Global Stock Selection Modelinng and Efficient Portfolio Construction and Management

JOHN B. GUERARD JR., HARRY MARKOWITZ, AND GANLIN XU

Expected returns on assets are not completely explained by using only historical means (and standard deviations). One can estimate models of expected return by using earnings expectations data, price momentum variables, and reported financial data. In this analysis, we construct and estimate a global stock selection model by using earnings expectations data, price momentum, and reported financial data for the period from January 1997 through December 2011. A composite value of earnings expectations information, value, and momentum factors is estimated for global stocks to identify potentially mispriced stocks. In addition, the regression weighting of factors enhances information coefficients relative to equally weighted factors. Analysts’ forecast and momentum variables dominate the regression-based composite model of expected returns. We create portfolios for the January 1997 through December 2011 period and simulate portfolio returns versus a set of global stock benchmark returns.

We begin with a review of the literature of stock selection models. In the following section, we discuss a composite model of stock selection. We then use the SunGard APT-based multifactor risk model to create efficient portfolios. Finally, we present and estimate the Data Mining Corrections test before offering our summary and conclusions.

A BRIEF LITERATURE REVIEW OF EXPECTED RETURNS MODELING AND STOCK SELECTION MODELS

There are many approaches to security valuation and the creation of expected returns. Early approaches to security analysis and stock selection involved the use of valuation techniques using reported earnings and other financial data. Graham and Dodd [1934] recommended that stocks be purchased on the basis of the price–earnings ratio (P/E). They suggested that no stock should be purchased if its P/E exceeded half the reciprocal of AAA yield. Graham and Dodd established the P/E criteria with other requirements. They also advocated the calculation of a security’s price-to-book ratio (P/B), but the P/B alone should not be used as a measure for stock selection.

The “low” P/E investment strategy is discussed in Williams [1938], the monograph that influenced Harry Markowitz and his thinking on portfolio construction. Basu [1977] reported evidence supporting the low P/E model. Academics often prefer to test the low P/E approach by testing its reciprocal, the “high E/P” approach. The high E/P approach specifically addresses the issue of negative earnings per share, which can confuse the low P/E test.

Bloch et al. [1993] developed and estimated an eight-factor model of expected
returns incorporating the low P/E strategy and similar strategies for book value, cash flow, and sales for stocks in the United States and Japan. Guerard et al. [2012] added price momentum and earnings forecasting to the Bloch et al., framework and found that price momentum and earnings forecasting accounted for 45% of the variable weights in the 10-factor USER Model. Bloch et al. [1993] tested the relative explanatory and predictive merits of alternative regression estimation procedures: ordinary least squares (OLS), robust regression using the Beaton–Tukey [1974] bi-square criterion to mitigate the impact of outliers, latent root to address the issue of multicollinearity (see Gunst et al. [1976]), and weighted latent root (WLRR), a combination of robust and latent root. The Guerard et al. [2012] USER model test substantiated the Bloch et al. [1993] approach, techniques, and conclusions. Further evidence on the anomalies appears in Levy [1999] and Dimson [1988].

THE GLOBAL EXPECTED RETURNS MODEL FOR STOCK SELECTION

Here we discuss issues of databases and the inclusion of variables in composite models to identify undervalued securities in a global stock universe. The database for this analysis is created by the use of all securities listed on the FactSet database during the period January 1997 through December 2011. We use the I/B/E/S database of forecasted earnings and require at least two analysts’ forecasts of each stock in a given month. There are a seemingly infinite number of financial variables that may be tested for statistical association with monthly security returns. Bloch et al. [1993] tested a set of fundamental variables for the U.S. during the 1975–1990 period. We initially test the effectiveness of the individual variables using the information coefficients (ICs), rather than the upper quintile excess returns or the excess returns of individual variable portfolio optimizations. The IC is the slope of the regression estimation in which ranked subsequent security returns are a linear function of the ranked financial strategy. The advantage of the IC approach is that the slope has a corresponding $t$-statistics that allows one to test the null hypothesis that the strategy is uncorrelated with subsequent returns. We refer to the IC test as a Level I test of portfolio construction and management. In developing a composite model, one seeks to combine variables that are statistically associated with subsequent returns. The individual variables tested in this study are as follows:

