Content

1. How to plug into Datastream
2. Finding what you are looking for
3. Download data for the stocks currently contained in a stock market index
4. Downloading data for a large number of stocks
5. Further important hints for working with Datastream
1. How to plug into Datastream

You can use Datastream through Excel via the Thomson Reuters Datastream add-in. Follow the instructions given in the Trading Room if you do not see the add-in in Excel.

In the following, we give examples on how to work with Datastream in Excel.

2. Finding what you are looking for

Requests conducted by students are typically *Time Series Requests*. From the *Datastream* ribbon in Excel, select *Time Series Request*.

A new window opens:

```
You need to determine the series (e.g. the DAX, BASF, German unemployment rate) and the type of data (e.g. price, depreciation, net income). For your request, you also need to specify the interval and the frequency. Series can be identified through their specific Datastream symbol code, the mnemonic. However, there are often alternative ways of identifying a series. For example, you can identify most German stocks by putting a “D” before the WKN (security identification number).
```
2.1. First example

You need the daily stock price of BASF AG over the years 2000 to 2007. In the Time Series Request window, click Find Series. The Datastream Navigator opens: By choosing the appropriate data category (e.g. Equities) you can narrow your search down by typing “basf” in the field to the right of the Explore button, then press Search. Results are shown immediately.

The result list contains many entries for BASF because BASF shares are traded on several exchanges and Datastream keeps individual records of different data sources. Just “BASF” leads you to the data from Deutsche Börse, the major German exchange. Click on the DS Mnemonic “D: BAS”, and the mnemonic will be pasted into the request window of Excel.

Search for the appropriate type of data using the Datatypes button. For example “UP” gives you the unadjusted price, the price that was actually quoted at the respective day in the past. Select a Start Date and an End Date as well as your Frequency. Have a look at the screenshot below:
Press *Submit* and the data will be written into the sheet. The upper left corner of the output range is the cell that was selected in the sheet before you opened the request window.
2.2 Second example

You need annual data on total assets of BASF starting in 2010. By clicking on the “Datatypes” button, you can search for the appropriate type of data, here WC02999 for “Total Assets”.

Select time period and frequency, press Submit and you obtain the following results:

The data is missing in 2019 and 2020 because the query was conducted in February 2020 when 2019 and 2020 results had not yet been reported.

2.3 Third example

You are looking for data on unemployment in Germany. In the Datastream Navigator, select the data category Economics and use the Advance Search button. You may want to type “unemployment” in the Name field, and “Germany” in the Market field.

Press Search Now and you will get a long list of different data in unemployment (e.g. differing in the data provider, or unemployment in various regions). Try to identify what you need. If the list is too long to go through, try to narrow down your search in the search criteria window of the Advanced Search in the Datastream Navigator.
3. Download data for the stocks currently contained in a stock market index

Select the *Time Series Request* window. Enter the Code for the index in *Series/Lists* field, e.g. "LDAXINDEX" for the DAX 30. Select the *Datatype* that you want, e.g. "RI" for total return index and select dates and frequencies as before. Tick "TS for each item in list". Press *Submit*.

The result should look similar to this:

![Time Series Request window](image)

How do I find out the index list code to make such a download? This is done with the *Datastream Navigator*. In the *Time Series Request* window, press the *Find Series* button. Datastream Navigator opens.
Select the data category \textit{Constituent lists} on the left side of Datastream Navigator and just type dax in the Search field (screenshot above) or click on \textit{Advanced Search} and search for \textit{dax} in \textit{All Fields} or in \textit{Name and Long Name} field (screenshot below).
4. Downloading data for a large number of stocks

Sometimes you need to compile a dataset with all stocks that are listed at a specific stock exchange as well as those that were listed at the exchange at some point in the past.

In this example, it is shown how to download all German stocks from the Xetra Exchange.

- Go directly to the Datastream Navigator.
- Firstly, choose a single category: Equities.
- Next search option is exchange – select Xetra.
- Then you can select market – select Germany.
- Under Type choose Equity to limit your search to common stocks. If you also want to include preferred stocks, simply click on Multiple and the new window opens. Then you see all possible options and you can select as many types as you want.
- If you do not choose between active and dead stocks under Activity, both of them will be selected.

Refining search on the left side of the Datastream Navigator automatically updates the results on the right side of the Datastream Navigator. You can see a long list with names, symbols and other information.

Next, you can see how to export your data into Excel.
You have now two possibilities to proceed. You press All (left blue arrow) and Datastream pastes all results to the Time Series Request window. Or you proceed with the Excel symbol (right blue arrow). The former method is straight forward but you are fixed to the results of this single search. Moreover, the number of characters that are allowed to be put in the Series/List field of Time Series Request window is limited to 8192 characters (you would not be able to obtain data for circa 2000 stocks at once). The latter method allows the combination of different search results and allows exporting much more data at once: search results with up to 9000 lines can be exported at once.

