Description of Price Indexes for New Single-Family Houses Under Construction

Introduction

The houses under construction indexes are a group of price indexes designed for use in deriving a constant dollar series from the current dollar series of single-family value put in place. The indexes are formed with data for houses built for sale, contractor-built houses, owner-built houses, and houses built for rent. All indexes apply to the construction cost of new houses that are under construction as defined for the single-family value put in place estimates. Construction cost excludes the value of land and other nonconstruction costs. See Value of Construction Put in Place methodology documentation for more information.

The under construction indexes are only computed at the national level.

The data used for computing these indexes are obtained from the U.S. Census Bureau's Survey of Construction. The survey collects information on the physical characteristics and prices of new single-family houses. This is done through monthly interviews with the builders or owners of a national sample of new houses.

The methodology used to compute the value put in place of new single-family houses influenced the formulation of the indexes for houses under construction. The value put in place of new single-family houses is computed from data collected in the Survey of Construction and evaluation studies performed in the past and is an indirect estimate of construction value.

Types of indexes

Laspeyres and Paasche type price indexes are computed for single-family houses under construction. From these a Fisher Ideal index is derived. The indexes are published as Laspeyres Constant Quality Index for New Single-Family Houses Under Construction and the Fisher Price Deflator Index for New Single-Family Houses Under Construction. We do not publish the Paasche type index. The following sections describe the various types of indexes and their computation.

Price index design – Laspeyres type indexes

The basic form of a Laspeyres type price index is:

$$\frac{\sum_i (p_{ti} * q_{0i})}{\sum_i (p_{0i} * q_{0i})}$$

Where the p_{0i}'s and p_{ti}'s are the prices in the base and current period, respectively, and the q_{0i}'s are the quantities in the base period. This represents the ratio of the current cost of the quantity of goods purchased in the base year to the cost in base year prices of the same quantity of goods. Notice that the denominator is the construction cost of the average base period house. To compute this index the *p's* must be derived from a regression model since we only collect a total house and land price.

Experience has shown that regression estimation of the price in the following multiplicative model is superior to estimation for the above additive model:

$$\frac{e^{\sum_i (d_{ti} * q_{0i})}}{e^{\sum_i (d_{0i} * q_{0i})}}$$

Where the d_{0i}'s and d_{ti}'s are price factors and the q_{oi}'s are the quantities in the base period.

It is necessary to obtain the *d's* and *q's* for various commodities to compute this index. We estimate from survey data the *q*'s (quantities) for commodities that we refer to as house characteristics. The *d's* – referred to as price factors – for the characteristics cannot be collected so we estimate them from a regression model. For this reason the sums in the above equation can be thought of as regressed values taken from a regression model. The regression models used to estimate the price factors are described in more detail below.

Forming strata also improves the regression estimates. We use five strata: four defined as the detached units in the four Census Regions and the fifth as all attached units. The Laspeyres price index at the national level is then:

$$\frac{e^{\sum_h w_h * \sum_i (d_{ti} * q_{0i})}}{e^{\sum_h w_h \sum_i (d_{0i} * q_{0i})}}$$

Where the first sum in each term is over the five strata and w_h is the proportionality factor for houses under construction in the stratum and the other terms are computed by stratum.

Paasche Type Price Index Computation

The basic form of a Paasche type price index is:

$$\frac{\sum_i (p_{t\,i} * q_{ti})}{\sum_i (p_{0i} * q_{ti})}$$

Where the p_{0i}'s and p_{ti}'s are the prices in the base and current period, respectively, and the q_{ti}'s are the quantities in the current period. This represents the ratio of the current cost of the quantity of goods purchased in the current period to the cost in base year prices of the same quantity of goods. It differs from a Laspeyres type index in that quantities in the Laspeyres index are base period quantities and in the Paasche index they are current period quantities. Notice that the numerator of the Paasche type index is the construction cost of the average current period house.

As with the Laspeyres index described in the previous section, the Paasche additive model shown above is replaced with a multiplicative model that is the weighted average of five strata. We estimate from survey data the quantities and price factors for commodities referred to as house characteristics.

