Working Paper
Series

FINRISK
National Centre of Competence in Research
Financial Valuation and Risk Management

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 390

Predicting Stock Price Movements: Regressions versus Economists

Paul Söderlind

First version: March 2007
Current version: March 2007

This research has been carried out within the NCCR FINRISK project on
"New Methods in Theoretical and Empirical Asset Pricing"

Predicting Stock Price Movements: Regressions versus Economists

Paul Söderlind*

29 March 2007

Abstract

The out-of-sample forecasting performance of traditional stock return models (dividend yield, t-bill rate, etc.) is compared with the forecasting performance of the Livingston survey. The results suggest that the survey forecasts are much like a "too large" forecasting model: poor performance and too sensitive to irrelevant information.

Keywords: Livingston survey, out-of-sample forecasts

JEL Classification Numbers: G12

1 Introduction

This paper studies two aspects of the Livingston survey on predicted equity prices: could this panel of economists predict price changes—and did they believe they could?

Return predictability is one of the most contested areas in financial economics. Recently, Goyal and Welch (2004) argue that the evidence of predictability of equity returns disappears when out-of-sample forecasts are considered. In a reply, Campbell and Thompson (2005) claim that there is still some out-of-sample predictability, provided we put reasonable restrictions on the estimated models.

*University of St. Gallen and CEPR. *Address:* SBF, University of St. Gallen, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Financial support by the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK) is gratefully acknowledged. I thank Paolo Giordani for comments; Jan Bernhard, Michael Fischer and Felix Moldenhauer for help with data.

The contribution of this paper is to study the forecasting performance of the Livingston survey. While this has been done several times before (see, for instance, Dokko and Edelstein (1989), Bondt (1991) and Pearce (1984)), this paper updates the evidence, relates the findings to the new literature of out-of-sample forecasting and investigates also other (than the forecasting performance) properties of the forecasts.

It is found that the the survey forecasts perform worse than a naive forecast (the historical mean) and that they share many properties with "too large" prediction models—in particular oversensitivity to recent data.

2 Data

Most of the data used in this paper are standard: various S&P indices, inflation rates, t-bill rates and dates of NBER recessions. See Appendix A for details. The survey data on predicted future stock index values is somewhat less known, so this section is devoted to presenting and discussing it.

The Livingston semi-annual survey collects forecasts of economists in industry, government, banking and academia. It was started in 1946 by the financial columnist Joseph Livingston, but since 1990 it has been administered by the Federal Reserve Bank of Philadelphia.

Since June 1952 the survey has asked about the predicted Standard & Poor (S&P) index level 6 and 12 months ahead in time: from June 1952 to June 1990 the survey asked about the S&P Industrials index, and from December 1990 and onwards it asked about the S&P 500 Composite. For further details, see Federal Reserve Bank of Philadelphia (2004) and Federal Reserve Bank of Philadelphia (2005).

This is a very interesting data set: it gives a long time series of predictions from a group of well informed economists. However, there are several issues of how to use the data.

First, while the Livingston data base contains information on individual answers, these series are typically too short to be useful for assessing the forecasting performance: some sort of pooling is necessary. I choose to use the median forecasts, that is, for each time period I find the median across the forecasters. (The median is preferred to the mean since it is less sensitive to outliers.) This provides a long data series, but it is likely to exaggerate the forecasting performance of a representative survey respondent.[1] The results should therefore be thought of as a (crude) upper boundary of the forecasting performance.

Second, it is actually fairly tricky to calculate implied expected capital gains from the survey data. (Expected capital gains are all we can hope for since the survey only asks about future index levels, not dividends.) The June and December surveys are sent out in late May and November respectively, and contain “base values” of the stock index, typically from mid May/November. The June survey asks for the index levels at the end of December the same year and June next year; the December survey asks for the index levels at the end of June and December next year. It would seem to be straightforward to calculate an expected capital gain by dividing the expected future level by the base value. Unfortunately, some of the base values are highly unreliable.[2]

An alternative approach (used by Pearce (1984)) is to replace the base value with the index level from the last day of the month before the survey. However, it is unclear if the respondents had access to the end-of-month values when filling out the forms. There are also indications that the predictions in December 1989 and June 1990 are for the S&P 500 rather than the S&P Industrials.[3]

In practice, the time series pattern on the expected gains is sensitive to this choice—and both alternatives produce some strange expected capital gains.[4] I will therefore combine the 6- and 12-month forecasts to calculate an implied expected growth rate over a 6-month horizon starting 6 months from now (as in Dokko and Edelstein (1989)). This avoids the problem with the base level, and also the potential change of the index in late 1989 (provided we are willing to assume that the respondents had the same beliefs about capital gains on the S&P Industrials and S&P 500).

