

What Works On Wall Street — APPENDIX

“As a rule, I always look for what others ignore.”

—Marshall McLuhan

Research Methodology

1. Data

Annual, quarterly, and monthly data from Standard & Poor’s Compustat Database and the Center for Research in Security Prices (CRSP) equity database. All Compustat data from 1963 through 2009 were downloaded to O’Shaughnessy Asset Management PCs from Factset while CRSP data from 1926 to 2009 were downloaded using the CRSPSift tool. Statistical Analysis Software (SAS) was used to perform all back-tests from 1926.

2. Time Horizon

We examined the CRSP database over the 84 years from 1926 to 2009 while Compustat was observed over 47 years from 1963 to 2009. Tests typically start on January 1, 1927 for CRSP and January 1, 1964 for Compustat as most factors require at least 1 year of data. Using Factset as a data aggregator, we were able to use monthly and quarterly data, allowing for time lags of 45 trading days. While in the case of annual data, we allowed for time lags of 90 trading days to avoid look-ahead bias.

3. Universe

We are cognizant of transaction costs and include only stocks that have sufficient liquidity to be purchased without excessively eroding returns. We set a market capitalization of \$200 million as a minimum (in 2007 dollars) for the “All Stocks” universe. To account for inflation, we deflated the current value of \$200 million back to 1926. These were the market capitalization minimums for the CRSP and Compustat database (Compustat data is used from June 30, 1967):

	Minimum Market Cap	
6/30/1927	\$	16,500,000
6/30/1937	\$	13,608,000
6/30/1947	\$	20,864,250
6/30/1957	\$	26,592,000
6/30/1967	\$	31,465,034
6/30/1977	\$	57,246,000
6/30/1987	\$	107,244,000
6/30/1997	\$	151,401,260
6/30/2007	\$	195,858,810
12/31/2009	\$	205,652,370

All stocks with a deflated market capitalization in excess of \$200 million are included and are designated “All Stocks” in the book. Many money managers use a large stocks and small stocks benchmark from which to select stocks. As a proxy for large stocks, we denoted all stocks with market capitalizations greater than the mean in any given month as “Large Stocks”. For small stocks, we denoted all stocks with market capitalizations less than the mean but greater than a deflated \$200 million in any month as “Small

Stocks”. A union of “Small Stocks” and “Large Stocks” makes up the “All Stocks” universe. Below is a snapshot of the “All Stocks,” “Large Stocks” and “Small Stocks” universes as of December 31, 2009:

	Number of Companies	Min. MCAP	Median MCAP	Avg. MCAP
All Stocks	2879	\$ 205,652,370	\$ 1,294,916,700	\$ 8,046,924,920
Large Stocks	523	\$ 8,061,664,600	\$ 19,748,096,000	\$ 36,810,272,300
Small Stocks	2356	\$ 205,652,370	\$ 871,474,700	\$ 1,661,852,470

4. Returns

When simulating back-tests, we form 12 separate portfolios or sleeves each year for each strategy, one every month, and average the results of each of the 12 sleeves to produce our portfolio return. The rebalance period for each sleeve remains one year.

Portfolio weights are calculated monthly as follows in SAS:

- w_{it} = Weight of stock i at the beginning of month t
- r_{it} = Total return of stock i for month t
- p_t = Portfolio total return for month t

Initial Weight of stock i at time t=0 (w_{i0}) = 1 (Portfolio is equally weighted)

$$w_{it} = w_{i(t-1)} \times r_{i(t-1)}$$

$$p_t = \frac{\sum(r_{it} \times w_{it})}{\sum w_{it}}$$

As an example, consider a 2 stock portfolio consisting of stock A and B that qualified for a high value composite screen. The below chart will illustrate the example:

	Stock A		Stock B	
	Return	Weight	Return	Weight
9/30/1991	1.014	1.00	1.04	1.00
10/31/1991	1.051	1.014	0.98	1.04
11/29/1991	1.054	1.066	1.02	1.019

Return of strategy at 11/29/1991 = $[(1.054 * 1.066 + 1.02 * 1.019) / (1.066 + 1.019)] - 1 = .037$ or 3.7%

We use the actual total return as calculated by Compustat’s total return function within Factset (MRET). Portfolio returns are calculated on a monthly basis and then rolled up to compute annualized returns. Portfolios are constructed using 25 or 50 stock baskets as well as by partitioning the universe into deciles.

5. Data Definitions

Here are the definitions of some data items used within the book, followed in parentheses by their Factset formula:

Sales: Net sales over the last 4 quarters, time-lagged by 45 days (COUNT4(CQ_SALES(R2 #1 L45D)))

Common Shares Outstanding: The net number of common shares outstanding as of month-end (CM_SHS(0))

Market Capitalization: Calculated by multiplying common shares outstanding by market price per share (SP_MCAP(0))

EBITDA: Represents the earnings before interest, taxes, depreciation and amortization over the last 4 quarters, lagged by 45 days (COUNT4(CQ_EBITDA(#1 L45D,R2)))

Free Cash Flow: Calculated as the sum of net income and depreciation/amortization expense minus changes in working capital and capital expenditures (COUNT4(CQ_CF_FREE(#1 L45D,R2)))

Dividend Yield: Determined by multiplying the latest dividend per share by the number of times it is paid in a year, relative to the market price per share (100*(MADR(0)/OCM_PRICE))

