



Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast
 المجلة
 الدولية
 للتنبؤ
 الحرة
 ورقة
 بحثية
 paper


Earnings forecasting in a global stock selection model and efficient portfolio construction and management

 John B. Guerard Jr.^{a,*}, Harry Markowitz^{a,b}, GanLin Xu^{a,c}
^a McKinley Capital Management, LLC, 3301 C Street, Suite 500, Anchorage, AK 99503, United States

^b Harry Markowitz & Company, San Diego, CA, United States

^c Guided Choice, San Diego, CA, United States

ARTICLE INFO

Keywords:

 Earnings Forecasting
 I/B/E/S
 Portfolio optimization
 Information ratio
 Efficient frontier
 Active returns

ABSTRACT

Stock selection models often use analysts' expectations, momentum, and fundamental data. We find support for composite modeling using these sources of data for global stocks during the period 1997–2011. We also find evidence to support the use of SunGard APT and Axioma multi-factor models for portfolio construction and risk control. Three levels of testing for stock selection and portfolio construction models are developed and estimated. We create portfolios for January 1997–December 2011. We report three conclusions: (1) analysts' forecast information was rewarded by the global market between January 1997 and December 2011; (2) analysts' forecasts can be combined with reported fundamental data, such as earnings, book value, cash flow and sales, and also with momentum, in a stock selection model for identifying mispriced securities; and (3) the portfolio returns of the multi-factor risk-controlled portfolios allow us to reject the null hypothesis for the data mining corrections test. The earnings forecasting variable dominates our composite model in terms of its impact on stock selection.

© 2014 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Expected returns on assets are a key input in the mean-variance portfolio selection process. One can estimate models of expected returns 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 these data for the period from January 1997 to December 2011. Earnings expectations information has been being rewarded in global stocks for the past fifteen years or so, and we expect it to continue to be the primary variable driving global

stocks. Despite the recent volatility of the momentum factor, momentum is still associated statistically with security returns, and can be used with other factors to rank stocks for purchase. A composite model of earnings expectations information, value, and momentum factors is estimated for global stocks in order to identify potentially mispriced stocks. In addition, the regression-weighting of factors enhanced the information coefficients relative to equally-weighted factors. Analysts' forecast and momentum variables are dominant in the regression-based composite model of expected returns. We create portfolios for the period January 1997–December 2011, and simulate portfolio returns which we compare with a set of global stock benchmark returns.

We begin with a review of the literature on stock selection models in Section 2. In Section 3, we discuss the testing of a composite model of stock selection, incorporating

* Corresponding author. Tel.: +1 907 563 4488.

E-mail addresses: jguerard@mckinleycapital.com (J.B. Guerard Jr.), hmarkowitz@mckinleycapital.com (H. Markowitz), gxu@mckinleycapital.com (G. Xu).

<http://dx.doi.org/10.1016/j.ijforecast.2014.10.003>

0169-2070/© 2014 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

earnings forecast information. We use an APT-based multi-factor risk model to create efficient portfolios in Section 4. In Section 5, we present and estimate the data mining corrections test. In Section 6, we discuss the relevance of the “alpha alignment factor” and show its relevance. Section 7 presents our summary and conclusions.

2. A literature review of expected returns modeling and stock selection models

There are many different approaches to security valuation and the creation of expected returns. One seeks to select expected returns inputs that are associated statistically with stock returns. The correlation coefficient between the strategy and the subsequent returns is referred to as the information coefficient, IC (Grinold & Kahn, 1999). The expected returns input normally consists of variables that are denoted anomalies, which can be used as inputs to the portfolio construction process in order to produce portfolios that outperform the market. The early approaches to security analysis and stock selection involved the use of valuation techniques that used reported earnings and other financial data. Graham, Dodd, and Cottle (1934) recommended that stocks be purchased on the basis of the price-to-earnings (P/E) ratio. They suggested that no stock should be purchased if its price-to-earnings ratio exceeded 1.5 times the P/E multiple of the market. Graham and Dodd established the P/E criteria, and it was then discussed by Williams (1938), who wrote the monograph that influenced Harry Markowitz’s thinking on portfolio construction. It is interesting that Graham and Dodd proposed the low P/E model at the height of the Great Depression. Basu (1977) reported evidence supporting the low P/E model. The recent literature on financial anomalies is summarized by Fama and French (2008) and Levy (1999).

There is an extensive body of literature on the impact of individual value ratios on the cross-section of stock returns. We go beyond using just one or two of the standard value ratios (EP and BP), and also include the cash-to-price ratio (CP) and/or the sales-to-price ratio (SP). The major papers on the combination of value ratios for the prediction of stock returns (including at least CP and/or SP) include those of Bloch, Guerard, Markowitz, Todd, and Xu (1993), Chan, Hamao, and Lakonishok (1991), Guerard, Rachev, and Shao (2013), Haugen and Baker (2010) and Lakonishok, Shleifer, and Vishny (1994).

Chan et al. (1991) used seemingly unrelated regressions (SUR) to model CAPM excess returns as functions of traditional fundamental variables such as earnings, book values and cash flows relative to price, denoted as EP, BP and CP. Moreover, size was measured as the natural logarithm of market capitalization (LS).¹ Betas were estimated simultaneously, and cross-sectional correlations of residuals were addressed. When fractal portfolios were constructed by sorting on the EP ratio, the highest EP quintile portfolio outperformed the lowest EP quintile portfolio, and the EP

effect was not statistically significant. The portfolios composed of and sorted by the highest BP and CP outperformed the portfolios composed of the lowest BP and CP stocks. In the authors’ multiple regressions, the size and book-to-market variables were positive and statistically significant. The EP coefficient was negative and statistically significant at the 10% level. Applying an adaptation of the Fama and MacBeth (1973) time series of portfolio cross-sections to the Japanese market produced negative and statistically significant coefficients on EP and size, but positive and statistically significant coefficients for the BP and CP variables. Chan et al. (1991, p. 1760) summarized their findings as follows: “The performance of the book-to-market ratio is especially noteworthy; this variable is the most important of the four variables investigated”.

