INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

Exploring new administrative data sources for the development of the Consumer Price Index:

The Portuguese experience with actual rentals for housing[1]

April 2018

Vítor Mendonça[2], Rui Evangelista[3]

Housing rents are one of the most important categories of the Portuguese CPI, representing almost 4% of the total household expenditure. Up until recently, the CPI covering actual rentals for housing was compiled based on information taken from an expensive monthly survey, which followed nearly 3,000 dwellings and whose underlying sampling frames were updated every 10 years with each housing and population Census. Following a 2017 transmission arrangement made between Statistics Portugal and the Portuguese Fiscal Authority, it was possible to access administrative data related to housing rental contracts and paid rents, which have, according to a legal act of March 2015, to be electronically registered by the majority of landlords. The access to these electronic receipts data, which covers nearly 350,000 of the estimated stock of 550,000 rented dwellings in Portugal, opened up the possibility of replacing the above-mentioned survey with a more comprehensive and up-to-date data source. In addition, it also provided the grounds for the empirical testing of alternative price index compilation methods (e.g., stratification and hedonic). This paper gives an overview of the challenges that were faced in the appropriation of the new data source, and presents some tentative results on the use of different price index compilation methods on 2016 and 2017 data. The results provided in this paper are not only of interest to all price index compilers engaged in the compilation of rental price indices but also to researchers and those interested in the recent evolution of the Portuguese rental market.

JEL Classification: C43, C81, E31, R21, R31

Keywords: Consumer price index, new data sources, rents

[1] The analyses, opinions and findings of this paper represent the views of the authors and are not necessarily those of Statistics Portugal. Comments are welcome.

[2] Statistics Portugal, head of the Consumer Price Index section, National Accounts Department; vitor.mendonca@ine.pt.

[3] Statistics Portugal, head of the Price Statistics unit, National Accounts Department; rui.evangelista@ine.pt.

Contents

1. Introduction

Housing rents are one of the most important components of the Portuguese Consumer Price Index (CPI), with its weight representing almost 4% of the total household final monetary consumption expenditure. Until the end of 2017, the rents index was compiled using price information taken from a monthly survey with a sample size of around 3,000 rented dwellings (Statistics Portugal, 2014). The index was stratified by region, size and type of rent (e.g., social or market rent) and followed a matched-sample framework in which monthly rent payments were paired with the information observed for the same housing units in December of each year [4].

Although in line with existing standards on the treatment of housing rentals in the CPI (see, *inter alia*, Eurostat, 2014), its compilation was based on a data source that suffered from three main shortcomings. The first one revolved around the frequency of the update of the underlying sampling frame of the survey, which was only to be carried out every ten years after the end of a population and housing Census [5]. While the rental market was well portrayed at the start of the use of a new sampling frame, the ability of the index to include additions to the stock of rented dwelling units (e.g., those that were previously own-occupied and that are now rented) is at best very limited [6]. A second disadvantage stemmed from the fact that rental units were not generally replaced after two consecutive months of missing rents information. In practice, this meant that the sample size of the paired information used to compile the index diminished over time. In addition, it also meant that the situations in which a rented dwelling stayed on the market for more than two consecutive months after the end of a rental contract were not captured by the index (i.e., cases in which typically the rent of the same dwelling is generally updated upwards). The third drawback was the cost of the survey, which employed substantial staff and financial resources.

Following a 2017 transmission arrangement made between Statistics Portugal and *Autoridade Tributária e Aduaneira* – the Portuguese Fiscal Authority (PFA), it was possible to access administrative data on housing rental contracts and investigate, in this way, the impact from changing from a survey to an administrative data source virtually covering the population of market rents [7]. This coverage stems from the fact that the data was administratively generated by electronic rent receipts, which were made obligatory by law from the beginning of 2015 onwards (MD, 2015).

The use of these data, when combined with other administrative sources already in use in the current production of other real estate price statistics [8], opened up the possibility of using different compilation methods and address two issues associated with the use of a matched-sample approach. The first one has to do with the fact that throughout a year, new rentals to the market will not contribute to the measurement of housing rentals, as they do not find a

[4] This approach can be characterised as a long-term chain index in which the sample of housing units are refreshed annually (see ILO et al., 2004: p. 171).
[5] The last Census was carried out in 2011.
[6] This could only be done purposively through the direct observation of the rental market by surveyors.
[7] Social or subsidized rents are not covered by this data source. However, this type of contracts only represents a small portion of the rental market. Based on 2011 Census data, it is possible to estimate the share of this market segment as 8 to 9% of the total rented dwelling stock. In terms of expenditure, according to the Portuguese 2015/2016 Household Budget Survey, social rents are estimated to have a share of less than 3% of the total expenditure on rentals.
[8] The Portuguese Commercial Property Price Index is an example (Gonçalves and Evangelista, 2017).

pair for comparison purposes in the December (price base) period. In a dynamic market (e.g., new rentals appearing on the market, prices increasing), the use of the matched sample framework may introduce some downward bias to the housing rentals CPI component [9]. The second reason refers to the "aging depreciation" effect, which has been estimated in Shimizu et al. (2015) for the Japanese CPI. In a matched-sample approach, even when all physical characteristics remain fixed (i.e., the size of the rented dwelling does not increase/decrease), the age of the dwelling increases. As such, it may be argued that the matched-sample approach does not take into account a correction factor for the depreciation of the housing structures and may be prone to downward bias.

