

The use of administrative data for Italian private housing rental prices index compilation

Alessandro Brunetti, Orietta Patacchia, Federico Polidoro¹

Abstract Since January 2022, the Italian CPI and HICP indices of housing rentals for private sector are calculated using administrative data. In particular, the new indices are based on the information contained in the database of rents, which the Tax Office has been providing to Istat since 2017. The database is used for a twofold purpose: firstly, the estimation of the weight for “Existing dwellings new to the household sector” for the Owner Occupied Housing Price index (OOHPI) compilation; secondly, the production of housing rentals consumer price index for private sector (excluding social rents) replacing traditional survey data collection as source of basic information. The main advantage of using administrative data to replace those coming from survey is the coverage of almost the entire private rental market thanks to the obligation of rent contract registration when the duration is greater than 30 days a year. This paper gives an overview of steps followed to obtain data useful for index calculation: the removal of buildings not used for housing purposes; the integration with statistical registers (such as the Statistical Register of active Enterprises, ASIA); the identification and treatment of duplicates detected in the same month; the detection of contracts regarding the whole apartment; the removal of outliers in terms of dwelling area and rent per square meter. In addition, the paper focuses on the stratification strategy and strata weight construction. Finally, method used to calculate rental price indices and the results obtained are reported, also comparing the new indicators with the indices based on the data derived from the traditional data collection.

Key words: administrative data, rents

¹ Alessandro Brunetti, Istat, albrunet@istat.it
Orietta Patacchia, Istat, patacchi@istat.it
Federico Polidoro, Istat, polidoro@istat.it

1. Introduction

Since January 2022 the Italian CPI and HICP indices of housing rentals index for private sector are calculated using administrative data. In particular, the new indices are based on the information contained in the database of rents held by the Real Estate Market Observatory of the Tax Office.

The new database is used for a twofold purpose: firstly, the estimation of the weight for “Existing dwellings new to the household sector” for the Owner Occupied Housing index (OOHI) compilation; secondly, the production of housing rentals consumer price index for private sector (excluding social rents) with the aim to replace traditional survey data collection as source of basic information. On the latter the paper focuses the attention.

The provision of the data base by Tax Office started in 2017 in the framework of the cooperation between Istat and the fiscal agencies of the Ministry of Economy and Finance (established in a formal agreement). Then, the Italian National Statistical Program, for each statistical survey, elaboration, research project, makes explicit the possible use of administrative data (coming also from other institutions), regulating their treatment and use for statistical aims, guaranteeing the protection of personal data information. The agreement between Istat and the Ministry of Economy and Finance and the Tax Office represents the basis for the transmission to Istat of the data related to acquisition of dwelling by households that allows, since 2012, the regular compilation of Italian HPI and then OOHPI.

The availability of information on private rentals prices from 2017 in the Tax Office data base, allowed Istat to analyze the characteristics of this data base, asking to improve its quality (in a very good environment of cooperation with the experts of the Tax Office) and testing the treatments necessary to use these data for the computation of CPI and HICP. From quarterly the provision is going to become monthly from the data of the second quarter of 2022 improving a lot the timeliness and frequency of the update of the basic data to produce the monthly indices.

Previously (until 2021) the data collection of the information related to prices of private rentals were collected locally by the Municipal Offices of Statistics either through data collectors interviewing monthly a sample of households or receiving the monthly updates of the prices of rentals from the data bases of big real estate owners or obtaining the monthly data from the local institutions in charge of managing social housing (the rental prices of the dwellings owned by public institutions are still collected at municipal level).

Indeed, the availability of the admin data through the data base of registered rentals allows a complete change in survey design. From a survey based on a sample Istat has moved to a survey based to almost the census of the statistical units to observe that meant a complete change in the data treatment in the methodology of calculation of the indices. Actually the data base does not include the entire universe of private rentals because it does not contain the information on the rentals agreed before 2017 but it could be considered widely representative given the characteristics of the private rental market and the average duration of the contracts. Moreover, in 2022 the data used have been those referred to provincial chief towns, having planned the extension of the indices coverage to the entire national territory since 2023.

Characteristics of the data base and analysis of the treatment of the basic information on one side and methodology adopted to compile the indices are the two main topics of this paper that stresses the new frontier for CPI and HICP represented by the use of admin data as further data source in the framework of the evolution of consumer price survey towards a multi-source and multi-technique survey.

