EPSAL

Equity Portfolio Statistical Analysis Language

EPSAL

COMS 4115 – Programming Languages and Translators

Spring 2009

Charles Neher cgn2103@columbia.edu

Due Date – 2/10/09

1. Introduction

The purpose of EPSAL (Equity Portfolio Statistical Analysis Language) is to provide a simple set of tools to analyze a time series set of data points for a data set of S&P 500 index stocks for a 60 month period. The S&P 500 represents approximately 70% of the value of the U.S. equity market. The listed companies are highly diverse; spanning every relevant portion of the U.S. economy[1].This language will be focused providing key words for the descriptive and inferential statistics in measuring an individual stock or against the index population . A population is defined as the entire set of S&P members but most queries will be focused on subsets. The descriptive statistics will focus on summarizing a set of numerical data into an informative presentation. The Inferential statistics cover algorithms to make forecast estimates and aggregate views on smaller subset in relation to the total population. [2]

An example of the features of this language:

- Performance calculation of index over varying periods
- Comparison of a initial stock against average of index
- Portfolio composition performance metrics – i.e. - % cash vs. index
- Regression to mean of individual stocks or group of stocks
- Weighted mean of a portfolio return
- Population Variance
- Population standard deviation

2. Data Set Details

These data sets consist of files which are archived together, one for each day. The individual files contain a record for each stock, organized as Ticker, Open, High, Low, Close, and Volume, delimited by commas.

Example

DATE	Ticker	Open	High	Low	Close	Volume
20080207	A	33.38	33.5	32.2	32.42	55187
20080208	A	32.19	32.32	31.81	32	54934
20080211	A	32	32	31.58	31.86	46008
20080212	A	32	32.3	31.79	31.85	38668
20080213	A	32.05	33.16	31.92	32.68	60191
20080214	A	32.68	32.825	31.05	31.54	66854
20080215	A	31.41	31.54	30.48	30.65	36729
20080219	A	30.94	31.1	30.67	30.93	39034
20080220	A	30.61	31.61	30.58	31.53	34482
20080221	A	31.58	31.97	31.46	31.61	39864

[1] http://www.fool.com/school/indices/sp500.htm
[2] Schwesser Study Notes – pg 159 – Statistical Concepts and Market Returns

20080222	A	31.43	31.58	30.78	31.41	22253
20080225	A	31.42	31.79	31.2	31.5	30575
20080226	A	31.41	31.91	31.24	31.87	65289
20080227	A	31.74	32.44	31.65	32.03	20699

Total data points for one year's data is 126,492 rows x 7 fields = 885,444

Close prices are adjusted for dividends and splits

3. Calculation Examples[3]

Arithmetic Return – Annual

AR = ((Price final – Price initial) / Price Initial) * 100

Arithmetic Average of Return

AVGR = 1/n (AR(1) + AR(2) +……..+AR(n))

AR – observed total return for a year

n = number of years

Variance of Returns var(R)

VARR = ∑[Rt – Avg (R)]^2 / (n-1)

4. Language Specification

Comments - /* Comment */

End of Statements - ;

Data Types

The data types are the individual calculations which will have input date range parameters

[3] Frank Fabozzi - / James Grant – Equity Portfolio Analysis – pg 45

AR – Arithmetic Return

AVGR – Arithmetic Return

SAMPM - sample mean

POPM – Population Mean

VARR – Variance of Return

WMEAN – Weighted Mean

MODE – Mode

GMEAN – Geometric Mean

POPVAR – Population Variance

POPSTDD- Population Standard deviation

Keywords

DELTA

FIND

CALC

IF

THEN

ELSE

RAND

PRINT

RETURN

FROM

TO

SNP

Code Example

```
/* Calculate Arithmetic Return of  Stock 'YHOO' from  2/7/2008 to 2/7/2009 */
D1 = 20080207;
D2= 20090207;
YahooReturn  = AR  YAHOO FROM D1 TO D2;
PRINT YahooRetrun  ;

/* Calculate Arithmetic Return of  Stock 'MSFT ' from  1/7/2009 to 2/7/2009 */
D1 = 20090107;
D2= 20090207;
MSFTReturn  = AR  MSFT FROM D1 TO D2;
PRINT MSFTRetrun  ;

/* Calculate Population Varaiance in S&P from  2/7/2005 to 2/7/2008 */
D1 = 20050207;
D2= 20080207;
PopVarince 3yrs  = POPVAR SNP D1 TO D2;
PRINT PopVarince 3yrs  ;
```