Paper 183-31

Compass Rose in Stock Markets: The Power of SASECRSP LIBNAME

Taras Zlupko, CRSP®, University of Chicago, Chicago, IL

ABSTRACT

The SASECRSP interface engine provides access to CRSP and Compustat financial databases directly from SAS®. The engine presents the data as native SAS datasets which can be processed as such by SAS. This significantly facilitates any kind of analysis when using these databases. The SASECRSP LIBNAME statement enables the user to specify database of interest. It also includes some additional options which allow to place restrictions on the chosen database. The SASECRSP LIBNAME statement options are a powerful tool which can significantly increase processing efficiency. This is illustrated by accessing returns data to detect the compass rose pattern in financial markets.

INTRODUCTION

Financial databases available from the Center for Research in Security Prices (CRSP) and Compustat are some of the best known and broadly used research databases of US stock markets and company fundamental data respectively. CRSP was founded in 1960 by the Graduate School of Business at the University of Chicago and is the first to collect stock market data on electronic media with data going back to 1925. Availability of CRSP databases is widely regarded as a milestone which greatly contributed to development of the modern theory of finance as well as the practice of financial securities investment.

Compustat started collecting company fundamental data in the mid 1960th. By the end of 1990th the CRSP/Compustat Merged Database (CCM) became available from CRSP. The key element of the merged data is CRSPLink. This allows combining CRSP and Compustat databases by matching unique CRSP security identifiers (PERMNO®) with Compustat company identifiers (GVKEY®). Therefore it possible to conduct seamless time series historic queries using both databases which opens up new opportunities in financial economics research. The merged database provides access to over 22,000 companies and a combined set of over 1,000 raw and derived data items.[1]

CRSP stock, indices, the CRSP/Compustat Merged Databases aw well as the Compustat databases are provided in the proprietary CRSPAccess format which makes the data accessible by the SASECRSP engine. Once accessed by the engine, the data is presented as native SAS datasets and can be processed by SAS data step and procedures without any additional transformations. As reported at SUGI30, the SASECRSP interface engine now features multiple keys to access CRSP stock databases, as well as access to CRSP index and index group data and the CRSP/Compustat Merged database.

THE SASECRSP LIBNAME STATEMENT

In order to start processing data in CRSPAccess format using SAS it is necessary to set-up the library. This is done by using standard LIBNAME statement which specifies SASECRSP as the compatibility engine, physical location of the data and a set identifier.

For example, the following sets up the library ‘dstk’ containing daily CRSP stock data with physical location of the data in directory c:\crspdstk\:

```
LIBNAME dstk sasecrsp 'c:\crspdstk\' setid=10;
```

The value of SETID depends on CRSPAccess database accessed such as daily or monthly stock database, index or index groups or CRSP/Compustat Merged Database.

[1] In addition to US Stock and Indices Databases and CRSP/Compustat Merged Database, the following is available from CRSP in native SAS data format: CRSP Survivor-Bias-Free US Mutual Fund Database, CRSP US Treasury Database and CRSP/Ziman Real Estate Data Series

The options available when setting-up the library allow restricting the data accessible through SASECRSP by multiple keys to particular universe of securities, companies, indices in index groups. Restrictions can be based on either predetermined universe defined by PERMNO, PERMCO®, CUSIP®, TICKER, GVKEY, INDNO® or some other keys. SASECRSP LIBNAME options also allow to limit the stock data accessed based on security industry affiliation and all databases can be restricted by date range of interest.

Following is the example of SASECRSP LIBNAME restricting the data access to one security for five-year period from January 01, 2001 to December 12, 2005:

```
LIBNAME ibm sasecrsp 'c:\crspdstk\'
              setid=10
                     permno=12490
                            range='20010101-20051230';
```

Therefore, SASECRSP LIBNAME not only makes the CRSPAccess databases immediately available for any SAS process but also allows to place a number of restrictions on data being accessed. The latter is important in increasing SASECRSP performance efficiency. This can be illustrated by accessing returns data to detect the compass rose pattern in stock markets.

