(https://wrds-web.wharton.upenn.edu/wrds/)

⌂ (/wrds/index.cfm) / Support (/wrds/support/index.cfm) / Data (/wrds/support/Data/index.cfm)
  / Using TAQ Data Efficiently (/wrds/support/Data/_005Using TAQ Data Efficiently/index.cfm)
  / Using TAQ Data Efficiently

Using TAQ Data Efficiently

The Trade and Quote (TAQ) database contains consolidated intraday transactions data (trades and quotes) for all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), Nasdaq National Market System (NMS) and SmallCap issues, as well as stocks traded on Arca (formerly Pacific Stock Exchange) and other regional exchanges.

I. Introduction

Since 1993, TAQ files (Consolidated Trade and Quote data) have been stored at WRDS in monthly SAS datasets (file names follow this format CTYYMM and CQYYMM, respectively). However, in 2007 and afterwards, the Quote data size has increased substantially to levels close to 1 TerraByte per month. In October 2008, the Quote dataset, cq0810.sas7bdat, occupies 840GB and its Index set, cq0810.sas7bndx, is 360GB.

Extracting and processing such large monthly TAQ datasets became very inefficient and caused system slowdown for all WRDS users, not just TAQ users. In part, the problem was generated by the size of the TAQ monthly files and by inefficient programming techniques in processing those datasets.

Top

II. TAQ Daily File Structure

As a response, from January 2009 forward, TAQ files on WRDS will not be stored in monthly blocks of data, and instead will be available only in daily datasets, as they are provided by NYSE. All trading day records are stored in a separate dataset. This natural daily structure, which separates data along trading day units, is expected to speed up TAQ processing, make maintenance easy and efficient, and most importantly enable efficient use of TAQ data by saving users' time and WRDS I/O and computing resources. The new data structure will make it more intuitive to program loops on a daily basis, and is expected to limit the impact of inefficient coding practices by reducing the size of the underlying datasets. The new daily datasets are also index efficient, as they have the trading date in the name of the dataset, which does away with the need for a simple or composite DATE Index. Please note that the date variable will remain in the daily datasets for our users' convenience when appending or analyzing multiple daily datasets. The daily Trade and Quote datasets are sorted by SYMBOL and TIME, and indexed only by SYMBOL.

From 2009 onward, only daily TAQ datasets will be generated on WRDS. Researchers that use the web queries will not see any change as the updated queries seamlessly extract information using any date range. The change will be visible only for interactive users who access TAQ data directly on the

WRDS server. We will also begin producing a backfill of daily files for previous years, starting in 2008 and going back in time. If you have any suggestions, feedback, or comments please let us know at wrds-support@wharton.upenn.edu (mailto:wrds-support@wharton.upenn.edu), with "TAQ Daily Structure" in the subject line.

III. Efficient Programming Techniques

We also provide a set of new TAQ research applications that deal with efficient SAS programming techniques applied to specific TAQ research issues. The first one is a SAS procedure that consolidates screening, merging, and calculation of statistics from the trades and quotes data into a single step. In the SAS community, such a Data Step is known as a "DOW Loop". The second application uses SAS Views. By utilizing SAS Data Views, one no longer needs to create, write, or transfer large temporary SAS files making the process more time and space efficient. <

In the next sections, we will provide an overview of the TAQ data, and document TAQ variables. Then, we will present WRDS guidelines for efficient programming using TAQ data. Sample SAS macros for data retrieval and processing (merging and analysis) will be explained in the following section. We conclude with reference to additional support materials and research applications on TAQ data on WRDS.

IV. WRDS Resources on Using TAQ Efficiently

We recommend new TAQ users to start with the WRDS Overview of TAQ data, which is a compilation of various documentations, manuals, and detailed variable definitions that we received from NYSE TAQ.

If you are an interactive user who analyzes TAQ data using PC SAS/Connect or Unix SAS, we recommend that you continue going through the remaining parts of the overview, mainly the "TAQ SAS Programming Issues" document. We would like to encourage our interactive users to pay close attention to Rule 2 in this document, which explains how users can employ SAS Views and speed up substantially the processing of their large TAQ jobs. The following Lee and Ready sample program provides an additional example on how to efficiently use SAS Views in implementing the Lee and Ready (1991) algorithm to infer trade directions from TAQ intraday data. The NBBO sample program also employs SAS Views in computing the National Best Bid and Offer (NBBO) for NYSE's Trade and Quote (TAQ) data.

An important documentation for our interactive SAS users is the research application on DOW Loop Approach. DOW Loop is another SAS programming concept that is very efficient in screening, merging, and analyzing large datasets in just one step. The purpose of this research application is to walk the reader through the DOW Loop implementation, so that WRDS users can understand how exactly the loop works and can make appropriate modifications to the program to suit their research needs.

- WRDS Overview of TAQ Data (_001WRDS Overview of TAQ Data.cfm)
    - TAQ SAS Programming Issues (_027TAQ SAS Programming Issues.pdf.cfm)
    - Location of TAQ Data on WRDS (_005Location of TAQ Data on WRDS.cfm)
    - TAQ Dataset List (/wrds/tools/variable.cfm?library_id=56)

- A Note on the Use of PROC SQL with TAQ data on WRDS (https://wrds-web.wharton.upenn.edu/wrds/research/applications/microstructure/note_proc_sql_taq/index.cfm)
- SAS DOW Loop Approach (https://wrds-web.wharton.upenn.edu/wrds/research/applications/microstructure/sas%20dow%20loop/)
- Lee-Ready using SAS Data Views (https://wrds-web.wharton.upenn.edu/wrds/research/applications/intraday/index.cfm)
  - Note on Matching Trades and Quotes in Presences of Multiple Quotes (https://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_026Note%20on%20Matching%20Trades%20and%20Quotes%20in%20Presences%20of%20Multiple%20Quotes.cfm)
  - TAQ Literature (_030TAQ Literature.cfm)
- NBBO Derivation using SAS Data Views (_028NBBO.cfm)
- TAQ FAQs (https://wrds-web.wharton.upenn.edu/wrds/support/Data/_005Using%20TAQ%20Data%20Efficiently/_035TAQ%20FAQs.cfm)

Please let us know your thoughts or if you have any questions at wrds-support@wharton.upenn.edu (mailto:support@wharton.upenn.edu?subject=WRDS Support).

Top

Table of Contents

(http://www.wharton.upenn.edu)

About WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/index.cfm)
WRDS FAQs (https://wrds-web.wharton.upenn.edu/wrds/about/WRDS FAQs.cfm)
WRDS News (https://wrds-web.wharton.upenn.edu/wrds/news/index.cfm)

3 Ways to use WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/three_ways_to_use_WRDS.cfm)
Account Types on WRDS (https://wrds-web.wharton.upenn.edu/wrds/about/Account Types.cfm)
Terms of Use (https://wrds-web.wharton.upenn.edu/wrds/about/terms.cfm)

Account Preferences (https://wrds-web.wharton.upenn.edu/wrds/mywrds/preferences.cfm)
Info / Support Request (https://wrds-web.wharton.upenn.edu/wrds/about/external_support_request.cfm)
Privacy Policy (https://wrds-web.wharton.upenn.edu/wrds/about/privacy.cfm)

WRDS Demo (https://wrds-web.wharton.upenn.edu/wrds/demo/)
Conference Calendar (https://wrds-web.wharton.upenn.edu/wrds/about/Conference Calendar.cfm)
Best Paper Awards (http://www.whartonwrds.com/best-paper-award-winners/)

Wharton Research Data Services