The main objective of this paper is to present Statistics Portugal's work regarding the use of this new administrative data source to compile alternative housing rental indices. In particular, the results of four different alternative methods are presented and compared to the survey-based rents index. In addition, a framework for the investigation of the impact of neglecting new rental contracts and "age depreciation" effects is also presented in the paper's appendix.

This paper is organized into three sections. Section two describes survey and administrative data sources and presents the alternative methods used to calculate the housing rental component of the CPI. Section three highlights and compares the results of proposed methods *vis-à-vis* the already compiled survey-based rents index. Finally, the last section presents the concluding remarks of this paper and points out the way forward.

2. Compiling the housing component of the Portuguese CPI

2.1. Building an administrative data source for the rents index

The data used for deriving alternative housing rent indices comprises of three main datasets: electronic rent receipts data, which Statistics Portugal receives on a weekly basis, all changes in rents contracts, which is received on a monthly basis, and dwelling characteristics data (area, age of dwelling, quality of the location, etc), which is obtained through property tax records (*Imposto Municipal sobre Imóveis*) sent by the PFA to Statistics Portugal on a monthly basis. This last database is already being used in the compilation of the Portuguese house and commercial property price indices.

The overall quality of the electronic rent receipts data is perceived as good. It comprehends nearly all rents, as landlords are obliged to process the rents receipts online through the fiscal authority website. The exceptions to this obligation are not many. For instance, landlords with low income gains from rents or who are aged 65 or more are out of the scope of the coverage of electronic rent receipts procedure. Social or subsidized rents are also excluded as these are rental agreements that are normally provided by municipalities or charitable institutions, which are not subject to the issuing of an electronic rent receipt.

In the Portuguese tax system, tenants have an incentive to declare the amount of paid rents in their annual income tax declaration, something that is checked against the information provided by landlords to the PFA. This constitutes a mechanism that prevents the

[9] Krsinich (2009) provides some evidence on this effect in a short-term chain index framework.

underdeclaration of paid/received rents and something that contributes to the quality of the data.

In order for the data to be used in a rents index, these three datasets are linked through a unique identifier for each dwelling. The link between rent receipts and dwellings is made using a rental contract number, which is possible to follow through time. However, this link is not guaranteed to always yield a one-to-one relationship:

- Some contracts cover more than one dwelling, for example, a contract comprising a dwelling and a separate garage. Since it is not possible to a) split the rent value between the two items, and b) guarantee that the two items will always be rented together, these contracts are not used in calculations;
- Some dwellings are covered by more than one contract. For example, a large dwelling might be rented by the room (e.g., for students) and some parts of the dwelling may be of common use (e.g., the kitchen). Since it is not possible to a) identify unequivocally each room and b) guarantee that all the rooms of the dwelling are always rented (in which case the sum of all the amounts paid for the whole dwelling could potentially be used), these contracts are not used in calculations.

Rent receipts that cover more than one month were also excluded, as well as multiple receipts on a single month (which could be due to mistakes, duplicates, or new contracts starting in the middle of a month). Moreover, only receipts that cover between 28 and 31 days were taken into account. Finally, dwellings that were not used for housing purposes were also dropped out from future index calculations. In addition, to remove extreme cases, the following restrictions were applied:

- Dwelling area must be between 10 and 1,000 square meters;
- Rent value must be less than 5,000 Euros;
- Rent per square meter must be less than 25 Euros per square meter.

As a result of the application of these rules, around 31% of the observations were excluded. For 2017, an average of around 285 thousand observations were available in the final dataset, with a total of 543 thousand dwellings having at least one receipt in the database. It must be noted that, each month, around 4% of the dwellings did not have a receipt in the previous month. However, most of these were previously rented at some period prior to the current month.

In 2017, 86.5% of the dwellings presented no rent change and 12.4% changed only once within the year (the legal framework on rents states that these can only be revised once per year, unless agreed otherwise by both parties or if there is a tenant change). As a result, only 1.1% of the dwellings had more than one rent change in 2017 – essentially new contracts for existing rented dwellings.

2.2. Rents index compilation methods

Four different rents indices were compiled for the 2016-2017 period [10], with the aim of improving the quality of the rents index and maximizing available administrative data. The comparison basis for these indices was the survey-based rents index already in production. One the major problems of the existing survey was that vacant dwellings would leave the sample after two months of non-reported price quotes, which meant that new contracts were not covered if the dwelling remained vacant for more than two months. All indices follow prices per square meter and use the Jevons (geometric) average as lower-level elementary index number formula.

The different price indices are described below (Methods A to D).

Method A is similar to the survey-based rents index. For the base period, it considers all dwellings that were rented at any moment in the previous year, using the last known rent value as the base price. Dwellings are removed from the sample if no rent receipt is available for two consecutive months (the last observed rent is replicated / carried-forward in the two periods [11]). In 2016, the sample is reduced from 244,174 to 132,276 dwellings (end of the year). New contracts are only taken into account if the dwelling is empty for less than two months between tenant changes.

Method B is similar to the previous method but, instead of dropping all rents information that remain with no observed rent value for more than two consecutive months, receipts can be used whenever the rented dwelling reappears on the market. In practice, a matched sample between December of the previous year and the current month is taken into account. As in method A, missing values are imputed for a maximum of two consecutive months. In 2016, the sample reduces from 244,174 to 183,006 dwellings (end of the year). This method increases the number of dwellings in the sample, and allows all new contracts to be considered in the calculations, as long as the dwellings were present in the rental market in the previous year. Rents for dwellings that are new to the rental market are not taken into account [12].