In the rest of the paper, I use these expected capital gains over a 6-month horizon, starting (approximately) 6 months from the survey date. I form one series of expectations

[1]There is good theoretical and empirical evidence that combining (averaging) forecasts typically reduces the forecast error variance (see, for instance, Bates and Granger (1969), Winkler and Makridakis (1983) and Batchelor and Dua (1995)).

[2]Consider the following two examples. First, the base value for the Dec 1999 survey is 1229.23, while the S&P 500 index was never below 1350 during Nov to mid Dec. Similarly, the base value for the Dec 2000 survey is 1429.40, which is the index value on 31 Oct 2000, whereas the 30 Nov value is 1314.95.

[3]The base values are 339 and 354 respectively, which corresponds well to the S&P values in mid Nov 1989 and May 1990. The predictions are for virtually unchanged levels. In contrast, the S&P Industrials was around 400 during this period.

[4]The evidence of average over- or under-prediction is not much affected by the choice of base value, however.

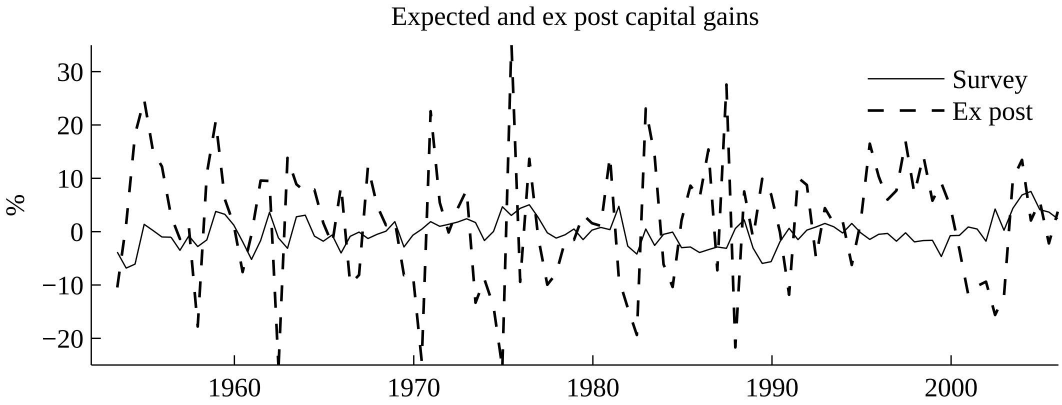

Figure 1: Expected and ex post capital gains on the combined S&P index in excess of a riskfree rate, %. This figure shows the expected and ex post capital gains (in excess of a riskfree rate) for the 6-month horizon. The survey forecasts were made 12 months earlier. The data is for the S&P Industrials for the period up to June 1990, and the S&P 500 after that.

by using the S&P Industrials for the early period (June 1952 to June 1990) and the S&P 500 for the late period (since December 1990)—and a similar series for the ex post capital gains. *Figure 1* shows these series: the ex post capital gains together with the survey predictions made 12 months earlier. As expected, the ex post data is much more volatile. It also seems as if the forecast performance is mixed.

3 Forecasting Performance

This section reports the forecasting performance of simple regression models (in-sample and out-of-sample) and of the Livingston survey.

The first column of *Table 1* summarises the in-sample predictability from using prediction equations—in terms of the traditional R^2. For instance, the first number says that using only the dividend yield as predictor gives an R^2 of around 2%. The other variables (returns, t-bill rate, the historical mean return calculated on a sample from 1926 to the month before the return, an indicator of NBER recessions and the inflation rate) are all worse. Using all these predictors at the same time gives a pretty impressive R^2 of 11%. (To avoid confusion, please notice that the historical mean is here used as a predictor in a regression equation, so the forecast is not the same as the historical mean.)

