

# General Notes on the Philadelphia Fed's Real-Time Data Set for Macroeconomists (RTDSM) – Selected Non-NIPA Variables

Last Updated: December 22, 2016

## I. Overview

This document contains general notes for a subset of the variables in the Philadelphia Fed's real-time data set. The selected variables, listed in Table 1 below, are the non-NIPA variables that we formerly included in the group called the *core variables*. In the recent reorganization of our web page for real-time data, we have dropped the distinction between core variables and noncore variables.

Table 1. Selected Non-NIPA Variables from the Philadelphia Fed's Real-Time Data Set

<b>Monetary and Financial</b>	<b>Price Level Indices</b>
M1 Money Stock	Consumer Price Index
M2 Money Stock	
Total Reserves	
Nonborrowed Reserves	<b>Labor Market</b>
Nonborrowed Reserves + Extended Credit	Unemployment Rate
Monetary Base	

A *real-time data set* shows the observations for a variable as those observations were revised over time. The Philadelphia Fed's real-time data set records snapshots, or vintages, of the data as they existed at various points in time in the past, before the data were fully revised. The *vintage date* is an important concept in a real-time data set: It refers to the date on which the data were available to the public. Our data sets, organized as Excel workbooks, provide all vintages for a particular variable in one Excel workbook.

In general, a real-time data set is organized with observation dates in the rows and vintage dates in the columns. This characterizes the construction of the Philadelphia Fed's data. When the data are organized this way, each column shows the entire time-series history of the variable that would have been available to someone at the vintage date shown in the column header. Thus, an analyst can easily track the revisions to an observation by moving horizontally across the columns. Indeed, as you move from one column to the next, two features of the data change. First, the new column lists any additional observations released by the government statistical agency. Second, the new column records any revisions to the previous observations. All of the non-NIPA variables listed in Table 1 have monthly observations, and the vintages are

collected quarterly. In particular, the observations shown in each vintage represent the data that were available in the middle of the quarter. We tried to time the vintages as close as possible to the 15<sup>th</sup> of the middle month of each quarter, and we do not incorporate the data of any release published after the 15<sup>th</sup>. For example, the data given in the 2000:M2 vintage would represent the data available as of mid-February 2000. Similarly, the vintage 2000:M5 represents data available as of mid-May 2000. Since moving from quarter to quarter introduces three new months of data, each column typically has three more observations than the previous column. In some cases, however, the timing of a data release may occur after the 15th, and thus some vintages may include only two additional observations.

## II. File Structure and Variable Names

Our real-time data files are stored as Excel worksheets, and, in general, there is one file for each variable. Each column shows all time-series observations available at a particular vintage date (which is listed in the column header). The file name, which also serves as the root name for the column headers, describes the real-time variable. For example, the file for the consumer price index is called `cpiQvMd.xlsx`. Each column header in the file (except the first, which gives the date of the time-series observation) follows the nomenclature given by `CPIyyQq`, where `yy` is a two-digit number indicating the vintage year, `Q` denotes Quarter, and `q` is a number indicating the vintage quarter. For example, the first column to contain a vintage of data for the consumer price index has the header `CPI94Q3`, indicating that this column contains the monthly observations available to researchers around August 15, 1994.

Table 2 shows the variable names used for the variables included in this documentation. We use these mnemonics to name our data files and as column headers in each file.

Table 2. Selected Non-NIPA Variables and Associated Naming Conventions in the Philadelphia Fed's Real-Time Data Set

Non-NIPA Variable	NAME
<b>Price Level Indices</b>	
Consumer Price Index	CPI
<b>Labor Market</b>	
Unemployment Rate	RUC
<b>Monetary and Financial</b>	
M1 Money Stock	M1
M2 Money Stock	M2
Total Reserves	TRBASA
Nonborrowed Reserves	NBRBASA
Nonborrowed Reserves + Extended Credit	NBRECBASA
Monetary Base	BASEBASA

### III. Our Methodology of Data Collection

Our real-time data set is organized around the following principle: *Each vintage should include the exact values of the observations that would have been known at the vintage date.* We construct our vintages using the observations that are listed in the publications of U.S. government statistical agencies, such as the Bureau of Labor Statistics and the Federal Reserve Board. The dates on which the agencies' reports were published correspond to the dates of our vintages. Thus, we have one report for each vintage. We begin the process of vintage collection by locating a "start-up" report containing the entire time series history that would have been available to someone at the first vintage date in our data set. (The first vintage date varies from variable to variable.)