Method C considers all the valid receipts in each month, irrespective of whether there is any previous information for each dwelling. This methods uses all of the available information, however changes in quality are not controlled. In effect, this method results in ratios of average of all available rents, and cannot be interpreted as a measure of pure price change. In 2016, the initial sample of 188,691 dwellings in January compares to 229,165 dwellings by the end of the year. All contracts, existing and new, as well as contracts for newly rented dwellings are considered.

Method D applies a monthly matched-sample method and compiles an index using a short-term chain index procedure. All dwellings that have a receipt for the current and the previous month are paired, and the rent change is linked to the previous month index. This method will only include new contracts in their second month of existence, where a price change is not

[10] For 2017, results are only presented until June, when the choice of the method was decided and the computations of the remaining methods stopped.

[11] Given the percentage of cases in which there is no price change (see previous section), this imputation technique is considered to give the best estimate for the missing rent information.

[12] The best way to assess the impact of neglecting these rents from the rental index is through the application of the hedonic method. See the appendix for more details on how this could be done in this context.

expected, unless the tenant change is immediate (i.e. there is no waiting period between contracts). This method tends to underestimate price changes resulting from new contracts, as the main price change is generally in the first month of the new contract. The initial sample of 156,015 dwellings with receipts in January 2016 and December 2015 compares to 180,273 paired dwellings between November and December 2016.

Additionally, we are preparing the evaluation of a hedonic rents index that includes all available information in its calculation process. It is the best approach to assess the shortcomings of matched-sampled methods. Due to the lack of time and the absence of a longer time series, it was not possible to derive empirical results. It was, however, possible to derive a conceptual framework on which future empirical work will be based (see the appendix).

Figure 1 provides an overview of the number of observations considered in the calculations of indices following Methods A to D.

Figure 1 – Number of observations considered in the calculations[13]

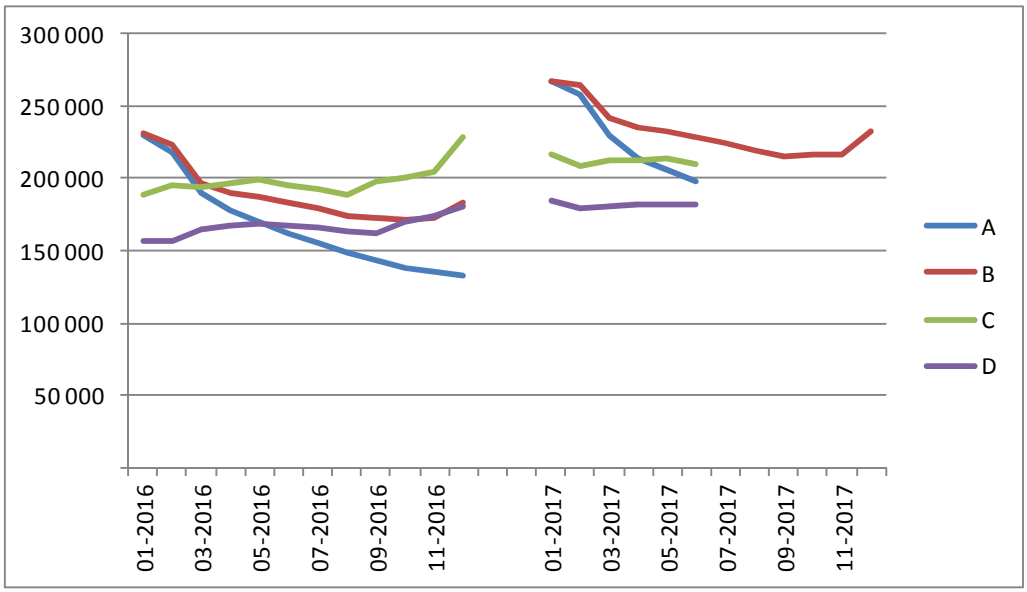


Table 1 of the appendix provides the sample sizes used in the calculations.

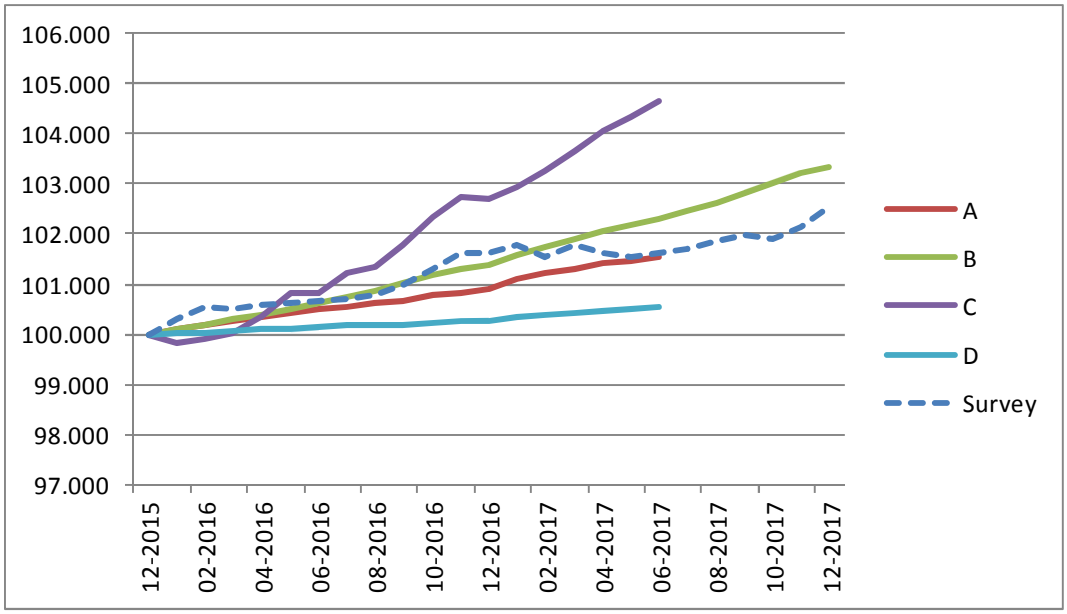
3. Comparison between survey-based and alternative rents indices

