Forecasting T-Bill Yields: Accuracy *versus* Profitability?

by

Roy Batchelor

City University Business School

July 1998

Abstract: *Several studies have found little relation between the accuracy of financial forecasts and the profitability of trading rules based on these forecasts. This paper uses a large panel of published forecasts of US Treasury bill yields to test this proposition, and a number of related hypotheses. In our panel, there are significant correlations between all accuracy measures and trading profits. We show that the paradoxical results of earlier studies appeared because they used much smaller samples of forecasts dominated by two unrepresentative members - the average "consensus" forecast (very accurate but only moderately profitable), and the spot-based "random walk" forecast (even more accurate but even less profitable).*

Acknowledgments: Versions of this paper were presented to the International Symposium on Forecasting, Stockholm, and the UK Money, Macro and Finance Research Group Conference, St. Andrews. The author would like to thank Bob Eggert, founder of the Blue Chip forecasting service, and Randell Moore, Editor of Blue Chip Financial Forecasts, for their cooperation in assembling the data used in the study; and Herman Stekler for valuable comments on the initial draft.

Keywords: Interest Rates, Forecasting, Financial Futures.
JEL Class No. : C52, E47

✉City University Business School,
Frobisher Crescent, Barbican, London EC2Y 8HB, UK

☎Tel (-44) 71-477 8733 💻Fax (-44) 71-477 8881 r.a.batchelor@city.ac.uk

This paper studies the accuracy of 13500 forecasts of 3-month Treasury bill yields made by a panel of professional US forecasters, the profitability of futures trading rules based on these forecasts, and the relationship between forecast accuracy and profitability.

Published forecasts of the 3-month T-bill yield have already been much analyzed, with consistent results. In terms of mean absolute and mean squared errors, average "consensus" forecasts of groups of professional forecasters are more accurate than the majority of individual forecasters. But even the consensus is less accurate than the "random walk" forecast, which suggests that the best predictor of future spot yields is simply the current spot yield (see for example Bowlin and Martin, 1975; Belongia, 1987; Hafer, Hein and MacDonald, 1992). Consensus forecasts are also less accurate than the forecasts implied by prices of nearby T-bill futures contracts (Friedman, 1979; Hafer and Hein, 1989; MacDonald and Hein, 1989). Some individual forecasters do outperform the consensus, and occasionally even the random walk. But Kolb and Stekler (1996) show that differences in accuracy across one representative group of US forecasters - contributors to the *Wall Street Journal* surveys - are not statistically significant, so any out-performance is more likely due to chance than skill.

For financial forecasts, monetary gain is at least as relevant as mean square error as a criterion for a successful forecast. If the process driving T-bill yields is close to a random walk, as the above results on forecaster accuracy suggest, Acar (1993, p92) has shown that the linear correlation between mean square error and trading returns will be significantly negative. Its exact value will depend on the relative variances of the forecast and target variable, and the trading rule adopted. For example, in the case where these variances are equal, and equal long/ short futures positions are taken according as the forecast yield is below/ above the yield implied by the T-bill futures price, we would expect a correlation of -0.56. A formal proof is given in the Appendix below.

The few published empirical studies of the profitability of financial forecasts have produced quite different results. Leitch and Tanner (1991, henceforth LT) showed that there was *no* significant correlation between the RMSE of a set of forecasts of the US T-bill yields in the 1980s, and the profitability of T-bill futures trading rules based on these forecasts. Comparable studies of the foreign exchange market have similarly found little correlation between the accuracy and the profitability of exchange rate forecasts (Boothe, 1983; Boothe and Glassman, 1987). Leitch and

Tanner (1995) find almost no correlation between the RMSE and the directional accuracy (which they equate to profitability) of US GNP and inflation forecasts.

Although influential, these papers suffer from several limitations which we aim to remedy here. The first is that they often use data with a short forecast horizon, an important consideration because most exploitable inefficiencies in futures markets occur in the more distant contracts. The much-researched *Goldsmith-Nagan Bond and Money Market Letter* (described in Friedman, 1979, and Dua, 1988) asks contributors only for 3- and 6-month-ahead interest rate forecasts, and the twice-yearly *Wall Street Journal* surveys of financial analysts used by Belongia (1987) and Kolb and Stekler (1996) asks for 6-month forecasts. This problem does not affect the LT study, however, since their source - the Commonwealth Research Institute's *Money Rate Report* - publishes forecasts up to 12 months ahead.

A further problem is that most studies look only at the performance of the consensus forecast and a small number of market-based or model-based alternatives. It is important to establish whether results generated from these small and incomplete panels generalize to the larger population of forecasts. This problem does affect the LT(1991) study, since they include only one individual forecaster alongside the consensus, the forward rate, the random walk, a naive extrapolative forecast, and an AR(2) model. The largest consistent panel of interest rate forecasters is the group of nine assembled by Belongia (1987). Kolb and Stekler (1996) use data from a much larger group, but the frequent changes in the *Wall Street Journal* contributors made it impossible for them to analyze individual track records.

The data used in our study come from another well-known but under-researched source, the monthly *Blue Chip Financial Forecasts* newsletter. From this we extract forecasts of 3-month Treasury bill yields for a panel of 25 individual forecasters who contributed regularly to the service. The forecasts were made in the years 1983-92, for the current and four successive quarters. This yields a database very much larger than those in the studies cited above, with more genuine *ex ante* forecasts, covering a longer time period and a greater variety of forecast horizons.

The first section below describes the data in detail. The second section looks at the accuracy of individual and consensus forecasts. The third section assesses the profitability of trading on the forecasts, and the relationship between accuracy and profitability.

Some of our findings support conventional wisdom. The spot rate, for example, is consistently more accurate than the consensus forecast, and almost all individual forecasts, at all horizons. However, on many other points we disagree. We find that, except at the very shortest horizon, there are significant differences in the accuracy of individual forecasters. Although the futures price for nearby contracts is indeed more accurate than the consensus forecast and most individual forecasters, this is not true for horizons of 6 months and more, so that the more distant futures contracts are not efficiently priced. As a result, buying and selling these T-bill futures according as the forecast yield is below or above that implied by the futures market generates excess profits to almost all forecasters. Finally, and contrary to the results of Leitch and Tanner (1991), the level of these profits proves to be significantly correlated across individuals not only with the directional accuracy of their forecasts, but also their root mean square errors. The concluding section shows that the apparently paradoxical results of LT, and Boothe (1983) and Boothe and Glassman (1987), are most likely due to the small and unrepresentative samples they used.

I. BLUE CHIP FINANCIAL FORECASTS

The Blue Chip Financial Forecasts service began in late 1982. Around the 23rd day of each month, the service conducts a telephone survey of the interest rate forecasts, and forecasts of related economic variables, made by a panel of economists and financial analysts in major US commercial banks, stockbrokers, insurance companies, corporations, and financial consultancies. The forecasts for each variable are tabulated and the consensus forecast calculated as the arithmetic average of the individual figures. The full results are mailed back to participants and subscribers on the 1st day of the following month, which we term the "survey month". Summary results are also regularly reported by the *Wall Street Journal.*

The 3-month Treasury bill yield (auction average, investment or bond equivalent yield) has been included in the survey from the outset. All forecasts are for quarterly averages and are made for five horizons, the current and four successive quarters. So in the survey months January, February and March, forecasts are for average interest rates in Q1, Q2, Q3 and Q4 of the current year, and Q1 of the following year. In April, May and June, forecasts are for Q2, Q3, and Q4 of the current year, and Q1 and Q2 of the following year; and so on. This means that each forecaster makes 15

predictions of the average T-bill yield in any particular quarter. Forecasts for the average T-bill yield in 1988Q1, for example, first appeared in the January 1987 newsletter, and were revised and published every month until March 1988.