Bloch et al. (1993) built fundamental-based stock selection models for Japanese and United States stocks. The investable stock universe was the first section, non-financial Tokyo Stock Exchange common stocks from January 1975 to December 1990 in Japan, and the 1000 largest market-capitalized common stocks from November 1975 to December 1990 in the United States. They found that a series of Markowitz (1952, 1959) mean-variance efficient portfolios using the higher EP values in Japan underperformed the universe benchmark, whereas the BP, CP, and SP (sales-to-price, or sales yield) variables outperformed the universe benchmark. For the United States, the optimized portfolios using the BP, CP, SP, and EP variables outperformed the U.S. S&P 500, providing support for the Graham and Dodd concept of using the relative rankings of value-focused fundamental ratios to select stocks.² Bloch et al. (1993) used relative ratios as well as current ratio values. Not only might an investor want to purchase a low P/E stock, one might also wish to purchase when the ratio is at a relatively low value compared to its historical value, in this case a low P/E relative to its average over the last five years. Bloch et al. (1993) estimated Eq. (1) in order to assess empirically the relative explanatory power of each of the eight value ratios in the equation:

$$TR = w_0 + w_1EP + w_2BP + w_3CP + w_4SP + w_5REP + w_6RBP + w_7RCP + w_8RSP + e_t. \quad (1)$$

Given concerns about both outlier distortion and multicollinearity, Bloch et al. (1993) tested the relative explanatory and predictive merits of alternative regression estimation procedures: OLS, robust regression using the Beaton and Tukey (1974) bi-square criterion to mitigate the impact of outliers, latent roots to address the issue of multicollinearity (see Gunst, Webster, & Mason, 1976), and

¹ Chan et al. (1991) define cash as the sum of earnings and depreciation, without explicit correction for other noncash revenue or expenses.

² One finds the Price/Earnings, Price/Book and Price/Sales ratios listed among the accounting anomalies by Levy (1999, p. 434). Levy also discusses the dividend yield as a (positive) stock anomaly. Malkiel (1996) cites evidence in support of buying low P/E, low P/B, and high D/P (dividend yield) stocks for a good performance, provided that the low P/E stocks have modest growth prospects (pp. 204–210). Malkiel speaks of a “double bonus”; that is, if growth occurs, earnings increase and the price-to-earnings multiple may increase, driving the price up even further. Of course, should growth fail to occur, both earnings and the P/E multiple may fall.

weighted latent roots, denoted WLRR, a combination of robust and latent roots. Bloch et al. (1993) used the estimated regression coefficients to construct a rolling horizon return forecast. The predicted returns and predictions of risk parameters were used as inputs for a mean–variance optimizer (see Markowitz, 1987) to create mean–variance efficient portfolios in financial markets in both Japan and the United States. Bloch et al. (1993) reported several results. First, they compared OLS and WLRR techniques, inputting the expected return forecasts produced by each method into a mean–variance optimizer. The WLRR-constructed composite model portfolio produced higher Sharpe ratios and geometric means than the OLS-constructed composite model portfolio in both Japan and the United States, indicating that controlling for both outliers and multicollinearity is important when using regression-estimated composite forecasts. Second, Bloch et al. (1993) quantified the survivor bias and found that it was not statistically significant in either Japan or the United States for the period tested. Third, they investigated period-to-period portfolio revision and found that tighter turnover and rebalancing triggers led to higher portfolio returns for value-based strategies. Finally, Markowitz and Xu (1994) developed a test for data mining. In addition to testing the hypothesis of data mining, the test can also be used to estimate and assess the expected differences between the best test model and the average of simulated policies.

In a thorough assessment of value versus growth in the United States, Lakonishok et al. (1994) examined the intersection of the Compustat and CRSP databases for annual portfolios for NYSE and AMEX common stocks, April 1963 to April 1990. Their value measures were three current value ratios: EP, BP and CP. Their growth measure was the five-year average annual growth of sales (GS). They performed three types of tests: a univariate ranking into annual decile portfolios for each of the four variables, bivariate rankings on CP (value) and GS (growth, glamour), and finally a multivariate regression adaptation of the Fama and MacBeth (1973) time series pooling of cross-sectional regressions. The univariate regression coefficient for GS was significantly negative. The EP, BP, and CP coefficients were all significantly positive. When Lakonishok et al. performed a multivariate regression using all four variables, they found significantly positive coefficients for BP and EP (but not CP), and significantly negative coefficients for GS. Lakonishok et al. (1994) concluded that buying out-of-favor value stocks outperformed growth (glamour) stocks during the period April 1968 to April 1990, that future growth was difficult to predict from past growth alone, that the actual future growth of the glamour stocks was much lower than past growth, relative to the growth of value stocks, and that the value strategies were not significantly riskier than growth (or ‘glamour’) strategies ex post.