Due to the short time length of the data, the comparison between survey-based and alternative rents indices was only possible for the period starting in January 2016 and ending in June 2017. Figure 2 presents the index results for the four methods described in the previous

[13] Methods A and B consider up to two consecutive months of imputed rent values. Methods C and D only consider observed rents. Receipts that were uploaded after the monthly CPI calculation (first week of the following month) were not taken into account in the calculations but can be used in the following month imputation, if the conditions for imputation are met.

section (base 100 = December 2015). The data used in this figure is available in Table 2 of the appendix.

Figure 2 – Indices for free market housing rentals



The next two figures present the monthly and annual rates of change of the methods. Tables 3 and 4 in the appendix provide the rates of change.

Figure 3 – Monthly rates of change for free market housing rentals

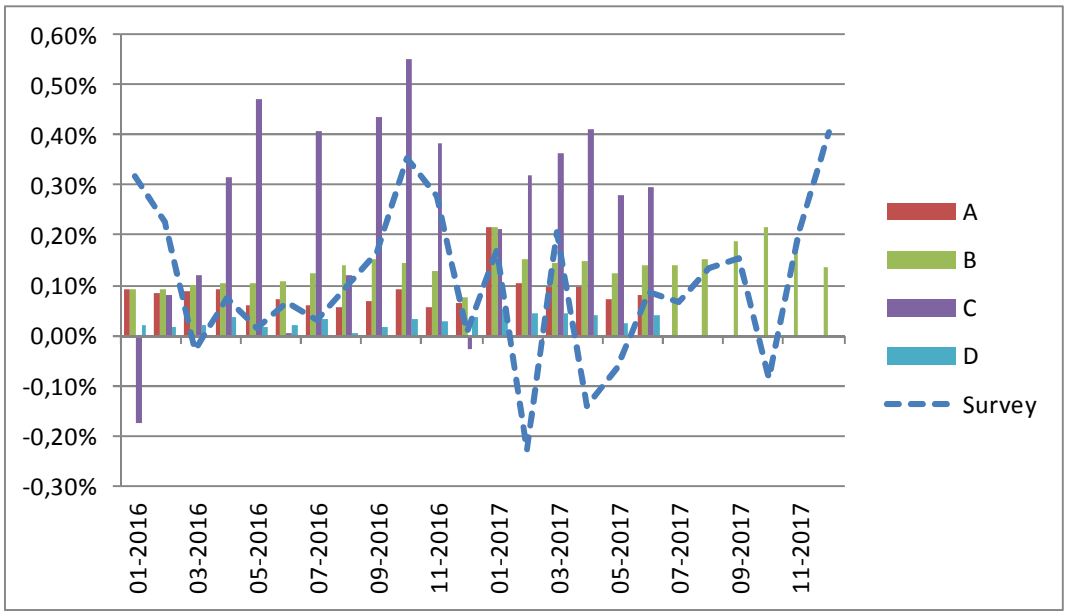
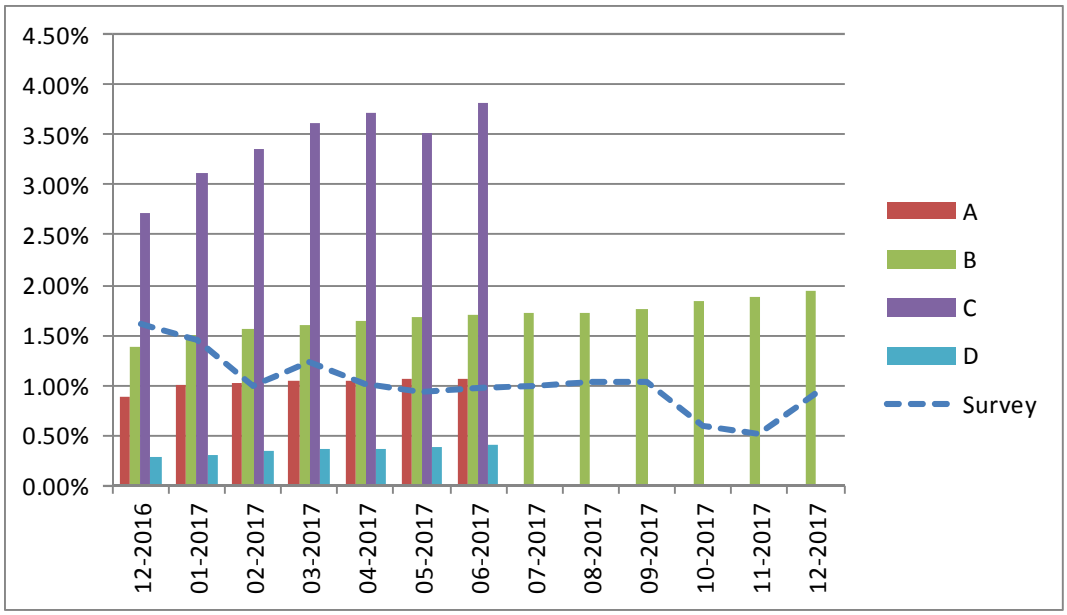


Figure 4 – Annual rates of change for free market housing rentals



Overall, the results are mostly in accordance with the *a priori* expectations.