The number of panelists has stayed at around 50 throughout the history of the service. But the composition of the panel has changed over time as some forecasters have retired and other have joined, and of course not all panelists provide forecasts in every month. This means that the longer the time period considered, the smaller will be any panel of regular contributors. In this study we confine our attention to the 123 monthly surveys in the period September 1982 - December 1992, which contained forecasts for the 36 target quarters in the period 1984Q1 - 1992Q4. This allows us to construct a panel of 25 regular forecasters, defined as those who (a) missed at most 6 of the monthly surveys, *and* (b) were individuals - in one case a team of two individuals - who contributed throughout the period, possibly changing institution (13 cases), *or* (c) were institutions who contributed throughout the period, and changed forecasting staff no more than twice (12 cases). Our raw data thus consist of 15 monthly forecasts x 36 target quarters x 25 forecasters = 13500 forecasts. The small number of missing values in this matrix have been proxied for each forecaster by linear interpolation between the forecasts made in neighboring months.

A frequent criticism of surveys such as the *Bond and Money Market Letter*, which publish only consensus forecasts, is that contributors have no reputational incentives to produce good forecasts. This is not the case with the Blue Chip panel, where individual forecasts are published and attributed. Since January 1993, forecasters names have been published alongside their forecasts. Previously the Blue Chip forecasters were identified by letter codes (such as A, AA, B, BB,...) which changed occasionally as forecasters changed affiliation and new institutions were substituted for old. A list of all contributors was also published each month and most, but not all, contributors agreed to have their current letter code printed alongside their name, so that their forecasts could be identified. Of our 25 regular forecasters only one sought anonymity, and even in this case it is possible to infer his identity since in several months he was the only anonymous contributor. The forecasters and their institutions are listed in Table 1. Batchelor and Dua (1992) test the rationality of these individual forecasters and forecasting institutions.

In order to generate rules for futures trading, we need to translate the quarterly average forecasts into point forecasts for the delivery days of the IMM 90-day Treasury bill futures contracts. These

fall on various dates in March, June, September and December each year, depending on the timing of the T-bill auction. The simplest approach is to assign the quarterly average forecasts to the mid-point of each quarter, and to make a linear interpolation between the average forecast for the quarter in which the futures contract is delivered, and the forecast for the immediately succeeding quarter. We also experimented with various nonlinear interpolation schemes, but these did not produce significant changes in the forecasts, and no changes in the crucial forecaster rankings used below. Since in each month the Blue Chip forecasters predict five quarters ahead, interpolation allows us to forecast the T-bill yields implied by the price of the nearby futures contract and the next three futures contracts, with effective horizons running from 1-month (e.g. the forecast of the current year MAR contract made on 23 Feb and published on 1 March of the current year) up to 12 months (e.g. the forecast of the current year MAR contract made on 23 March of the previous year, and published on 1 April of the previous year). The data set then reduces to 12 monthly forecasts x 9 years x 4 futures contracts = 432 forecasts for each of the 25 forecasters, making in all 10800 forecasts for futures delivery dates.

II. ACCURACY

Two conventional criteria are used to judge the accuracy of these forecasts. One is directional accuracy (DA), measured by the proportion of forecasts which correctly predict the direction of change in the T-bill yield between the forecast date and the futures delivery date, when spot and futures yields converge. The other error metric is the root mean square error (RMSE) of the forecasts. We also computed the mean absolute error (MAE), but forecaster ranks by MAE were virtually identical to those by RMSE, and so are not reported here.

Benchmarks

We compare the performance of individual forecasters on these measures with a number of benchmarks. One is the consensus forecast, defined here as the arithmetic average of the 25 individual forecasts. This differs a little from the consensus forecasts published by the Blue Chip service, since our panel includes only the most regular forecasters, just over half the total respondents in a typical month. In terms of overall accuracy there is no significant difference

between the two consensus measures, and our main concern here is to assess the effect of aggregation on forecast performance.

The other benchmarks are the current spot rate, and the rate implied by current futures prices for the relevant target month. In the context of the Blue Chip survey, there are two interpretations of what should count as a "current" interest rate. One possibility is to use rates on the 23rd of the month preceding the publication of the survey, or the most recent trading day before, since these rates are observable when the *forecaster* makes his/ her prediction. The other possibility is to use rates from the 2nd day of the publication month, or the nearest trading day thereafter, since these are observable at the time a *user* of the Blue Chip service receives the forecasts. We report results for both sets of spot and futures market rates below, where they are denoted SPOT-23, FUT-23 and SPOT+2, FUT+2 respectively. The futures market forecast is defined as 100 *minus* the price of the IMM 90-day T-bill futures contract for the relevant target month.

Figure 1 shows spot (= futures) yields at delivery dates in the years 1981-95, along with the 6-month ahead consensus forecasts, and the high and low individual forecasts in the panel. Our data begin just after the period of high and highly volatile US short term interest rates caused by the temporary switch to monetary base targeting by the Fed in the early 1980s. This in itself may cause our results to differ from those of studies using the Goldsmith Nagan data (1977 onwards) and the Wall Street Journal and Commonwealth Research Institute data (1982 onwards). The consensus forecasts appear to be anchored on the spot rates of the survey month, so during years when the T-bill yield is increasing (e.g. 1988-9) there are persistent underpredictions, and when the T-bill yield is decreasing (1990-2) persistent overpredictions. The spread of individual forecasts around the consensus tends to be widest when yields are relatively high (1984-5 and 1989-90) and narrowest when yields are low (1991-2).

Consensus forecasts

Table 2 compares the directional accuracy of the consensus forecast with the directional accuracy of forecasts from the spot and futures markets. DA-23 forecasts a rise in T-bill yields if the consensus forecast is higher than SPOT-23, the actual T-bill yield at the time the individual forecasts were made. DA+2 forecasts a rise if the consensus is higher than SPOT+2, the yield at

the time the forecasts were published. FUT-23 and FUT+2 predict that the yield will rise/ fall according as the implied futures yield is higher or lower than the contemporaneous spot yields SPOT-23 and SPOT+2. Finally, ΔSPOT+2 is an extrapolative forecast based the change in the spot rate in the (roughly 10 day) period between the forecast preparation date and the publication date. The T-bill yield is forecast to rise or fall according as SPOT+2 is higher or lower than SPOT-23. This is an arbitrary rule based only on the availability of the figures in our data set, and with no data snooping.

The table shows that on average over all horizons, *the consensus forecast gets the direction of change in interest rates correct less than half the time* (48-49%). This compares unfavorably with the futures market, which is correct 56-57% of the time, and even less favorably with the extrapolation of the spot rate over the 10 days before publication, which gets the direction of change correct 61% of the time. The consensus does worst at predicting the very short, one month, changes in T-bill yields, improves for 2- to 6-month horizons, then deteriorates again at longer horizons. At no horizon does the consensus forecast outperform the simple spot rate extrapolation, but for 2- to 6- month ahead predictions its performance is comparable to the futures market.

Table 3 compares the root mean square error of the consensus forecast with market-based alternatives. Here SPOT-23 and SPOT+2 are the levels of the spot T-bill yield on the preparation and publication dates of the forecasts, and FUT-23 and FUT+2 the corresponding yields implied by futures prices on these days. Both measures tell the same story. The accuracy of the consensus forecast declines steadily with the forecast horizon. Consistent with the studies cited above, at all horizons the consensus forecasts are dominated by the naive forecast that T-bill yields will stay unchanged.

Except at the very shortest 1-month horizon, the futures market forecasts are also less accurate than the no-change spot-based forecast. At horizons of up to 6 months - that is for futures forecasts based on the two nearest contracts - the futures market is generally more accurate than the consensus forecast. However, for horizons above six months - that is, for forecasts based on the more distant futures contracts - the situation is reversed, and the consensus forecasts are more accurate than the futures market. This inefficiency in the futures market may reflect thin trading in the more distant contracts.

Individual forecasts

The first column of Table 4 ranks individual forecasters according to their RMSE averaged over all horizons, and shows the directional accuracy statistics DA+2 for each. Individual forecasters/ institutions are identified by letter codes. There is a wide disparity in the performance of the forecasters, the average errors of the worst being over 50% higher than the average errors of the best. There is a weak but positive correlation between RMSE ranks and directional accuracy ranks (rank correlation = 0.43). Some forecasters, like GG and Y, get the direction of change in T-bill yields correct around 60% of the time, but are relatively bad at predicting the size of the change. Others, like B and O, rarely make bad point forecasts, but more often than not get the direction of change wrong.