- \( EP = \) earnings per share/price per share;
- \( BP = \) book value per share/price per share;
- \( CP = \) cash flow per share/price per share;
- \( SP = \) sales per share/price per share;
- \( PM = \) price momentum as \( \frac{Price_{t-1}}{Price_{t-12}} \);
- \( FEP1 = \) one-year-ahead forecast earnings per share/price per share;
- \( FEP2 = \) two-year-ahead forecast earnings per share/price per share;
- \( RV1 = \) one-year-ahead forecast earnings per share monthly revision/price per share;
- \( RV2 = \) two-year-ahead forecast earnings per share monthly revision/price per share;
- \( BR1 = \) one-year-ahead forecast earnings per share monthly breadth;
- \( BR2 = \) two-year-ahead forecast earnings per share monthly breadth;
- \( CTEF = \) equally-weighted \( FEP1, FEP2, BR1, BR2, RV1, \) and \( RV2 \).

How does one develop and estimate a stock selection model? One can survey the academic and practitioner literature, as we have done, and calculate information coefficients for the equity universe within which one seeks to manage assets. Exhibit 1 reports the average information coefficients of several important variables in this study.

Strong support exists for the earnings expectations variables and fundamental variables (particularly earnings and cash flow). An objective examination of the reported ICs leads one to identify \( CTEF, EP, \) and \( FEP \) as leading variables for inclusion in stock selection models.

| Exhibit 1 |

Information Coefficients of FSGLER Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>1997–2011 IC (t)</th>
<th>2003–2011 IC (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( EP )</td>
<td>0.037 (6.21)</td>
<td>0.029 (4.75)</td>
</tr>
<tr>
<td>( BP )</td>
<td>0.008 (0.11)</td>
<td>0.002 (0.22)</td>
</tr>
<tr>
<td>( FEP )</td>
<td>0.040 (6.67)</td>
<td>0.032 (5.85)</td>
</tr>
<tr>
<td>( CTEF )</td>
<td>0.035 (10.42)</td>
<td>0.033 (8.68)</td>
</tr>
<tr>
<td>( EWC )</td>
<td>0.028 (6.07)</td>
<td>0.027 (6.87)</td>
</tr>
<tr>
<td>( GLER )</td>
<td>0.045 (6.98)</td>
<td>0.036 (5.41)</td>
</tr>
</tbody>
</table>
Bloch et al. [1993] used the first eight factors as reported in Equation (1) to model expected return. If we add price momentum, calculated as the price at \( t - 1 \) divided by the price at \( t - 12 \) (months), \((PM)\) and the consensus analysts’ earnings forecasts and revisions variable \((CTEF)\), to the stock selection model, we can estimate an expanded stock selection model to use as an input to an optimization analysis. The PM model reflects the findings of Brush [2001, 2007]. The CTEF variable incorporates the spirit of earnings revisions as in Elton et al. [1981]. Guerard et al. [2012] estimated this model, denoted as USER, with US stocks during the 1997–2007 period. The GLER model builds upon the anomalies research of Graham, Dodd, and Cottle [1962], Latane, Tuttle, and Jones [1975], and Haugen and Baker [2010], Dimson [1988], Lakonishok, Shleifer, and Vishny [1994]. The global stock selection model estimated in this study, denoted as Global Expected Returns (GLER), is as follows:

\[
TR_{it} = a_0 + a_1 EP + a_2 BP + a_3 CP + a_4 SP + a_5 REP + a_6 RBP + a_7 RCP + a_8 RSP + a_9 CTEF + a_{10} PM + e_t 
\]

where

\[
EP = \frac{\text{earnings per share}}{\text{price per share}} = \text{earnings–price ratio};
\]

\[
BP = \frac{\text{book value per share}}{\text{price per share}} = \text{book–price ratio};
\]