Method C leads to indices that are clearly above all the others. This is easily explained by the lack of consistency of the monthly sample. The rental market in Portugal has been expanding rapidly and prices for new rentals are increasing. This method emphasizes the impact of new rentals, which are directly compared to the rental stock. Since no sample or quality adjustments are performed, prices for higher-quality dwellings are being compared directly with prices of all the remaining rentals. Despite including all contracts, new and existing, as well as for newly constructed dwellings, this method does not measures pure price changes. The use of the hedonic methodology would allow to maximize the use of the data and produce constant-quality price indices.

Method D, on the other hand, yields the lowest indices. Since only matching receipts are considered, and since the legislation only allows for rent increases after 12 months, most of the monthly matching observations present no price change. In fact, the legislation limits the annual rent increases to the previous year's CPI (excluding housing). For 2017, this limit was very low (0.54%). Using this method, the annual rate for June 2017 was 0.39%, which is consistent with the expected result (it should be noted that there is a higher share of rent updates in the first trimester of each year, for historical reasons). The performance of this method, which reflects the exclusion of new contracts (where major rent changes are usual), make it as unsuitable for CPI purposes.

Method A, is the method that is more comparable to the survey-based rents index and, unsurprisingly leads to similar results. This method has the advantage of being of straightforward implementation. However, it did not take full advantage of the vast increase in sample size, as dwellings that exited the sample would not reappear until the following year. Since it became clear that most new contracts reappear after a period where the dwelling is not rented (or at least no receipt has yet been uploaded), we felt that improvements could be made to the method.

Method B, which takes into account all new contracts for any previously rented dwelling, constitutes the best compromise between the compilation of a constant-quality rents index and the inclusion of as much new rental contracts as possible in the compilation of the CPI. This method generates indices than are slightly higher than the survey-based baseline index and were considered, based on the analysis made, the best option to substitute the survey-based index.

4. Final considerations and the way forward

Despite the relatively short period of time available for the investigation of the potentialities of the new administrative data source (the first dataset was received in March 2017), it is possible to draw a few conclusions regarding its use in the Portuguese CPI. First, the new data source strengthens the reliability of produced indices since it eliminates sample variability and limits the impact of any dwelling-specific significant rent change. Second, the administrative data source allows for the compilation of methods that, while capable of addressing quality change in a simple and straightforward way, are also able to address the market dynamism of the rental market better than the previously compiled rents index. An example is given by Method B, which is able to take into account price changes from situation in which a new tenant comes into a dwelling that was left unoccupied (after the end of a contract) for at least two consecutive months. With the survey-based method, these new contracts were seldom considered in the calculations. Finally, the use of this new data source would reduce drastically the costs associated with the direct observation of rents (as this can be restricted to the collection of social or subsidized rents).

Following this analysis, Statistics Portugal decided to start the compilation of a new rents index based on Method B [14]. The first three months of the implementation of the new approach have shown an average monthly price increase of around 0.25% (a figure that is not much different from CPI past records on this CPI component). Despite preliminary analysis suggesting the contrary, it will be necessary to keep on monitoring whether the non-inclusion of rents that are not transmitted in time to be integrated in the calculation of the index impact significantly on results [15].

Further future developments would cover the questions that were not possible to be tackled empirically given the limited time and the relatively short period of the series available for research. In particular, the framework that is presented in the appendix for assessing the likely impact of not taking into account new entrants to the market and “aging depreciation” is going to be tested empirically.

[14] Statistics Portugal has also decided to use this new data source to calculate and disseminate median rent statistics at local level (Statistics Portugal, 2017).

[15] The rents data of month t is normally sent by the PFA around the end of the first week of the following month. However, for bureaucratic or data processing reasons, some rents information is sent in posterior transmissions of data (and thus not incorporated in the calculations of the CPI).

Appendices

Table 1 – Sample size by compilation method

	Method [16]			
	A	B	C	D
01-2016	231,197	231,197	188,691	156,015
02-2016	218,335	222,551	195,139	156,744
03-2016	190,135	197,134	194,567	164,974
04-2016	177,976	189,747	196,663	166,996
05-2016	169,563	186,717	198,902	168,300
06-2016	161,885	182,784	194,714	167,865
07-2016	155,476	179,063	192,100	165,327
08-2016	148,936	173,647	188,108	162,856
09-2016	143,076	172,769	197,819	162,570
10-2016	138,495	171,903	200,381	170,089
11-2016	134,834	172,329	204,095	173,640
12-2016	132,138	183,006	229,165	180,273
01-2017	266,523	266,523	216,646	184,051
02-2017	257,476	263,942	209,120	179,211
03-2017	230,021	242,502	212,854	180,714
04-2017	213,847	234,525	212,901	182,462
05-2017	205,681	232,953	213,399	182,028
06-2017	197,387	228,811	210,281	181,789
07-2017		224,194		
08-2017		218,726		
09-2017		215,135		
10-2017		216,182		
11-2017		217,208		
12-2017		232,783		
	Monthly Average			
2016	166,687	188,571	198,362	166,304
2017	228,429	232,790	212,534	181,709

[16] Method A and Method B include imputed values. The remaining methods do not include imputed values. Receipts that were uploaded after the monthly CPI calculation (first week of the following month) are not taken into account in the calculation but can be used in the following month imputation, if the conditions for imputation are met.