THE COMPASS ROSE

The compass rose in stock returns was first discussed by Crack and Ledoit (1996). They showed that if, after removing outliers, daily stock returns are plotted against themselves with a lag of one day, the chart appear to resemble a pattern of evenly spaced lines radiating from the origin which they term a 'compass rose'. Conditions of compass rose existence and possible implications of this pattern in different financial markets using different asset classes and different data frequency have been discussed in a number of papers since then (e.g., see Kramer and Runde (1997), Szpiro (1998), Wang and Wang (2002), Cai, Hudson, and Keasey (2003), Lee, Mathur, and Gleason(2005)).

Empirical tests for the presence of the compass rose pattern in securities markets can be conducted using daily CRSP US Stock Database. Enhanced version of this database which contains daily frequency of data going back to December 1925 became available from CRSP for the first time in January of 2006. This database is unparalleled resource of its kind and is used in this paper.

For illustrative purpose, we will assume that there is only one stock of interest -- International Business Machines Corp., ticker IBM. International Business Machines Corp. stock is the one trading for the entire 1925-2005 period. Daily returns of the IBM stock can be accessed on either full CRSP database or the database restricted to the IBM stock only using the options available in SASECRSP LIBNAME. We will do so both ways monitoring the time required to retrieve the data of interest.

I. The library is set up for entire database.

I.1. Set up the library for CRSP daily stock database.

```
LIBNAME dstk sasecrsp 'c:\crspdstk\'
              setid=10;
```

I.2. Extract daily returns and convert date from CRSPAccess Date to SAS Date format.

```
data ibmret_full;
       set dstk.retx;
              where permno=12490;
                     date=crspdcsd(caldt);format date date9.;
                            rename retx=return;
run;
```

Excerpt from the log file follows:

```
WARNING: Defaulting to selecting all PERMNOs in CRSP or CCM database.
NOTE: There were 21269 observations read from the data set DSTK.RETX.
             WHERE permno=12490;
NOTE: The data set WORK.IBMRET_FULL has 21269 observations and 4 variables.
NOTE: DATA statement used (Total process time):
         real time           4:45.25
         cpu time            3:34.54
```

II. The database is restricted to IBM stock only.

II.1 Set up the library for CRSP daily stock database restricted to IBM stock only.

```
LIBNAME ibm sasecrsp 'c:\crspdstk\'
            setid=10
            permno=12490;
```

II.2. Extract daily returns and convert date from CRSPAccess Date to SAS Date format as in I.2.

```
data ibmret_restr;
      set ibm.retx;
            where permno=12490;
                  date=crspdcsd(caldt);format date date9.;
                        rename retx=return;
run;
```

Excerpt from the log file follows:

```
WARNING: Defaulting to selecting all PERMNOs in CRSP or CCM database.
NOTE: There were 21269 observations read from the data set IBM.RETX.
            WHERE permno=12490;
NOTE: The data set WORK.IBMRET_RESTR has 21269 observations and 4 variables.
NOTE: DATA statement used (Total process time):
        real time           0.18 seconds
        cpu time            0.12 seconds
```

The datasets ibmret_full and ibmret_restr created above are identical. However, as the excerpts from the log file indicate, it took significantly less time to create ibmret_restr on the database restricted to one stock which was set up using the relevant SASECRSP LIBNAME option than ibmret_full.

Figure 1 shows the time series of IBM stock returns using the datasets created above.

