Chapter 16

Validation of Housing Data

The International Comparison Program (ICP) collects housing data through both its rental survey and its dwelling services questionnaire (as described in chapter 9).

For ICP 2011, the Global Office prepared a form on rental specifications for the rental survey (annex A), which ideally is conducted twice in the benchmark year to obtain annual average rents. The form breaks down dwelling types into 64 categories by applying eight criteria, including dwelling type, size, utilities, and age information (see chapter 9 for details).

All participating economies also were required to provide the dwelling services questionnaire (annex B). It was used to collect information for the volume method, which indirectly computes purchasing power parities (PPPs) by dividing the ratios of the volumes of dwelling services into their expenditure relatives. The questionnaire asks for number of dwelling units, rooms, and occupants, as well as information on utilities by dwelling type, construction type, and location. Table 16.1 is a summary of the types of data collected for the reference year.

If the user cost method is applied, the relevant information on housing expenditures is necessary, as specified in chapter 9.

Because dwelling services can be measured using two methods, the validation processes for the methods differ. The rental information is basically price data that share similar characteristics with other surveys, including household consumption, machinery and equipment, and compensation of government employees. As a result, the validation procedure largely follows the same steps, such as validation table analysis, that the other price data require. For the volume of dwelling data, the relevancy of the data had to be verified with simple arithmetic checks and by knowledge of the housing market and economic structure. Therefore, coordination among regions and the Global Office or World Bank is critical in securing data quality.

The general validation has three stages: intraeconomy validation, intereconomy validation, and global validation.

Table 16.1 Housing Data Collected for Reference Year, Rental Survey and Dwelling Services Questionnaire, ICP 2011

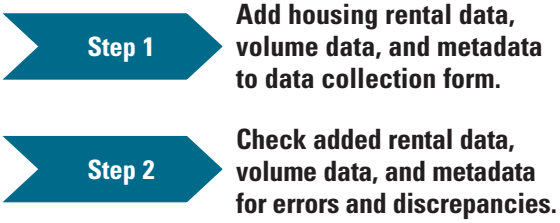
	Rental method	Volume method
Survey form	Rental survey	Dwelling services questionnaire
For	64 dwelling categories specified globally by • Dwelling type (e.g., villa/single-family house) • Size • Electricity • Inside water • Private toilet • Air-conditioning or central heating • Structure age	All dwellings broken down by • Construction type (modern or traditional) • Location of dwelling (urban/rural)
Data	• Yearly rent (local currency unit) • Location (urban/rural) • Comments	• Number of dwelling units • Number of rooms • Usable surface area (1,000 m^3) • Number of occupants • Land area occupied by dwellings (1,000 m^3) • Number of dwelling units with electricity/inside water/private toilet/central heating/air-conditioning • Percentage of dwelling units (rented/owner-occupied)

Source: ICP, http://icp.worldbank.org/.

Intra-economy Validation

During the intra-economy validation stage, the national coordinating agencies (NCAs) validate both the rental survey and the dwelling services questionnaire data to ensure the quality of data using the following steps.

Initial Data Validation



The rental prices, volume data, and metadata are checked for any errors or discrepancies. The NCAs first confirm whether prices are based on market rents. Subsidized rents such as those paid by employees living in dwellings owned by their employers or by tenants in low-rent public housing are not market rents and should not be reported on the questionnaire. In addition to rents, the NCAs check whether the weights in the total rental stock by dwelling type are correctly reported.