As expected from the theory of forecast combination (Bates and Granger, 1969; Batchelor and Dua, 1994) the consensus forecast is better in terms of RMSE than the great majority - 21 out of 25 - of the individual forecasters. However, in terms of directional accuracy the consensus does much less well, outperforming only 15 of the individual forecasters.

Again consistent with the studies cited earlier, the random walk forecast SPOT+2 performs exceptionally well. The spot rate is superior to *all* of the individual forecasters in terms of RMSE. The simple extrapolation of the past 10-days change in the spot rate, ΔSPOT+2, also outperforms all but one of the individual forecasters as a predictor of the direction of change in T-bill yields. The futures market performs less well on average, outperforming only 16 of the individual forecasters in terms of RMSE and only 19 in terms of directional accuracy. The two nearby contracts, however, do much better, outperforming all of the individual forecasters at horizons up to 3 months, and all but 4 at horizons of 4-6 months.

These general features of the individual forecasts raise a number of questions. Are the differences that we observe statistically significant? Are they consistent over forecast horizons? Are they consistent over time?

While some forecasters have clearly been more accurate than others, it is not straightforward to make inferences about whether these differences are statistically significant. One possibility is to conduct an analysis of variance on the squared forecast errors made by forecaster i at time t (e_{it}^2).

However, since the variance of the forecast errors may have changed over time, a conventional F-test cannot be used. A nonparametric alternative, due to Friedman (1937) and implemented in Batchelor (1990), is to map the e_{it}^2 into ranks r_{it} showing the rank of forecaster i among all errors made in forecasting at time t. Under the null hypothesis that there is no difference among the n participants in the survey, then the sum of individual i's ranks over T forecasts (r_i) will have expected value T(n+1)/2 and variance Tn(n+1)/12. Hence the statistic

$$f = \sum_{i=1}^{n} \frac{\{ri - T(n+1)/2\}^2}{Tn(n+1)/12} \qquad (1)$$

will be asymptotically distributed as χ^2_{n-1}. Generalizations to unbalanced panel data by Prentice (1979) and Skillings and Mack (1981) were used by MacDonald and Marsh (1996) and Kolb and Stekler (1996) respectively to analyze differences among forecasters. We have computed (1) for accuracy ranks based on squared errors, and also on directional accuracy. In the latter case, in any one survey a number of forecasters (say k) will be correct and the rest (25-k) wrong, and to implement the test the correct forecasters are assigned the average tied rank k/2 and the incorrect forecasters the average rank (25+k)/2.

Table 5 shows values of this statistic at each monthly horizon, for horizons grouped by futures contract (1-3, 4-6, etc.), and for ranks pooled over all horizons. Looking at individual months, there are no differences significant at the 5% level in directional accuracy ranks, but there are significant differences in squared error ranks at all horizons except 2- and 3-months ahead. Pooling over horizons makes *all* the test statistics highly significant, suggesting that there is some consistency in the relative performance of forecasters at different horizons.

A close look at the correlations between RMSE ranks across forecast horizons shows that rankings for the 1-month ahead forecast are only weakly correlated with rankings for longer term forecasts. But for all other horizons there are significant positive correlations of accuracy ranks across forecasters. For example, the correlation between 4-6 month and 6-9 month accuracy ranks is 0.91, and the correlation between 4-6 month accuracy ranks and 10-12 month ranks 0.73. We can conclude that for all but the 1-month horizon, there are consistent and statistically significant differences in forecast accuracy across individual forecasters.

To establish whether these differences in accuracy are stable over time, we divide the sample into forecasts made in the early years 1984-87 and the later years 1988-91, and compute rankings by directional accuracy and squared errors for each subperiod. Taking all the forecasters together, there appears to be little relation between accuracy rankings in the two subsamples - the rank correlations are +0.33 for directional accuracy and +0.09 for RMSE.

These inconsistencies in relative performance turn out to be largely due to the effects of some institutions changing their forecaster. If we look at the rankings of the 13 *individuals* who produced forecasts all through the period, the correlations across periods are very much higher (0.80 for directional accuracy, 0.48 for RMSE). On the other hand, a few of the 12 *institutions* which changed forecaster experienced dramatic changes in forecast accuracy, not always for the better, so that their rank correlations between the early and late years are -0.24 for directional accuracy and -0.14 for RMSE. This strengthens our conclusion that there are significant differences in performance across forecasting organizations, and suggests further that expertise is specific to individual forecasters.

III. PROFITABILITY

The implication of any interest rate forecast for trading profits depends on what trading rule is used. In this paper we adopt a very simple and intuitively appealing rule - that the user takes a long or short position in one T-bill futures contract according as the forecast for the contract delivery date is below or above the yield implied by the current futures price. This is the strategy assumed by Booth (1983) and Boothe and Glassman (1987), and is Rule B in Leitch and Tanner (1991). It has the rationale that if the forecast is correct, the futures price must show a trend rise or fall so as to converge on the forecast cash market yield at delivery, and these price trends will make profits for the long or short investor respectively.

Other trading rules are of course possible. For example, the user might follow a filter rule, only taking a position if the forecast yield differs from the yield implied by the futures price by more than some threshold; or the user might take a larger position if the forecast is very different from the implied yield. Rule A in Leitch and Tanner (1991) is to buy and sell futures according as the forecast yield is below or above the spot (rather than the futures) yield. However, unlike Rule B this

does not have a recognizable theoretical basis, though Leitch and Tanner claim that the "general pattern [of profits] is very similar across trading rules" (p585).

In addition to specifying the trading rule, an assumption must be made about when the futures position is taken and when it is unwound. Given the timing of the Blue Chip forecasts, we have considered two dates at which futures positions might be taken - the 23rd day of the month preceding the survey month, the day at which the forecast is made, and the 2nd of the publication month, the day the forecast is received by subscribers. Profits in the pre-release period between these two dates can be regarded as "inside" profits, available to the forecasters but not the subscribers to the Blue Chip service. Since forecasts are revised each month, we assume positions are rolled over or reversed on a monthly basis, either on the 23rd or 2nd of the month. In the case of the 1-month ahead forecast, the position is unwound within the month, on the delivery date of the contract.

The profits distribution

Forecasts are made for four futures delivery dates, and we assume that a one-contract position is taken in all four futures contracts. This trading strategy sets an upper limit on the profits which could be earned by a forecaster who had perfect foresight with respect to the direction of change of futures prices. Figure 2 shows the profits “available” in each forecast month measured in basis points, with an indication of whether the ideal trader should have been short or long. Changes from month to month in the prices of futures contracts with different delivery dates are highly correlated, and in almost all months it would have been optimal to go either short or long in *all four* futures contracts. Futures prices rose in rather more months than they fell, reflecting the general downward trend in T-bill yields over the 1980s, and the optimal trading strategy requires a long position in exactly 60% of months.

A maximum of 14500bps could have been earned from our trading strategy. The value of one basis point for the IMM 90-day T-bill contract is $25, so the maximum profit is $362500. These profits are highly concentrated in a small number of months. About 25% of the profits could have been earned by getting the direction of change in the market right in 10% of months, about 50% could have been earned in 25% of months, and about 75% in 50% of months. This is important because it means that the ability to predict major market moves is as valuable as the ability to predict its level

or direction of change, and there is no necessary connection between directional accuracy and profitability. A forecaster could perform poorly most of the time, but provided he/ she was correct in a high proportion of the most profitable months, could still earn high profits by following our trading rule. Available profits are also somewhat concentrated in the more distant of the four futures contracts involved in the trading strategy - the nearby contract contributes only 18% of overall profit, the next 22%, and the two most distant contracts about 30% each.

Profits by forecaster

Table 4 summarizes the profits made by following each forecaster, assuming trading on the 2nd of the publication month. There are at least six interesting features of the exercise.

First, there is no advantage to trading early, on the 23rd of the preceding month, when the forecasts are prepared. If anything the profits made by trading ahead of publication are lower, so there is no evidence that there are "insider" profits to be made in this market, or that the futures market responds ahead of the publication of the forecasts. Although the rankings of some forecasters are a little different depending on which date is used to execute the futures trades, the general pattern is consistent. The top five forecasters (Y, D, G, C, GG) stay the same, as do the bottom two (L, XX). We have not therefore reported the profits from these trades in detail on the Table, since they do not affect our argument.