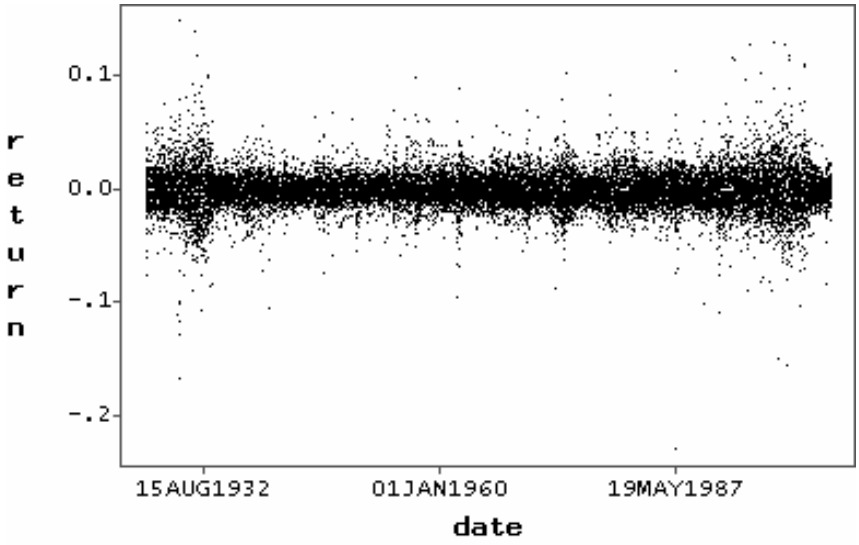

Figure 1. Daily returns of IBM stock from 1925 to 2005.

Having accessed returns on either full or restricted database, the remaining processing is the same. After filtering for outliers and creating lagged returns, the plot of daily returns of the IBM stock against their one period lags is presented in Figure 2 below.

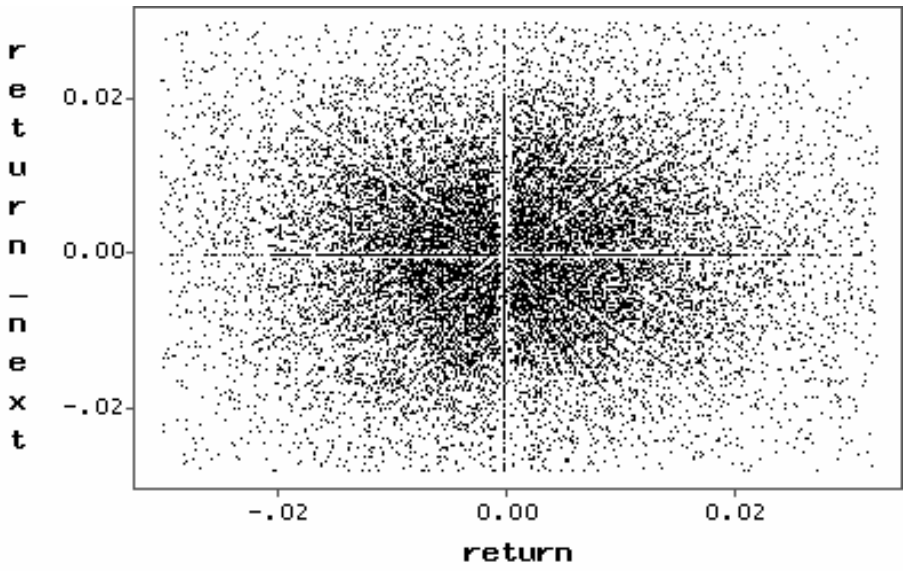

Figure 2. Phase portrait of daily returns of IBM stock, 1925-2005.

The horizontal axis in Figure 2 shows return on IBM stock on a given day. The vertical axis shows return on the next day. The data span the period from December 1925 to December 2005. The compass rose pattern is clearly visible in Figure 2.

A number of conditions have been discussed in the literature for the compass rose pattern to appear and some implications of the pattern suggested. Crack and Ledoit (1996) identify three necessary and sufficient conditions that the series of a stock price must satisfy in order for the compass rose pattern to exist as follows:

- Price changes are small relative to the price level
- Changes in price occur in small discreet jumps
- Price level varies over a relatively wide range

Findings by Kramer and Runde (1997) and Szpiro (1998) provide further support to the idea that discreteness is the determining factor in the appearance of the compass rose pattern in the phase portrait. Cai, Hudson, and Keasey (2003) examine the frequency of trading on the compass rose formation. Lee, Mathur, and Gleason (2005) find the tick/volatility ratio a determinant of the pattern.

In the light of the above, one may want to produce the phase portrait of stock returns to explore the evidence of compass rose formation during different time periods. In particular, it may be appropriate to conduct some analysis using the data from pre- and post- New York Stock Exchange decimalization. The decimalization was introduced on NYSE in 2000. Therefore, we will use two five year periods of IBM stock returns. One period ranging from 1991 to 1995 and another period ranging from 2001 to 2005. The assumption is that the data has to be accessed all anew.