2. Administrative data source description and treatment

2.1 Main characteristics of the Tax Office data base and preliminary selection of the statistical units

The database of rents has been provided quarterly to Istat since 2017. It contains information about all new rental contracts registered in the reference quarter referring to all type of properties (residential

and non-residential) and it is provided 35/40 days after the end of the reference quarter. The provision of the data occurs in the context of a Memorandum of Understanding between the Ministry of Economy and Finance, Fiscal Agencies and Istat, aimed at the management and development of the statistical information exchange system. Database of rents covers the entire national territory but for properties located in the autonomous provinces of Bolzano and Trento some information are missing (in particular dwelling area in terms of squared metres and the number of rooms).

The new database combines information belonging to various data sources. In particular, the starting point are data collected from the model used to register rental agreements and to communicate any extension, transfers or resolutions (Registration Leases Properties model, RLI). Then, cadastral information and the so called OMI zones² are added. In more details the data base of rentals provides information about: the type of property in term of cadastral categories; the type of the contract; the size in terms of square meters and/or number of rooms; the location of the property (cadastral code; cadastral information as map sheet and number; OMI zone; address); the value of rental price; the start date of occupancy, the end date of occupancy and sometimes the date of the early termination (information useful to follow the life of the contract); information about landlords and tenants including fiscal code; the tax system in particular the option for a flat-rate tax³.

² OMI (the acronym is derived from Osservatorio del Mercato Immobiliare, Observatory of Real Estate Market) zones are territorial areas homogeneous in terms of main characteristics of the real estate market. The definition provided by the Observatory is the following "Portion of the territorial strip that reflects a homogeneous sector of the local real estate market, in which there is a substantial uniformity of appreciation for economic and socio-environmental conditions. This uniformity is translated into homogeneity of the values of market of real estate units included in a range with one deviation between minimum and maximum value, as a rule, not more than 50%." The OMI zones are almost 27,500 (to have a term of comparison, Italian municipalities are less than 8,000).

³ To better understand the main characteristics of the flat-tax regime, look at paragraph 2.2

Administrative source is treated to become useful for the statistical aims related to HICP/CPI estimations. In particular, properties not used for housing purposes are removed. Moreover, the integration with statistical registers such as the Statistical Register of active Enterprises (ASIA) through VAT code of the property owners allows the identification of the owner type in order to distinguish between private and social contracts. Finally, only contracts where the property owner is a physical or a legal person and the tenant is a household are used for inflation estimation (around 84-87% of the yearly number of new rental contracts); also only contracts involving a whole apartment belonging to residential categories are selected. Table 1 summarizes main characteristics of the database of rents and the results of first selections.

Table 1. Recap of the main characteristics of the database of rents

Data source	Database of new rental contracts (Tax Office)			
Data availability	From 2017			
Geographical coverage	Entire national territory (special case: Trentino)			
Frequency	Quarterly (monthly from the second quarter of 2022)			
Reference period	Quarter (month from the second quarter of 2022)			
Timeliness	35/40 days after the end of the reference quarter (it will be reduced with the switch to the monthly provision of the data)			
Type of property	Residential	A/1 – Luxury dwellings A/2 – Well-finished (civil) dwellings A/3 – Economic dwellings A/4 – Popular dwellings A/5 – Ultra-popular dwellings A/6 – Rural dwellings A/7 – Detached houses A/8 – Villas A/9 – Castles, palaces of great artistic and historical value A/11 – Traditional local dwellings and accommodations		
	Non-residential (out of scope)			
Type of contract	Free contracts or agreed fee contracts; long or short term contract			
New contracts (thousands)	Year			
	2017	2018	2019	2020
	1,682	1,670	1,752	1,545
Target population (thousands)	1,417 (84%)	1,407 (84%)	1,486 (85%)	1,345 (87%)
Selected contracts (thousands)	813 (48%)	814 (48%)	844 (48%)	769 (50%)
Main variables	Dwelling location including OMI zones; size; rent			

2.2 Data cleaning

After the operations aforementioned to select the statistical units, data cleaning processes are carried out. In particular:

- dwellings rented for less the 32 days are excluded as a proxy of short-stay accommodation;
- duplicates in the same month and in the same quarter have been detected and removed (for example, contracts regarding the same dwellings with the same start date of occupancy but different key registration probably due to mistakes during documents compilation: the last one is used for calculation);
- ambiguous contracts have been separated (for example, if the start date of occupancy and the date of the early termination of a contract are coincident or nearly coincident);
- outliers in terms of dwelling area and rent per square meter were dropped out from calculation (rent must be more than 0.5 euros per square meter and less than 110 euros per square meter).

As a result, around 8-10% of the observations was excluded.