\[
CP = \frac{\text{cash flow per share}}{\text{price per share}} = \text{cash flow–price ratio};
\]

\[
SP = \frac{\text{net sales per share}}{\text{price per share}} = \text{sales–price ratio};
\]

\[
REP = \frac{\text{current EP ratio}}{\text{average EP ratio over the past five years}};
\]

\[
RBP = \frac{\text{current BP ratio}}{\text{average BP ratio over the past five years}};
\]

\[
RCP = \frac{\text{current CP ratio}}{\text{average CP ratio over the past five years}};
\]

\[
RSP = \frac{\text{current SP ratio}}{\text{average SP ratio over the past five years}};
\]

\[
CTEF = \text{consensus earnings per share I/B/E/S forecast, revisions, and breadth,}
\]

\[
PM = \text{price momentum; and}
\]

\[
e = \text{randomly distributed error term.}
\]

The GLER model is estimated using WLRR analysis on Equation (1) to identify variables statistically significant at the 10% level; uses the normalized coefficients as weights; and averages the variable weights over the past 12 months. The 12-month smoothing is consistent with the four-quarter smoothing in Bloch et al. [1993]. We refer to the GLER model estimated with FactSet data as the FSGLER model to avoid confusion with the GLER model estimated with the Wharton Research Data Services Global Compustat database in Anand and Gultekin [forthcoming 2014], which they refer to as WRDS GLER. In terms of ICs, the use of the WLRR procedure produces the highest IC for the models during the 1997-2011 time periods, shown in Exhibit 1. The WLRR technique produces the largest and most statistically significant IC, a result consistent with the previously noted studies. The \( t \)-statistics on the composite model exceed the \( t \)-statistics of its components. The purpose of a composite security valuation model is to identify the determinants of security returns and produce a statistically significant out-of-sample ranking metric of total returns.

The EP and BP variables are significant in explaining returns; however, the majority of the forecast performance is attributable to other model variables, namely the relative earnings-to-price, the relative cash-to-price, relative sales-to-price, and earnings forecast variables. Exhibit 2 reports that the consensus earnings forecasting variable, CTEF, dominates the top/bottom (one- and three-) decile spreads.

The GLER model can be input into an optimization system to create optimized Global portfolios, just as Guerard et al. [2012] created US stock portfolios.

**EFFICIENT APT PORTFOLIO CONSTRUCTION**

Portfolio construction and management, as formulated in Markowitz [1959], seeks to identify the efficient frontier, the point at which the portfolio return is maximized for a given level of risk or, equivalently, portfolio risk is minimized for a given level of portfolio return. The portfolio expected return, denoted by \( E(R_p) \), is calculated by taking the sum of the security weight multiplied by its respective expected return:

\[
E(R_p) = \sum_{i=1}^{N} w_i E(R_i) 
\]

The portfolio standard deviation is the sum of the weighted securities’ covariance:
where $N$ is the number of candidate securities; $w_i$ is the weight for security $i$ such that $\sum w_i = 1$, indicating that the portfolio is fully invested; and $E(R_i)$ is the expected return for security $i$. The authors assume that the reader has read several times the foundations of portfolio selection found in Markowitz [1959].

We introduced the reader to the Markowitz [1952 and 1959] model in which investors are compensated for bearing the total risk of the portfolio. Implicit in the development of modern portfolio theory, such as the capital asset pricing model (CAPM) by Sharpe [1964],Lintner [1965a], and Mossin [1966], is the concept that investors are compensated for bearing systematic or market risk, not total risk. Systematic risk is measured by a stock’s beta. Beta is the slope of the market model in which the stock return is regressed as a function of the market return. An investor is not compensated for bearing risk that may be diversified away from the portfolio but rather is compensated for portfolio risk as measured by factor exposures.

