1. Purpose of the statistical system of "Real Estate Prices", the publication objective

Since 1997 the Ministry of Finance of the CR and the Czech Statistical Office have been cooperating on the formation of a system for the monitoring of real estate prices in the Czech Republic.

The purpose of the system being is to provide information on the distribution of the price level by real estate type, location, and by other decisive factors as well as development in this distribution in time. The system purpose is not to replace so-called price maps, which local authorities form for their purposes yet to provide global information at the macroeconomic level.

This publication also does not strive to present such system in its full. Besides the providing of information on the real estate prices dependence on various determining factors, it also presents reasons for the selection of such factors in the given data source. One of the objectives is to find the maximum information content of the current data status. The Publication follows to another one of "Prices of Types of Real Estate Monitored in 2007 – 2009".

2. Data sources, frequency

The statement for stamp duty land tax (SDLT) is the data source here. The CZSO shall receive these data encrypted once per month. The data have been collected since February 1999. They are delayed by 6 months, on average, after the date of sale or submission of the SDLT form. The database of price information was established at Internal Revenue Offices by the processing of data from submitted SDLT forms in 1998 (according to the Decree No. 279/1997 Sb.).

The main advantage of this administrative source is it is based on real, actually paid prices (admitted in SDLT returns). It is a nationwide and regular data flow on prices of transactions on the real estate market.

Internal Revenue Offices enter into the database selected data on sales of real estate from SDLT forms and the files created are then supplied to the CZSO. These are types of real estate as follows: buildings and halls, family houses, recreational cottages and houses, recreational chalets and garden huts, garages, wells, flats and non-dwelling areas, construction plots, agricultural land, forests land, other plots and land, and forest stands.

3. Variables monitored and derived, homogeneity

Each (complete) real estate sold is usually formed of partial real estates of various types (e.g. buildings and construction plots). Location is reflected in the appraisal value as calculated by the valuation surveyor according to the decree on property valuation. The appraisal value of the whole, complete real estate is a sum of appraisal values of partial real estates. The purchase price of a complete real estate is the recorded price, which the real estate was sold for. The purchase price of a partial real estate is subsequently derived proportionally to the portion of the appraisal value of the partial real estate of the total appraisal value. The unit price (appraisal value or purchase price) is the appropriate price for one measuring unit of the corresponding type of real estate (for 1 m^2 or 1 m^3, etc.).

The basic task is to determine the average unit price of a certain type of real estate in the area or band limited by parameters set (for instance, region, extent of wear, etc). For this purpose all sales of complete real estates in this area or band, which contain a partial real estate of the surveyed type, are searched. Rules for the suitable sampling of real estates in the area or band surveyed require homogeneity is defined as a portion (percentage) of the value of a partial real estate of the surveyed type in the whole sale of the given complete real estate.

The selection of real estates for the determination of the average purchase unit price of a certain type of real estate shall be carried out in three steps. First, only those complete real estates are selected, for which purchase price and appraisal value are not extremely different. In the second step, only those complete real estates are taken out of this set, for which homogeneity of the type surveyed is higher than a fixed limit for this type of real estate. In the third step, improbable extremes, mostly generated by errors in data entering, are eradicated from the remaining data. Then the arithmetic average of the remaining purchase unit prices of the type of real estate surveyed in the given area is taken as the average purchase unit price of the given type of real estate in the area.

4. Explaining and explained variables, basic principles and their determination

The priority explained variable in tables, which consists the subject matter of this publication, is the average purchase unit price. Then as a derivative, the price index, which is the ratio of two such average prices for two different time periods. The subset of data, which are available in the aforementioned data sources, shall be selected as explaining variables. The determining factors shall be clearly territory and time aspect as a priority user's request for the real estate price statistics. Territorial breakdown inside the publication corresponds to the state as at 1 January 2011.

Further explaining determining factors must meet the condition that there is a statistically provable dependence of the explained variable on these factors and, simultaneously, the factors should be mutually independent, as much as possible.

The extent of wear (expressed as percentage - data from the assessor) was selected from explaining factors available. The dependence of the average sales unit price on the extent of wear is very conclusive. Other explaining factor applied, which is virtually independent on all of the previously mentioned, is the size of municipality (population from the Territory Analytical Materials for 2008) where the real estate surveyed is located.

5. Selection of types of real estates

For this publication merely those real estates available in the source database were selected, for which there are enough number of price data classified by various combinations of explaining factors determined above. Moreover, the aforementioned limiting effect of homogeneity for certain types of real estate shall be taken into account. For instance, various types of land/plots form mostly minority portions of prices in the sale of the complete real estate. Their prices are then more or less determined by other type of real estate, which they are sold with, as a family house or a building, for example. The applicable number of price data for such plots is then substantially reduced.