The global specifications provide size, age, and other specifications such as the existence of an air-conditioning unit (see box 16.1). Within the same type of dwelling, it can be assumed in most

BOX **16.1**

Example of Validation within the Same Dwelling Type, ICP 2011

Dwelling type	Size (m^2)	Air-conditioning or central heating	Structure age	ABC Republic (ABC dollars)
Villa/single-family house	120–180	No	>5 years	5,040
Villa/single-family house	120–180	Yes	>5 years	5,800
Villa/single-family house	120–180	No	<5 years	6,800
Villa/single-family house	120–180	Yes	<5 years	7,860
Villa/single-family house	180–240	No	>5 years	6,400
Villa/single-family house	180–240	Yes	>5 years	7,120
Villa/single-family house	180–240	No	<5 years	7,640
Villa/single-family house	180–240	Yes	<5 years	7,720

In this case, the rental prices within the same size are plausible. However, when they are compared with a different size under the same dwelling type, the better option in size, 180–240 square meters, has a lower rental price than a smaller unit of 120–180, even though the other criteria in the specifications are the same. Therefore the price should be flagged for further investigation because of the possibility of a price error or a product error. If there is an acceptable reason for the difference, such as location or quality of the building, the price should not be removed or edited.

cases that a smaller living unit is less expensive than a larger one, a unit in an older structure is less expensive than one in a newer structure, a unit without air-conditioning is less expensive than one with air-conditioning, and so forth. In other words, keeping other elements constant, when a better option (larger, newer, more equipped, etc.) has less expensive rentals, it should be flagged for further checking and vice versa. The need for a further check may not hold because of possible differences in the elements not described in the specifications (e.g., an older building can be located in a more valuable area). However, this check provides a list of rental prices that should be flagged for checking.

Step 4 Check that rental data are geographically plausible.

In most economies, urban areas have higher housing costs than rural areas because of higher demand. Therefore, when data for urban and rural areas are provided, the rental prices should be compared and checked between urban and rural areas for the same dwelling type.

More broadly, the location information on urban or rural should be carefully examined at the intra-economy level and reported to the regional coordinating agencies (RCAs) in order to conduct intereconomy comparisons properly. When urban and rural data are available, they should be averaged to obtain national average rental prices. In doing so, weighted average is recommended, and the weights are supposed to be the number of dwellings in each location. If that information is not available, population could be used as approximate weights.

Step 5 Check that volume data are plausible in comparison with other data sources.

Comparisons with data from other official statistics such as a census survey are one of the first tests for the volume data (see box 16.2).

The basic arithmetic checks with other official statistics would include:

- Ratio of [total number of dwelling units] to [population]
- Ratio of [total number of dwelling units] to [number of households]
- Ratio of [total number of occupants] to [population].

Usually, the total number of dwelling units is lower than the population number; it is closer to the number of households. The total number of occupants should be close to the population number. If any unusual ratios are observed, they should be flagged for further investigation. However, there are exceptional cases for any of the tests depending on the economy's situation. Thus the NCA's or other expert's knowledge of the housing market should be fully utilized in the validation process. For example, in ICP 2011 exceptional cases were observed where the ratio of the total number of dwellings to population exceeded 1—for example, on some small islands with many resort houses.

Step 6 **Check that volume data are plausible in relation to information within the questionnaire.**

The relevancy of the volume data also should be checked in relation to the other indicators in the dwelling service questionnaire. The basic tests include:

- Sum of the total correct within each category (all dwellings/type of construction/location of dwellings)
- Sum of the total the same across all three categories
- Ratio of [number of rooms] to [number of dwelling units]
- Ratio of [units with air-conditioning/central heating] to [number of dwelling units]
- Ratio of [units with private toilet] to [number of units with inside water].

The number of rooms per unit cannot be less than 1, and the number of units with a private toilet cannot be larger than that with inside water because a private toilet is generally available when inside water is available. These general

BOX **16.2**

Example of Validation of Volume Data with Population, ICP 2011

Number of dwellings/population	25%
Number of occupants/population	385%

In this example, even though the ratio of the number of dwellings to population is plausible, the number of occupants provided does not fall in the same range as population, and thus it should be flagged for further investigation because of the possibility of an error.

Number of dwellings/population	79%
Number of occupants/population	385%

However, coverage of other official statistics needs to be carefully examined as well. In this example, both the number of occupants and the number of dwellings are too high compared with the population. It is thus possible that the source of the population data is based on a smaller area than the housing information.