III. The library is set up for entire database.

III.1. Set up the library for CRSP daily stock database.

```
LIBNAME dstk sasecrsp 'c:\crspdstk\'
            setid=10;
```

III.2. Extract daily returns and convert date from CRSPAccess Date to SAS Date format.
Restrict the data to the 1991-1995 period.

```
data ibmret_full;
      set dstk.retx;
            where permno=12490;
                  date=crspdcsd(caldt);format date date9.;
                        if '01Jan1991'd<=date<='31Dec1995'd;
run;
```

Excerpt from the log file follows:

```
WARNING: Defaulting to selecting all PERMNOs in CRSP or CCM database.
NOTE: There were 21269 observations read from the data set DSTK.RETX.
             WHERE permno=12490;
NOTE: The data set WORK.IBMRET_FULL has 1264 observations and 4 variables.
NOTE: DATA statement used (Total process time):
         real time           4:44.15
         cpu time            3:34.17
```

IV. The database is restricted to IBM stock and the date range of interest only.

IV.1 Set up the library for CRSP daily stock database restricted to IBM stock and 1991-1995 period only.

```
LIBNAME ibm sasecrsp 'c:\crspdstk\'
              setid=10
              permno=12490
              range = '19910101-19951231';
```

IV.2. Extract daily returns and convert date from CRSPAccess Date to SAS Date format as in I.2.

```
data ibmret_restr;
       set ibm.retx;
              where permno=12490;
                     date=crspdcsd(caldt);format date date9.;
                            rename retx=return;
run;
```

Excerpt from the log file follows:

```
WARNING: Defaulting to selecting all PERMNOs in CRSP or CCM database.
NOTE: There were 1264 observations read from the data set IBM.RETX.
             WHERE permno=12490;
NOTE: The data set WORK.IBMRET_RESTR has 1264 observations and 4 variables.
NOTE: DATA statement used (Total process time):
         real time           0.11 seconds
         cpu time            0.04 seconds
```

The datasets ibmret_full and ibmret_restr created in steps III.2 and IV.2 are identical and contain IBM stock returns from January 1991 to December 1995. However, the time to access the data of interest by setting up the restricted database using the SASECRSP LIBNAME options is significantly shorter than when data is accessed using unrestricted database.

After the IBM daily stock returns during 2001-2005 are accessed in the same manner, the phase portraits can be plotted for both periods of interest. These are presented in Figure 3 and Figure 4 below.

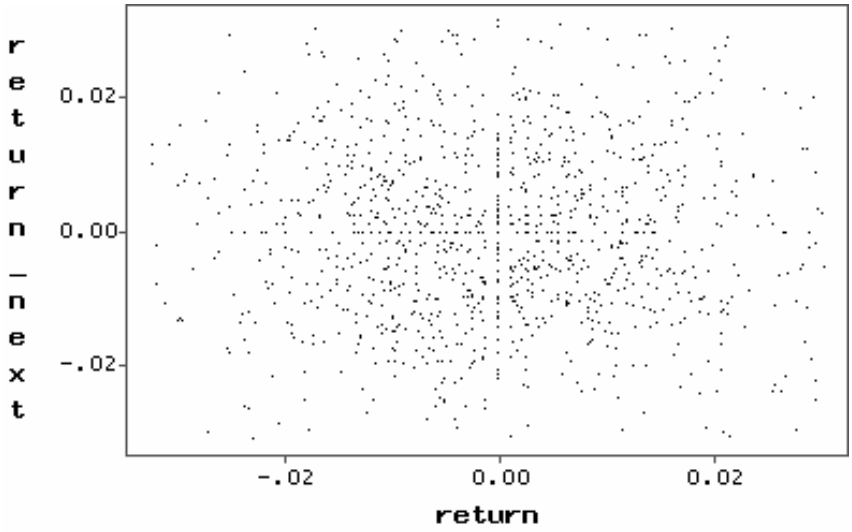

Figure 3. Phase portrait of daily returns of IBM stock, 1991-1995.

