The Journal of
FIXED INCOME
VOLUME 12 NUMBER 3 www.iijfi.com DECEMBER 2002

Quantitative Forecasting Models and Active Diversification for International Bonds

ANTTI ILMANEN AND RAFEY SAYOOD

Quantitative Forecasting Models and Active Diversification for International Bonds

Antti Ilmanen and Rafey Sayood

Antti Ilmanen
is a managing director in European Fixed Income Strategy at Citigroup, London, UK.
antti.ilmanen@citigroup.com

Rafey Sayood
is a managing director in Fixed Income Research at Citigroup, London, UK.
rafey.s.sayood@citigroup.com

Extensive empirical evidence documents relatively consistent, if modest, predictability in excess bond returns and excess currency returns (see Bilson [1993] and Ilmanen [1995, 1997]). Meanwhile, theoretical analysis shows that the ability of active investors to add value on a risk-adjusted basis is proportional to their forecasting skill and the breadth or diversity of their active positions (see Grinold and Kahn [1995] and Lee [2000]). This key insight—sometimes called the fundamental law of active management—has motivated us to extend empirical forecasting exercises to new sorts of trades, such as curve steepness positioning and currency-hedged cross-country spread trading.

We review the performance of increasingly complex yet quite straightforward and transparent trading strategies. We first use single indicators to predict specific trades. We then pool these indicators into a multipredictor forecasting model for each trade, and finally diversify across several trades.

The success of these quantitative trading strategies rests on the twin pillars of the limited forecastability of returns and diversification across strategies. In predicting risky asset returns, beating 50/50 odds consistently is difficult, as most track records attest. Some systematic predictors, however, achieve 55% to 60% accuracy, or hit rates, in predicting the monthly performance of various cross-country, directional, and curve trades. We can magnify this small edge by combining several strategies into a composite portfolio with a smoother performance over time. In our backtests, the mere diversification effect turns these individually near-60% strategies into an active portfolio of trades that is profitable (beats the benchmark) about 70% of the time.

Here is a major lesson also for discretionary portfolio managers. Risk-taking along several dimensions can reduce active risk (tracking error) and result in more consistent outperformance versus benchmark. We would argue that breadth is an easier way to improve risk-adjusted performance than depth. Many macro-oriented bond investors waste the "free lunch" of active diversification by concentrating their risk excessively and inefficiently on bond market direction.

Data mining bias clouds any predictability findings. Active search for anomalies and trading opportunities in historical data can give rise to spurious or at least exaggerated findings. We try to mitigate this bias—eliminating it is virtually impossible—by requiring a reasonable economic logic and a relatively robust historical performance from each predictor. It should also help that we keep simplicity and transparency as guiding principles in development of the forecasting models. Finally, true out-of-sample experiments are always the best test.[1]

So what economic logic then constrains us from going on too blatant fishing expeditions? The academic literature on return predictability explains empirical regularities either by a rational time-varying risk premium or by market inefficiencies (often driven by systematic behavioral biases). For most trades, useful predictors include carry and value indicators that may be linked to required risk premiums as well as momentum and underreaction patterns that may be linked to behavioral biases.

I. PERFORMANCE OF QUANTITATIVE TRADING MODELS

Market-Directional and Curve Steepness Trades

Our bond market timing model is an extension of work in Ilmanen [1997]. There he shows that steep curve, high real yield, recent equity weakness, and positive bond market momentum are bullish indicators for next month's excess returns of U.S. Treasuries (over one-month money). Here we examine the predictability of the monthly excess returns of German Bunds.

We add two more predictors to the original four-predictor set: the recent trend in the Commodity Research Bureau (CRB) index, and the change in the trade-weighted exchange rate. Falling commodity prices and appreciating exchange rates signal disinflationary pressures, and should boost bond returns both contemporaneously and in the near future (given the observation that mild underreaction effects occur in many asset markets).

We graph the cumulative profits of self-financed long-short strategies since 1992, paying little attention to average profits because volatilities vary across trade types and because position sizes (and thus profits and their volatility) are scaleable. The consistency of performance is more important; it is numerically measured by the Sharpe ratio and visually captured by the smoothness of the upward slope in cumulative profit lines. These graphs show succinctly each strategy's risk-adjusted performance, including a visual summary on the frequency of underperformance and depth of drawdowns.[2]

Note that real-money managers and investors can overlay the self-financed strategies as overweight and underweight positions on top of their benchmark. Thus, they can view the average profit, volatility, and Sharpe ratio as alpha, tracking error, and information ratio, respectively.[3]

Exhibits 1A and 1B focus on market-timing rules. Exhibit 1A shows the performance of trading rules that buy or sell seven- to ten-year Bunds versus cash each month, depending on whether an individual predictor is above or below its historical (previous-decade) average. All predictors except for real yield are profitable; the value-oriented real yield strategy cumulated large losses by being short between 1995 and 1998 when Bund yields just kept rallying despite ever-lower real yields.[4] Trading based on last month's currency change (disinflation indicator) and on lagged Bund return (momentum indicator) produces the best profits.

Exhibit 1B shows the performance of the regression-based market-timing model. The model uses all six indicators to predict next month's relative performance of Bunds versus cash.[5]

The 1/-1 strategy buys or sells one unit of Bunds depending on the sign of the forecast. The scaled strategy takes into account the size of the forecast, not just its sign. For comparison, we also plot the 1/-1 strategy performance for U.S. Treasuries, which indicates higher returns but mainly due to higher profit and loss volatility.

Overall, these timing strategies were successful in capturing last decade's bull market trend, but they did not escape the bear markets in 1999 or in late 2001.

Exhibits 2A and 2B focus on the performance of duration-neutral Bund yield curve flattening trades. These trades involve selling a short-maturity bond, buying a duration-neutral (smaller) amount of a longer bond, and parking the remaining cash in one-month deposits. We predict the return on such flatteners, as opposed to simply the change in yield curve steepness, because the former measure includes the carry of the trade and not just its capital gains.