Because of the difference in source data, the number of occupants and population cannot be exactly the same, although it should be within an acceptable range. Usually, the gap is within the range of ±10 percent. The ratio of total number of dwelling units to population differs for each economy, but usually falls in the range of 15–40 percent.

When comparing data, the scale of the units used (such as thousands, millions, etc.) should be carefully checked to ensure they are comparable. In ICP 2011, some data were provided at the wrong scale, but this can be avoided with the checks just described.

rules work as a useful check to validate the relevancy of the provided data. The same applies to the tests with population data, although these do not necessarily hold if there is a plausible reason. For example, the ratio of units with central heating would be nil in tropical countries, but it would be very high in cold regions.

Step 7 **Check that rental, volume, and expenditure data are generally in line.**

Theoretically, an economy's expenditure for basic heading 110411.1, actual and imputed rentals for housing, in the national accounts can be approximated using rental price data and volume information by type of construction. More specifically, the housing expenditure should match the total of the average rental prices for a villa/single-family house and attached house/row house multiplied by the housing volume for modern construction houses; a studio/one-bedroom/two-bedroom apartment multiplied by the housing volume for modern construction flats; and typical/traditional dwellings multiplied by the housing volume for traditional construction.

In reality, because of the differences in the systems of data collection used for expenditures versus rental and volume, as well as infeasibility issues,[1] the systems produce only rough sketches of the relationships. Thus those data would not yield any exact matches in numerical value. However, the calculation with simple arithmetic averages should still give rough approximations, and if the expenditure from the national accounts data is too large or too small, something could be wrong with the data. If so, the NCA should check the breakdown figures to determine the source of the problem.

Step 8 **Check that data are temporally plausible by comparing them with those from the previous ICP round.**

As mentioned earlier, the rental survey should be conducted twice in a benchmark year, and a temporal analysis between the data from the two surveys is highly recommended to check the plausibility of results. In doing so, attention should be paid to seasonality as well as inflation between the two survey periods because in some cases there is a tendency for higher yearly rents to be reported during the peak season. Yearly rents are reported for each survey, taken as an average from each survey to obtain the annual average rents.

If dwelling services data are available from previous rounds of the ICP, the NCAs are encouraged to carry out temporal analysis between ICP rounds to check the plausibility of the new information.

If possible, the volume data also should be compared with data from earlier rounds of the ICP. Temporal analysis helps the NCAs verify the validity of the gathered information. Data from the latest census and other surveys can be utilized for validation.

Step 9 **Analyze price data and metadata for flagged cases.**

Finalization of Data

Initial data validation → **Finalization of data**

Step 1 **Confirm that rental data, volume data, and metadata have been intra-economy validated.**

Step 2 **Submit rental data, volume data, and metadata to the regional coordinating agency.**

Intereconomy Validation

Intra-economy validation → **Intereconomy validation** → **Global validation**

After receiving all data and metadata from the NCAs, the RCA conducts the regional-level validation. Like validation of the household consumption survey, this level of validation is

a collective process involving the RCA and a group of economies. Rental data collected in economies in the region should be checked for their accuracy and comparability. Even though the RCA leads the process, the active involvement of the NCAs is essential because their cooperation is required to investigate the data when the RCA finds any potential problems.

The RCAs carry out steps similar to those carried out during the intra-economy validation stage. This section focuses on the steps unique to the intereconomy validation process.

Initial Data Validation

Add rental data, volume data, and metadata to the data validation tool.

Once economies have submitted the national annual rents for each dwelling type applicable to the local situation, the RCAs ensure that the economies have followed the agreed-on global or regionally modified specifications.

Convert rental prices in local currency into common currency using the annual average exchange rate.

After the completeness of the data and metadata has been verified, rents are then converted from the local currency units (LCUs) into a common currency.