Guerard [2012], Guerard et al. [2012], Wormald and van der Merwe [2012], and Guerard et al. [2013] demonstrate and report the effectiveness of a multifactor model and optimization system, the SunGard APT system, in portfolio construction and management. The APT model, documented in the APT Analytics Guide [APT, 2011], uses a 20-factor beta model of covariance based on 3.5 years of weekly stock return data. The SunGard APT Model follows the Ross [1976] factor theory, but APT estimates at least 20 orthogonal factors. Saxena and Stubbs [2012] showed the effectiveness of a Statistical Risk Model using the Axioma system. The trade-off curves in Guerard [2012] were created by varying lambda, a measure of risk aversion, as a portfolio decision variable. Exhibit 3 shows the tradeoff curves of selected variables.

As lambda rises, the expected return of the portfolio rises and the number securities in the portfolio declines. The creation of portfolios with a multifactor model and the generation of excess returns is referred to as the Level II test of portfolio construction and management.

The GLER Model risk–return frontier substantiates the effectiveness of the USER analysis in global markets. Fundamental data modeling, earnings forecasting, and momentum strategies have been rewarded in global markets during the 1997–2011 period. For a similar period, 1999 through 2009, Deng and Min [forthcoming 2014] report that the GLER portfolio simulation information ratios (IRs) and Sharpe ratios (ShRs) exceed the USER portfolio statistics. It has paid to be a global investor.

**A FURTHER TEST OF DATA MINING CORRECTIONS**

Markowitz and Xu [1994] demonstrated the naïve practice of Wall Street to cut historical back-tested excess returns in half and not test for the statistical significance
of the excess returns. Let us trace the development of the Markowitz and Xu model and estimate the data mining corrections estimator for a series of US expected return models.

Let $\text{GM}_b$ be the back-tested geometric mean of the “best” historical simulation during $T$ periods. Markowitz and Xu [1994] work with the logarithm of the geometric mean.

$$g_b = \log_e (1 + \text{GM}_b)$$ (4)

The Markowitz and Xu data mining corrections (DMC) test assumes that the $T$-period historical returns were identically and independently distributed (i.i.d.), and that future returns are drawn from the same population (also i.i.d.). Because we test many models, not just the best model, the best geometric mean is no longer the best unbiased estimate of the true, underlying population $g_b$.

Markowitz and Xu [1994] set $y_a$ as the logarithm of one plus the return for the $i^{th}$ portfolio selection model in period $t$. In their Model I and II, $y_a$ is

$$y_a = \mu_i + z_t + \varepsilon_{it}$$ (5)

where $\mu_i$ is the model effect, $z_t$ is the period effect, and $\varepsilon_{it}$ is a random deviation. These effects are all assumed to be uncorrelated. Readers are referred back to the original article for the correlated $\varepsilon_{it}$ case. Finally, the appropriate estimate of $\mu_i$ is not the average return

$$\bar{r} = \frac{\sum_{t=1}^{T} y_{it}}{T}$$ (6)

but rather

$$\hat{\mu}_i = \bar{r} + \beta (\bar{r} - \bar{r})$$ (7)
where $\bar{r} = \Sigma_{n=1}^{n} \bar{r}/n$ is the grand mean of methods and $\hat{\beta} = \sum_{i=1}^{n} \hat{r}_i/\Sigma_i r_i$ is the regression coefficient of $\hat{r}$ as a function of $r$, such that $0 < \hat{\beta} < 1$. In other words, the best estimate of model effect $\hat{r}$ is its sample estimate regressed back to the average estimate (the grand average).

The Markowitz–Xu DMC test does not use a “holdout period,” because they can be routinely data-mined as well. That is, one can vary the estimation and holdout periods to generate the desired conclusion. Markowitz and Xu [1994] tested the DPOS strategies in Bloch et al. [1993], and reported a Model II $\hat{\beta}$ of 0.59, which was statistically significant—that is, approximately 59% of the excess returns could be expected to continue. Alternative portfolio models were created using the factors discussed in Equation (1) as tilt factors. The GLER variable analysis passes the DMC test criteria for the US market, indicating that the stock selection and portfolio construction methodologies produce superior returns that are not due to chance. The application of the Markowitz–Xu [1994] DMC test is referred to as a Level III test of portfolio construction and management.