Since curve steepness is crucially driven by monetary policy, we explore various monetary policy indicators as predictors. Central banks prefer gradual rate policy adjustments, and we find that recent rate increases predict future rate increases and curve flattening. Beyond this rear-view mirror indicator, faster growth and rising inflation point to policy tightening pressures. Indeed, we find that such pressures predict future curve flattening. Finally, we include carry and value (mean reversion) indicators for the curve trade.

The duration-neutral curve positions exhibit such a low inherent P/L volatility that the Y-axis scales in Exhibit 2 are narrower than those in Exhibit 1. Exhibit 2A focuses on 2-10 flatteners or steepeners where the monthly position depends on whether the predictor is above or below its past average.

Monetary policy pressure indicators—equity market momentum, business confidence momentum, and inflation momentum—are the best predictors of next month's performance. Following the central bank's recent policy moves also turns out to be a profitable strategy. Most portfolio managers' logic of putting on curve positions based on monetary policy prospects—flatteners in a tightening

EXHIBIT 1A
Cumulative Performance of Timing Strategies Using Individual Predictors

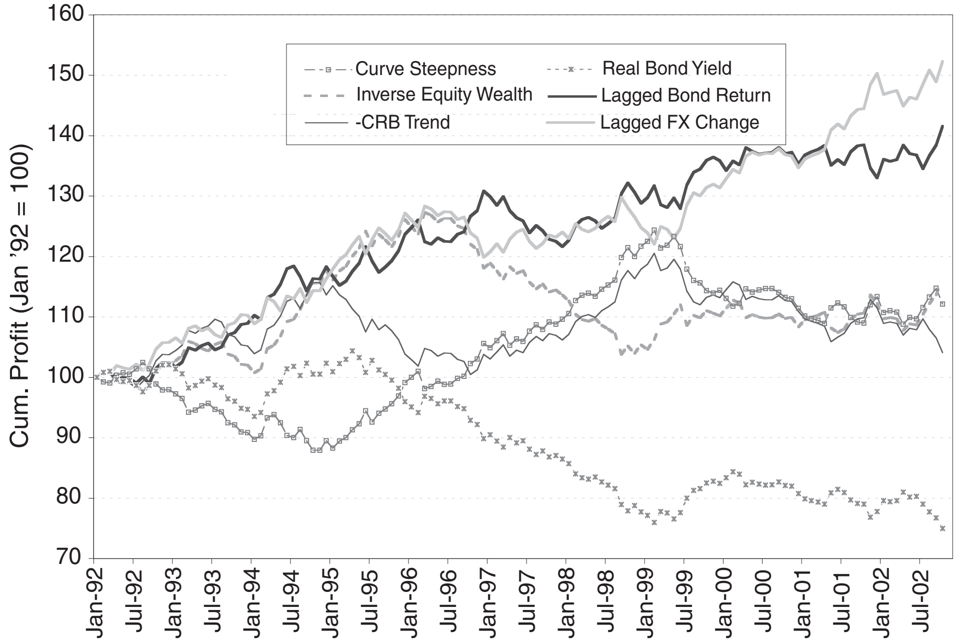
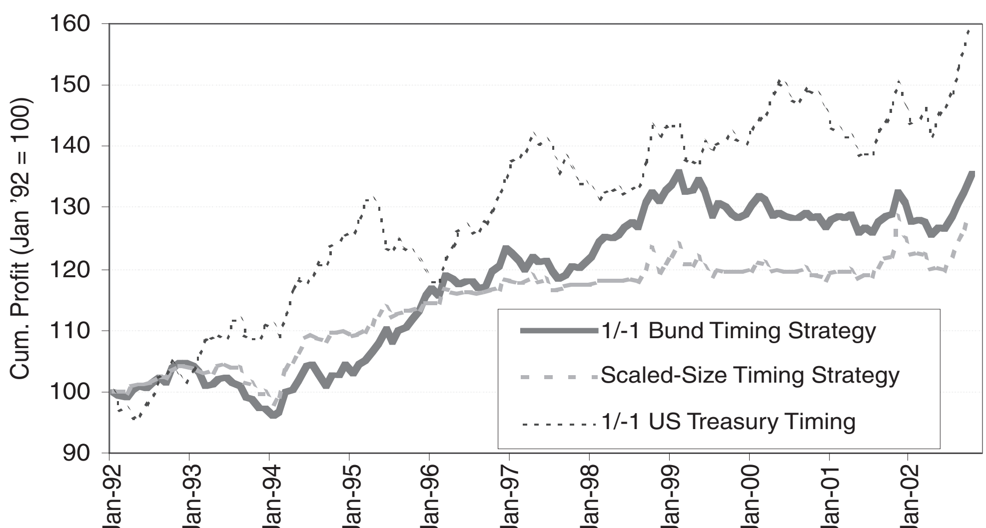


EXHIBIT 1B
Cumulative Performance of Timing Strategies Combining Six Predictors



environment and steepeners in an easing environment—would have worked well if the available data had been used systematically. Rolling yield (carry) and steepness (mean reversion) variables do not add value as single predictors—but they turn out to be useful parts of the regression-based forecasting model when included with the monetary policy pressure indicators (and with significantly positive coefficients).