Check that rental and volume data are plausible within an economy.

Although this step is similar to the validation carried out at the economy level, it is important and beneficial that this step be carried out by the RCA, which is knowledgeable about the entire region.

Check that rental and volume data are plausible across economies.

Now that the data are based on the reference quantity and expressed in a common currency, the rental prices can be compared across economies (see box 16.3). Even when the data are plausible in the intra-economy validation, conducting basic checks across economies before proceeding to statistical tests is beneficial. The RCAs need to employ their knowledge of the economic structure and situation of each economy in their region.

Volume data also have to be validated, taking into consideration the regional context. Each region would have some common housing features, and thus flagging the outliers in ratios checked for the intra-economy validation would be highly beneficial. For example, the ratio of the number of rooms to the number of dwelling units would be similar across a region because of the similarity in construction methods, or the ratio of units with air-conditioning or central heating to the number of dwelling units would be close in a region because of the similarity in climate. Therefore, it is crucial to compare those ratios across economies to find outliers to be flagged for checking.

The RCAs ensure that economies have provided the relevant information to the extent possible and that the data are comparable across economies. The evaluation of completeness of quality indicators is critical at this point because the ICP seeks to compute quality-adjusted volume measures for more accurate comparisons.

Once the RCAs confirm that the quality indicators are provided along with the volume measures, they calculate the quality-adjusted volume, as described in chapter 9. Quality-adjusted volume together with expenditure data will produce indirect PPPs across economies, which allow the RCAs to check the plausibility of the collected information.

Validation Table Analysis

BOX **16.3**

Example of Validation of Volume Data across Economies, ICP 2011

	Economy A	Economy B	Economy C	Economy D	Economy E
Number of dwellings/population	20%	9%	36%	28%	23%
Number of occupants/population	102%	87%	86%	105%	98%

In this example, the ratio of the number of dwellings to population for most of the economies in this region is 20–30 percent. However, economy B has a lower ratio, 9 percent, which is less than half that of other economies. Therefore, it should be flagged for further investigation because of the possibility of a data error. It also should be checked with other data from the economy to identify any possible data issues. If there is an acceptable reason for the difference, such as culturally economy B has a larger family size than other economies in the same region, the data should not be removed or edited.

Just as in the household consumption survey validation process, analyses using tables such as the Quaranta and Dikhanov make it possible to conduct detailed comparisons. Analytical tables provide indexes for more objective validation such as the coefficient of variation and PPP-ratio. Thus the validation steps described in chapter 15 for the household consumption survey should be taken. Temporal analyses using data from the previous round of the ICP and indexes such as price level indexes and PPPs, if feasible, also help the RCAs examine the trend of the housing market and validate it against other rental market studies and knowledge.

Once validation at the intereconomy level is complete, it is recommended that, for validation purposes, regional PPPs be computed for the basic heading actual and imputed rentals for housing. By estimating regional PPPs, the RCAs can derive real housing volumes by dividing expenditures by PPPs. In doing so, it is possible to check whether the rental data provided and the estimated volume measures are plausible. After the RCAs have applied their geographical knowledge, any significant outlier should be confirmed with the NCAs.

Finalization of Data

After confirmation that the rental data, volume data, and metadata have been intereconomy validated, the RCA should submit the data to the Development Data Group at the World Bank for the global-level validation process across regions.

Global Validation

Validation at the global level is the same as the regional validation but on a larger scale. Similar to the process for the validation of household consumption survey data, the active involvement of the NCAs and RCAs is crucial to this final validation stage. The Development Data Group carries out validation in the same manner as the RCAs to ensure consistency as well as comparability across regions and economies.