Bloch et al. (1993) wrote their manuscript in 1991. At the time of the original estimation of Eq. (1), the International Institutional Estimation Brokerage Service (I/B/E/S) was only four years old, having started in 1987, and did not have sufficient data for model building and testing. The original international database consisted primarily of

earnings forecasts (means, medians and standard deviations) of large (in terms of market capitalization) securities. The domestic I/B/E/S database was created in 1976, and included a well-established body of literature on earnings forecasting analysis, testing, and portfolio construction, published by Bruce and Epstein (1994). The Bruce and Epstein volume reprinted a study reporting the ineffectiveness of earnings forecasting in creating portfolios that could generate excess returns over the period 1972–1976 (Elton, Gruber, & Gultekin, 1981). However, in the same study, Elton et al. (1981) reported the effectiveness of earnings forecast revisions in creating portfolios that are capable of generating excess returns. The Bruce and Epstein volume also reprinted a study by Hawkins, Chamberlain, and Daniel (1984) which reported large excess returns for domestic stocks, which have the largest positive monthly earnings revisions for the period 1975–1980. Wheeler (1994) developed and tested a United States-only stock strategy in which analyst forecast revision breadth, defined as the number of upward forecast revisions less the number of downward forecast revisions, divided by the total number of estimates, was the criterion for stock selection. Wheeler found statistically significant excess returns from the breadth strategy. Thus, earnings forecasts per share, earnings forecast revisions, and earnings forecast breadth had all been documented by 1994. Guerard, Gultekin, and Stone (1997) created a composite forecasting variable consisting of consensus analysts’ forecasts, forecast revisions and the breadth variables, which they referred to as a proprietary growth variable, PRGR, and reported that the composite earnings variable, when added to Eq. (1) as a ninth variable, averaged a relative weight of 33%. This result complements that of Lakonishok et al. (1994) in showing that rank-ordered portfolio returns have significant value and growth components. Guerard (1997) reported the dominance of the (same) consensus earnings efficiency variable, referred to as CTEF, relative to analysts’ revisions, forecasted earnings yields, and breadth in generating excess returns. Guerard and Mark (2003) tested analysts’ revisions, forecasted earnings yields, breadth, and CTEF using a multi-factor risk model, the Barra USE2 risk model, see Grinold and Kahn (1999) and Rudd and Clasing (1982). Guerard and Mark (2003) reported that forecasted earnings per share produced positive, but statistically insignificant, excess returns (total active returns), in the Barra system. Asset selection was negative for United States stocks in the 1990–2001 Frank Russell stock universe using forecasted earnings yields. Similarly, Guerard and Mark (2003) reported that forecast earnings revisions produced positive, but statistically insignificant, excess returns that were positively (but not statistically) associated with earnings yields. Breadth and CTEF were positively and statistically associated with earnings yields and smaller-sized stocks (another Barra risk model factor), but produced statistically significant excess returns and asset selection. Moreover, the CTEF produced larger excess returns and better asset selection than forecasted earnings yields, revisions, or breadth. Clearly, earnings forecasts were associated with United States stock returns and should be added to Eq. (1). For an excellent summary of the earnings forecasting literature, the reader is referred to Ramnath, Rock, and Shane (2008).

Momentum investing was being studied by academics at about the same time that earnings forecasting studies were being published. [Arnott \(1979\)](#) and [Brush and Boles \(1983\)](#) found statistically significant power in the relative strength. Brush and Boles' analysis was particularly valuable, because it reported that the short-term monthly price momentum model, taking the price at time $t - 1$ divided by the price 12 months ago, $t - 12$, was associated with total returns. Brush and Boles found that beta adjustments enhanced the predictive power slightly in the six- to twelve-month periods. [Fama and French's \(1992, 1995, 2008\)](#) price momentum variable uses the price two months ago divided by the price twelve months ago, thus avoiding the well-known return or residual reversal effect. We refer to it as FFPM. Fama and French's studies find significant stock price anomalies, and [Brush \(2007\)](#) and [Korajczyk and Sadka \(2004\)](#) confirm the findings after considering transaction costs. [Lesmond, Schill, and Zhou \(2004\)](#) found that price momentum returns did not exceed the transaction costs. The vast majority of studies have found that the use of three-, six-, and twelve-month price momentum variables, often defined as intermediate-term momentum variables, is associated statistically significantly with excess returns. Momentum is associated closely with excess returns in the academic literature, and should be added to Eq. (1). Further evidence on the anomalies is provided by [Levy \(1999\)](#). [Guerard, Xu, and Gultekin \(2012\)](#) added the [Guerard et al. \(1997\)](#) composite earnings forecasting variable CTEF and the Fama and French FFPM variable to the equation, to create a ten-factor stock selection model for the United States expected returns, which they referred to as the USER model.

$$TR_{t+1} = a_0 + a_1EP_t + a_2BP_t + a_3CP_t + a_4SP_t + a_5REP_t + a_6RBP_t + a_7RCP_t + a_8RSP_t + a_9CTEF_t + a_{10}PM_t + e_t, \quad (2)$$

where:

EP = [earnings per share]/[price per share]
= earnings-price ratio;

BP = [book value per share]/[price per share]
= book-price ratio;

CP = [cash flow per share]/[price per share]
= cash flow-price ratio;

SP = [net sales per share]/[price per share]
= sales-price ratio;

REP = [current EP ratio]/[average EP ratio over the past five years];

RBP = [current BP ratio]/[average BP ratio over the past five years];

RCP = [current CP ratio]/[average CP ratio over the past five years];

RSP = [current SP ratio]/[average SP ratio over the past five years];

CTEF = consensus earnings-per-share I/B/E/S forecast, revisions and breadth;

PM = price momentum; and

e = randomly distributed error term.

Table 1
Information coefficients of FSGLER variables.