EXHIBIT 2A

Cumulative Performance of Curve Trading Strategies Using Individual Predictors

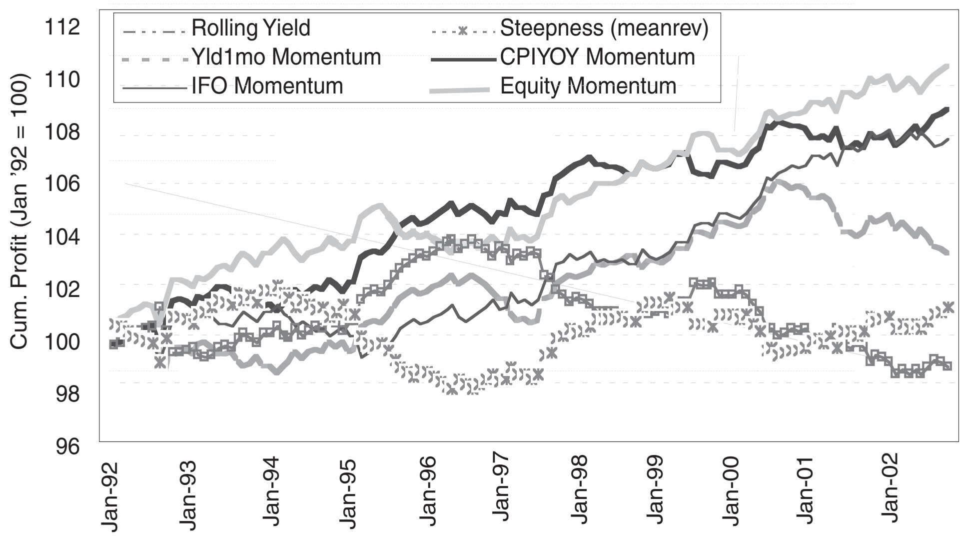


EXHIBIT 2B

Cumulative Performance of Curve Trading Strategies Using Six Predictors

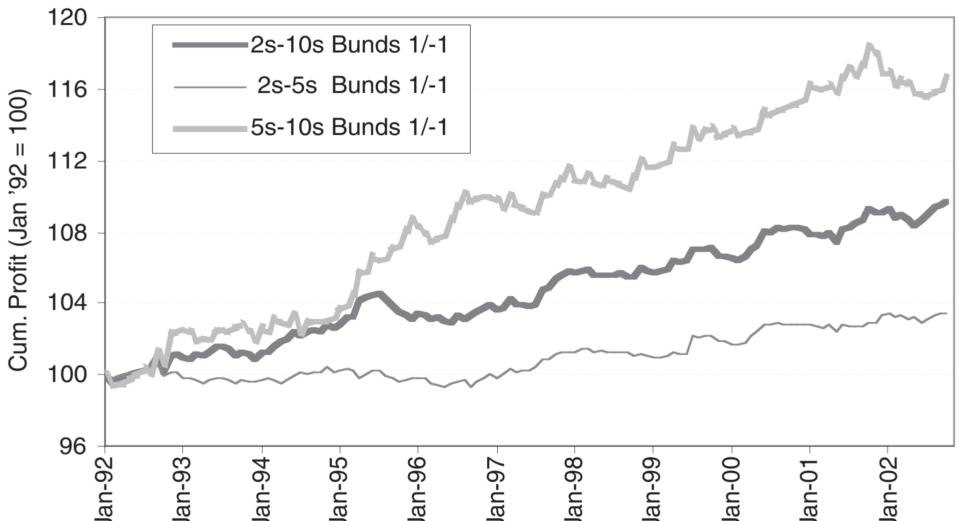


Exhibit 2B shows that the regression-based models produced relatively consistent outperformance over the past decade. We show results for several maturities. Incidentally, the fact that the 5-10 curve trade produces much higher profits than the other curve trades largely reflects the higher duration (and thus P/L volatility) in this trade; its Sharpe ratio is only modestly higher than those of the other curve trades.

Asset Allocation Trades

Exhibits 3-5 display the performance of the six-country global asset allocation models for hedged seven- to ten-year bonds, two-year swaps, and currencies. The six markets are Germany (European Monetary Union proxy), the U.K., Sweden, the U.S., Canada, and Japan. Each month, we rank the markets depending on indi-

Exhibit 3A

Cumulative Performance of Global 10-Year Bond Allocation Strategies Using Four Individual Predictors

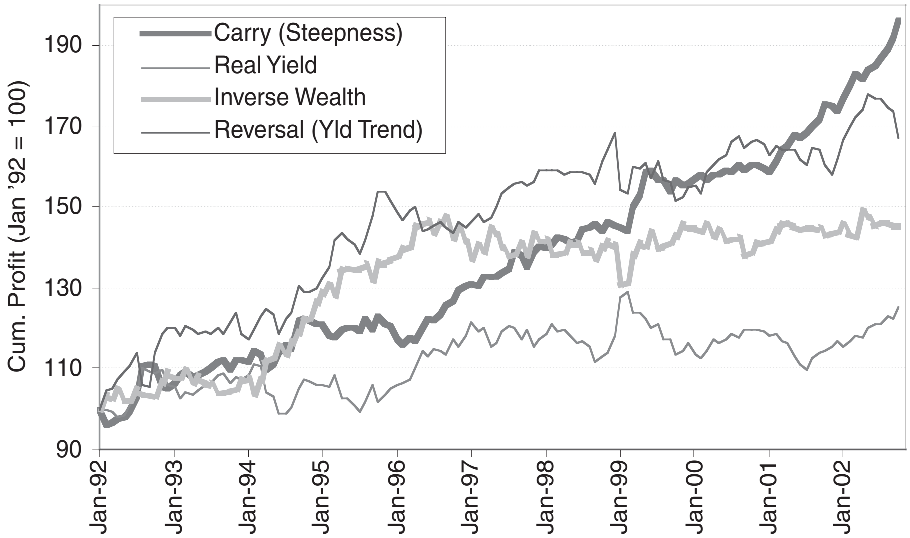
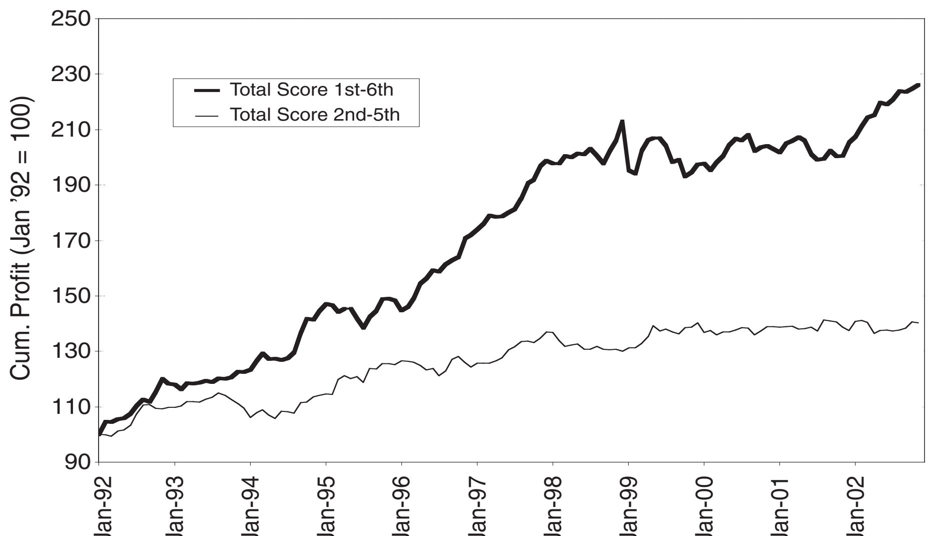