The GLER variable has a DMC coefficient of 0.74 and is highly statistically significant, having an F-value of 1.9. Thus, one could expect 74% of the excess returns of the GLER model relative to the average return to continue. More importantly, the GLER model produced a higher geometric mean than did an average model geometric mean that could have been used to manage an equity portfolio in the US equity market during the period January 1997 through December 2011. The GLER Model passes all three levels of hypothesis testing: (1) The model and its components have statistically significant information coefficients, (2) the strategy produces excess returns after transaction costs, and (3) the strategy produces a significantly higher geometric mean than the average model that could have been used to manage assets in the universe.

CONCLUSIONS

Investing with analysts’ expectations, fundamental data, and momentum variables is a good investment strategy over the long run. Stock selection models often use momentum, analysts’ expectations, and fundamental data. We find support for composite modeling using these sources of data. We find additional evidence to support the use of APT multifactor models for portfolio construction and risk control. We develop and estimate three levels of testing for stock selection and portfolio construction. The uses of multifactor risk-controlled portfolio returns allow us to reject the data mining corrections test null hypothesis. The anomalies literature can be applied in real-world global portfolio construction.

ENDNOTE

The authors are grateful to Martin J. Gruber, who refereed the technical version of this study, entitled “Earnings Forecasting in A Global Stock Selection Model and Efficient Portfolio Construction and Management.” The authors benefited from research conversations with Anureet Saxena, Eli Krauklis, Laurence Wormald, and Bernell Stone. Any errors remaining are the responsibility of the authors.

REFERENCES


To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.

Disclaimer
McKinley Capital Management, LLC (“McKinley Capital”) is a registered investment adviser under the Securities and Exchange Commission Investment Advisers Act of 1940. One or multiple authors may be currently employed by McKinley Capital. However, the firm may not be held liable for comments made outside the scope of this presentation. The material provided herein has been prepared for institutional clients and financially knowledgeable and sophisticated high net-worth individuals. This material may contain confidential and/or proprietary information and may only be relied upon for this report. The data is unaudited and may not correspond to similar calculations provided in other reports. This is not an offer to purchase or sell any security or service, is not reflective of composite or individual portfolio ownership and may not be relied upon for investment purposes. Investments and commentary were based on information available at the time and are subject to change without notice. Any positive comments regarding specific securities may no longer be applicable and should not be relied up for investment purposes. No one security is profitable all the time and there is always the possibility of selling it at a loss. No one formula or set of data will be 100% reliable or profitable at any given period. Past performance is not indicative of future returns. Investments are subject to immediate change without notice.

Because McKinley Capital’s investment process is proprietary, composite returns and individual client returns may at various times materially differ from stated benchmarks. The author(s) may utilize a combination of firm specific and general publicly available sources and resources to prepare material or prove content. Charts, graphs and other visual presentations and text information are derived from internal, proprietary, and/or service vendor technology sources and/or may have been extracted from other firm data bases. As a result, the tabulation of certain reports may not precisely match other published data. Data may have originated from various sources including but not limited to Bloomberg, Clarifi, MSCI/Barra, Russell Indices, and/or other systems and programs. Please refer to the MSCI/Barra, Russell Investments, and FTSE websites for complete details on all related indices. Future investments may be made under different economic conditions, in different securities and using different investment strategies. Global investing also carries additional risks and/or costs including but not limited to, political, economic, financial market, currency exchange, liquidity, accounting, and trading capability risks. Shorting and derivatives may materially increase overall risk and negatively affect returns in the portfolio.

Investors must consider total costs including management fees, expenses, brokerage commissions, custodial services, and individual tax considerations when estimating a potential return. Actual investment advisory fees incurred by institutional and high net-worth clients may vary. A fee schedule for McKinley Capital is described in Form ADV Part II, which can be obtained from McKinley Capital Management, LLC, 3301 C Street, Suite 500, Anchorage, Alaska 99503, 1.907.563.4488, or via the firm’s website at www.mckinleycapital.com. All information is believed to be correct but accuracy cannot be guaranteed. Please review the references section of this report for further details on locating complete source information.