Table 2: Cleaning procedures, some results (*data in thousand*)

	Year			
	2017	2018	2019	2020
New contracts	1,682	1,670	1,752	1,545
Cleaned data	647 (38%)	700 (42%)	730 (42%)	655 (42%)

2.3 Obtaining final price of the rent calculation

Database of rental prices provides information about rentals on monthly or annually base. The final price used for index compilation takes into account the registration tax eventually due, depending on whether the lessor chooses or not the tax system known as "flat-rate tax" (in Italian "cedolare secca"), which entered in force on the 7th of April 2011 with the Legislative Decree of 14 March 2011, n. 23 concerning "Provisions of Municipal Fiscal Federalism". The above mentioned Legislative Decree introduces a new voluntary rental income taxation system. This flat-tax system, which is alternative to the ordinary one (that calculates the amount of tax due using the marginal rate that the owner has on her/his personal income), is

reserved for individuals who have the ownership or other real right of use of rented dwellings; so it has effects only in the calculation of index for actual rentals paid by tenants to private lessors. The new taxation system is voluntary and therefore its application depends on the lessor option that has to be chosen by the lessor at the registration time of the rent agreement, communicating it to the tenant. The grant to this taxation system carries takes place for the entire duration of the contract, unless revocation. It allows the lessor to have a preferential taxation (flat-tax) on the income received from the rental, also with effects for the tenant who is no longer required the payment of the registration tax of the rent agreement. As counterpart for this fiscal advantages, it entails the suspension of lessor right to ask for an updating of rent, even if it is foreseen in the contract.

If the lessor does not option for flat-rate tax, the registration tax on rent agreement is also added to the rent provided by Tax Office. The tax payment is annual. If the contract lasts several years, it is possible to pay either the tax due for the entire duration at the time of registration or the tax each year within 30 days of the expiry date of the previous annuity. The registration tax depends on the type of the contract. For ordinary contracts it is 2% of the annual rent with a minimum tax of 67 euros for the first year (for subsidized contracts the 2% registration tax is calculated on 70% of the annual rent). The lessor and the tenant are liable for the payment in equal parts. In conclusion, hypothesizing the annual payment mode and the share-out in equal parts of registration tax between lessor and lessee, the amount of registration tax, calculated as 1% of annual rent (or 70% of annual rent for subsidized contracts), has been added to the rent (for the first registration the amount cannot be less than euro 67 euros).

An ad-hoc procedure has been implemented to deal with collected prices referring to contracts whose fees need to be updated according to the consumer prices for blue and white-collar worker households index (FOI). In particular, a SAS procedure is used to detect rents to be updated yearly (mainly long contracts when the owner does not apply the option for flat-rate tax) and to calculate the new rent taking into account the inflation index calculated by Istat.

3. Compiling the rent index for Italian HICP and CPI

3.1 Stratification

After the selection of the statistical units, the cleaning steps implemented and the setup of the rules to update the rentals related to contracts where the flat-tax regime is not fixed, the price indices of rentals for HICP and CPI, are calculated using a stratification approach to control for quality changes in the composition of rented properties. In theory, the more detailed the stratification, the more the index controls for changes in the characteristics of rented properties covered by the index. Anyway, too many strata could have a very small number of observations within some of them and in some case to empty strata.

In general, for the stratification purposes, different combinations of variables have been hypothesized in order to choose the best ones in terms of representativeness and number of observations in the cells. As a first step, on the basis of the analyses conducted, categories A2, A3 and A4 (civil, economic and popular) were selected, which cover approximately 95% in terms of rentals. Furthermore, given that, as aforementioned, in 2022 the choice was to limit the coverage to provincial chief towns, a distinction has been implemented between two groups of chief towns: the biggest ones (in terms of demographic size) have been broken down by groups of OMI zones (macro areas that are available only for 11 towns⁴), the others by OMI segments⁵.

In the case of provincial capitals with macro-areas (11 municipalities), the stratification takes into account the following variables: macro-area, type of contract (ordinary; subsidized), market segment (long and transitory for ordinary market; agreed and transitory for subsidized market) and surface (a subdivision into

⁴ The 11 towns for which macro areas are available are the following: Bari, Bologna, Catanzaro, Firenze, Genova, Milano, Napoli, Palermo, Roma, Torino, Venezia.

⁵ An OMI segment is an aggregation of contiguous homogeneous zones. It represents a territorial area with specific geographical location in the municipality and reflects a consolidated urban location. Each municipal territory is divided into the following segments: Central, Semi-central, Peripheral, Suburban, Extra-urban.

classes based on quartiles has been adopted in order to guarantee a minimum number of observations per cell equal to 25% of the observations). An aggregation at a higher level is carried out if the number of observations is deemed insufficient (for example if, for a certain macro-area, the number of observations for the facilitated transitional contracts relating to a certain surface class is insufficient, these are aggregated so that the elementary layer refers to facilitated transitional contracts, without distinction by class of surface). The number of strata, making reference to the third quarter of 2021, is equal to 1057. Of these, 7% have a number of monthly observations exceeding 1000; 57% a number of monthly observations between 101 and 1000; 17% between 51 and 100; 13% between 31 and 50; 5% between 11 and 30.