Second, the profit performance of the consensus forecast is rather worse in terms of profitability than it is in terms of accuracy. The consensus profitability ranking (11/27) is comparable to its directional accuracy rank (12/28) but substantially lower than its RMSE rank (5/28).

Third, the performance of trading rules based on the random walk forecasts SPOT+2 is very much worse than their accuracy suggests. These forecasts were top-ranked (1/28) in terms of RMSE. However, they are close to bottom-ranked (24/27) in terms of profitability. Trading rules which assume that futures yields tend move away from their current level and towards the current spot rate do not appear to be viable. Our results are more consistent with the expectations theory of futures yields, which suggests that the spot rate converges on the futures rather than vice versa.

Fourth, although the optimal strategy is to take more long positions than short, there is in practice no relation between the proportion of long positions taken by forecasters and the profits they make. Following the most profitable forecaster Y would have meant going long in 71% of months; following the second most profitable forecaster D meant going long in only 46% of months. For the group as a whole the rank correlation between profitability and percentage long is only 0.08. This underlines the fact that the heavy-tailed distribution of price changes, which concentrates most available profits in a small number of months, means that identifying these months when major market moves occur is more important for profitability than getting the direction of change right on average over many months.

Fifth, the relatively profitable forecasters do not run excessive risks. In the table we have calculated the Sharpe ratio for each forecaster, by dividing the mean monthly profit for each forecaster by the standard deviation of these monthly profits. Although standard deviations do differ across forecasters, the differences are small relative to differences in profits, and the ranking of forecasters by Sharpe ratio is identical to their ranking by profitability. Figures for maximum drawdown, the largest loss sustained in one losing run by each forecaster, also revealed no systematic differences in risk across forecasters.

Sixth, although our calculations do not factor in transactions costs, their incorporation reduces profits a little but does not change the fact that almost all forecasters make positive profits, and has no impact on profitability rankings. Leitch and Tanner (1991) reckon that over a 3-month period, the typical holding period for one of our forecasters, a trader would incur a round trip brokerage fee of between $10 and $75, and about $40 opportunity loss on initial margin - in total around 2-4 bps on the 90-day T-bill contract. Given an average of 16 roll-overs each year, over a 9 year period a typical forecaster would incur costs of under 600bps, well below the profits achieved by all but the very worst forecaster.

We found earlier that accuracy was fairly consistent over time for individual forecasters, but not for institutions with changing forecast staff. This is also true of profitability, though less strongly. The rank correlation of profitability between the periods 1984-7 and 1988-92 is +0.26 for the 13 individual forecasters, but -0.23 for the 12 institutions.

We also found earlier that differences in accuracy across forecasters were fairly consistent across horizons. The same is *not* true for profitability. The rank correlation of forecasters according to profits made at the 1-3 month and 4-6 month horizons (i.e. from the two nearby futures contracts) is positive, at 0.46. But rank correlations by profitability at longer horizons are not significant. Some forecasters are consistently successful (Y) or unsuccessful (XX) at all horizons. But some make profits from short term forecasts but not from long (EE), and others are only profitable at long horizons (L). As a result, the correlation between ranks according to 1-3 month profits and 9-12 month profits is only -0.09.

This finding is important because the figures in Table 4 for profitability averaged over horizons conceal a great disparity in the profits from trading in the nearby and more distant contracts. Profits from the nearby contract, based on 1-3 month ahead forecasts, contribute only 5% of the total profits earned by our forecasters, and profits from the next contract, based on 4-6 month ahead forecasts, account for 15% of total profits. In contrast, trading in the more distant futures contracts, based on 7-9 and 10-12 month T-bill yield forecasts, accounts for 31% and 49% of total profits respectively. The proportion of profits actually coming from the most distant contract is much greater than its share of "available profits" (around 30%). This strong result could have been anticipated, given the relative inaccuracy of the distant futures prices in predicting spot yields, and the greater profits available from the larger price movements in the more distant contracts. It implies that those forecasters who make good forecasts at long horizons will be most profitable overall.

A final question is whether the profits made by our panel are significantly higher than could be achieved by chance. Because available profits are not normally distributed, it is difficult to provide an analytical answer to this question. We have instead conducted a simulation experiment in which a trader is assumed to take long or short positions in all contracts each month with fixed probabilities. Since the optimal strategy would involve 60% long positions we assume that the probability of going long in any month is 0.6. Figure 3 compares the distribution of profits from 5000 simulations of this strategy with the actual distribution of profits for our 25 forecasters. The panel makes profits much greater than could be achieved by chance. The very worst forecaster makes more than the average of the simulations (660bps). Of our 25 forecasters, 17 exceed the 90th percentile of the simulation distribution (2900bps) and 10 exceed the 95th percentile (3500bps). The top 10 forecasters also outperform the results of going long in *every* month

(3384bps). And if the forecasters are compared with the results of trading at random with a 0.5 probability of taking a long position each month, all but one outperform the 95th percentile of profits (1700bps).

IV. ACCURACY *versus* PROFITABILITY?

Figure 4 shows the relationship between overall profitability, with trading taking place on the 2nd day of the month the survey is published, and one measure of forecast accuracy (RMSE). The Figure shows that, although the relation is far from perfect, in general *more profits are earned by the more accurate forecasters*. Table 6 shows rank correlations between profitability and all accuracy measures. The first row of the Table shows that on average over all horizons there is a significant positive correlation between profitability and directional accuracy ranks (0.66), and between profitability and RMSE ranks (0.52). Positive rank correlations between profitability and both DA+2 and RMSE are also found in the subsamples of individual forecasters and forecasting institutions.

The strong correlation between profitability and directional accuracy is not unexpected. Many authors effectively equate directional accuracy - "market timing" - with trading value (Merton, 1981; Hendriksson and Merton, 1981; Leitch and Tanner, 1995), though Cumby and Modest (1987) do emphasise that profits also depend on the incidence of "big hits". Acar (1993, p92) has shown formally that if the target variable follows a random walk - a reasonable approximation for T-bill yields given the performance of the no-change forecast - the linear correlation between trading returns and directional accuracy is $\sqrt{(2/\pi)} = 0.80$. A proof is given in the Appendix. The figure of +0.60 from our sample of 25 forecasters, and the figure of 0.84 from the six forecasts in Leitch and Tanner (1991, Table 2,), are therefore both of the expected order of magnitude.

The corresponding theoretical linear correlation between profitability and squared errors lies around $-1/\sqrt{\pi} = -0.56$, the exact value depending on the relative variance of the forecast and target variables (see Appendix). The linear correlation of -0.63 between profits and RMSEs found in our sample of 25 forecasters is therefore consistent with this theoretical value. But Leitch and Tanner (1991, Table 2) find a perversely positive correlation of +0.21 between profitability and RMSE for

traders following their Rule B. Their much-cited result is at odds with theory and our empirical findings, and merits closer inspection.

Table 6 shows that there is a tendency for rank correlations between directional accuracy and profitability to be high at short horizons and low at long, whereas rank correlations between profits and RMSE are much higher at the longer forecast horizons, and not significant at short horizons. This is unlikely to explain any discrepancies between our study and LT, however, since their forecast horizon, though a little shorter than ours (12 rather than 15 months), nonetheless lets them simulate trading in the first three, and sometimes the first four, futures contracts.

The LT data start in 1980 and end in 1987. Table 6 reports rank correlations for our 25 forecasters for forecasts made in the years 1984-87 and 1988-91, the former overlapping the LT sample. All of the correlations remain significantly positive in both subsamples, though there is some slight evidence that the rank correlation of profitability with RMSE is lower in the early period (0.47) than in the later (0.69), and vice versa for directional accuracy. So differences in the characteristics of the two data periods might have contributed in a small way to the differences in the results.