Table 2 – Indices for free market housing rentals

	Method				
	Survey-based	**A**	**B**	**C**	**D**
12-2015	100.000	100.000	100.000	100.000	100.000
01-2016	100.319	100.094	100.094	99.825	100.021
02-2016	100.547	100.179	100.186	99.903	100.036
03-2016	100.516	100.267	100.286	100.022	100.056
04-2016	100.590	100.359	100.392	100.339	100.094
05-2016	100.604	100.418	100.494	100.812	100.109
06-2016	100.669	100.489	100.604	100.815	100.132
07-2016	100.699	100.548	100.730	101.224	100.164
08-2016	100.796	100.603	100.871	101.344	100.166
09-2016	100.962	100.671	101.024	101.786	100.183
10-2016	101.319	100.762	101.170	102.347	100.214
11-2016	101.601	100.818	101.300	102.740	100.244
12-2016	101.608	100.884	101.378	102.714	100.279
01-2017	101.782	101.100	101.595	102.933	100.332
02-2017	101.550	101.204	101.748	103.262	100.375
03-2017	101.760	101.303	101.896	103.637	100.419
04-2017	101.618	101.401	102.046	104.065	100.458
05-2017	101.551	101.476	102.174	104.354	100.484
06-2017	101.639	101.556	102.316	104.664	100.526
07-2017	101.705		102.460		
08-2017	101.841		102.617		
09-2017	101.998		102.811		
10-2017	101.911		103.031		
11-2017	102.122		103.211		
12-2017	102.539		103.353		

Table 3 – Monthly rates of change for free market housing rentals

	Method				
	Survey-based	**A**	**B**	**C**	**D**
01-2016	0.32%	0.09%	0.09%	-0.17%	0.02%
02-2016	0.23%	0.09%	0.09%	0.08%	0.01%
03-2016	-0.03%	0.09%	0.10%	0.12%	0.02%
04-2016	0.07%	0.09%	0.11%	0.32%	0.04%
05-2016	0.01%	0.06%	0.10%	0.47%	0.02%
06-2016	0.06%	0.07%	0.11%	0.00%	0.02%
07-2016	0.03%	0.06%	0.12%	0.41%	0.03%
08-2016	0.10%	0.05%	0.14%	0.12%	0.00%
09-2016	0.17%	0.07%	0.15%	0.44%	0.02%
10-2016	0.35%	0.09%	0.14%	0.55%	0.03%
11-2016	0.28%	0.05%	0.13%	0.38%	0.03%
12-2016	0.01%	0.07%	0.08%	-0.03%	0.04%
01-2017	0.17%	0.21%	0.21%	0.21%	0.05%
02-2017	-0.23%	0.10%	0.15%	0.32%	0.04%
03-2017	0.21%	0.10%	0.15%	0.36%	0.04%
04-2017	-0.14%	0.10%	0.15%	0.41%	0.04%
05-2017	-0.07%	0.07%	0.13%	0.28%	0.03%
06-2017	0.09%	0.08%	0.14%	0.30%	0.04%
07-2017	0.07%		0.14%		
08-2017	0.13%		0.15%		
09-2017	0.15%		0.19%		
10-2017	-0.08%		0.21%		
11-2017	0.21%		0.18%		
12-2017	0.41%		0.14%		
Coefficient of Variation	148.9%	41.1%	25.6%	77.0%	44.2%

Table 4 – Annual rates of change for free market housing rentals

	Method				
	Survey-based	**A**	**B**	**C**	**D**
12-2016	1.61%	0.88%	1.38%	2.71%	0.28%
01-2017	1.46%	1.01%	1.50%	3.11%	0.31%
02-2017	1.00%	1.02%	1.56%	3.36%	0.34%
03-2017	1.24%	1.03%	1.61%	3.61%	0.36%
04-2017	1.02%	1.04%	1.65%	3.71%	0.36%
05-2017	0.94%	1.05%	1.67%	3.51%	0.37%
06-2017	0.96%	1.06%	1.70%	3.82%	0.39%
07-2017	1.00%		1.72%		
08-2017	1.04%		1.73%		
09-2017	1.03%		1.77%		
10-2017	0.58%		1.84%		
11-2017	0.51%		1.89%		
12-2017	0.92%		1.95%		

Assessing the impact of neglecting new rental contracts in the housing rentals component of the CPI

The starting point for the analysis of the impact of neglecting new (and disappearing) rents in the rental component of the CPI is an unadjusted Jevons index [17] between current (*t*) and base period (*0*), which can be defined as:

$$\ddot{I}_{t,0} = \frac{exp\left[\frac{1}{n_t}\sum n_t \, ln(r_{i,t})\right]}{exp\left[\frac{1}{n_0}\sum n_0 \, ln(r_{i,0})\right]} \qquad (1),$$

where $r_{i,t}$ stands for the observed rent in dwelling *i* in period t, n_0 is the total number of observed rents in base period (note, due to rents "births" and "deaths", n_0 is usually not equal to n_t).