Buyback Yield: Represents share repurchases by a firm over the past 12 months, calculated as the percentage change in shares outstanding (100*(1-CM_SHS(#1)/CM_SHS(#1-12)))

6. Formulas

All formulas use the above items as well as common ranking and averaging techniques. Here are the definitions:

Average: Established using the AVG function in PROC SQL within SAS. The function calculates the average value of an item or expression over a set. Here is an example of the AVG function being applied to market capitalization:

```
PROC SQL;
SELECT caldt, avg(mcap) as avg_mcap
FROM research.research
GROUP BY caldt
RUN;
```

Percentile Rank: A percentile rank of 50 for a particular data item indicates that 50% of the data is ranked above and 50% below. This is accomplished using the PROC RANK function within SAS. This function determines the percentile rank(1-100) of an item within a data set by month. Here is an example of the PROC RANK function being applied to market capitalization:

```
PROC RANK DATA=reslocal.temp_subset GROUPS=100 OUT=reslocal.temp_subset;
BY period; VAR mcap; RANKS pct_mcap ; RUN;
```

Ordinal Rank: Such as the top 50 by dividend yield or the top 50 by sales-to-price is accomplished using the PROC RANK function within SAS. Here is an example of the PROC RANK function being applied to sales-to-price:

```
PROC RANK DATA=reslocal.temp_subset TIES=low OUT=reslocal.temp_subset;
BY period; VAR sp; RANKS high; RUN;
```

Sample Composite Value Factor Calculation

When combining groups of factors we assign a percentile ranking for each stock in the universe by item based on the percentile ranking methodology discussed above. If a value is missing for a factor, we assign a neutral rank of 50. Once all factors are ranked, we sum across all factors and assign the stocks to deciles based upon their overall cumulative ranking on all factors. Those with the highest

scores are assigned to decile one while those with the lowest scores are assigned to decile ten. Here is an example of how the composite value factor is computed in SAS:

```
data strategy.wvows; set strategy.wvows;
  comp_value =      sum(0,coalesce(pct_cfp_free,50),
                    coalesce(pct_ebev,50),
                    coalesce(pct_ep,50),
                    coalesce(pct_sp,50),
                    coalesce(pct_bp,50),
                    coalesce(pct_sh_yld,50)
                    );
Run;
```

In addition to return, we are concerned with secondary or ancillary statistics to gain more insight into a strategy. EnCorr Analyzer was used to generate the values for the various statistics. Here are the definitions for the most prominent metrics used in the book, followed by the formula used to compute them:

Standard Deviation of Return

The extent to which observations in a data series differ with the average return for the entire series. The larger the standard deviation, the “riskier” the strategy.

$$\text{Standard Deviation} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2}$$

Semi-Standard Deviation of Return Below Zero (Downside Risk)

Focuses on the portion of risk that is to the left of all observations below zero return. It essentially focuses on downside risk and the *lower* this number is, the lower the risk of the strategy when stock prices are falling.

$$\text{Semi-Standard Deviation} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \bar{r})^2 \text{ where } r_i < 0}$$

T-Statistic

Measures how likely it is that results are due to chance. Typically, a T-Statistic of +/-1.96 (where there are at least 20 observations) indicates a statistically significant selection return at the 95 percent level of confidence. The higher the T-Statistic, the less likely the results were due to chance.

$$t = \frac{\hat{\mu} - \mu}{\frac{\hat{\sigma}}{N}}$$

where: $\hat{\mu}$ = sample mean
 μ = population mean
 $\hat{\sigma}$ = sample standard deviation
 N = sample size

Beta

Beta theoretically measures how closely a security or portfolio of securities perform relative the market as a whole. A stock or portfolio with a Beta of 1 would suggest that it would have similar risk

and returns to the market as a whole, whereas a security or a portfolio with a Beta of 1.4 would be considered much more risky and have higher return potential than the market as a whole.

$$\text{Beta} = \frac{\text{Cov}(r_p, r_m)}{\sigma_m^2}$$

where: r_p = portfolio return
 r_m = market return
 σ_m^2 = variance of the market
 $\text{Cov}(r_p, r_m)$ = covariance of portfolio return to market return

Sharpe Ratio

Measures the reward to risk of a strategy, with higher numbers indicating better risk-adjusted returns.

$$\text{SR} = \frac{R - R_f}{\sigma}$$

where: R = asset return
 R_f = risk free rate of return
 σ = standard deviation of return

Treynor Ratio

Measures the reward to systematic risk of a strategy, with higher numbers indicating better risk-adjusted returns.

$$\text{TR} = \frac{R - R_f}{\beta}$$

where: R = asset return
 R_f = risk free rate of return
 β = portfolio beta

Sortino Ratio

Measures the reward to downside-risk of a strategy, with higher numbers indicating better risk-adjusted returns.

$$\text{SOR} = \frac{R - T}{DR}$$

where: R = asset return
 T = target rate of return
 DR = downside risk

7. Taxes, Commissions, and Market Impact Costs

These are not included, but a real-time use of the strategies reveals that market impact and commissions are minimal. Taxes would reduce the returns according to the tax rate faced. Because the portfolio sleeves are rebalanced annually, a taxable investor would pay at the 15% capital gains tax rate on all gains provided he or she sells the appropriate tax lot and would presumably sell all losses for short-term capital losses to be reported against ordinary income.