(ICs)		
Variable	1997–2011 IC (t)	2003–2011 IC (t)
EP	0.037 (6.21)	0.029 (4.75)
FEP	0.040 (6.67)	0.032 (5.85)
CTEF	0.035 (10.42)	0.033 (8.68)
GLER	0.045 (6.98)	0.036 (5.41)

The USER model produced highly statistically significant active returns and stock selections. Moreover, the USER model also passed the [Markowitz and Xu \(1994\)](#) data mining corrections (DMC) test, indicating that the USER return was statistically different from the average of the approximately 21 models that were tested. We will employ the DMC test in Section 5.

[Guerard et al. \(2013\)](#) estimated Eq. (2) for all global stocks included in the FactSet database over the period January 1997–December 2011. They referred to the global expected returns model as the GLER model. The GLER model produced highly statistically significant active returns and better stock selections than the USER model over the corresponding period. That is, global stock selection models outperformed domestic stock selection models. Thus, United States investors should prefer global portfolios in order to maximize portfolio returns.

3. Building and testing stock selection models

How does one develop, estimate, and test a global stock selection model? We use the [Guerard et al. \(2013\)](#) database of global stocks included on the FactSet database during the period January 1997–December 2011. The number of stocks grows to approximately 16,000 during the period 1997–2011. [Guerard et al.'s \(2013\)](#) universe was restricted to global stocks that were covered by at least two analysts, which reduced the number to approximately 7000–8000 stocks. 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. Information coefficients, ICs, are estimated by ranking strategies and the subsequent monthly returns. We estimate one-month IC in this analysis. There is strong support for the earnings expectations variables and the fundamental variables (particularly earnings and cash flow). An objective examination of the reported ICs, as shown in [Table 1](#), leads one to identify CTEF, EP, and CP as leading variables for inclusion in stock selection models.

The results support the low P/E (high earnings yield) approach to value investing advocated by [Graham et al. \(1934\)](#) and [Graham, Dodd, and Cottle \(1962\)](#), and validate the cross-sectional return anomaly found by [Basu \(1977\)](#). They marginally (only in 2003–2011) support [Fama and French's \(1992, 1995\)](#) finding that the book-to-market ratio is an important variable for explaining the cross-section of security returns. While the EP and BP variables are significant in explaining returns, the majority of the forecasting performance is attributable to the other model variables, namely the relative earnings-to-price, relative cash-to-price, relative sales-to-price, and earnings forecast

Table 2

Top/bottom decile spreads of FSGLER variables.

Variable	Top 3 decile spreads (t)	Annualized top 3 decile spreads	Top decile spreads (t)	Annualized top decile spreads
1997–2011				
EP	0.42% (1.66)	4.43%	0.20% (0.66)	1.40%
BP	0.50 (1.67)	5.21	0.96 (1.45)	9.54
FEP	0.50 (1.95)	5.38	0.54 (2.02)	5.11
CTEF	0.72 (5.85)	8.85	1.16 (7.27)	14.54
EWC	0.70 (3.67)	8.37	1.06 (3.55)	12.36
GLER	1.12 (4.54)	13.55	1.48 (3.67)	17.19
2003–2011				
EP	0.23% (0.80)	2.267	0.25% (1.08)	2.65
BP	0.80 (1.78)	8.74	0.47 (1.57)	5.21
FEP	0.38 (1.33)	4.12	0.23 (1.39)	2.66
CTEF	1.07 (6.48)	13.37	0.58 (4.88)	7.12
EWC	0.97 (5.79)	12.07	0.60 (1.83)	7.41
GLER	1.33 (3.98)	16.32	0.96 (4.59)	11.78

variables. A regression weighting of the GLER factors, using the weighted principal components regression discussed by Guerard et al. (2013), produces a higher IC than the equally-weighted ten-factor model, denoted EWC. The consensus earnings forecasting variable, CTEF, dominates the top/bottom (one and three) decile spreads.³ See Table 2.

The regression weighting model produces higher decile spreads than the EWC model. We refer to the IC and decile spread tests as a Level I model of portfolio statistical significance. The GLER model can be an input into an optimization system for creating optimized portfolios.

4. Efficient APT portfolio construction

The mean–variance (MV59) portfolio construction and management can be summarized as:

$$\text{minimize } w^T C w - \lambda \mu^T w, \quad (3)$$

where μ is the expected return vector, C is the variance–covariance matrix, w is the portfolio weights, and λ is the risk–return tradeoff parameter. The estimation of C is usually done by a multifactor model, in which the individual stock return R_j of security j at time t , dropping the subscript t for time, may be written like this:

$$R_j = \sum_{k=1}^K \beta_{jk} \tilde{f}_k + \tilde{e}_j. \quad (4)$$

The nonfactor, or asset-specific, return on security j , \tilde{e}_j , is the residual risk of the security after removing the estimated impacts of the K factors. The term \tilde{f}_k is the rate of return on factor k . The factor model simplifies the C as the

sum of the systematic risk covariance and diagonal specific variances,

$$C = \beta C_{f,f} \beta' + \Sigma. \quad (5)$$

Accordingly, the portfolio risk is decomposed into the systematic risk and specific risk

$$\begin{aligned} \sigma_p^2 &= w' \beta C_{f,f} \beta' w + w' \Sigma w \\ &= \sigma_{\beta P}^2 + \sigma_{SP}^2. \end{aligned} \quad (6)$$