To be more precise, our methodology is as follows. Because we are interested in real-time data, we are unable to use the historical data as they now exist. Such data reflect revisions and perhaps variable definitions that would not have been known at each of our vintage dates. Instead, for each variable, we search for a hard-copy source that contains a deep historical time series. We refer to such a source as a *deep-history report*. For most of the variables in our data sets, we were able to find at least one deep-history report per year, which we used to construct a

benchmark vintage. Deep-history reports are usually published shortly after the data are subject to a benchmark revision. In a benchmark revision, the entire historical time series is subject to change. This contrasts with the normal month-to-month revisions, in which only the most recent observations are changed. Between deep-history reports, and thus between benchmark vintages, we use *high-frequency reports* — such as the Board of Governors’ weekly H.6, for M1 and M2, and weekly H.3, for reserves data — to update and extend the most recent observations of the previous vintage, in order to construct the next vintage in the quarterly sequence. This procedure delivers a quarterly sequence of real-time vintages, each of which contains only the data that would have been available at the vintage date<sup>1</sup>.

To illustrate the procedure, consider our methodology for collecting M1 and M2. In the mid-1970s, the Board of Governors began to issue an internal document called Money Stock Revisions. This document contains an entire time series for all official measures of the money stock in use at the time. The publication was usually released early in the year, and we used it to construct our benchmark vintages. Nonbenchmark vintages were constructed from the benchmark by adding additional observations to the benchmark, as reported in the Board’s weekly H.6 release, and renaming the vintage.

In some cases, we were unable to locate a deep-history report for a particular variable and thus did not include observations for that variable in the vintage. We adopted a very conservative approach in this project: If there were any questions about the real-time availability of the data, we chose not to use such data in constructing our data sets. This conservative approach comes at a cost because, in some cases, it produces gaps of missing data. Sometimes these gaps persist across consecutive vintages. We leave to you the task of eliminating the gaps in the manner most appropriate for your research project. We have, however, attempted to document when the gaps occur, on a variable-by-variable basis. For details, see the documentation for each variable.

#### **IV. Discussion of the Last Observation in Each Vintage**

It is natural to think that the last monthly observation in each vintage should be that for the month before the month of the vintage date. For example, one might expect that the last monthly observation in the column labeled “65Q4” (a vintage date of November 1965) would be that for October 1965. Indeed, that is the case for the unemployment rate. But that is not the case for other variables. The key distinction is whether the last monthly observation is available (that is, released by the government agency that compiles the data) before or after the 15<sup>th</sup> of the middle month of each quarter. In the 1990s, for example, we know that the unemployment report was usually released on the first Friday of each month and that it contained the observation for the preceding month. (We collected all vintages for unemployment under the assumption that the same was true prior to the 1990s.) Likewise, data on reserves for the month prior to the month of the vintage date are generally available on or before the 15<sup>th</sup>. Such is not the case for the monetary aggregates (M1 and M2): For these variables, the observation for the month before

---

<sup>1</sup> The only exception to this methodology is our real-time data on the CPI. For this variable, all our vintages were collected in real-time in electronic files. Thus, for this variable, we did not need to locate hard-copy statistical reports.

the month of the vintage date is sometimes available and sometimes not. Mostly, the last observation for the CPI is not available by the 15<sup>th</sup>.

To take a recent example, consider the CPI report of February 2002, which contained a CPI observation for January. That report was released by the BLS on February 20, too late to be incorporated in the column CPI02Q1. However, the CPI report of May 2002, containing an observation for April, was released on May 15, just meeting our cut-off date. Thus, the column CPI02Q2 contains an observation for April (and possible revisions to the observations prior to April).

## V. Quality of the Data

In our judgment, the data in RTDSM are of high quality. We believe that each vintage accurately represents the exact data that would have been available at the vintage date. We have also taken steps to minimize our own data-entry errors. Undoubtedly, some errors remain, and users should examine the data carefully for outliers we may have overlooked.

Questions about the data should be addressed to:

*Tom Stark*  
*Research Department*  
*Federal Reserve Bank of Philadelphia*  
*Ten Independence Mall*  
*Philadelphia, PA 19106-1574*  
*Tel: (215) 574-6436*  
*E-mail: Tom.Stark@phil.frb.org*