Therefore solely those types of real estate, where a relatively sufficient number of price data having homogeneity higher than the required limit, were selected for this publication as follows:

Order	Type of real estate	Band of the extent of wear	Minimum value of homogeneity
1.	RD – single-dwelling family houses,	⟨0; 10⟩	60%
		(10; 50⟩	60%
		(50; 75⟩	70%
		(75; 100⟩	75%
2.	BY – dwellings (flats)	⟨0; 100⟩	70%
3.	DO – buildings – multi-dwelling buildings	⟨0; 100⟩	40%
4.	SP – construction plots	X	10%

6. Determination of bands of the extent of wear

Bands of the extent of wear had to be determined in order to compile the real estate classification by the extent of wear. Besides the breakdown into groups by 10% each, which is, however, too fine if combined with a breakdown by other factors and thus cause the population is too disintegrated, specific wider bands were determined for each type of real estate. The determination was based on the trend of average unit price of the given real estate type depending on its extent of wear. The groups were established the way that differences of unit prices inside are relatively low, differences among the groups are high, and none of the groups is too large or too small. This process resulted in the following bands of the extent of wear (as percentage) used in majority of analyses as well in tables published:

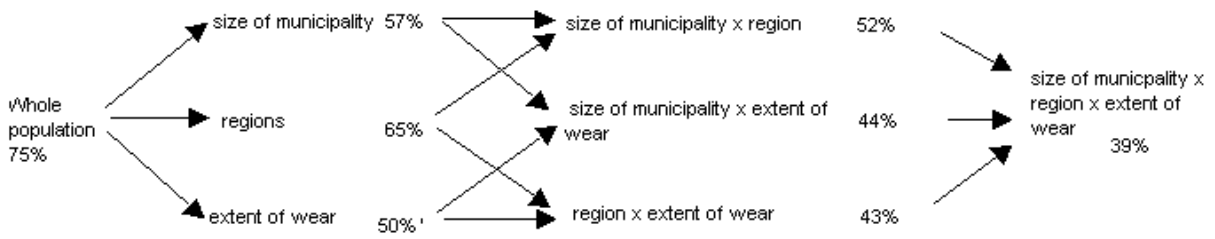
Type of real estate	Band of the extent of wear			
	1	2	3	4
RD – family houses	⟨0; 10⟩	(10; 50⟩	(50; 75⟩	(75; 100⟩
BY – flats	⟨0; 5⟩	(5; 20⟩	(20; 45⟩	(45; 100⟩
DO – multi-dwelling buildings	⟨0; 25⟩	(25; 65⟩	(65; 100⟩	

7. Analysis by means of coefficient of variation of groups by determining factor

Now, knowing the respective factors determining the unit price what remains is to determine their optimal combination for the publishing tables. Subdividing the population by each factor much reduces the number of prices in each of the groups (cross-sections) formed this way.

Therefore, effects of all the factors at the same time cannot be investigated to be able to get representative results. Moreover, the four explaining variables established by us cannot by far explain the whole variability of the unit price. The unit price results from many other factors unknown to us. If a sufficient number of price data remains in each group (cross-section), formed by subdividing of the population by the established factors, we may hope, that these effects, unknown to us, will be partly eliminated.

Our optimising procedure consists in how fast the average coefficient of variation (the ratio of standard deviation and average) of unit prices of the group decreases following to each subsequent subdividing of the population by a further (newly added) explaining factor. The procedure is best depicted in the scheme of the analysis carried out for family houses (data for 2007-2009).



The scheme above shows, for instance, that the most effective way is to subdivide the population into four groups by the extent of wear, because this reduces the relative variance of prices from 75% to 50% and the generated groups have approximately the same number of price quotations. The subdividing procedure must be halted in a certain moment when the number of price data is not representative in a certain portion of the groups. The aforementioned analysis was carried out for all selected types of real estate, while explaining factors (as the region and district for region, for example) were applied in different alternatives. This resulted in an optimum decision, which of the tables could be published.

8. Time dependence –price index

Besides the determination of the dependence of the average unit prices on various explaining factors, it is also important to determine the price development in time in the form of the price index. This index should be of the Laspeyres type with fixed base weights in order to be compatible with other price statistics. And the most important is it should be adjusted for quality changes over time that means it should be a pure price index.

Because of the character of the price data acquired, the shortest time unit is a quarter of year. Then there is a specific issue occurring for construction plots, where there is enough data, yet their variability is by far the highest of all four types of real estate and the surveying of the plot unit price in each quarter of the year keeping the same quality is of extreme difficulty.