If the investor is more concerned about tracking a particular benchmark, the mean–variance optimization in Eq. (3) can be reformulated as a mean–variance tracking error at risk (MVTaR) optimization:

$$\text{minimize } (w - w_b)^T C (w - w_b) - \lambda \mu^T (w - w_b), \quad (7)$$

where w_b is the weight vector of the benchmark. One can also add equal active weighing constraints (EAW):

$$|w_j - w_{(b)j}| \leq x, \quad \text{for all } j. \quad (8)$$

The MVTaR with constraints in Eq. (8) will be referred to as EAWxTaR. The total tracking error can be decomposed into the systematic tracking error and the specific tracking error:

$$\begin{aligned} \sigma_{pTE}^2 &= (w - w_b)' \beta C_{f,f} \beta' (w - w_b) \\ &\quad + (w - w_b)' \Sigma (w - w_b) \\ &= \sigma_{\beta PTE}^2 + \sigma_{SPTE}^2. \end{aligned} \quad (9)$$

Multi-factor risk models evolved in the works of King (1966), Rosenberg (1974), Ross (1976), and Ross and Roll (1980). The different choices of factors lead to different risk model products. The domestic Barra risk model based on company fundamental data (like BP and size, as discussed in Section 2) was developed by Rosenberg (1974) and Rosenberg and Marathe (1979), and was discussed thoroughly by Grinold and Kahn (1999) and Rudd and Clasing (1982). The Barra attribution analysis is used in this analysis in order to determine the statistical significance of stock selection. The Sunguard APT model, developed by Blin and Bender, followed the Roll factor theory, and estimated more than 20 orthogonal factors based on 3.5 years of weekly stock returns data. Blin, Bender, and Guerard

³ Asness, Moskowitz, and Pedersen (2013) and Berger, Israel, and Moskowitz (2009) find that an equally-weighted composite model of 50/50 value (as measured by the book value) and momentum is effective for modeling equity returns. Larry Fisher was a long-time resident of Chatham, NJ, and lived one street away from the primary author of this study. During his 1991 CRSP seminar presentation, Fisher told Guerard that the value-based strategy of Eq. (1) was merely a test of the CRSP momentum strategy. The estimation of Eq. (2) addresses the answer to Professor Fisher's observation more completely.

Table 3
Efficient frontier of the global stock selection model with various portfolio optimization techniques 1999–2011.

APT risk model							
Earnings model or component	Mean–variance methodology	Lambda	Annualized return	Standard deviation	Sharpe ratio	Information ratio	Tracking error
GLER	M59	1000	15.84	24.97	0.590	0.78	13.11
		500	16.34	24.85	0.590	0.82	12.08
		200	16.37	24.38	0.610	0.85	12.68
		100	15.90	24.61	0.580	0.81	12.66
		5	10.11	19.36	0.440	0.51	8.81
Benchmark			5.59		0.240		
GLER	TaR	1000	16.10	21.93	0.660	0.94	11.18
		500	15.91	21.99	0.651	0.90	11.44
		200	16.09	20.95	0.691	0.97	10.83
		100	14.18	21.24	0.591	0.77	11.23
		5	8.51	20.03	0.344	0.33	8.75
GLER	EAW2aR2	1000	14.80	21.96	0.600	0.94	11.07
		500	14.30	21.65	0.590	0.80	10.87
		200	14.15	20.92	0.600	0.85	10.04
		100	13.49	20.82	0.570	0.80	9.84
		5	10.77	20.79	0.440	0.43	12.18

(1997) and Guerard (2012) demonstrated the effectiveness of Sungard APT systems in portfolio construction and management.

Let us examine the use of the Sungard APT model for creating monthly portfolios using the GLER stock selection model with the assumption of a 150-basis point transaction cost each way and an 8% monthly turnover for the 1997–2011 timeframe. We use Guerard et al.'s (2013) global FactSet universe, consisting of stocks covered by at least two analysts. As lambda rises, the expected return of the portfolio rises and the number of securities in the portfolio declines.

An increase in lambda serves to produce portfolios with higher geometric means (GM), Sharpe ratios (ShR), and information ratios (IRs). If one seeks to maximize the geometric mean of a portfolio, consistent with Latane (1959) and Markowitz (1976), then one should employ a lambda of at least 200.⁴ An examination of Table 3 shows that the Sharpe ratio is maximized by a lambda of 200, as was reported by Guerard et al. (2013). We show three sets of results in Table 3, generated by (1) MV59, (2) a practitioner's variation of MVTaR that weights the APT systematic risk at three times the importance of the specific risk, and (3) EAW2TaR, where x in the constraints equation (Eq. (8)) is set to two. The MV59, MVTaR, and EAW2TaR portfolio methodologies all show that one should never use a lambda of less than 100 if the asset manager seeks to maximize any portfolio measurement criterion. The efficient frontiers of the MV59, MVTaR, and EAW2TaR portfolios report substantial excess returns for any given level of risk. Moreover, for the GLER expected returns model, the

MVTaR optimization technique produces higher geometric means, Sharpe ratios, and information ratios than the MV59 and EAW2TaR techniques. We run a large (seemingly infinite) set of portfolio efficient frontiers, varying the ratio of systematic risk to total risk,⁵ and find that the tracking error at risk formulation is an optimal solution for the GLER data, at least for this specific time frame. However, we remind readers that there is an infinite set of portfolios that lie on or near the efficient frontier.

