

**Notes on the Philadelphia Fed's  
Real-Time Data Set for Macroeconomists (RTDSM)**

*New Privately Owned Housing Units Started*

## **I. General Comments**

This file provides documentation for the Philadelphia Fed's real-time data set of the total number of new privately owned housing units started, sometimes abbreviated to "housing starts." The data set consists of multiple worksheets, each containing the real-time monthly observations that would have been available to someone at the point of time (the *vintage* date) given in the column headers of the worksheets. These headers follow the nomenclature given by HSTARTSyyMm, where HSTARTS denotes total housing starts, yy is a two-digit number representing the vintage year, M denotes the word "month," and m represents the vintage month. For example, the observations available in June 1989 are those given in the column HSTARTS89M6. The last observation in this column is that for May 1989, since that was the last observation listed in the *Current Construction Reports: Housing Starts (C20)* report of May 1989, which is released in June. The time series observations within a vintage are labeled as yyyy:mm, where yyyy is a four-digit number representing the year of the observation, and mm is a two-digit number representing the month of the observation. For example, the observation for May 1989 is labeled 1989:05. All data are monthly, seasonally adjusted, and expressed in thousands of housing units started.

This is one of several variables in the RTDSM for which we have collected monthly vintages. Data vintages from February 1968 to December 1974 are in the file **HStarts1.xls**; vintages from January 1975 to December 1989 are in the file **HStarts2.xls**; vintages from January 1990 to October 1998 are in the file **HStarts3.xls**; and, vintages from November 1998 to present are in the file **HStarts4.xls**. All the files are Excel worksheets.

## **II. Methodology**

Our methodology for collecting real-time observations on housing starts is identical to that described for the other variables in RTDSM. We begin, with the vintage of April 1970, by locating a hard copy deep-history report containing all the time series observations that would have been available to someone in April 1970. Subsequent vintages are then added from high-frequency reports, which contain a much more limited span of observations (usually the last 12 to 24 months). As we move from one vintage to the next, two things happen. First, we obtain an additional month of data. This is the initial release of housing starts for that month. Second, we incorporate any revisions to observations common to both vintages. Our high-frequency source is *Current Construction Reports: Housing Starts (C20)*, produced by the United States Bureau of the Census. We continue in this fashion until the Census Bureau releases a benchmark revision to the data. Since such revisions generally affect more observations than are reported in *Current Construction Reports* (usually the last three years, in recent benchmark revisions), we must find

another deep-history report. When we are unable to locate a deep-history report listing all observations, our vintages contain only those observations listed in our high-frequency source.<sup>1</sup>

We obtain deep-history reports from a variety of sources (described in a table below), and not all reports begin with the same observation date. Thus, because we are careful to include only those observations that we are sure were available in real time, some vintages have different starting observation dates. The following table shows the first observation available in each vintage:

*Table 1. First Observation, By Vintage*

<b>Vintages</b>	<b>First Observation Date</b>
February 1968 to March 1970	Only the last 12 to 13 monthly observations are included.
April 1970 to January 1972	1947:01
February 1972 to January 1973	1969:01
February 1973 to March 2000	1947:01
April 2000 to present	1959:01

As mentioned previously, we obtain deep-history reports from a variety of sources. The following table lists the vintage dates reflecting benchmark revisions, the range of observations affected by the revisions, and the source of our deep-history report.

---

<sup>1</sup> For this reason, our vintages covering the period February 1968 to March 1970 include only the last year's worth of monthly observations.

Table 2. Benchmark Revisions

<b>Vintage of Benchmark Revision</b>	<b>Starting Point of Revised Observations</b>	<b>Deep-History Source</b>
April 1970	Startup	<i>Business Conditions Digest, June 1970</i>
February 1972	1969:01	<i>Business Conditions Digest, June 1971</i>
February 1973	1970:01	<i>Business Conditions Digest, July 1973</i>
February 1974	1971:01	<i>Business Conditions Digest, March 1974</i>
February 1975	1972:01	<i>Business Conditions Digest, April 1975</i>
February 1976	1973:01	<i>Business Conditions Digest, July 1976</i>
February 1977	1974:07	<i>Business Conditions Digest, April 1977</i>
February 1978	1975:01	<i>Business Conditions Digest, June 1978</i>
February 1979	1976:01	<i>Business Conditions Digest, May 1979</i>
February 1980	1977:01	<i>Business Conditions Digest, March 1980</i>
February 1981	1978:01	<i>Business Conditions Digest, March 1981</i>
February 1982	1979:01	<i>Business Conditions Digest, March 1982</i>
February 1983	1980:01	<i>Business Conditions Digest, June 1983</i>
February 1984	1981:01	<i>Business Conditions Digest, April 1984</i>
February 1985	1982:01	<i>Business Conditions Digest, March 1985</i>
February 1986	1983:01	<i>Business Conditions Digest, June 1986</i>
February 1987	1984:01	<i>Business Conditions Digest, March 1987</i>
February 1988	1985:01	<i>Business Conditions Digest, May 1988</i>