In order to meet the aforementioned requirement for fixed quality of the type of real estate surveyed over time a qualitative adjustment shall be applied. Because the index is based on the regional breakdown, that means observes the price development for all possible cross-combinations of region x size of municipality, it is necessary to carry out a qualitative adjustment for differences in other known determining factors (extent of wear, size of real

estate) from one quarter of the year to another. In other words the combination region x size of municipality is fixed, what is changed is the period (in the quarter of year interval) and the extent of wear and size of real estate must be kept constant.

The method applied for family houses and flats uses regression analysis. The regression model assumes the price of the real estate surveyed depends on the period, size category of municipality, extent of wear, and the size of real estate. The whole surveyed set of the real estate surveyed (as family houses, aggregated data for 2007-2009, for instance) is subdivided into subgroups defined by a cross-combination of the size of municipality and the extent of wear group. Then the aforementioned regression model is determined in every subgroup formed this way. It will provide a statistical dependence of the price on the extent of wear and the "size" of the family houses in the subgroup. First, the average ("standard") level of the extent of wear is determined in each of the subgroups. Then every price of the given real estate in this subgroup can be "converted" to appropriate "standard" in this subgroup (i.e. one-floor house without a cellar and of average extent of wear), that means to carry out the required quantitative adjustment prior the subsequent price comparison.

The model applied for family houses assumes a linear dependence of the logarithm of the unit price on the period, size category of municipality (variables have discrete values) and the extent of wear (variables have continuous values). For flats the logarithm of the unit price depends on the period and the fact, if the respective flat is in a bricked or panel multi-dwelling building (variables have discrete values), and on its size (variables have continuous values). A quadratic form, or any other non-linear form of explaining variables was excluded for statistical reasons. The necessary calculations were performed for multi-dwelling buildings yet it has often come out that a better way to use prices is to keep them not adjusted by the aforementioned method, because regression models are not always statistically provable and adjusted prices demonstrate often even higher variability than the non-adjusted ones.

A special approach is used for construction plots because analogical explaining variables are not available. The main factor (and in fact the only one known) is the location. Therefore the whole set of construction plots was subdivided into respective cadastral districts. Based on the data for 2008-2011 a "reference price" was determined for each cadastral district. Then the ratio of these prices for two different cadastral districts determines the "qualitative relation" of these two cadastral districts. The qualitative adjustment can be then carried out on the basis of this relation. This method is relatively effective; variability of "adjusted prices" has been substantially reduced.

The year 2005 (whole-year period) was chosen as the price and index base period to construct the index. In 2008 real estate price indices were reviewed and new, more realistic weightings were created for all types of real estate indices. Weights for construction plot index and multi-dwelling buildings index are determined on the basis of relative ratios of sums of absolute appraisal values of real estates falling into the given categories for 2004 – 2006 (in order to get more robust weights). Weights for the overall index have been determined the same way. Weights for flats and family houses are solved as a compromise – a combination of outputs from sums of absolute appraisal values and data from the Census on the numbers of family houses and flats (these data are not available for other types of real estate).

9. Determination of the set of publication tables

The analyses resulted in the determination of the extent of published tables. One of the aims of this publication is also to establish limits of the information contained within the data set investigated, that is to determine the real estate unit price by means of certain chosen factors. The maximum possible population is represented by the set of family houses, which is relatively large, relations are provable, and is relatively homogeneous in terms of price. For other types of real estate solely specific subsets could be selected because the aforementioned characteristics are absent. Construction plots are a special case because the factor of the extent of wear is missing here, which is usually the most significant one for other types of real estate. Therefore there are partially special modifications of otherwise standard types of tables given for construction plots.

Table "Overview of Publishing Tables" provides the complete overview. In its left section one can find the table type marked by number, in the middle section there are determining factors given in tables, and the right section gives page number, on which one can find corresponding table for every type of real estate, if there is any.

10. Information content of tables, interpretation

Each table includes the average purchase unit price of the given type of real estate for the breakdown by established factor. These are always simple arithmetic averages. The number of transfers, which the price was obtained from, is given as information on reliability and representativeness, whenever possible. In case there are less than three data, the price is not given yet the numbers are. Furthermore, the coefficient of variation of the set, which the average price was calculated for, is given as a measure of data accuracy and level of homogeneity of the given group, whenever possible. In some cases quantiles are given as well. These are important namely for construction plots of high price variability. Average values of factors, which in the given table are not applied as determining ones (as the extent of wear, for instance), or possibly other quantities (as the size of real estate given in appropriate measuring units, for example) are given in some cases, too. In every case these are simple arithmetic averages of the group.

11. Estimates of price indexes

As a result of efforts to make the publication more up-to-date we also provide the estimates of price indexes of flats and family houses for 1st and 2nd half years of 2011, which are to be published on the CZSO website every half year (expected dates of publishing are 16 January 2012 and 16 July 2012, respectively).