Exhibit 3B

Cumulative Performance of Global 10-Year Bond Allocation Strategies Using Total Score Rankings



vidual predictors or on their total scores. Each country's total score equals the average rank on four predictors, and thus summarizes information in several predictors.[6]

Once we have established the usefulness of certain predictors in duration trading within one country, it is natural to explore duration trading across countries (that is, currency-hedged cross-country bond trading strategies) using similar predictors. The bond allocation strategies rely on the tendency of hedged bonds to outperform their international peers if they have a steep curve, a high real yield, a weak equity market, or a recently underperforming bond market.

The first three patterns are consistent with the market-timing evidence. While most assets exhibit a trend continuation bias, we find that relative bond market movements exhibit a reversal tendency—apparent evidence of mean reversion patterns in cross-country spreads. That is, recent outperformers tend to lag subsequently, while recent laggards tend to catch up.

Exhibit 3A shows that the carry-oriented curve strategy of buying hedged bonds in the country with the

EXHIBIT 4A

Cumulative Performance of Global 2-Year Swap Allocation Strategies Using Four Individual Predictors

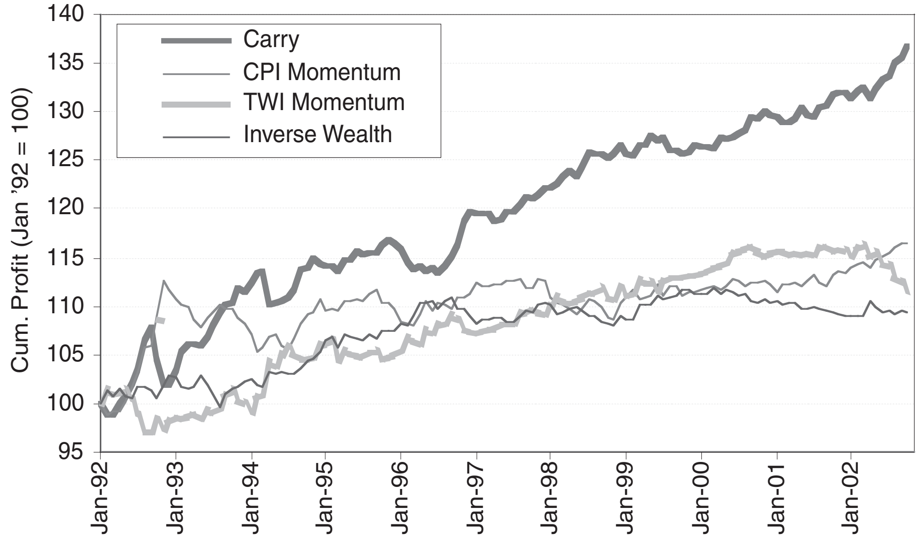
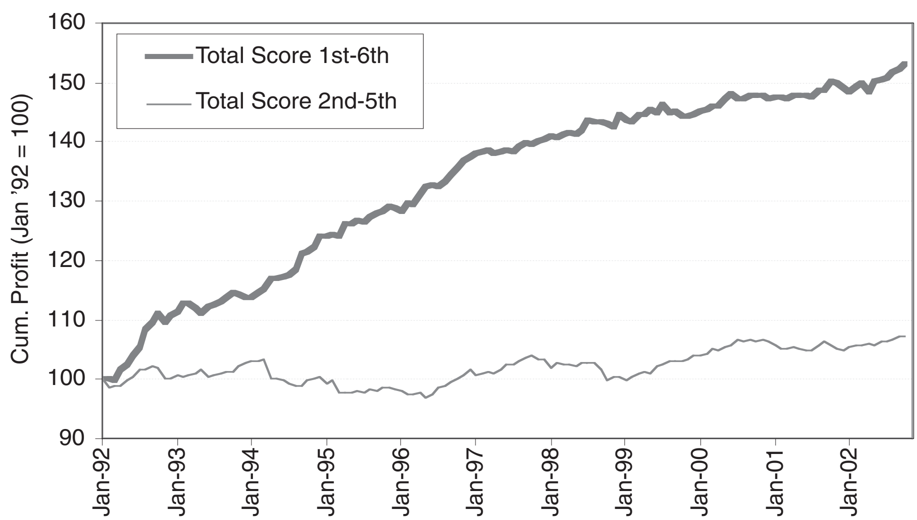


EXHIBIT 4B

Cumulative Performance of Global 2-Year Swap Allocation Strategies Using Total Score Rankings



steepest curve and selling bonds in the country with the flattest curve produces stellar profits year after year. The Sharpe ratio (1.1) is the best among all the single-indicator strategies, and its performance has only improved over time. The trading rules based on the other three indicators have stalled in recent years (Sharpe ratios near 0.2 in the past five years, compared to 1.8 for the carry strategy).