By far the most likely explanation for the discrepancy between our correlations and those in LT lies in the composition and size of the samples used. LT include the random walk forecast and the consensus forecast in their sample. As we have seen, the random walk forecast is very accurate but not very profitable, while the consensus forecast is fairly accurate but only averagely profitable. The random walk forecast is a clear outlier on Figure 4. When the consensus forecast and the SPOT+2 forecast are added to our sample, the rank correlations between accuracy and profitability do indeed fall, as shown in the next-to-last row of Table 6, although all remain positive. The linear correlations of profitability with directional accuracy and RMSE also fall, to +0.55 and -0.51 respectively, but again remain significant and correctly signed.

In addition to the random walk and consensus forecasts, LT use only four other forecasts. There is a danger that such a small sample will be dominated by the inconsistent performance of the random walk. With our larger data set it is possible to quantify this risk. We have looked exhaustively at all the ${}_4C_{25}$ = 12650 subsamples of [four forecasters + random walk (SPOT+2) + consensus] which could be drawn from our sample of 25 forecasters, and computed the resulting correlations between accuracy and profitability for these subsamples.

Figure 5 shows the distribution of correlations between profitability and directional accuracy, and between profitability and RMSE. The distribution of directional accuracy correlations is unimodal, with a median of +0.15. The distribution has quite a wide spread, however, and the probability of finding a negative correlation in a sample of [four forecasters + random walk + consensus] is over +0.33. Our large-sample estimate of 0.51 is around the upper 95th percentile of the distribution. The LT figure of +0.84 for the correlation between directional accuracy and profitability under Rule B is extraordinarily high, and higher than *any* of the correlations found in our subsamples.

The distribution of small-sample correlations between profitability and RMSE on Figure 5 is multi-modal and has a very large spread. The median correlation is -0.15 and the distribution is negatively skewed. But the probability of finding a perverse positive correlation from a sample of [four forecasters + random walk + consensus] is significantly positive, almost 0.40. The weak positive correlation of +0.21 found by LT is well within the 95th percentile of this distribution, around the 75th percentile. On the other hand, our "true" large sample estimate of -0.55 is well below the median, around the 25th percentile. We can conclude that any inferences about correlations between accuracy and profitability drawn from samples of the size and type used by Leitch and Tanner (1991) will be extremely fragile.

V. CONCLUDING COMMENTS

In this study we have shown that there are significant differences across professional forecasters in the accuracy of short term interest rate forecasts, and that these differences in accuracy translate into differences in the profitability of trading rules based on the forecasts.

Although commonsensical, both findings go against the grain of previously published research. For example, Batchelor (1990) found no persistent differences in the accuracy of professional forecasts of major macroeconomic variables such as GNP, inflation and unemployment. It is a little paradoxical that significant differences in expertise should be found in interest rate forecasting, an area where forecastability is much more limited.

We have established that the perverse correlations between accuracy and profitability found by Leitch and Tanner (1991) reflect the fact that, given the time series characteristics of the T-bill

yield, a sample of six forecasters is too small to produce reliable inferences. This small sample bias is aggravated by the fact that the random walk forecast, while generally the most accurate, produces relatively low levels of profits if used to determine futures positions.

It seems reasonable to conjecture that this kind of bias also explains the low or perverse correlations between the accuracy and profitability found in the small sets of exchange rate forecasts studied by Boothe (1984) and Boothe and Glassman (1987). In a parallel study, MacDonald and Marsh (1996) use a larger group of over 25 foreign exchange forecasters, and also find that the random walk forecast is generally the most accurate, but far from the most profitable, forecast. Their data yields a variety of RMSE-profit correlations depending on the currency studied, but these correlations are - like our own - almost invariably significantly negative.

REFERENCES

Acar, E., 1993, *Economic Evaluation of Financial Forecasting*, PhD. Thesis (unpublished), City University Business School.

Batchelor, R., 1990, 'All forecasters are equal', *Journal of Business and Economic Statistics*, 8, 143-4.

Batchelor, R., and P. Dua, 1992, 'Testing the rationality of panel interest rate forecasts', *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 27-32.

Batchelor, R. and P. Dua, 1994, 'Forecaster diversity and the benefits of combining forecasts', *Management Science*, 24, 331-340.

Bowlin, O. D. and J. D. Martin, 1975, 'Extrapolation of yields over the short run: forecasts or folly?', *Journal of Monetary Economics*, 10, 475-88.

Boothe, P., 1983, 'Speculative profit opportunities in the Canadian foreign exchange market', *Canadian Journal of Economics*, 16, 4, 603-11.

Boothe, P., and D. Glassman, 1987, 'Comparing exchange rate forecasting models: accuracy versus profitability', *International Journal of Forecasting*, 3, 65-79.

Belongia, M. T., 1987, 'Predicting interest rates: a comparison of professional and market-based forecasts', *Federal Reserve Bank of St. Louis Review*, 69, 9-15.

Cumby, R. E., and D. M. Modest, 1987, 'Testing market timing ability: a framework for forecast evaluation', *Journal of Financial Economics*, 19, 169-89.

Dua, P., 1988, 'Multiperiod forecasts of interest rates', *Journal of Business and Economic Statistics*, 6, 381-4.

Friedman, M., 1937, 'The use of ranks to avoid the assumption of normality implicit in analysis of variance', *Journal of the American Statistical Association*, 32, 675-701.

Friedman, B. M., 1979, 'Interest rate expectations versus forward rates: evidence from an expectations survey', *Journal of Finance*, 34, 965-73.

Hafer, R. W. and S. E. Hein, 1989, 'Comparing futures and survey forecasts of near-term Treasury bill rates', *Federal Reserve Bank of St Louis Review*, 71, 33-42.

Hafer, R. W., S. E. Hein and S. S. MacDonald, 1992, 'Market and survey forecasts of the three-month Treasury bill rate', *Journal of Business*, 65, 1, 123-138.

Hendriksson, R. D., and R, C. Merton, 1981, 'On market timing and investment performance II: statistical procedures for evaluating forecast skills', *Journal of Business*,

Kolb, R. A. and H. O. Stekler, 1996, 'How well do analysts forecast interest rates?', *Journal of Forecasting*, 15, 385-94.

Leitch, G. and J. E. Tanner, 1991, 'Economic forecast evaluation: profits versus the conventional error measures', *American Economic Review*, 81, 3, 580-590.

Leitch, G. and J. E. Tanner, 1995, Professional economic forecasts: are they worth their costs?, *Journal of Forecasting*, 14, 143-57.

MacDonald, S. S. and Hein, S. E., 1989, 'Futures rates and forward rates as predictors of near-term Treasury bill rates', Journal of Futures Markets, 9, 249-62.

MacDonald, R. and I. W. Marsh, 1996, 'Currency forecasters are heterogeneous: confirmation and consequences', *Journal of International Money and Finance*, 15, 5, 665-685.

Merton, R. C., 1981, 'On market timing and investment performance I: an equilibrium theory of value for market forecasts', *Journal of Business*, 54, 513-33.

Prentice, M. J., 1979, On the problem of *m* incomplete rankings, *Biometrika*, 66, 167-170.

Skillings, J. H. and G. A. Mack, 1981, On the use of Friedman-type statistics in balanced and unbalanced block designs, *Technometrics*, 23, 171-7.

APPENDIX: Theoretical correlations between profits and error measures.

In the special case where futures prices follow a random walk without drift, it is possible to derive analytical expressions for correlations between profits and conventional measures of forecast error.

Suppose the actual changes in futures prices between all successive time periods t-1 and t, denoted p_t , are independent normal i.i.d. variables with distribution $N(0, \sigma^2)$. We denote a forecast formed at t-1 for the change in futures price between t-1 and t as f_t , and assume these are unbiased and homoscedastic, so that $f_t \sim N(0, \tau^2)$. The profit from a position in the futures market at t-1 is $x_t = \lambda_t p_t$ where $\lambda_t = +1$ or -1 according as the position is long or short. Under the trading rule in the paper, the trader will go long if the forecast price change is less than the difference between the forward (= futures) price and spot price at t-1.

Theorem 1. Profits follow the same distribution as the underlying price changes.