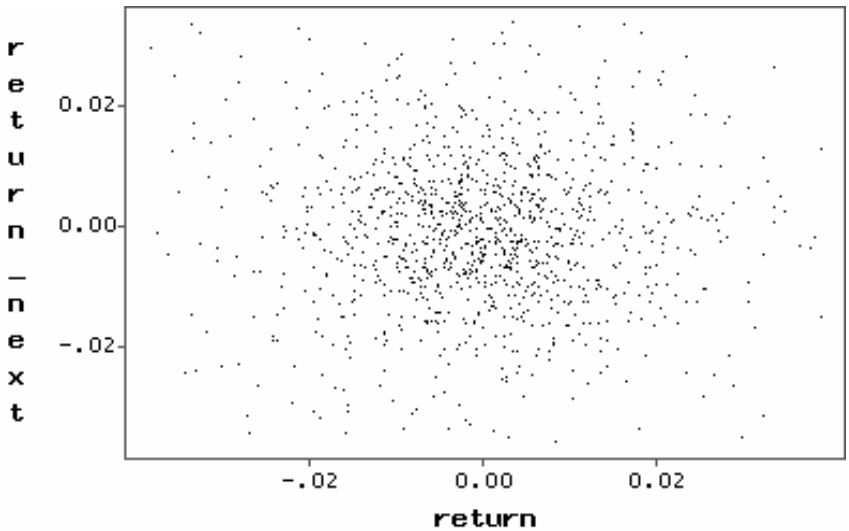

Figure 4. Phase portrait of daily returns of IBM stock, 2001-2005.

The evidence of a compass rose pattern of the IBM stock daily returns during the 1991-1995 period presented in Figure 3 is week but pronounced. As Figure 4 shows, the pattern is not evident during the 2001-2005 period. This may indicate that introduction of decimalization on the New York Stock Exchange influenced the stock returns compass rose pattern formation.

CONCLUSION

Being indispensable when accessing financial databases available from CRSP using SAS, SASECRSP engine provides tools to increase the efficiency of data processing. This paper discussed some of the options available in the SASECRSP LIBNAME statement and their use from program execution efficiency point of view.

The paper provided some examples of accessing daily stock returns data from CRSP expanded daily database and presented phase portraits of the IBM stock in order to detect the presence of the compass rose pattern. The pattern can be clearly seen in daily returns over the 1925-2005 period. It is somewhat pronounced during the 1991-1995 period and not evident in the 2001-2005 period. This may indicate that the compass rose pattern formation is influenced by market decimalization when daily frequency data is used.

The examples presented in the paper show that when accessing returns of one stock the processing time differs significantly for unrestricted and restricted CRSP database. These differences will be compounded when multiple references to database are required in the course of program execution based on either varying universe of securities or time ranges or both. Higher efficiency can be achieved by referencing restricted database using options available in the SASECRSP LIBNAME statement.

The data access modalities presented in this paper show that the SASECRSP LIBNAME statement options are essential and powerful tool when working with SASECRSP engine in accessing CRSP databases. Their use allows higher productivity of data processing and increased efficiency when conducting any kind of financial research using CRSP. It will be of interest to test the SASECRSP engine and the SASECRSP LIBNAME statement options performance more rigorously in the context of new financial applications development.

REFERENCES

Cai, C., Hudson, R. and K.Keasey. 2003. “Trading Frequency and the Compass Rose”. *Applied Economic Letters* 10:511-517.

Crack, T.F., and O.Ledoit. 1996. “Robust Structure Without Predictability: The “Compass Rose” Pattern of the Stock Market”. *Journal of Finance* 51:751-762.

Kramer, W. and R.Runde. 1997. “Chaos and the Compass Rose”. *Economics Letters* 54:113-118.

Lee, I.C., Mathur, I. and K.C.Gleason. 2005. "The Tick/Volatility Ratio as a Determinant of the Compass Rose Pattern". *The European Journal of Finance* 11:93-109.

Szpiro, G.C. 1998. "Tick size, the compass rose and market nanostructure". *Journal of Banking and Finance* 22:1559–1569.

Wang, H. and C. Wang. 2002. "Visibility of the Compass Rose in Financial Asset Returns: A Quantitative Study". *Journal of Banking and Finance* 26:1099–1111.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Taras Zlupko
CRSP, University of Chicago
105 West Adams Street, Suite 1700
Chicago, IL 60603
Work Phone: (312)263-6654
E-mail: Taras.Zlupko@crsp.ChicagoGSB.edu