Exhibit 3B shows that when we pool the information in four predictors into total scores (each country's average rank across the four criteria), the strategy of buying the top-ranked market against selling the bottom-ranked market has produced relatively consistent outperformance. The strategy has exceptional results in 1996-1997, after which its performance stabilizes until the beginning of 2002. The second-best strategy (buying the second-ranked market against selling the fifth-ranked) also has been profitable but much less so than the first-best strategy.

Exhibits 4A and 4B review the performance of the global asset allocation indicators for two-year swaps. The results are virtually identical if we use currency-hedged one to three-year government bonds instead.

EXHIBIT 5A

Cumulative Performance of Global Currency Allocation Strategies Using Four Individual Predictors

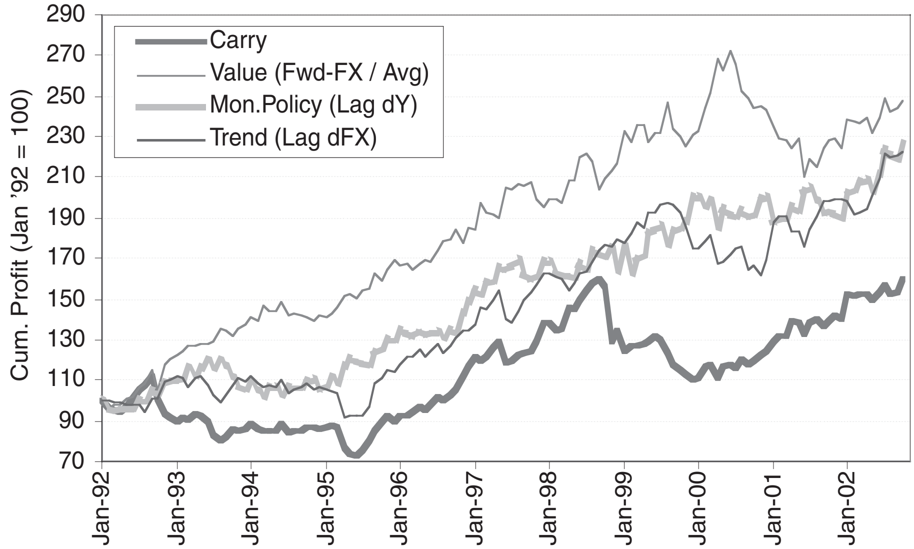
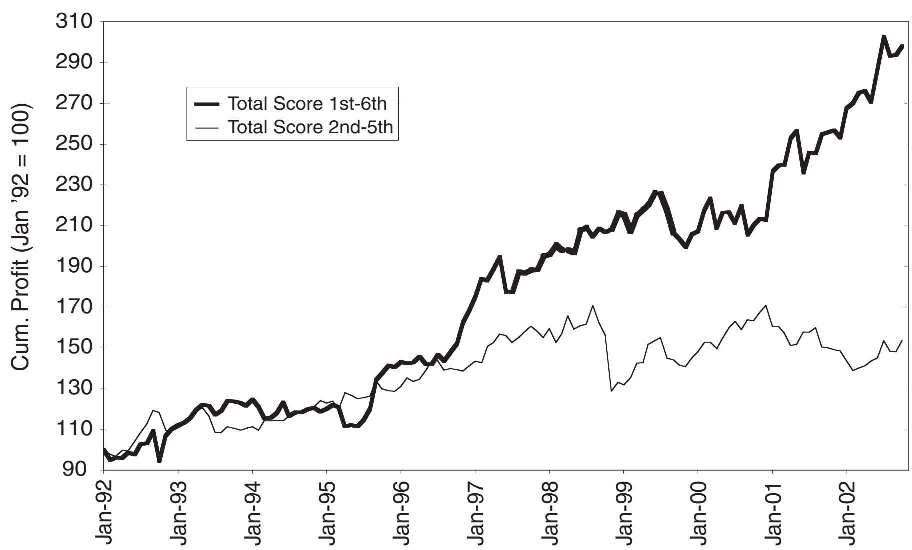


EXHIBIT 5B

Cumulative Performance of Global Currency Allocation Strategies Using Total Score Rankings



It is natural to use similar indicators as in the ten-year bond allocation, but we made some changes, given our knowledge that front-end carry strategies offer especially consistent profits and that monetary policy is a particularly important driver of short-maturity rates. Indeed, the carry-oriented strategy based on front-end steepness has had by far the best performance, while the other indicators that involve receiving 2s in countries where inflation/growth pressures favor monetary policy easing have added some value.

The first total score strategy that combines information in the four indicators (with a double weight for carry) has been an extremely consistent outperformer. The large difference compared to the performance of the second-best total score strategy suggests that the superb consistency may be partly sample-specific.

The intuition behind the total score strategy is appealing. The ranking favors countries where markets discount monetary policy tightening (steep yield curve between one month and two years), while other evidence—falling inflation, appreciating currency, and weak equities—suggests that the central bank should *not* tighten the policy.

Moving from interest rate risk to currency risk, we buy and sell one currency of the six, depending on four predictors and the combined total score. Monthly currency returns include capital gains from exchange rate moves as well as carry income from one-month deposits.

Exhibit 5A shows that all currency allocation strategies produce large profits but with significant volatility (note that the Y-axis scale is wider than in Exhibits 3 and 4). Carry-seeking and trend-following strategies are the best-known currency trading strategies, but we also include a value strategy (buy currencies whose ten-year forward exchange rates are cheap versus the historical average, and sell similarly rich currencies) and a policy tracking strategy (buy rising-rate currencies and sell falling-rate currencies).