Given the random walk assumption, for an unbiased forecaster $Prob(\lambda_{t-1} = +1) = Prob(\lambda_{t-1} = -1) = 0.5$. Then the characteristic function of the distribution of x_t can be rewritten as

$$\begin{aligned} C_x(z) &= E\{\exp(-iz\lambda_t p_t)\} \\ &= Prob(\lambda_t = +1).E\{\exp(-izp_t)\} + Prob(\lambda_t = -1).E\{\exp(+izp_t)\} \\ &= 0.5\, C_p(-z) + 0.5\, C_p(+z) = C_p(z), \end{aligned}$$

where $C_p(+z)$ is the characteristic function of p_t. Hence $x_t \sim N(0, \sigma^2)$.

Theorem 2. The linear correlation of profits with directional accuracy is $\sqrt{(2/\pi)}$.

Directional accuracy is defined as $DA_t = +1$ or-1 according as $f_t . p_t > 0$ or $f_t . p_t < 0$. Given the random walk assumption, directional accuracy DA_t follows a binomial distribution with expected value $E(DA_t) = Prob(f_t>0, p_t>0) + Prob(f_t<0, p_t<0) = 0$, and standard deviation $S(DA_t) = 0.5$.

Then $Corr(DA_t, x_t) = Cov(DA_t, x_t) / \{S(DA_t)S(x_t)\} = Cov(DA_t, x_t)/ 0.5\sigma$, and

$$Cov(DA_t, x_t) = E(DA_t . x_t) = (+1) . \int_{f_t>0} \int_{p_t>0} f_t.p_t\, \phi(f)\, \phi(p)\, df\, dp + (-1) . \int_{f_t<0} \int_{p_t<0} f_t.p_t\, \phi(f)\, \phi(p)\, df\, dp$$

$$= 0.5 \{ \int_{p_t>0} p_t \ \phi(p)\, dp + \int_{p_t<0} - p_t \ \phi(p)\, dp \}$$

Each of the above integrals is $\sigma/\sqrt{(2\pi)}$. Hence Corr $(DA_t, x_t) = \{\sigma/\sqrt{(2\pi)}\} / 0.5\sigma = \sqrt{(2/\pi)}$.

<u>Theorem 3.</u> <u>The linear correlation of profits with squared error is $-(2/\sqrt{\pi}).\ \rho/(1+\rho)$, where $\rho = \tau^2/\sigma^2$ is the ratio of the forecast variance to the target variance.</u>

Given the random walk assumption, p_t and f_t are uncorrelated, so the forecast error $e_t = (p_t - f_t)$ is $N(0, \sigma^2 + \tau^2)$. The expected value of the squared error SE_t is therefore $E(SE_t) = E(e_t^2) = \sigma^2 + \tau^2$. The variance of the squared error is $V(SE_t) = E(e_t^4) - \{ E(e_t^2)^2\} = 3(\sigma^2 + \tau^2)^2 - (\sigma^2 + \tau^2)^2 = 2(\sigma^2 + \tau^2)^2$, so the standard deviation of squared error is $S(SE_t) = \sqrt{2}(\sigma^2 + \tau^2).\ \tau^2/\sigma^2$

Then the correlation between squared errors and profits is

Corr $(SE_t, x_t) = Cov(SE_t, x_t) / \{S(SE_t)S(x_t)\} = Cov (SE_t, x_t)/\{\sqrt{2}(\sigma^2 + \tau^2).\sigma\}$

and

$Cov (SE_t, x_t) = E(SE_t \ .\ x_t) = E\{(p_t - f_t)^2 \ .\ \lambda_t p_t\} = E\{(p_t^2 - 2p_tf_t + f_t^2) \ .\ \lambda_t p_t\}$

$= E(\lambda_t p_t^3) - 2E(\lambda_t f_t p_t^2) + E(\lambda_t f_t^2 p_t)$.

Since λ and p are independent, and p is normal, the first term $E(\lambda_t p_t^3) = E(\lambda_t).E(p_t^3) = E(\lambda_t).0 = 0$. Similarly, the third term can be simplified to $E(\lambda_t f_t^2 p_t) = E(\lambda_t f_t^2).E(p_t) = E(\lambda_t f_t^2).0 = 0$. Hence,

$Cov (SE_t, x_t) = - 2E(\lambda_t f_t p_t^2) = -2E(\lambda_t f_t).E(p_t^2)$

$$= -2\{ \int_{f_t>0} f_t \ \phi(f)\, df - \int_{f_t<0} f_t \ \phi(f)\, df \}.\sigma^2$$

Since each integral is $\tau/\sqrt{(2\pi)}$, Cov $(SE_t, x_t) = -2.\{\tau.\sqrt{(2/\pi)}\}\sigma^2$

Hence, Corr $(SE_t, x_t) = \{-2.\sqrt{(2/\pi)}.\tau.\sigma^2\}/\{\sqrt{2}(\sigma^2 + \tau^2).\sigma\} = -(2/\sqrt{\pi}).\ \rho/(1+\rho)$, where $\rho = \tau^2/\sigma^2$. When the variances of forecast and actual are equal, $\rho = 1$ and Corr $(SE_t, x_t) = -1/\sqrt{\pi} = -0.5642$.

Table 1. Blue Chip Financial Forecasts: Panel of Regular Forecasters.

Institution	Forecaster (at 12/92)	Previous Institutions	Previous Forecasters
Standard and Poors Corporation	David M. Blitzer		
Aubrey G. Lanston & Co.	David M. Jones		
Bankers Trust, New York	Jay N. Woodworth		
Manufacturers Hanover Trust	Irwin L. Keller		
Oliver Jones and Associates	Oliver Jones		
US Trust Company NY	Thomas W. Synott III		
Metropolitan Insurance Companies	Robert H. Vatter		
Prudential Insurance Co. of America	Michael W. Keran		
Nelson Banking/ Finance Economics Inc.	Richard Nelson	Chemical Bank, SUNY, FHLB	
Moskowitz Capital Consulting Inc.	Arnold X. Moskowitz	Dean Witter Reynolds, County Natwest Bank	
FES Corporation	Herbert E. Neil	Harris Trust and Savings Bank	
DePrince and Associates	Albert E. DePrince Jr.	Marine Midland Bank	
C. J. Lawrence Inc.	Edward E. Yardeni, Deborah Johnson	Prudential Bache Securities	
First Interstate Bank	Lynn Reaser		A. James Meigs, Jerry Jordan
Sanford C. Bernstein & Co.	Giulio Martini		David A. Levine
Chase Manhattan Bank	Henry Willmore		Douglas L. Bendt, Greg Hoelscher
Valley National Bank	John Lucking		Elliott D. Pollack
Wells Fargo Bank	Gary Schossberg, Mark Green		Harold C. Nathan
National Association of Realtors	John A. Tuccillo		Jack Carlson
Merrill Lynch NY	Donald Straszheim		Robert A. Schwartz, Nancy Vanden Houten
US Chamber of Commerce	William MacReynolds		Ronald D. Utt, Grace Ortiz
First National Bank of Chicago	James E. Annable Jr.		Roy E. Moor
LaSalle National Bank	Carl R. Tannenbaum		William W. Tongue
Corestates Financial Coproration	Carol A. Leisenring	Philadelphia National Bank	A. Gilbert Heebner
First Fidelity Bankcorp	Joel L. Naroff	The Fidelity Bank	Mickey D. Levy

Table 2. Directional Accuracy of Consensus and Market Forecasts.

Horizon (months)	Consensus DA-23	DA+2	Market DSPOT+2	FUT-23	FUT+2
1	0.47	0.33	0.83	0.69	0.78
2	0.59	0.53	0.71	0.65	0.59
3	0.61	0.58	0.72	0.69	0.56
4-6	0.50	0.53	0.62	0.62	0.59
7-9	0.42	0.46	0.55	0.50	0.49
10-12	0.46	0.49	0.53	0.49	0.53
All	**0.48**	**0.49**	**0.61**	**0.57**	**0.56**

Table 3. RMSE of Consensus and Market Forecasts.