Annex A

Rental Survey Questionnaire, ICP 2011

Economy name		Global specifications				
Dwelling code	Dwelling type	Size (m^2)	Approximate size (sq. ft.)	Reference size (m^2)	Approx. reference size (sq. ft.)	Electricity
1104111.011	Villa/single-family house	120–180	1,300–1,950	150	1,600	Yes
1104111.021	Villa/single-family house	120–180	1,300–1,950	150	1,600	Yes
1104111.031	Villa/single-family house	120–180	1,300–1,950	150	1,600	Yes
1104111.041	Villa/single-family house	120–180	1,300–1,950	150	1,600	Yes
1104111.051	Villa/single-family house	180–240	1,950–2,600	210	2,300	Yes
1104111.061	Villa/single-family house	180–240	1,950–2,600	210	2,300	Yes
1104111.071	Villa/single-family house	180–240	1,950–2,600	210	2,300	Yes
1104111.081	Villa/single-family house	180–240	1,950–2,600	210	2,300	Yes
1104111.091	Villa/single-family house	240–360	2,600–3,900	300	3,300	Yes
1104111.101	Villa/single-family house	240–360	2,600–3,900	300	3,300	Yes
1104111.111	Villa/single-family house	240–360	2,600–3,900	300	3,300	Yes
1104111.121	Villa/single-family house	240–360	2,600–3900	300	3,300	Yes
1104111.131	Villa/single-family house	360–460	3,900–5,000	400	4,300	Yes
1104111.141	Villa/single-family house	360–460	3,900–5,000	400	4,300	Yes
1104111.151	Villa/single-family house	360–460	3,900–5,000	400	4,300	Yes
1104111.161	Villa/single-family house	360–460	3,900–5,000	400	4,300	Yes
1104111.171	Attached house/row house	80–120	850–1,300	100	1,000	Yes
1104111.181	Attached house/row house	80–120	850–1,300	100	1,000	Yes
1104111.191	Attached house/row house	80–120	850–1,300	100	1,000	Yes
1104111.201	Attached house/row house	80–120	850–1,300	100	1,000	Yes
1104111.211	Attached house/row house	120–180	1,300–1,950	150	1,600	Yes
1104111.221	Attached house/row house	120–180	1,300–1,950	150	1,600	Yes
1104111.231	Attached house/row house	120–180	1,300–1,950	150	1,600	Yes
1104111.241	Attached house/row house	120–180	1,300–1,950	150	1,600	Yes
1104111.251	Attached house/row house	180–240	1,950–2,600	210	2,200	Yes
1104111.261	Attached house/row house	180–240	1,950–2,600	210	2,200	Yes
1104111.271	Attached house/row house	180–240	1,950–2,600	210	2,200	Yes
1104111.281	Attached house/row house	180–240	1,950–2,600	210	2200	Yes
1104111.291	Studio apartment	15–35	160–380	25	270	Yes
1104111.301	Studio apartment	15–35	160–380	25	270	Yes
1104111.311	Studio apartment	15–35	160–380	25	270	Yes
1104111.321	Studio apartment	15–35	160–380	25	270	Yes
1104111.331	Studio apartment	35–60	380–650	45	480	Yes
1104111.341	Studio apartment	35–60	380–650	45	480	Yes
1104111.351	Studio apartment	35–60	380–650	45	480	Yes
1104111.361	Studio apartment	35–60	380–650	45	480	Yes

					Observations		
Inside water	Private toilet	Private kitchen	Air-conditioning or central heating	Structure age	Yearly rent (LCU)	Location (urban/rural)	Comments
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	< 5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			