The value strategy performs best, except for a deep drawdown in 2000 (losses from prematurely overweighting the undervalued euro). The trend and policy strategies also produce reasonably consistent profits. The carry strategy experienced sharp losses between 1992-1995 and 1998-1999, but had great runs in 1995-1998 and 2000-2001, when the low-yielding yen depreciated.

The first total score strategy in Exhibit 5B exhibits surprisingly small drawdowns, and is the only strategy with double-digit annual average returns since 1992. The second-best strategy performs initially as well as the first one, but has lagged in 2001-2002.

II. DIVERSIFYING ACROSS STRATEGIES—KEY TO CONSISTENCY

We have shown each predictor's usefulness as a trading signal on a stand-alone basis, and then pooled information in several predictors into a regression forecast or total score ranking. The final step is to combine several trades into a composite strategy so as to achieve diversification gains. Here we take the simplest possible approaches in constructing the active portfolio of trades—equal nominal weights or equal P/L volatilities.

Pooling information across indicators does boost profit consistency, but only modestly. Among the single-indicator strategies in Exhibits 1A through 5A, the typical (median) hit rate is 57%, and the Sharpe ratio is 0.55. For the strategies in Exhibits 1B through 5B that summarize several predictors' information into a forecasting model or a total score ranking, the typical hit rate rises to 60% and the Sharpe ratio to 0.8. Exhibit 6 summarizes the comparisons.

The advantage of diversifying across several strategies appears much greater. The hit rates rise to between 60% and 70%, and Sharpe ratios rise to between 0.9 and 1.7. (Among the single-indicator strategies, only the carry-oriented bond and swap allocation strategies across countries achieve a Sharpe ratio of 1.0.)

Exhibits 7A and 7B graph the performance of various composite trading strategies, displaying in general much smoother P/L cumulation than in Exhibits 1-5. In Exhibit 7A, the two currency allocation strategies give the highest cumulative returns, and the three Bund curve steepness trading strategies the lowest returns. The apparent difference in performance mainly reflects the inherently high volatility of currency trading and low volatility of curve trading. Interestingly, both composites have similar Sharpe ratios (near 1.0).

Exhibit 7B focuses on two broad composites that we highlight as *flagship strategies*. The "old" flagship strategy of buying top-two bond markets and top-two currencies against the bottom two (according to the total score rankings) cumulates the higher returns. This four-trade composite is less diversified and more volatile than desired, however, because equal-money weighting makes the composite's P/L uncomfortably highly correlated with the volatile dollar-yen currency carry trade. Indeed, this composite has a lower Sharpe ratio than our "new" flagship strategy (1.4 versus 1.7), despite higher absolute returns (that reflect a focus on inherently high-volatility strategies).

The new flagship composite includes eight active strategies each month. The strategies are sized so that all have similar P/L volatilities.[7]

1. A market-timing position—long- or short-duration position based on the regression model signal.
2. A Bund curve flattener or steepener position between 2s and 10s depending on the regression model signal.
3. A euro swap curve flattener or steepener position between 5s and 10s depending on the signal from a similar regression model.
4. and 5. Two cross-country trades between currency-hedged ten-year bonds depending on the total score ranking.
6. and 7. Two cross-country trades between currencies depending the total score ranking.
8. A cross-country trade in two-year swaps based on the total score ranking.

EXHIBIT 6

Forecasting Performance Statistics, January 1992-September 2002

BUND MARKET TIMING

	Steepness	Real Yield	Inv.Wealth	BondLag	CRB Mom	FX Lag	1/-1 Bund	Scaled-Size Bund	1/-1 Treasury
Average P/L	1.17	-2.58	1.50	3.34	0.48	4.02	2.93	2.46	4.68
P/L volatility	4.57	4.52	4.56	4.48	4.58	4.44	4.51	3.97	5.75
Sharpe ratio	0.26	-0.57	0.33	0.74	0.10	0.91	0.65	0.62	0.81
Hit rate	0.54	0.40	0.53	0.60	0.49	0.57	0.57	0.57	0.62

BUND 2s-10s STEEPNESS TRADING

	Carry/Roll	Steepness	Rate Mom	CPI Mom	IFO Mom	Equity Mom	2s-10s 1/-1	2s-5s 1/-1	5s-10s 1/-1
Average P/L	-0.12	0.11	0.30	0.81	0.71	0.95	0.87	0.31	1.44
P/L volatility	1.06	1.06	1.06	1.04	1.04	1.03	1.03	0.72	1.52
Sharpe ratio	-0.11	0.10	0.28	0.78	0.68	0.93	0.84	0.43	0.95
Hit rate	0.50	0.50	0.56	0.59	0.57	0.62	0.60	0.57	0.60

GLOBAL (Hedged 10-Year) BOND ALLOCATION

	Carry	Real Yield	Inv.Wealth	Reversal	Total Score Ranking 1st-6th	Total Score Ranking 2nd-5th
Average P/L	6.44	2.37	3.70	5.09	7.75	3.28
P/L volatility	5.74	7.37	6.36	7.67	6.14	4.77
Sharpe ratio	1.12	0.32	0.58	0.66	1.26	0.69
Hit rate	0.64	0.53	0.60	0.58	0.67	0.57

GLOBAL CURRENCY ALLOCATION

	Carry	Value	Mon.Pol.	Trend	Total Score Ranking 1st-6th	Total Score Ranking 2nd-5th
Average P/L	5.29	9.09	8.17	8.13	11.01	4.65
P/L volatility	13.45	11.25	10.52	11.60	12.77	11.30
Sharpe ratio	0.39	0.81	0.78	0.70	0.86	0.41
Hit rate	0.63	0.58	0.59	0.62	0.65	0.57