Table 2 (continued). Benchmark Revisions

<b>Vintage of Benchmark Revision</b>	<b>Starting Point of Revised Observations</b>	<b>Deep-History Source</b>
February 1989	1986:01	<i>Business Conditions Digest, April 1989</i>
February 1990	1987:01	<i>Current Construction Reports: Housing Starts, C20-9001</i>
February 1991	1988:01	<i>Current Construction Reports: Housing Starts, C20-9101</i>
February 1992	1989:01	<i>Current Construction Reports: Housing Starts, C20/92-1</i>
February 1993	1990:01	<i>Current Construction Reports: Housing Starts, C20/93-1</i>
February 1994	1991:01	<i>Current Construction Reports: Housing Starts, C20/94-1</i>
February 1995	1992:01	<i>Current Construction Reports: Housing Starts, C20/95-1</i>
February 1996	1993:01	<i>Current Construction Reports: Housing Starts, C20/96-1 &amp; Philadelphia Fed Internal Records</i>
February 1997	1994:01	<i>Current Construction Reports: Housing Starts, C20/97-1 &amp; Philadelphia Fed Internal Records</i>
February 1998	1995:01	<i>Current Construction Reports: Housing Starts, C20/98-1</i>
February 1999	1996:01	<i>Collected in real time by the Philadelphia Fed.</i>
February 2000	1997:01	<i>Collected in real time by the Philadelphia Fed.</i>
May 2001	1998:01	<i>Collected in real time by the Philadelphia Fed.</i>
May 2002	1999:01	<i>Collected in real time by the Philadelphia Fed.</i>
May 2003	2000:01	<i>Collected in real time by the Philadelphia Fed.</i>
May 2004	2001:01	<i>Collected in real time by the Philadelphia Fed.</i>
May 2005	2002:01	<i>Collected in real time by the Philadelphia Fed.</i>

There are a few methodological points to consider in understanding precisely how we incorporate a benchmark revision into a new vintage. First, the Census Bureau's C20 report always provides a brief discussion of the revision process, including the range of observations affected by the revision. When possible, we do not rely on this discussion when deciding on the range of observations to carry over from the preceding vintage. In many cases, our deep-history reports (listed in Table 2) contain observations dated prior to the date that the C20 says is the first date affected by a benchmark revision. In these cases, we use the full array of observations from the new deep-history report, not just the range that the C20 asserts is the range affected by the benchmark revision.

Second, in some cases our deep-history report was published on a date after the vintage date, suggesting the possibility that some of the observations in that report may not have been available on the vintage date. (Such a possibility could occur if the Census Bureau released the benchmark-revised observations over time, rather than all at once—which sometimes occurs for other variables, for example, when the BEA releases a benchmark revision to the NIPA.) For the housing data, we treat this as a publication lag, not a lag in the availability of the revised observations. Thus, in the interest of maintaining as deep a history as possible in each vintage, we decided on the following policy: Unless we detected evidence in the C20's discussion of the benchmark revision that the revised values would be released in a staggered form, we assumed all revised values were released at the same time. We detected no evidence to the contrary in any benchmark revision.

Third, there is one potential problem with the preceding policy: When the deep-history report is published a number of months after the date of the benchmark revision, the tail-end observations in the report can reflect normal month-to-month revisions that would not have been known on the date of the benchmark revision. We do not incorporate these observations in the vintage of the benchmark revision. Rather, we take the last 12 observations from those listed in the edition of *Current Construction Reports: Housing Starts* in which the benchmark revision occurred. For example, in constructing the vintage of February 1984 (see table 2), we used the *Business Conditions Digest* of April 1984 for observations from 1981:M1 to 1982:M12 and the C20 of January 1984 for observations from 1983:M1 to 1984:01. (Observations prior to 1981:M1 were taken from the vintage of January 1984 and double-checked against the deep history provided in the *Business Conditions Digest*.)