Horizon (months)	Consensus DA-23	Market SPOT-23	SPOT+2	FUT-23	FUT+2
1	0.14	0.09	0.06	0.06	0.06
2	0.29	0.32	0.24	0.28	0.20
3	0.48	0.50	0.44	0.58	0.50
4-6	0.85	0.75	0.74	0.74	1.01
7-9	1.71	1.44	1.42	2.19	2.15
10-12	2.57	2.28	2.28	3.43	3.35
All	**1.36**	**1.19**	**1.17**	**1.67**	**1.69**

Table 4. Forecast Accuracy and Profitability by Individual Forecaster: average over all horizons

Forecaster	DA+2	*Rank*	RMSE	*Rank*	Profits	*Rank*	Sharpe	% Long
D	0.59	*2*	1.21	*1*	4503	*2*	0.23	0.46
U	0.55	*6*	1.24	*2*	3257	*11*	0.16	0.56
O	0.47	*13*	1.31	*3*	2965	*15*	0.15	0.59
W	0.48	*11*	1.42	*4*	3503	*10*	0.18	0.51
B	0.44	*17*	1.43	*5*	3049	*12*	0.15	0.56
C	0.57	*5*	1.47	*6*	4119	*4*	0.21	0.49
KK	0.51	*8*	1.47	*7*	3907	*6*	0.20	0.55
Q	0.49	*10*	1.61	*8*	3859	*7*	0.19	0.53
R	0.58	*3*	1.69	*9*	3825	*8*	0.19	0.48
GG	0.64	*1*	1.73	*10*	3943	*5*	0.20	0.74
CC	0.52	*7*	1.76	*11*	3015	*13*	0.15	0.42
HH	0.44	*18*	1.77	*12*	2485	*20*	0.12	0.59
OO	0.50	*9*	1.80	*13*	2919	*16*	0.15	0.59
V	0.38	*24*	1.80	*14*	2509	*18*	0.12	0.51
AA	0.44	*15*	1.84	*15*	2509	*19*	0.12	0.49
DD	0.47	*12*	1.91	*16*	2079	*23*	0.10	0.56
G	0.46	*14*	1.92	*17*	4189	*3*	0.21	0.50
F	0.40	*23*	1.92	*18*	2991	*14*	0.15	0.51
Y	0.57	*4*	1.93	*19*	5089	*1*	0.26	0.71
A	0.40	*22*	2.07	*20*	2369	*22*	0.12	0.44
M	0.32	*25*	2.09	*21*	2845	*17*	0.14	0.53
EE	0.43	*19*	2.13	*22*	2481	*21*	0.12	0.48
FF	0.41	*21*	2.46	*23*	3565	*9*	0.18	0.34
L	0.44	*16*	2.67	*24*	1807	*24*	0.09	0.32
XX	0.42	*20*	3.09	*25*	247	*25*	0.01	0.52
Consensus	0.49	*(12)*	1.36	*(5)*	3447	*(11)*	0.17	0.51
SPOT+2			1.17	*(1)*	2349	*(24)*	0.12	0.56
DSPOT	0.61	*(1)*						
Fut+2	0.56	*(7)*	1.69	*(11)*				

Notes: DA+2 is the directional accuracy of forecast changes in rates, relative to the spot rate on the 2nd day of the month of publication of the Blue Chip Financial Forecasts newsletter. **RMSE** is the root mean square forecast error. **Profits** are measured in basis points, summed over all months and contracts. **Sharpe** is the Sharpe ratio, of the mean monthly profit to the monthly standard deviation of profits. **% Long** is the percentage of all trades which involved buying a 90-day T-bill futures contract. Figures in *italics* are forecaster ranks among the 25 forecasters (1=best, 25=worst). Figures in parentheses are ranks of the consensus, spot and forward rate forecasts in the larger groups of 25+3 = 28 or 25+2 = 27 forecasts.

Table 5. Analysis of Variance by Ranks: Test Statistics

Horizon (months)	DA+2	RMSE
1	*29.69*	57.71
2	*18.23*	*31.79*
3	*20.10*	*25.15*
4	*26.67*	48.11
5	*32.44*	53.96
6	*18.92*	52.29
7	*30.32*	62.51
8	*25.64*	62.47
9	*32.40*	56.14
10	*32.04*	61.46
11	*25.47*	59.29
12	*34.74*	58.92
1-3	51.76	79.82
4-6	65.45	139.17
7-9	76.62	171.41
10-12	77.91	164.59
All	220.54	449.12

Note: Based on ranks of 25 forecasters. These statistics should be compared with the critical values of the χ^2 distribution with 24 degrees of freedom, which at the 5% and 1% levels are 36.4 and 43.0 respectively. Cases where there are *no* significant differences at the 5% significance level are shown in italics.

Table 6. Rank Correlations of Accuracy and Profitability

Forecasters	Horizon	Years	Correlation of Profits with: DA+2	RMSE
25	All	All	0.65	0.52
Linear Correlation			*0.60*	*-0.63*
13 Individuals	All	All	0.76	0.45
12 Institutions	All	All	0.53	0.60
25	All	1984-87	0.61	0.47
25	All	1988-91	0.43	0.69
25	1-3 mths	All	0.69	0.16
25	4-6 mths	All	0.36	0.18
25	7-9 mths	All	0.30	0.27
25	10-12 mths	All	0.14	0.38
25 + SPOT + Consensus	All	All	0.52	0.38
Linear Correlation			*0.51*	*-0.55*

