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. 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.

An example of the features of this language:

- Performance calculation of index over varying periods
- Comparison of an 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

<table>
<thead>
<tr>
<th>DATE</th>
<th>Ticker</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>20080207</td>
<td>A</td>
<td>33.38</td>
<td>33.5</td>
<td>32.2</td>
<td>32.42</td>
<td>55187</td>
</tr>
<tr>
<td>20080208</td>
<td>A</td>
<td>32.19</td>
<td>32.32</td>
<td>31.81</td>
<td>32</td>
<td>54934</td>
</tr>
<tr>
<td>20080211</td>
<td>A</td>
<td>32</td>
<td>32</td>
<td>31.58</td>
<td>31.86</td>
<td>46008</td>
</tr>
<tr>
<td>20080212</td>
<td>A</td>
<td>32</td>
<td>32.3</td>
<td>31.79</td>
<td>31.92</td>
<td>54934</td>
</tr>
<tr>
<td>20080213</td>
<td>A</td>
<td>32.05</td>
<td>33.16</td>
<td>31.92</td>
<td>32.68</td>
<td>60191</td>
</tr>
<tr>
<td>20080214</td>
<td>A</td>
<td>32.68</td>
<td>32.82</td>
<td>31.05</td>
<td>31.54</td>
<td>66854</td>
</tr>
<tr>
<td>20080215</td>
<td>A</td>
<td>31.41</td>
<td>31.54</td>
<td>30.48</td>
<td>30.65</td>
<td>36729</td>
</tr>
<tr>
<td>20080219</td>
<td>A</td>
<td>30.94</td>
<td>31.1</td>
<td>30.67</td>
<td>30.93</td>
<td>39034</td>
</tr>
<tr>
<td>20080220</td>
<td>A</td>
<td>30.61</td>
<td>31.61</td>
<td>30.58</td>
<td>31.53</td>
<td>34482</td>
</tr>
<tr>
<td>20080221</td>
<td>A</td>
<td>31.58</td>
<td>31.97</td>
<td>31.46</td>
<td>31.61</td>
<td>39864</td>
</tr>
</tbody>
</table>

---

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

**Arithmetic Return – Annual**

\[ AR = \frac{(Price \ final - Price \ initial)}{Price \ Initial} \times 100 \]

**Arithmetic Average of Return**

\[ AVGR = \frac{1}{n} (AR(1) + AR(2) + \ldots + AR(n)) \]

\( AR \) – observed total return for a year

\( n \) = number of years

**Variance of Returns var(R)**

\[ VARR = \frac{\sum (R_t - Avg \ (R))^2}{(n-1)} \]

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
SAMP – 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  ;