In the case of provincial capitals without macro-areas, the stratification is made on the basis of: the aggregate segment (central and semi-central; peripheral, suburban and rural); the type of contract (ordinary; facilitated); the surface (always on the basis of quartiles). An aggregation at a higher level is carried out if the number of observations is deemed insufficient. There are 4 elementary layers with an aggregate segment only; 23 are the strata with an aggregate segment and type of contract; 1380 with also surface class. The number of strata, making reference to the third quarter of 2021, at the end, is equal to 1407. Of these 0.2% have a number of monthly observations higher than 1000; 57.9% a number of monthly observations between 101 and 1000; 20% between 51 and 100; 10.9% between 31 and 50; 11% between 11 and 30.

The stratification determined on the basis of the data of year t-1 is kept constant for the calculation of the indices relating to year t. In total the strata identified are 2,464.

3.2 Compilation of the indices

The approach chosen for rentals price index compilation depends on data available and on the results of analysis. In particular, the same methodology used for other products in the basket is followed but the

set of internal weights has been constructed in a different way. In particular:

- for each provincial chief town, the rental price index is calculated as a weighted arithmetic means of strata indices whose weights are calculated from admin data base (transaction values). Elementary indices (strata indices) are calculated as the ratio between the geometric mean of rental prices of all the contracts identified in each in stratum that are in force in the current period and the geometric mean of rental prices of all the contracts identified in each in stratum that were in force in the base period (December of previous year; in 2022 it was the last quarter of 2021)⁶. In general, for the computation of the index for the month (m) of the year (t) dwellings regarding contracts started in the period from January 2017 to $m-1$ and not expired in the month m have been included together with the new contracts signed and entered into force in month m . We could say that it is adopted a combination of panel and cross approach to the final selection of the statistical units. Then, the annual rent is calculated and the price per square meter is considered. Likewise, about weight calculation, contracts started from 2017 and not expired in December of the year ($t-1$) together with those entered into force in December of year $t-1$, have been included; then from monthly datasets. Weights are updated yearly and are estimated on the basis of the same data base used for processing price information. In particular, for indices for the year (t) the weight calculation is based on data of the year ($t-1$);

- Having compiled the indices at level of provincial chief town, for each region, rental prices indices are calculated as a weighted arithmetic means of indices of provincial chief towns using weights calculated on the basis of the population living in rented houses⁷.

⁶ For the computation of the index for the month (m) of the year (t) dwellings regarding contracts started from 2017 and not expired in December of the year ($t-1$) have been included.

⁷ This was the choice carried out coherently with the weights system of the other elementary aggregates in the Italian HICP/CPI basket. It is in progress the revision of this approach.

3.3 Special case: the region of Trentino Alto Adige

Database of rents covers the entire national territory but for properties located in the autonomous provinces of Bolzano and Trento (both belonging to Italian region called Trentino-Alto Adige) some information are missing (in particular dwelling area and the number of rooms). Therefore, index compilation for this region is based on an exclusively panel approach.

Concerning stratification, only the type and the duration of the contract is taken into account (free long or short term contract; agreed fee long or short term contract).

Concerning the estimation of the indices, the panel of dwellings selected in the period of reference of the basis is kept stable, with the addition of possible dwelling for which rental contract entry into force over the year. The observations that could be used are classified in 3 cases:

Case A: the rental price of the dwelling is available in both the base and the current period.

Case B: the rental price of the dwelling is available in the base period not in the current one (for example the contract is finished and no information about new contract is available). In this case the price of the rental of this dwelling is imputed by multiplying the prices in the reference period (December) by the geometric mean of the price relatives of the remaining matched items in the stratum. In practice, the series is continued in the database by including the imputed price.