Guerard et al. (2013) reported that the portfolio active returns, or excess returns, of 1093 basis points were highly statistically significant (with a t -statistic of 3.85), and consisted of factor contributions (701 basis points, with a t -statistic of 3.31) and specific returns (391 basis points, with a t -statistic of 2.07) over the period 1999–2011. One refers to the decomposition of excess returns as an "attribution analysis", see Grinold and Kahn (1999). Most often, attribution analysis is created using a fundamental data-based risk model. Guerard et al. (2013) used the Axioma worldwide fundamental risk model, AX-WW2.1, or FUND, which consists of exchange rate sensitivity, growth, leverage, liquidity, medium-term momentum, short-term momentum, size, value, and volatility. The GLER model factor returns consisted of the medium-term momentum (478 basis points with a t -statistic of 6.95), value (157 basis points with a t -statistic of 5.44), and growth (77 basis points with a t -statistic of 5.21). The fundamental risk model most commonly referenced and used is the MSCI Barra GEM3 (global equity model); see Menchero, Morozov, and Shepard (2010) and Rudd and Clasing (1982) for descriptions of Barra risk model estimations. One can create an infinite number of portfolios that beat the market, once one has a statistically significant expected return input, and preferably one that is not correlated with risk factors. Wormald and van der Merwe (2012) and Shao, Rachev, and Mu (2015) are alternative formulations of optimization that

⁴ The authors believe that the use of lambdas that are less than the levels that maximize the geometric mean or the Sharpe ratio is due to investors' preferences. The authors prefer to maximize the GM and ShR criteria, even if the tracking errors are larger than those of enhanced-index strategies. They refer to the use of larger lambdas as "gird your loins" and/or "put on your big boy investment pants" strategies. One of the authors, John Guerard, refers to the use of lambdas less than 75 as being for "wimpy index huggers".

⁵ Readers may request information regarding the set of additional trade-off curves and analyses from the corresponding author, John Guerard.

are very useful in creating efficient frontiers using earnings forecasting data. We refer to the creation of portfolios with a multi-factor model and the generation of the efficient frontier as a Level II test of portfolio construction.

5. A further test of data mining corrections

In the practical world of Wall Street, it is conventional wisdom to cut your historical backtested excess returns in half; that is, if your backtested excess return (the portfolio geometric mean return less the geometric mean of the benchmark) was 6%, or 600 basis points, an investment manager and/or a client might expect 3% excess returns in the future. How do we justify this cutoff?

Markowitz and Xu (1994) proposed three statistical models for estimating the cutoff, which are close to half.

In particular, model II assumes that the modeler tests N models for T periods. Let y_{it} be the logarithm of one plus the return for the i th portfolio selection model in period t , with the form

$$y_{it} = \mu_i + z_t + \varepsilon_{it}, \quad (10)$$

where μ_i is a model effect, z_t is a period effect, and ε_{it} is a random deviation. The random deviation of the return, ε_{it} , has a zero mean and is uncorrelated with the model effect, μ_i and other model random effects. Then, the best linear estimate of the unknown μ_i is

$$\hat{\mu}_i = E\mu + \beta (\bar{r}_i - E\mu) \quad (11)$$

$$\beta = \frac{\text{cov}(\bar{r}_i, \mu)}{\text{Var}(\bar{r}_i)}, \quad (12)$$

where β is the regression coefficient of μ_i as a function of \bar{r}_i , the sample mean of method i , and $E\mu$ is the expected average performance of all models. In practice, $E\mu$ is estimated by the sample grand average

$$\bar{r} = \sum_{i=1}^N \bar{r}_i / n. \quad (13)$$

The Markowitz–Xu DMC test does not use a “holdout period”, as it can be data mined routinely 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 of Bloch et al. (1993), and reported a β of 0.59, which is 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 Eq. (2) as tilt factors. The GLER variable analysis passes the data mining corrections test criterion for the global market, indicating that the stock selection and portfolio construction methodologies produce superior returns that cannot be due to chance. The GLER variable has a data mining corrections coefficient of 0.74, and is highly statistically significant, having a F -value of 1.9. Thus, one could expect 74% of the excess returns on the GLER model relative to the average return to be continued. More importantly, the GLER model produced a higher geometric mean than an average model that could have been used to manage an equity portfolio in the global equity market over the period January 1997–December 2011.

The Markowitz and Xu (1994) data mining corrections test is referred to as a Level III test. 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 considered.

6. The alpha alignment factor: an application to global earnings forecasting

Several practitioners have decided to perform a “post-mortem” analysis of mean–variance portfolios, attempted to understand the reasons for the deviation of ex-post performances from ex-ante targets, and used their analysis to suggest enhancements to mean–variance optimization inputs, in order to overcome the discrepancy. Lee and Stefek (2008) and Saxena and Stubbs (2012) define this as a factor alignment problem (FAP), which arises as a result of the complex interactions between the factors used for forecasting expected returns, risks and constraints. While predicting expected returns is exclusively a forward-looking activity, risk prediction focuses on explaining the cross-sectional variability of returns, mostly by using historical data. Expected-return modelers are interested in the first moment of the equity return process, while risk modelers focus on the second moments. These differences in ultimate goals inevitably introduce different factors for expected returns and risks. Even for the “same” factors, expected-return and risk modelers may choose different definitions for good reasons. Saxena and Stubbs (2012) reported that the earning-to-price (E/P) and book-to-price (B/P) ratios used in the USER Model and the Axiom Risk Model have average misalignment coefficients of 72% and 68%, respectively. While expected-return and risk models are indispensable components of any active strategy, there is also a third component, namely the set of constraints that is used to build a portfolio. Constraints play an important role in determining the composition of the optimal portfolio. Equal active weighting (EWA) constraints are one type of constraint. Most real-life quantitative strategies have other constraints that model desirable characteristics of the optimal portfolio. For example, a client may be reluctant to invest in stocks that benefit from alcohol, tobacco or gambling activities on ethical grounds, or may constrain their portfolio turnover so as to reduce their tax burden.

