adq_sta_20	-0,268***	-0,172	-	-0,568
Adq_sta_27	-	-	-	-
Adq_sta_40	-0,100	-	-	-
Adq_soc_comercial	-	0,024	-	-
Adq_soc_nfin_npt	0,311***	-	-	-
Adq_fjr_32	-	-	-	-
Adq_fjr_34	-	-0,08	-	-
Ngran_estb_adq_pt	-	-	-	-
Npeq_estb_adq_pt	-	-	-	0,499
Adq_atv_com_peq	-	-	-	-
Adq_atv_serv_peq	-	-	-0,643***	-

Adq_atv_ind_peq	-	-	-	0,174
Ln_VVN_ativo_adq	-	-	0,066***	0,102
Adq_vvn_ativo_pos	0,068**	0,177***	-	-
Adq_VVN_1	-	-	-	-
Adq_VVN_2	-0,083***	-	0,220***	0,055**
Adq_VVN_3	-	-	-	-
Sqrt_NPS_adq	0,007***	0,011*	-	0,028***
Adq_NPS_rem_1	-	-	-	-
Adq_NPS_rem_2	-	-	-	-
Ln_CPS_adq	0,083***	0,053	-	-
Adq_CPS_1	-	-	-	-
Adq_CPS_2	-	-	-	0,089**
Adq_CPS_3	-	-	-	-
N	202 468	86 380	236 128	49 612
Adjusted R^2	0,511	0,459	0,493	0,376

Table 8:Specification per strata. 0,6 repeatability

Sector	***Included CAE's[20]***
Services	55, 56, 68, 69, 722, 743, 749
Industry	412, 432
Retail Commerce	47
Wholesale Commerce	46

Table 9:CAE's per sector

[20] When CAE are represented by just two digits they englobe all sub-CAE, when represented by three digits they only refer to that specific CAE.