GLOBAL 2-Year SWAP ALLOCATION

	Carry	CPI Mom	TWI Mom	Inv.Wealth	Total Score Ranking 1st-6th	Total Score Ranking 2nd-5th
Average P/L	2.96	1.45	1.04	0.85	4.00	0.69
P/L volatility	2.82	2.75	2.32	2.00	2.24	2.24
Sharpe ratio	1.05	0.53	0.45	0.42	1.78	0.31
Hit rate	0.65	0.54	0.56	0.53	0.68	0.60

COMPOSITE STRATEGIES

	2 Timing	3 Curve	2 Bond Alloc	2 FX Alloc	2 Swap Alloc.	Flagships New 8	Flagships Old 4
Average P/L	3.80	0.88	5.71	7.95	2.47	4.10	6.78
P/L volatility	4.22	0.92	4.30	8.38	1.60	2.38	4.97
Sharpe ratio	0.90	0.96	1.33	0.95	1.55	1.72	1.36
Hit rate	0.60	0.61	0.66	0.67	0.69	0.67	0.68

CRB: Commodity Research Bureau.
IDF: Institut für Wirtschaftsforschung.
Mom: Momentum.

Both broad composites have provided smooth profit growth over the past decade, making money virtually every year. The 1998-1999 period—reflecting the Russian/LTCM crisis and Fed tightening/pre-Y2K fears—was the most challenging for these strategies, while the period from mid-1995 to mid-1998 was the best.

Exhibit 7A
Cumulative Performance of Composite Trading Strategies

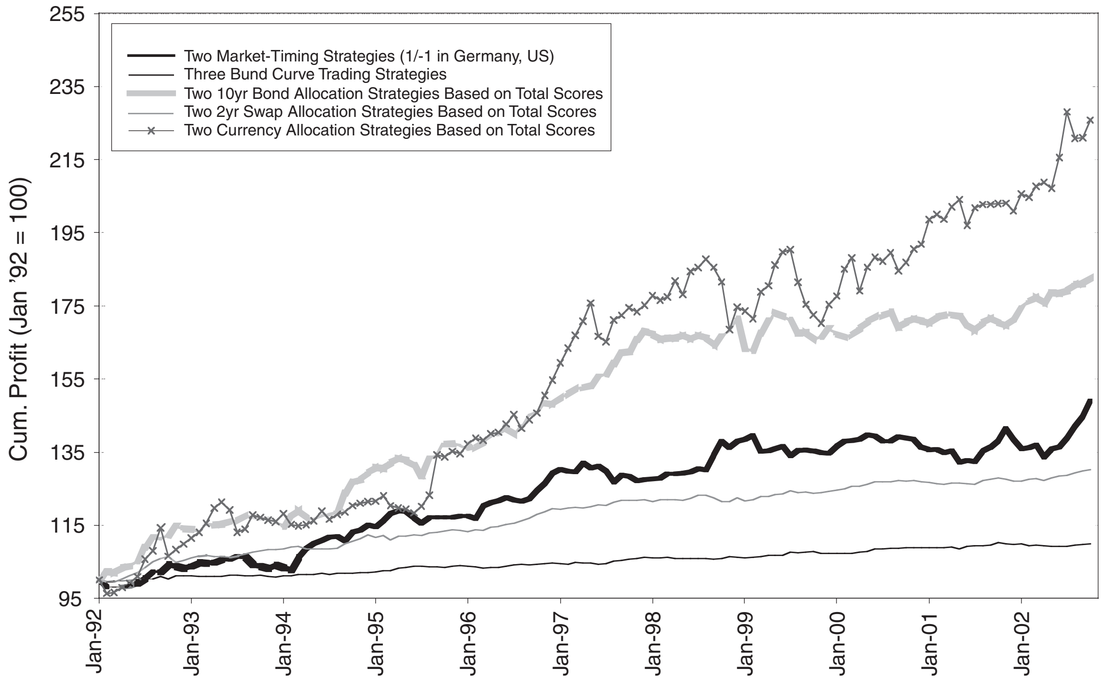
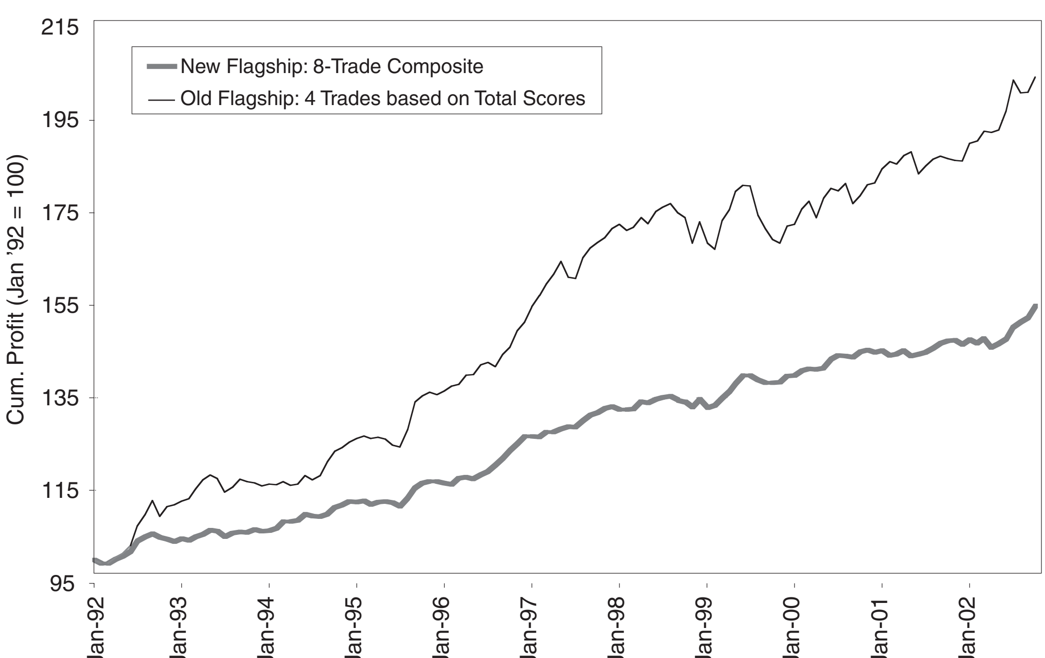


Exhibit 7B
Cumulative Performance of Diversified Composite Trading Strategies



III. WHAT CAN BE DONE BETTER?

Over the years, we have explored several ways to improve the performance of quantitative strategies. Although various regularities have not yet been arbitraged away, it is essential for future success to try constantly to sharpen one's investment edge. Here are some avenues.