Expression (1) can be defined as a "value" index, which we are interested to split between the traditional "price" and "volume" components and among the (A), (B) and (C) components. To make this separation, it is necessary to transform equation (1) in the following way:

$$\ddot{I}_{t,0} = \underbrace{\frac{exp\left[\frac{1}{n_t}\sum n_A \, ln(r_{i,t})\right]}{exp\left[\frac{1}{n_0}\sum n_A \, ln(r_{i,0})\right]}}_{(A)} \cdot \underbrace{\frac{1}{exp\left[\frac{1}{n_0}\sum n_B \, ln(r_{i,0})\right]}}_{(B)} \cdot \underbrace{\frac{exp\left[\frac{1}{n_t}\sum n_C \, ln(r_{i,t})\right]}{1}}_{(C)} \qquad (2),$$

where *(A), (B)* and *(C)* represent, respectively, the subsample of rents that are present in both periods (i.e., the matched-sample), those that are present in the base period (i.e., disappearing rents), and those available only in current period (i.e., new rents).

It follows from above that $n_0 = n_A + n_B$ and $n_t = n_A + n_C$. With a simple transformation, (2) can be written as:

$$\ddot{I}_{t,0} = \frac{exp\left[\frac{1}{n_A}\sum n_A \, ln(r_{i,t})\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A \, ln(r_{i,0})\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{1}{exp\left[\frac{1}{n_B}\sum n_B \, ln(r_{i,0})\right]^{\left(\frac{n_B}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_C}\sum n_C \, ln(r_{i,t})\right]^{\left(\frac{n_C}{n_t}\right)}}{1} \qquad (3),$$

Making use of an OLS estimator property for each of the subsamples, $E\left[ln(r_{i,0})\right] = E\left[\widehat{ln(r_{i,0})}\right]$, where $\widehat{ln(r_{i,0})} = x_{i,0}\hat{\beta}_0$, it is possible to make a correspondence between $\widehat{ln(r_{i,0})}$ and $x_{i,0}\hat{\beta}_{0;f}$, $f = (A, B, C)$ in (3), so that:

$$\ddot{I}_{t,0} = \frac{exp\left[\frac{1}{n_A}\sum n_A \, x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A \, x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{1}{exp\left[\frac{1}{n_B}\sum n_B \, x_{i,0}\widehat{\beta}_{0;B}\right]^{\left(\frac{n_B}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_C}\sum n_C \, x_{i,t}\widehat{\beta}_{t;C}\right]^{\left(\frac{n_C}{n_t}\right)}}{1} \qquad (4).$$

[17] The housing rental component of the Portuguese CPI is based on the Jevons formula.

Moreover, making some arrangements, it is possible to extend (*B*) into:

$$\frac{exp\left[\frac{1}{n_B}\sum n_B\, x_{i,0}\widehat{\beta}_t\right]^{\left(\frac{n_B}{n_0}\right)}}{exp\left[\frac{1}{n_B}\sum n_B\, x_{i,0}\widehat{\beta}_{0;B}\right]^{\left(\frac{n_B}{n_0}\right)}}\cdot\frac{1}{exp\left[\frac{1}{n_B}\sum n_B\, x_{i,0}\widehat{\beta}_t\right]^{\left(\frac{n_B}{n_0}\right)}}, i\ \in B$$

and *(C)* into: (5),

$$\frac{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_{t;C}\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_0\right]^{\left(\frac{n_C}{n_t}\right)}}\cdot\frac{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_0\right]^{\left(\frac{n_C}{n_t}\right)}}{1}, i\ \in C$$

It should be noted that, whereas $\hat{\beta}_{0;B}$ and $\hat{\beta}_{t;C}$ are estimated for the *(B)* and *(C)* subsamples, $\hat{\beta}_t$ and $\hat{\beta}_0$ are calculated with all the available data in periods *t* and *0*, respectively. Using the transformations illustrated in (5), the unadjusted price index (4) becomes:

$$\underbrace{\frac{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}}}_{(a)}\cdot\underbrace{\frac{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_0\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_B}\sum n_B\, x_{i,0}\widehat{\beta}_{0;B}\right]^{\left(\frac{n_B}{n_0}\right)}}}_{(b)}\cdot\underbrace{\frac{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_{t;C}\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_0\right]^{\left(\frac{n_C}{n_t}\right)}}}_{(c)} \qquad (6).$$

At this stage, it is possible to derive hedonic price rental price indices that take into account the marginal contribution of new rentals. To do this, it is necessary to transform (a) in (6) in the following manner:

$$\frac{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_t}\right)}}\cdot\frac{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \qquad (7).$$

Substituting (7) into (6), and focusing on matched and new entrant rentals - i.e., on *(c)* of (6) -, it is possible to derive a constant-quality price index as:

$$I^P_{t,0} = \frac{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum n_A\, x_{i,t}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_t}\right)}}\cdot\frac{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_{t;C}\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_C}\sum n_C\, x_{i,t}\widehat{\beta}_0\right]^{\left(\frac{n_C}{n_t}\right)}} \qquad (8).$$

This is “Paasche-type” index, where the focus in on period *t* and on new market entrants. Interestingly, (8) makes clear that total price change from period *0* to *t* is a weighted average

of the price behaviour observed in matched and new rentals market segments ($\frac{n_A}{n_t} + \frac{n_C}{n_t} = 1$). Moreover, looking back to what is left from (7) and (6), it is possible to have:

$$\frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,t}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_C}\sum_{n_C} x_{i,t}\widehat{\beta}_{0}\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_B}\sum_{n_B} x_{i,0}\widehat{\beta}_{0;B}\right]^{\left(\frac{n_B}{n_0}\right)}} \qquad (9),$$

which can be interpreted as a "Laspeyres"-type "volume" index, which measures the change in the value index that is explained by changes in the characteristics of rented dwellings.

