Definitions, Limitations, and Miscellaneous Statistical Methodology

Definitions and Limitations of the Data

A house is defined as sold when a sales contract is signed or a deposit is accepted regardless of the stage of construction. The month of sale refers to the contract or deposit date. For the purposes of the under construction indexes and for the one family value put in place estimate, a house is under construction in the month it is started and in the eleven following months. A house is considered started in the month excavation begins for the foundation.

The reported data for each house in the sample for the Survey of Construction are edited before use in the index computation. If the price or any characteristic is not reported, that sample case is rejected. 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.

Our price index is computed from actual transaction prices of new houses. Not included are any amenities the builder provides to the buyer that are not included in the initial price. For example, the price of a two-car garage added to a sales contract may not be included in the initial price.

The prices we estimate for our index computations may be influenced by characteristics of workmanship, materials, and mechanical equipment, which are not measured. Hence, it should be kept in mind that the price indexes in this report only account for such quality characteristics insofar as they may be correlated with the characteristics actually used. These characteristics account for from 60 to 80 percent of the variation in the logarithm of the prices.

Since the price index applies to the total sales price or the total sales price less nonconstruction costs, it covers cost of labor, materials, direct and indirect selling expenses, and the seller's profits and it may cover land costs. The index is thus conceptually broader in coverage than a cost index. Reflecting the price, the price index is affected by all factors which influence movement of house prices: both supply factors such as wage rates, material costs and productivity, and demand factors such as demographic changes, income, and availability of mortgage money.

Sampling Error

Sampling error reflects the fact that only a particular sample was surveyed rather than the entire population. The price index in a given period is calculated from a particular sample of new houses. If a separate index number were calculated from each of all possible

samples of identical size that could have been selected, using the particular procedure for calculating the index that is used for single-family houses, each of these numbers would differ from one another. The standard error, or sampling error, of a survey estimate is a measure of the variation among the estimates from all possible samples and, thus, is a measure of the precision with which an estimate from a particular sample approximates the average from all possible samples. The relative standard error equals the standard error divided by the estimated value to which it refers.

For the Constant Quality (Laspeyres) Price Indexes of New One-Family Houses Sold Including Value of Lot, the relative standard error of the annual index for the United States is 0.5 percent. The relative standard errors for the quarterly index as well as for the Midwest, South, and West regions' annual indexes are about 1.0 percent. The Northeast annual index has a relative standard error of about 2.0 percent.

For the Constant Quality (Laspeyres) and Price Deflator (Fisher) Price Indexes of New One-Family Houses Under Construction the relative standard errors are 0.5 percent for the monthly index and 0.4 percent for the annual index.

The sample estimate and an estimate of its relative standard error allow us to construct interval estimates with prescribed confidence that the interval includes the average result of all possible samples with the same size and design. A 90% confidence interval is defined to be from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate. If all possible samples were selected and surveyed under essentially the same conditions and all the respective 90% confidence intervals were generated, then approximately one-tenth would not include this average estimate.

For example, the 1992 annual price index for the Constant Quality (Laspeyres) Price Indexes of New One-Family Houses Sold was 56.4. Multiplying 56.4 by the relative standard error of 0.3%, we obtain 0.5 as the standard error. To obtain a 90% confidence interval, multiply 0.3 by 1.6, yielding limits of 55.5 and 56.9 (56.4 plus or minus 0.5). The average estimate of this annual price index may or may not be contained in this computed interval; but in 9 out of 10 samples, the interval calculated in this manner will contain the average estimate from all possible samples.

Nonsampling Error

As calculated for this report, the estimated relative standard error measures certain nonsampling errors, but does not measure any systematic biases in the data. Bias is the difference, averaged over all possible samples with the same size and design, between the estimates and the true value being estimated. Nonsampling errors for the Survey of Construction can be attributed to many sources: inability to obtain information about all cases in the sample, definitional difficulties, differences in interpretation of questions, inability or unwillingness of respondents to provide correct information, and errors made in processing the data. Nonsampling errors for the price index can result from excluding important characteristics like the quality of building materials from the regression, high correlation among regression characteristics, and use of an improper regression model.

These nonsampling errors also occur in complete censuses. We believe that most of the important response and operational errors are controlled in the course of reviewing the data for reasonableness and consistency. The regression model was chosen to minimize the amount of nonsampling error associated with the price index.