Fisher Type Price Index Computation

A Fisher Ideal index is formed by computing the square root of the product of compatible Laspeyres and Paasche type indexes.

Regression Models

The Survey of Construction collects the price including land and various characteristics for a sample of new houses under construction. Factors are applied to the price of houses sold and to the contract price of contractor-built houses to derive the construction cost.

For the purposes of the under construction indexes, the definitions of under construction and construction cost conform to the definition used in the single-family value put in place estimates. A house is under construction in the month it is started and in the eleven following months regardless of its month of completion. A house is considered started in the month excavation begins for the foundation.

To compute an index, price factors (*d's*) must be derived for a suitable set of available characteristics. Price factors are derived from the construction cost and characteristics data by weighted multiple linear regression. The weights used in the regression are the probability of selection of the houses in the Survey of Construction adjusted for the following considerations:

1. Houses built for sale are weighted to also represent houses built for rent
2. Contractor-built houses are weighted to also represent owner-built houses
3. Houses are weighted by the phasing pattern factor used in the calculation of the single-family value put in place estimate so they impact the index in proportion to their contribution to value put in place.

A more thorough description of the computation of the single-family value put in place estimates appears with the Value of Construction Put in Place methodology documentation.

Available characteristics of the sample of houses under construction are evaluated in order to determine a suitable set to use for index calculations. Since consumer preferences change over time, any set of characteristics is useful for a limited number of years. At irregular intervals, one set of characteristics is replaced by a new set. When a new set of characteristics is introduced into the index computations, the index for prior periods may be recomputed or the index may be re-based. See the section of documentation titled History of Single-Family Price Index Methodology for a complete discussion of re-basing procedures.

Presently there are five separate regression models used to calculate the price indexes: one model for detached units in each census region and one model for all attached units. Each of these models is designed to measure the contributions of important physical and geographic characteristics to the prices of new houses.

The characteristics used in each model except for floor area are divided into categories. For example, each house is classified by whether it has less than three bedrooms, three bedrooms, or more than three bedrooms; whether it has no garage or a carport, a one or two car garage, or a garage for three or more cars; etc. Each category is treated qualitatively in that a value of "1" indicates that the house has that characteristic and "0" indicates that the house does not have it. One category from each of the qualitative characteristics must be omitted to avoid an over determined system. The price and floor area are treated quantitatively – the logarithms of the actual values are used directly in the model.

The same set of characteristics is used for both the Laspeyres type indexes and the Paasche type indexes.

Tables B-1 (detached houses) and B-2 (attached houses) show by stratum the characteristics used in the regression models and the categories associated with them. These tables also show the base year average quantities (*q's*) used in the calculation of the Laspeyres type indexes. Quantities for the Paasche type indexes change each time period and are not shown.

A resistant regression procedure is used that diminishes the impact of unusual observations by down-weighting the data for those observations.

Since the regression does not include all of the characteristics that explain price variability and because the characteristics are interdependent, the estimated regression coefficients should not be regarded as estimates of the true price factors. The estimated regression coefficients are not shown.

Stratum Weights

The base year for the indexes is currently 2005. The weights (levels of activity in 2005) used to combine the five stratum indexes to form the U.S. Census Bureau Houses Under Construction Fixed-Weighted Index are shown in Table B-3.

Weights used to combine the strata indexes for the Paasche type index change each time period and are not shown.

Annual Indexes

Due to the fact that houses under construction may not be sold and have sales price data, the under construction index is subject to continual revision. In order to reduce the number of revisions and avoid problems of consistency between the monthly and annual indexes, the number of revisions to the data is now restricted. Restricting the number of revisions to monthly data greatly complicates the accumulation of data for an annual index. As a result, annual indexes are no longer computed from an annual set of data; they are derived from the monthly indexes.

The annual indexes shown in 1999 and later years are the geometric means of the monthly indexes for those years. For years prior to 1999, the annual indexes were computed from an annual set of data and were subject to repeated revision.