Figure 1. Actual v. Forecast T-bill Yields, 6-month Horizon

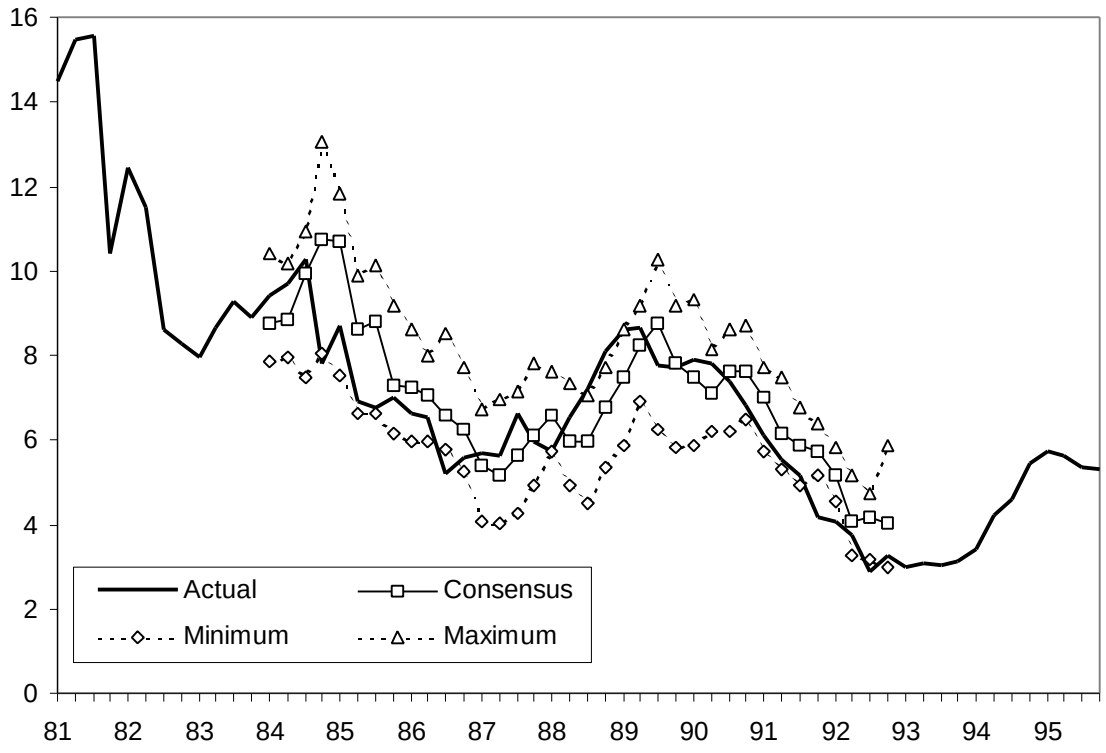

Figure 2. Available Profits in T-bill Futures Market

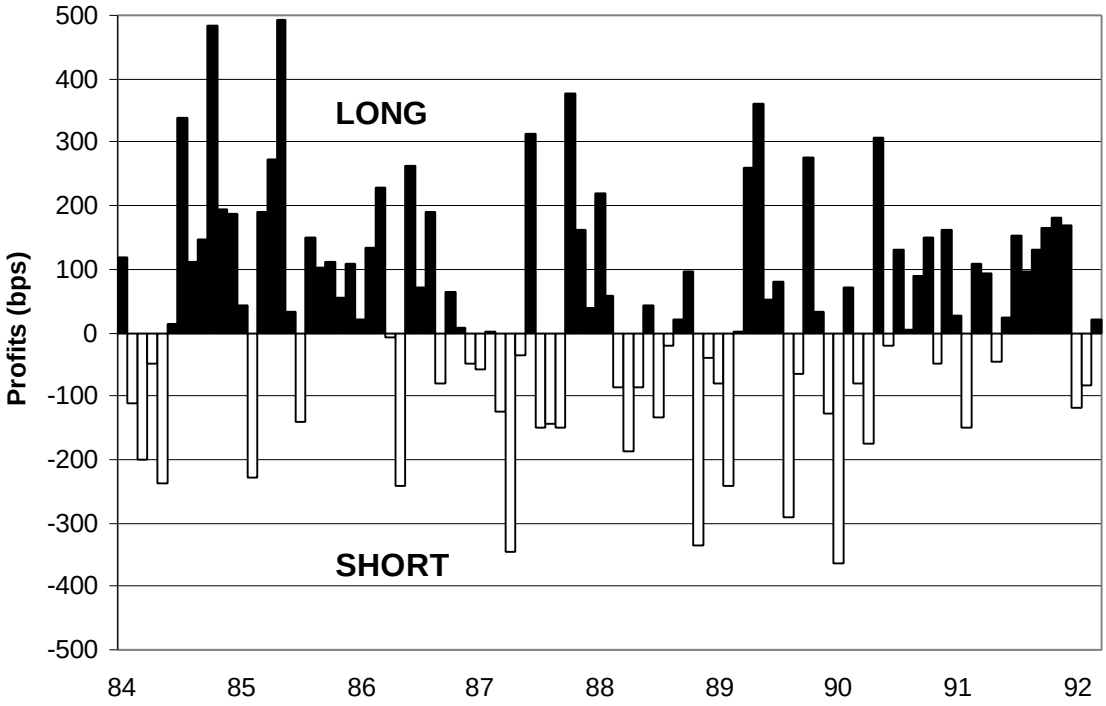

Figure 3. Actual v Simulated Profits in T-bill Futures Market
P(long) = 0.60, 5000 trials

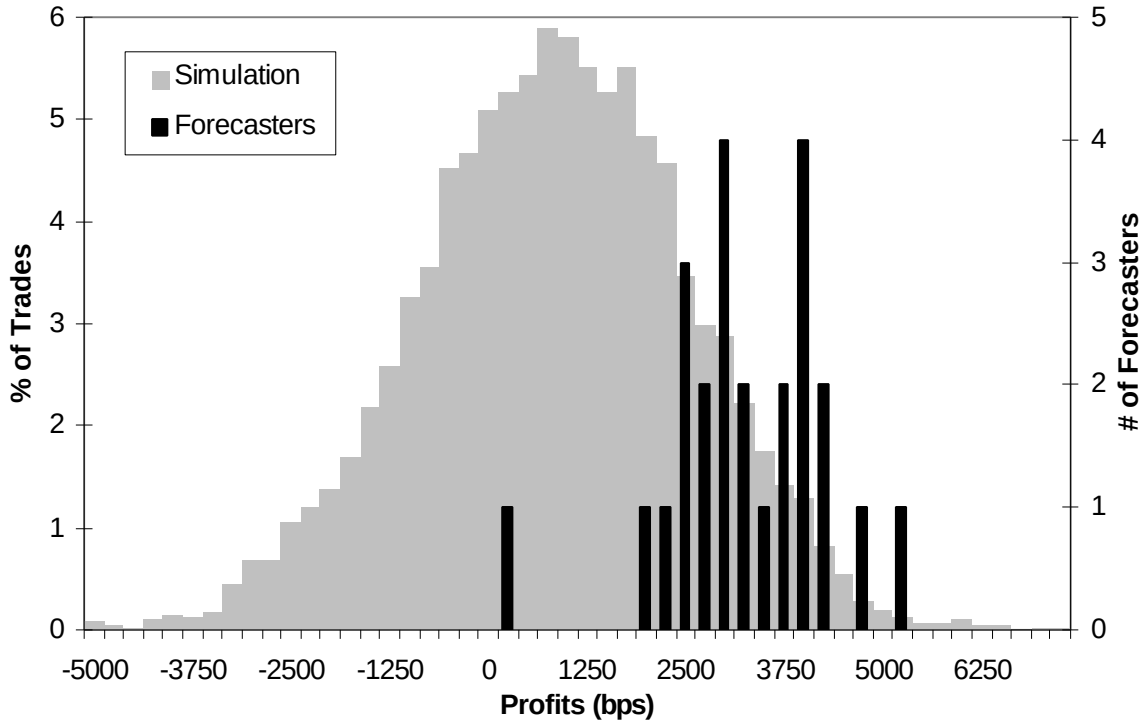

Figure 4. Accuracy versus Profitability, all horizons

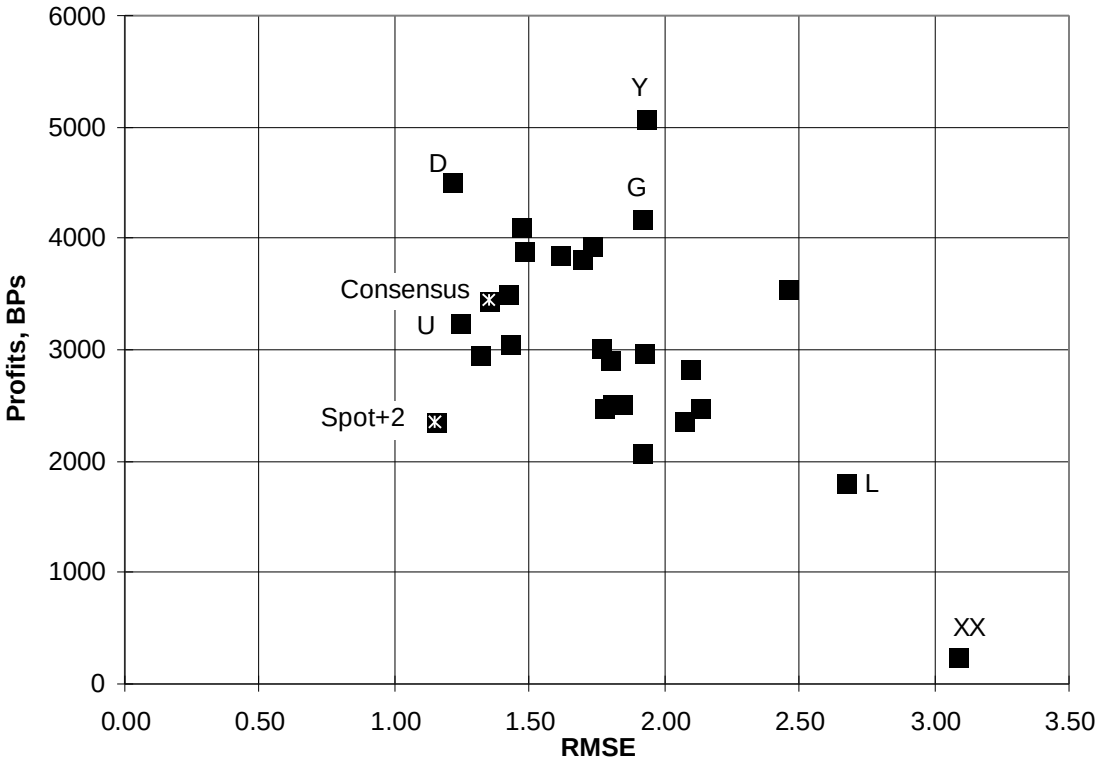

Figure 5. Distribution of Small-Sample Correlations between Accuracy and Profitability