### **III. Exact Release Dates**

We are unable to pinpoint the exact days on which the Census Bureau released its housing reports. However, in recent years, the Census Bureau has been releasing its reports during the third week of each month. At present, we do not know if the Census Bureau followed this pattern of release dates prior to the 1990s.

#### IV. Relationship between Monthly Employment Vintages & RTDSM Quarterly Vintages

In early June 1999, the Philadelphia Fed released its real-time data set, consisting of quarterly vintages of quarterly observations of NIPA variables [nominal and real GNP (GDP) and associated output-price indices and some broad components of real output, such as components of consumption expenditures and investment, real exports and imports, and government purchases] and quarterly vintages of monthly observations of non-NIPA variables [M1, M2, reserves, the unemployment rate, and the CPI]. These quarterly vintages contain the data available as of the 15<sup>th</sup> of the middle month of the quarter

If we assume the Census Bureau's housing report was always released during or before the third week of the month, the following table shows how to merge the quarterly vintages of the RTDSM data set with the monthly vintages of the data set on new privately owned housing units started—that is, how to merge the two to form a consistent quarterly information set.

*Table 3. Merging Quarterly Vintages from the RTDSM Real-Time Data Set with the Monthly Vintages of New Privately Owned Housing Units Started*

<b>Quarterly RTDSM Vintages</b>	<b>Monthly Vintages of Housing Starts (column header)</b>
yyyy:Q1	HSTARTSyyM2
yyyy:Q2	HSTARTSyyM5
yyyy:Q3	HSTARTSyyM8
yyyy:Q4	HSTARTSyyM11

#### V. Our Methodology for Incorporating Corrections to Errors

Occasionally, the Census Bureau's *Current Construction Reports* may contain errors in the reported data. In general, we do not know whether such errors are due to computational or typographical mistakes. When the Census Bureau discovers such errors, these are reported, along with the corrections, in a later edition. Our policy on incorporating this new information is as follows: When we discover a C20 that contains corrections to previously published data, we incorporate such corrections into that vintage. Subsequent vintages reflect the correction as well. On the premise that the corrections would not have been known at the time of previous vintages, *we do not adjust the observations in previous vintages.*

## **VI. Quality of the Data**

In our judgment, these data on housing starts are of high quality. We believe each vintage accurately represents the exact data that would have been available to someone at the vintage date. We have taken steps to minimize our own data-entry errors. In particular, all vintages have been checked at least twice for errors. Benchmark vintages have been checked three times. Some subtle errors possibly remain, and users should examine the data carefully for outliers that we may have overlooked.

Questions about the data should be addressed to:

Tom Stark  
Macroeconomic Database and Policy Support Manager  
Research Department  
Federal Reserve Bank of Philadelphia  
Ten Independence Mall  
Philadelphia, PA 19106-1574  
E-mail: [tom.stark@phil.frb.org](mailto:tom.stark@phil.frb.org)

## HSTARTSyyMm— New Privately Owned Housing Units Started

1. Thousands of private housing units started, seasonally adjusted, monthly.
2. First Vintage: February 1968
3. First Observation: Varies by vintage (see Table 1)
4. High-Frequency Source: Census Bureau's *Current Construction Report: Housing Starts (C20)*, and sometimes *Economic Indicators* (Council of Economic Advisers)
5. Deep-History Source: Varies by benchmark revision (see Table 2)
6. Vintage Date Format: HSTARTSyyMm  
yy = year of vintage; m = month of vintage, 1, 2,...,10,11,12
7. Observation Date Format: yyyy:mm  
yyyy = year of observation; mm = month of observation

### *Special Notes*

1. **HSTARTS68M2 to HSTARTS70M3—Short Time-Series History.** Because we were unable to locate deep-history reports over this vintage span, these vintages include only the last 12 to 13 monthly observations that were listed in the Census Bureau's *Current Construction Report: Housing Starts (C20)*.
2. **HSTARTS95M11—Federal Government Shutdown Delays Release of Data.** Due to the shutdown of the federal government, the release for this vintage occurred later in the month than usual.
3. **HSTARTS95M12—Federal Government Shutdown Postpones Release of Data.** Due to the shutdown of the federal government, the Census department did not release its report for this vintage. Therefore, the observations in this vintage are taken from the vintage for November.
4. **HSTARTS96M1—Federal Government Shutdown Delays Release of Observation for 1995:M12.** Due to the shutdown of the federal government, the last observation in this vintage is that for 1995:M11.