Details of the Regression Models and Model Characteristics

The one-family price indexes are formed from log-log multiple linear regression models.

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 logarithm of the actual values are used directly in the model.

Tables A-1 (detached houses) and A-2 (attached houses) show by stratum the characteristics used in the regression models and the categories associated with them for the Constant Quality (Laspeyres) Price Indexes of New One-Family Houses Sold Including Value of Lot. These tables also show the base year average quantities (*q's*) used in the calculation of the 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 for the Constant Quality (Laspeyres) Price Index of New One-Family Houses Under Construction. These tables also show the base year average quantities (*q's*) used in the calculation of the indexes. The same characteristics apply to the Price Deflator (Fisher) Index of New One-Family Houses Under Construction.

A resistant regression procedure is used which incorporates Tukey's biweight. Resistant regression significantly reduces the influence on the model of houses with unusual characteristics, price, or location by reducing the sample weight of each such case. In this way a case with an extreme value resulting from a legitimate, unusual situation or incorrect reporting or processing has a reduced impact upon the index.

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. These characteristics account for from 60 to 80 percent of the variation in the logarithm of the prices.