The naïve application of the optimizations in Eqs. (3) and (7) has the unintended effect of magnifying the sources of misalignment. The optimized portfolio underestimates the unknown systematic risk of the portion of the expected returns that is not aligned with the risk model. Consequently, it overloads the portion of the expected return that is uncorrelated with the risk factors. The empirical results in a test-bed of real-life active portfolios based on client data show clearly that the above-mentioned unknown systematic risk is a significant portion of the overall systematic risk, and should be addressed accordingly.

Sivaramakrishnan and Stubbs (2013) proposed the creation of a custom risk model by combining the factors used in both the expected-return and risk models, which does not address the factor alignment problem that is due to constraints. Saxena and Stubbs (2012) proposed that the risk variance–covariance matrix C be augmented with additional auxiliary factors in order to complete the risk model. The augmented risk model has the form of

$$C_{new} = C + \sigma_{\alpha}^2 \underline{\alpha} \cdot \underline{\alpha}' + \sigma_{\gamma}^2 \underline{\gamma} \cdot \underline{\gamma}', \quad (14)$$

where $\underline{\alpha}$ is the alpha alignment factor (AAF), σ_{α} is the estimated systematic risk of $\underline{\alpha}$, $\underline{\gamma}$ is the auxiliary factor for constraints, and σ_{γ} is the estimated systematic risk of $\underline{\gamma}$. The alpha alignment factor α is the uncorrelated portion of the uncorrelated expected-return model, i.e., the orthogonal component, with risk model factors.

Saxena and Stubbs (2012) applied their AAF methodology to the USER model, running a monthly backtest based on the above strategy over the time period 2001–2009 for various tracking error values of σ chosen from {4%, 5%...8%}. For each value of σ , the backtests were run on two setups, which were identical in all respects except one, namely that only the second setup used the AAF methodology ($\sigma_{\alpha} = 20\%$). Axioma's fundamental medium-horizon risk model (US2AxiomaMH) is used to model the active risk constraints. Saxena and Stubbs (2012) analyzed the time series of misalignment coefficients of alpha, implied alpha and the optimal portfolio, and found that almost 40%–60% of the alpha is not aligned with the risk factors. The alignment characteristics of the implied alpha are much better than those of the alpha. Among other things, this implies that the constraints of the above strategy, especially the long-only constraints, play a proactive role in containing the misalignment issue. In addition, not only do the orthogonal components of both the alpha and the implied alpha have systematic risk, but the magnitude of the systematic risk is comparable to that of the systematic risk associated with a median risk factor in US2AxiomaMH. Saxena and Stubbs (2012) showed the predicted and realized active risks for various risk target levels, and noted the significant downward bias in risk prediction when the AAF methodology is not employed.⁶ The realized risk-return frontier demonstrates that not only does using the AAF methodology improve the accuracy of risk prediction, it also moves the ex-post frontier upwards, thereby giving ex-post performance improvements. In other words, the AAF approach recognizes the possibility of missing systematic risk factors and makes amends to the greatest extent that is possible without a complete recalibration of the risk model that accounts for the latent systematic risk in alpha

factors explicitly. In the process of doing so, the AAF approach not only improves the accuracy of risk prediction, but also makes up for the lack of efficiency in the optimal portfolios.

We re-examine the FactSet-based GLER database and test the usefulness of the alpha alignment factor in two applications. First, we create GLER portfolios using the Axioma world-wide statistically-based risk model and the Axioma world-wide fundamentally-based risk model, discussed in the attribution analysis.⁷ Guerard (2013) created efficient frontiers using both of the Axioma risk models, and found that the statistically-based Axioma risk model, STAT, produced higher geometric means, Sharpe ratios, and information ratios than the Axioma fundamental risk model, FUND. We report a larger set of tracking error optimizations in Table 4, with the same result. An examination of Table 4 reveals that the geometric means and Sharpe ratios increase with the targeted tracking errors; however, the information ratios are higher in the lower tracking error range of 3%–6%, with at least 200 stocks, on average, in the optimal portfolios. We find that statistically-based risk models using principal components, such as Sungard APT and Axioma, produce more efficient trade-off curves than fundamentally-based risk models using our variables.

The application of the AAF, chosen initially to be 50%, is shown in Table 4. We find that risk is underestimated substantially at higher targeted tracking errors, with the AAF producing higher Sharpe ratios and information ratios in both fundamental and statistical risk model tests, particularly in the 7%–10% targeted tracking error range.⁸

In the second AAF application, we construct portfolios with CTEF as the expected-return model and the Axioma world-wide statistical risk (STAT) and fundamental risk (FUND) models under conditions identical to those of the GLER model. We report that portfolios constructed using the STAT model dominate those constructed using the FUND model, see Table 5 and footnote 9. The STAT model procedure increases the number of securities in the optimal portfolios substantially, see Table 5. The CTEF variables

⁷ The Axioma world-wide equity risk factor model (AX-WW2.1) seeks to forecast the medium-horizon risk, or risk 3–6 months ahead. The equity risk factor, or fundamental, model uses nine style factors: exchange rate sensitivity, growth (historical earnings and sales growth), leverage (debt-to-assets), liquidity (one-month trading volume divided by market capitalization), medium-term momentum (cumulative returns of the past year, excluding the previous month), short-term momentum (last-month return), size (natural logarithm of issuer market capitalization), value (book-to-price and earnings-to-price ratios), and volatility (three-month average of absolute returns divided by the cross-sectional standard deviation). The Axioma statistical risk model estimates 20 factors using principal components.