- *Smarter Indicator Weightings.* Surprisingly, the smarter weighting path does not appear very promising. The simplest ways to combine information have typically given as good long-run performance as smarter ways (e.g., weighting indicators by signal strength or past performance). Estimation risk may be the culprit. Smarter approaches may try to squeeze more information out of the limited predictability than there is; conversely, simpler approaches may give more robust signals.
- *Adding New Predictors.* It is not easy to find variables that enhance the performance of the forecasting models, but one can keep trying.
- *Refining the Regularities.* Second-generation models assess when individual predictive relations that work, on average, are more likely to excel or fail. The best-known regime effect is the tendency of carry strategies to underperform when market risk aversion indicators are high or rising. One can also try to identify structural breaks that end specific regularities. The search for such refinements, however, adds to data mining risks.
- *Improving Breadth by New Trading Rules*. Adding new dimensions of risk-taking is perhaps the most promising path for further enhancing risk-adjusted returns, given the evidence on diversification gains.
- *Smarter Ways of Combining Trades.* Instead of assigning equal P/L volatility to each trade, we could take into account correlations across trades and measure each trade's contribution to active portfolio risk. Moreover, we could take into account expected return differentials across trades (again, based on signal strength or forecasting track record). This is the intuition behind the logic of many optimizers' construction of portfolios of active trades—trading off marginal alpha and marginal tracking error. While the theory is appealing, we again find little advantage to smarter approaches (over using equal P/L volatilities), perhaps because of estimation risk.

ENDNOTES

This article is largely based on research reports written for Schroder Salomon Smith Barney by Antti Ilmanen and Rafey Sayood between 1998 and 2002.

[1]What is perceived to be out-of-sample often is in the eye of the beholder. We have used actual trading capital since early 2001 for the eight-trade composite we describe. This is a pure out-of-sample exercise, and the performance has been in line with back-testing results, with a Sharpe ratio well over 1.0. In addition, all the timing and curve strategies are based on quasi- out-of-sample regressions, in the sense that new regressions are estimated each month according to predictors that are available at the beginning of the month. Thus, investors could realistically have known each predictor's weight in the regression, as well as the predicted next-period return, if they had prespecified the same predictor set.

[2]No trading costs are deducted but they are small, given the liquidity of the traded assets and relatively infrequent trading.

[3]The constraints many real-money investors face can have some performance implications. When spread strategies involve underweighting smaller markets, such as Canada or Sweden, a short sales restriction implies that such markets can be underweighted only up to their index weight. Leverage restrictions mean that spread strategies that have inherently low volatility cannot be scaled up to have profit and loss volatility similar to inherently high-volatility trades. These constraints matter in practice only if the investor's tracking error target is relatively aggressive.

[4]Despite negative evidence in the pairwise analysis, real yield may have a positive role when used in conjunction with other predictors in our forecasting model. Indeed, while the simple correlation between real yield and subsequent excess bond return is typically negative during our sample period, the partial correlation in the six-predictor forecasting regression is consistently positive.

[5]Actually, we run these regressions for the three G-3 markets together while constraining the coefficients to be equal across countries. Using so-called seemingly unrelated regression estimation makes the regression coefficients more stable over time, and keeps the coefficient signs positive. The estimation is repeated each month using the previous decade's data. The fitted value of the estimation is the one-step-ahead forecast of the excess Bund return; global regression coefficients are multiplied by current German indicator values.

[6]With one exception (the carry for two-year swap allocation), we assign virtually equal weights to the indicators. As a tie-breaker rule that ensures an unambiguous total score ranking each month, we give slightly greater weights to indicators that performed better in the past. If we assign exactly equal weights to indicators, two countries would occasionally have the same total scores.

[7]When we combine different strategies into the composite portfolio of equal-volatility trades, we scale down the size of the inherently more volatile currency trades and scale up the size of the less volatile curve trades. For example, our currency allocation trades involve half the nominal amount in bond allocation trades and less than a tenth of the amount in the curve steepener trades. Here all strategies are sized to have about the same 6% annual P/L volatility as the global bond allocation trades. While the volatility-weighted eight-trade composite has the same average profit as the average of eight individual trades (after volatility scaling), the composite's volatility is less than half the average volatility of the eight individual trades. The resulting doubling of the Sharpe ratio is the beauty of diversification—and a sign that the strategies are not highly correlated with one another.

REFERENCES

Bilson, John. "Value, Yield and Trend: A Composite Forecasting Approach to Foreign Exchange Trading." In Andrew Gitlin, ed., *Strategic Currency Investing*. Chicago: Probus Publishing, 1993, pp. 366-403.

Grinold, Richard, and Ronald Kahn. *Active Portfolio Management*. New York: McGraw-Hill, 1995.

Ilmanen, Antti. "Forecasting U.S. Bond Returns." *The Journal of Fixed Income,* 7 (June 1997), pp. 22-37.

——. "Time-Varying Expected Returns in International Bond Markets." *Journal of Finance*, 50 (June 1995), pp. 481-506.

Lee, Wai. *Theory and Methodology of Tactical Asset Allocation*. New Hope, PA: Frank J. Fabozzi Associates, 2000.