Click on the Excel symbol and you can download the data. Close the navigator and request window and open the downloaded file.

Select the column with Symbols and copy it (the highlighted column in the above example). Now open a new Excel file. From the Thomson Reuters Datastream ribbon choose New Request Table.

In the Series Lookup column, paste your Symbols or refer to Symbols when they are stored outside the request table.
Now we discuss the filling of the remaining fields. Let us start with the left part. “Yes” in the **Update** column is obvious. “TS” is for the time series request because we want to download a time series for each code. The **Select Format** tells Datastream whether to display series names, dates, etc. and whether the data belonging to one code should be written in one row or one column. Say we would like to arrange the data such that the data for one stock are in one row. Then choose a “T” for transpose because by default a series is displayed in a column. The row title “R” in this case is the name of the series – we would like to display it for each series, which is why there is an “R” everywhere. The column titles “C” are the dates. We need these dates only once because they will be the same for each series if we choose the same start and end date for each series. So we add the “C” only for the first series. Also note that you get a selection of possible choices by clicking on the grey buttons.

Now to the right part. Choose your type of data, e.g. “RI” for total return index. Choose **Start Date** and **End Date**. Choose **Freq** (Frequency). Now the trickier part: Choose the **Data Destination**. I have inserted a new worksheet in the meantime, which I named “D”.

The data for the first series starts in D1A1; it spreads over two rows because it includes the dates. For the other series, we need just one row, which is why we then have D1A3, D1A4 and so on. Note that it is very easy to fill this down. Assume that you have already filled down the frequency cells. Now select two data destination cells which differ by one row, e.g. the ones with D1A3 and D1A4. Double click on the lower right corner and the cells will be filled down in the way you want it to be.

Then press **Process Table** and the data will be downloaded.

**A few tips for working with such data:**

- You will see a lot of cases where you do not get any data. This is not surprising with such a download. Some series will not have data in the period chosen by you; some may not have data for the type of data chosen by you, etc.

- If a stock stops trading on an exchange, there are no more prices and you may expect that Datastream has no values for the return series “RI”. This is not the case. Datastream rather keeps the value from the last trading day in its database and displays it as the current one. So you need to find a way of dealing with such cases. In other words: if the values for one series do not change from some time on until the end, you should discard those repeated values or set them to missing (e.g. to “NA”). Write a macro to do this job for you. (see also 5. c)).

Do not set these data points (or missing data in general) to zero. In some cases, this may be equivalent to discarding them, but in other cases it may introduce big errors.
5. Further important hints for working with Datastream

a) **Date conventions**

In finance and many other fields, the convention is to work with end-of-period values for monthly and yearly intervals. When working with monthly data, for example, you should therefore use month-end values. This can be accomplished by entering month-end dates into the date fields, as was done in the examples above.

With weekly data, it is more customary to select a date from the middle of the week. When combining different data sources, make sure that they are not based on a different convention.

b) **Historical index constituents**

If you use the standard constituent lists to download data for an index, you get data for the current constituents. For many empirical analyses, however, you need the historical index constituents as well (e.g. if you want to test how a strategy that selects 10 stocks out of the 30 DAX stocks would have performed in the years 2009-2019).

For some indices and dates, Datastream provides you with historical constituents. Codes are typically codes for the current list plus the historical date. If you do not get the necessary information from Datastream, you need to compile it from other sources.

c) **Identifying non-trading days**

Datastream usually exports data for exchange holidays even though there is no trade on such days. For some empirical applications, you need to highlight and purge the data of these days. There are several possible solutions.

- When you want to get for example total return index data for BASF, type “D:BAS” in the Series/List field and “X(RI)*IF#(X(P#S),NNA,ONE)” in the Datatype/Expression field of the Time Series Request window. You will get the total return index of BASF, but it will be negative when there is a non-trading day.
- Find a reliable source of trading days, e.g. https://www.xetra.com/xetra-en/newsroom/trading-calendar/trading-calendar-archive, or take the dates from research data files (e.g. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html )

d) **Which price/ index type to use**

For most purposes such as determining volatilities or investment performance, use the total return index (RI). UP (unadjusted price) is the price as it was historically obtained on the exchange. Use it to implement stock price restrictions (e.g. exclude all stocks with a stock price below one dollar). P (price) is the adjusted price, which takes stock splits and similar corporate actions into account, but does not adjust for dividends. It is Datastream’s default datatype but it should not be yours. If you use it instead of the total return index, make sure that you provide good reasons for doing so.

e) **Data quality**

A discussion of Datastream data quality is given in:


This literature suggests a number of screens to deal with quality problems, which you should also consider for your analysis.