table continues next page

Annex A *(Continued)*

Economy name		Global specifications				
Dwelling code	Dwelling type	Size (m^2)	Approximate size (sq. ft.)	Reference size (m^2)	Approx. reference size (sq. ft.)	Electricity
1104111.371	One-bedroom apartment	40–60	430–650	50	540	Yes
1104111.381	One-bedroom apartment	40–60	430–650	50	540	Yes
1104111.391	One-bedroom apartment	40–60	430–650	50	540	Yes
1104111.401	One-bedroom apartment	40–60	430–650	50	540	Yes
1104111.411	One-bedroom apartment	60–80	650–850	70	750	Yes
1104111.421	One-bedroom apartment	60–80	650–850	70	750	Yes
1104111.431	One-bedroom apartment	60–80	650–850	70	750	Yes
1104111.441	One-bedroom apartment	60–80	650–850	70	750	Yes
1104111.451	Two-bedroom apartment	60–80	540–850	70	750	Yes
1104111.461	Two-bedroom apartment	60–80	540–850	70	750	Yes
1104111.471	Two-bedroom apartment	60–80	540–850	70	750	Yes
1104111.481	Two-bedroom apartment	60–80	540–850	70	750	Yes
1104111.491	Two-bedroom apartment	80–120	850–1,300	100	1,000	Yes
1104111.501	Two-bedroom apartment	80–120	850–1,300	100	1,000	Yes
1104111.511	Two-bedroom apartment	80–120	850–1,300	100	1,000	Yes
1104111.521	Two-bedroom apartment	80–120	850–1,300	100	1000	Yes
1104111.531	Typical/traditional dwelling	25–75	270–800	50	540	Yes
1104111.541	Typical/traditional dwelling	25–75	270–800	50	540	Yes
1104111.551	Typical/traditional dwelling	25–75	270–800	50	540	Yes
1104111.561	Typical/traditional dwelling	25–75	270–800	50	540	Yes
1104111.571	Typical/traditional dwelling	25–75	270–800	50	540	No
1104111.581	Typical/traditional dwelling	25–75	270–800	50	540	No
1104111.591	Typical/traditional dwelling	80–120	850–1,300	100	1,000	Yes
1104111.601	Typical/traditional dwelling	80–120	850–1,300	100	1,000	Yes
1104111.611	Typical/traditional dwelling	80–120	850–1,300	100	1,000	Yes
1104111.621	Typical/traditional dwelling	80–120	850–1,300	100	1,000	Yes
1104111.631	Typical/traditional dwelling	80–120	850–1,300	100	1,000	No
1104111.641	Typical/traditional dwelling	80–120	850–1,300	100	1,000	No

Source: ICP, http://icp.worldbank.org/.

					Observations		
Inside water	Private toilet	Private kitchen	Air-conditioning or central heating	Structure age	Yearly rent (LCU)	Location (urban/rural)	Comments
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	Yes	<5 years			
Yes	Yes	Yes	Yes	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	No	No	No	<5 years			
Yes	No	No	No	>5 years			
No	No	No	No	<5 years			
No	No	No	No	>5 years			
Yes	Yes	Yes	No	<5 years			
Yes	Yes	Yes	No	>5 years			
Yes	No	No	No	<5 years			
Yes	No	No	No	>5 years			
No	No	No	No	<5 years			
No	No	No	No	>5 years			

Annex B
Dwelling Services Questionnaire, ICP 2011

FORM A. ICP DWELLING SERVICES QUESTIONNAIRE: VOLUME OF HOUSING IN 2011

Country:

Year (If data are not available for 2011):

	All dwellings			Type of construction				Location of dwellings			
	Houses	Flats	Total	Modern Construction		Traditional Construction	Total	Urban areas		Rural	Total
				Houses	Flats			Large urban	Small urban		
Number of dwellings units ('000s)											
Number of rooms ('000s)											
Usable surface area in thousand square meters (Specify other unit ____)											
Number of occupants ('000s)											
Land area occupied by dwellings in thousand square meters (Specify other unit ____)											
Number of dwellings units with :											
Electricity ('000s)											
Inside water ('000s)											
Private toilet ('000s)											
Central heating											
Air-conditioning											
Percent of dwellings units:											
Rented											
Owner occupied											

Source: ICP, http://icp.worldbank.org/.

NOTE

1. Because of the infeasibility of implementation, the NCAs are not asked to determine the volume of each type of dwelling. Thus the precise average rental prices for each type are not available.