⁸ Beheshti (2015) estimated alpha alignment factor models for 20% for CTEF, and found that AAF portfolios dominated non-AAF portfolios over 85% of the portfolio permutations. Moreover, in an examination of global portfolios using WRDS data, referred to by Xia, Min, and Deng (2015), Markowitz suggested graphing the Sharpe ratios and information ratios in order to identify possible “optimal” AAF levels. The resulting graph suggested that AAF levels of 20% and 40% were virtually identical in predictive power; all AAF levels (10, 20, 30,...90) produced better results than on-AAF portfolios. The AAF level of 50% reported for these data produces information ratios that are virtually identical to those from the AAF level of 20% reported in a follow-up study using ITG cost curves for portfolio construction over the period 2001–2011.

⁶ The bias statistic shown is a statistical metric that is used to measure the accuracy of risk prediction; if the ex-ante risk prediction is unbiased, then the bias statistic should be close to 1.0. Clearly, the bias statistics obtained without the aid of the AAF methodology are significantly above the 95% confidence interval, which shows that the downward bias in the risk prediction of optimized portfolios is statistically significant. The AAF methodology recognizes the possibility of inadequate systematic risk estimation, and guides the optimizer to avoid taking excessive unintended bets.

Table 4

Axioma WRDS global data.

GLER model												
Simulation period: Jan. 1999–Dec. 2011												
Transactions costs: 150 basis points each way, respectively												
Return model	Risk model	Tracking error	No AAF					AAF				
			Sharpe ratio	Information ratio	Ann. active return	Ann. active risk	N	Sharpe ratio	Information ratio	Ann. active return	Ann. active risk	N
GLER	STAT	4	0.448	1.247	8.72	6.99	216	0.290	1.159	4.79	4.14	516
		5	0.511	1.119	10.52	8.77	204	0.359	1.230	6.37	5.18	442
		6	0.516	1.089	11.02	10.12	188	0.397	1.145	7.43	6.49	383
		7	0.552	1.074	12.29	11.44	185	0.464	1.179	9.09	7.71	340
		8	0.605	1.111	14.14	12.73	177	0.532	1.236	10.94	8.86	304
	FUND	4	0.286	0.882	4.97	5.63	221	0.230	1.009	3.53	3.50	488
		5	0.320	0.841	5.84	6.94	199	0.269	0.971	4.45	4.59	414
		6	0.356	0.827	6.91	6.91	196	0.306	0.952	5.39	5.66	357
		7	0.414	0.885	8.45	8.45	188	0.344	0.946	6.36	6.72	318
		8	0.427	0.845	8.99	8.99	182	0.407	1.012	7.97	7.88	291

Table 5

Global earnings forecasting.

IBES global universe with at least two analysts											
Axioma statistical risk model											
Earnings model or component	Risk model	Targeted tracking error	Annualized return	Standard deviation	Active return	Active risk	Sharpe ratio	Information ratio	Number stocks		
CTEF	Fundamental	3	6.02	18.63	4.74	3.21	0.323	1.478	1423		
		4	7.64	19.03	6.36	4.16	0.402	1.578	1647		
		5	7.85	19.77	6.57	4.98	0.397	1.321	1737		
		6	9.78	20.15	8.50	5.78	0.485	1.470	1884		
		7	10.56	20.74	9.28	6.51	0.509	1.426	1978		
		8	11.71	20.64	10.43	6.90	0.567	1.511	2035		
		9	12.45	21.02	11.17	7.21	0.593	1.550	2097		
		10	12.82	20.90	11.55	7.67	0.614	1.504	2108		
		CTEF	Statistical	3	8.75	19.47	7.47	4.60	0.449	1.624	1694
				4	9.85	20.16	8.57	5.62	0.489	1.526	1740
5	11.21			20.79	9.93	6.26	0.539	1.586	1909		
6	13.67			21.04	12.39	6.84	0.650	1.817	1956		
7	14.72			21.36	13.44	7.97	0.689	1.687	2072		
8	14.81			21.22	13.53	8.44	0.698	1.602	2112		
9	16.35			21.30	15.07	8.55	0.768	1.764	2129		
10	16.11			21.05	14.83	8.56	0.765	1.733	2133		
CTEF	Statistical	3	5.69	18.28	4.41	3.39	0.312	1.302	129		
		4	7.09	18.55	5.81	4.43	0.382	1.312	73		
		5	7.48	18.22	6.20	4.61	0.411	1.343	70		
		6	8.60	18.87	7.32	5.48	0.456	1.335	70		
		7	9.13	19.21	7.85	5.73	0.474	1.370	70		
		8	11.50	19.04	10.22	6.98	0.604	1.464	70		
		9	10.79	19.15	9.51	6.54	0.563	1.454	70		
		10	10.16	19.59	8.88	7.30	0.519	1.216	70		

require more stocks than the GLER model in the Axioma simulations. We limit the number of securities to only 70 stocks each month and obtain a more investable solution that is still consistent with the risk-return tradeoff. As the tracking errors rise, the Sharpe ratios generally also rise. The information ratios tend to support the creation of portfolios with the lower tracking errors of the GLER model.

One notes an extremely import result in Table 5 with regard to the constraint on the number of stocks: the active risk of the holdings-constrained CTEF portfolios is less than the targeted tracking errors. Thus, the AAF procedure should not be used. The optimal CTEF portfolio construction technique uses the Axioma statistical risk model with constraints on the number of holdings. The Sharpe ratios

and information ratios on the CTEF model in Table 5 exceed those on the GLER model. The consensus earnings efficiency, CTEF, has dominated stock returns in global markets over the past 12 years. Earnings forecasting has been rewarded in global markets.

7. Conclusions

Investing based on analysts' expectations, fundamental data, and momentum variables is a good investment strategy in 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, as well as evidence supporting the use