It is important to note that (9) will only be a good proxy of "volume" if the following assumption - $\beta_{0;A} \cong \beta_{0;B} \cong \beta_0$ -, holds. This might be a reasonable assumption, as dwellings are goods that are not subject to rapid technological change and to drastic changes in tastes (e.g., people may now value bigger houses than 20 years ago, but this "valuation" is not likely to change from one monthly to another). However, if base period average shadow prices (i.e., the "betas") for all rented properties are substantially different from those available in both periods or that are no longer available in period *t*, then the "value" partition between "price" and "volume" might not be valid.

This analysis can be easily extended to focus on rental "exits". To do this, one would need to develop (a) in (6) as:

$$\frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_0}\right)}}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_0}\right)}},$$

which leads to the following "Laspeyres-type" constant quality index:

$$I_{t,0}^{L} = \frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_0}\right)}}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{0;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_B}\sum_{n_B} x_{i,0}\widehat{\beta}_{t}\right]^{\left(\frac{n_B}{n_0}\right)}}{exp\left[\frac{1}{n_B}\sum_{n_B} x_{i,0}\widehat{\beta}_{0;B}\right]^{\left(\frac{n_B}{n_0}\right)}} \qquad (10).$$

The focus of equation (10) is on base period and on the impact of rental "exits". In this case, the "volume" index is:

$$\frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,t}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_t}\right)}}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,0}\widehat{\beta}_{t;A}\right]^{\left(\frac{n_A}{n_0}\right)}} \cdot \frac{exp\left[\frac{1}{n_C}\sum_{n_C} x_{i,t}\widehat{\beta}_{t;C}\right]^{\left(\frac{n_C}{n_t}\right)}}{exp\left[\frac{1}{n_B}\sum_{n_B} x_{i,0}\widehat{\beta}_{t}\right]^{\left(\frac{n_B}{n_0}\right)}} \qquad (11),$$

where the assumption to now to hold is $\beta_{t;A} \cong \beta_{t;C} \cong \beta_t$. This might be a less credible assumption than the one that was done for the base period, especially for periods in which the market is "booming" (i.e., there are a lot of new entrants in the market, which may be very different from the rental units that are already on the market).

One way of investigating whether the matched-sample sample rental index is biased is to compare the Paasche-type price index (8) with a matched-sample index [18]:

$$I_{t,0}^{ms} = \frac{exp\left[\frac{1}{n_A}\sum_{n_t} ln(r_{i,t})\right]}{exp\left[\frac{1}{n_A}\sum_{n_t} ln(r_{i,0})\right]} \qquad (12).$$

What it is clear from above is that the degree to which the matched-sample index may differ from (8) or (10) has to do with two factors. The first one is the importance of new entrants or "exits" on the total transactions (i.e., the $\frac{n_B}{n_0}$ and $\frac{n_C}{n_t}$). The second factor is the degree to which price change in the new or disappearing rental markets deviates from that found in the subsample of all rentals that are matched in the two periods.

Moreover, to assess the possible impact of neglecting age effects on the rental index, (12) would be compared with the following index:

$$I_{t,0}^{Age} = \frac{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,t}\widehat{\beta}_{t;A}\right]}{exp\left[\frac{1}{n_A}\sum_{n_A} x_{i,t}\widehat{\beta}_{0;A}\right]} \qquad (13).$$

where naturally age is included in the hedonic model as an explanatory variable. In this context, the differences between (13) and (12) may be attributed to the fact that the former takes into account age depreciation and the latter does not.

[18] This comparison could be done with (10) as well or, more interestingly, with the geometric average of (8) and (10).

References

Eurostat (2014). *Harmonised Index on Consumer Prices – Recommendations: Treatment of actual rentals in the HICP*. Version of August 2014, TFQI09.2014-02. http://ec.europa.eu/eurostat/documents/272892/7048317/HPI-HICP+Recommendations+on+the+treatment+of+rents+-+June+2015

Gonçalves, I. and Evangelista, R. (2017). *A transactions-based commercial property price index for Portugal*. Paper presented at the 15th Ottawa Group Meting.

ILO, IMF, OECD, Eurostat and World Bank (2004). *Consumer Price Index Manual: Theory and Practice*. ILO, Geneva. http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_331153.pdf

Ministerial Decree (2015). *Portaria nº 98-A/2015 de 31 de Março*. https://dre.pt/web/guest/pesquisa/-/search/66896468/details/normal?l=1.

Shimizu, C., Imai, S. and Diewert, E. (2015). *Housing Rent and Japanese CPI. Bias from Nominal Rigidity of Rents*. Paper presented at the 14th Ottawa Group Meeting. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t2s6p20_pre.pdf

Statistics Portugal (2014). *Inquérito às rendas de habitação: Documento metodológico*. http://smi.ine.pt/UploadFile/Download/1773.

Statistics Portugal (2018). *House rental statistics at local level*. Press release. March 2018. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=314754447&DESTAQUESmodo=2.

Krsinich, F. (2009). *Using hedonic regression to assess the housing rentals component of the New Zealand Consumers Price Index*. Paper presented at the New Zealand Association of Economists 50th Anniversary conference.