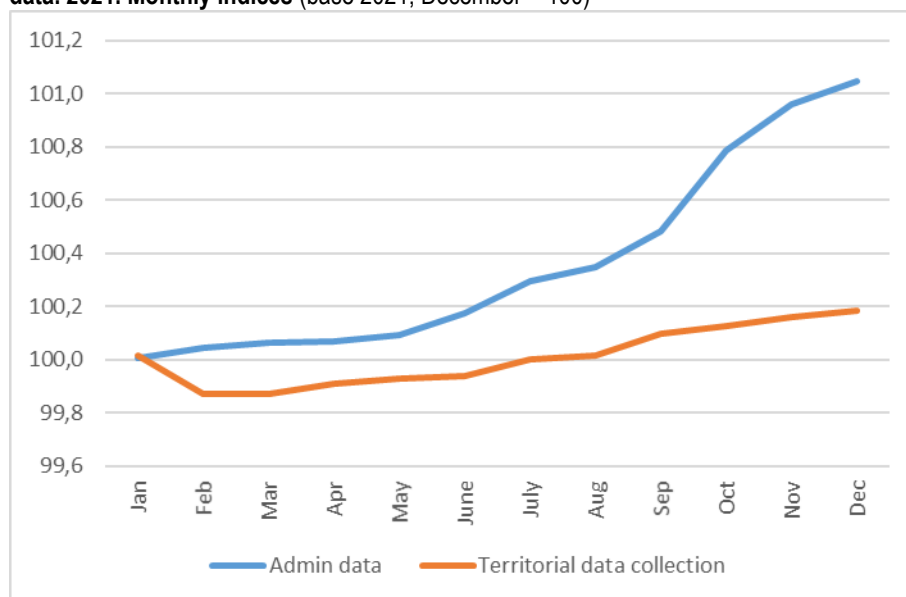
Case C: new contract in the current period. The price is missing in the base reference period and this price is imputed using the back rate of change of the index in the stratum under the hypothesis that the index of price rental of that contract would followed the evolution of the rental index of the stratum in the previous months.

4. Comparison between indices from administrative data and indices from survey (year 2021)

For 2021 it was possible to calculate the indices of private rental prices and compare these indices with the indices calculated using the information coming from the territorial data collection and that were used to estimate both Italian HICP and national CPI. Figure 1 shows the two indices and although the short period considered, it is possible to do some preliminary remarks:

- In December 2021, the m/m-12 rate of change of the index based on the data collected at territorial level is equal to 0.2%, that of the index compiled on the basis of admin data is equal to 1.0%;
- Therefore, given that in the Euro-area the HICP of actual rentals registered a rate of change equal to 1.2% in the same time span, it seems that the new data base for Italian index perform better in terms of detection of the temporal evolution of such important component of the HICP/CPI basket;
- The main reason of this improvement could be attributed to the previously unknown possibility of taking into account in the monthly estimation of this index of the new contract and not only of those already established in base reference period.
- The issue of rentals of secondary dwelling (indices of ECOICOP subclass 04.1.2.1 is not calculated yet) is still open and its solution could produce a further improvement and refinement of the actual rentals indices adding the coverage of this ECOICOP subclass to the Italian HICP/CPI estimation.

Figure 1. HICP/CPI indices of private rental prices compiled using territorial and admin data. 2021. Monthly indices (base 2021, December = 100)



5. Concluding remarks and next steps in the development of actual rentals price indices

The latter evidences analyzed shortly in the previous paragraph confirms how much the availability of private rental prices administrative data base of Tax office has represented an important improvement in the estimation of Italian consumer price indices of actual rentals for housing and of inflation in general.

The improvement obtained has to be attributed to different reasons related to the characteristics of this data base:

- The coverage in both territorial and temporal terms that give the possibility of considering all the municipalities, all the different types of contracts and either of the contracts already into force and those newly signed;

- The possibility of defining strata that are enough homogenous in terms of characteristic of the rental real estate market, allowing the monthly comparison of like with like;
- The variables contained in the data base that allow to identify correctly the prices to be used in the compilation of the indices.

As emerged in the previous paragraphs this new data source brings also some drawback, thereof some are close to be overcome. The two main drawbacks are the frequency (quarterly until the first quarter of 2022) and the timeliness (the contracts signed in month *m* are available about 35/40 days later). The first of the two drawbacks is on the way to be solved, the second one should be drastically reduced together with the monthly provision of the data that should reduce the delay to 15 days, thanks to the sound and productive cooperation with the experts of the Real Estate Observatory of the Tax Office.

Finally, the new data base opens the following perspectives:

- ✓ Covering all the municipalities in the Italian territory (not only the provincial chief towns as done until now);
- ✓ Exploring the possibility to cover ECOICOP subclass 04.1.2.1 Actual rentals paid by tenants for secondary residences;
- ✓ Improving the estimation of imputed rentals at least in the frame of HBS (given that this component is not considered neither in the Italian national CPIs);
- ✓ Using the information of this data base to improve the estimation of the thresholds of absolute poverty in Italy thereof one component is related to housing.

References

To be completed