The second column of Table 1 shows the out-of-sample evidence—in terms of the

	In-sample	Out-of-sample	Out-of-sample*
Regression with the following predictor:			
Dividend yield	2.2	−11.1	0.4
Returns	0.2	0.1	0.3
T-bill rate	0.7	−4.5	1.2
Historical mean	0.6	0.3	0.3
Recession	1.6	−0.3	−0.3
Inflation	1.3	0.3	0.3
All above	11.2	−12.9	−6.5
The following forecast:			
Zero		−1.3	−1.3
Survey		−7.7	

Table 1: R^2 from forecasting 6-month excess capital gains with different predictors, %. The table shows results for the 6-month horizon starting 6 months ahead in time. The out-of-sample R^2 is relative to the historical mean. The out-of-sample* restricts all predictions to be non-negative. See Figure 1 for details on the data.

out-of-sample R^2 (as in Campbell and Thompson (2005)). This measure compares the mean squared error of the prediction model and from using historical average return as the forecast.[5] To calculate the out-of-sample forecasts, the prediction model is estimated on the sample from 1926:1 to period t, and the observed predictor in t is then used to generate a forecast for the 6-month period starting 6 months after period t (just like the survey). This is repeated for all June and December months from 1952 to 2005. The historical mean is estimated in a similar recursive way.

The results in the second column of Table 1 indicate very little out-of-sample predictability: several variables (including the dividend yield) perform much worse than the historical mean—and the model using all predictors is worst. The latter is a common finding in the forecasting literature: large models often suffer from in-sample overfitting and will therefore have poor out-of-sample performance.

The third column of Table 1 also shows out-of-sample evidence, but where the predictions are replaced by zero if they are negative (similat to Campbell and Thompson

[5]The "out-of-sample R^2" in Campbell and Thompson (2005) is $R^2_{OS} = 1 - \sum_{t=s}^{T}(r_t - \hat{r}_t)^2 / \sum_{t=s}^{T}(r_t - \bar{r}_t)^2$, where s is the first period with an out-of-sample forecast, r_t is the return in t, $\hat{r}_t$ is the model forecast of the return in t and $\bar{r}_t$ is the average return for the sample from 1 to $t-1$.

(2005)).[6] This improves the performance quite a bit: most predictors get slightly positive R^2 values, but the large model (using all predictors) is still very poor: the R^2 is -6.5%.

The last two lines of Table 1 reports the out-of-sample R^2 for a zero forecast and the Livingston survey forecasts. Interestingly, In contrast, always predicting a zero capital gain is not such a terrible idea since the R^2 is -1.3%, but the survey is really poor: the R^2 is almost -8% which is even worse than the large model with non-negative predictions.[7]

Similarly, a classical test of forecast unbiasedness (regressing the ex post values on a constant and the forecasts, not shown in the table) gives an annualised intercept around 3.5% and a slope coefficient of -0.1. They are significantly (at the 5% level) different from 0 and 1 respectively. This means that the survey forecasts underestimated the average ex post capital gain and moved in the wrong direction.[8]

To summarise, the Livingston survey is a much worse forecaster (out-of-sample) than the most common one-variable prediction equations—the performance is as bad as a typical "too large" forecasting model. Since I am using the median forecast from the survey, this should be considered as an upper boundary of the forecasting performance of a randomly picked survey participant.[9]

Figure 2 illustrates the time profile of the out-of-sample predictability. The figure plots the out-of-sample R^2 calculated since 1952: the first sample is for 1952–1961 and the last is for 1952–2005 and therefore coincides with the second column in Table 1. Most of the variables have a fairly even performance over time, but the dividend yield only worked well in the early 1970s (as shown by a positive slope of the curve). Predicting zero seems to have been a terrible idea during the 1950s, but a good one 1965–1975, and as good as using the historical mean thereafter. The survey forecasts were poor for most of the period, except the late 1960s to 1980.

The general impression from Figure 2 is that the evidence of weak forecasting performance reported earlier is fairly robust across subsamples.

[6] Other non-linear restrictions, for instance, using the max of the prediction and the negative of the recent dividend yield, give similar results.

[7] This is similar to the results in Bondt (1991), who finds that professional forecasters have virtually zero forecasting performance.

[8] This stands in contrast to the findings by Dokko and Edelstein (1989).

[9] In contrast, much evidence shows that the Livingston survey performs well in forecasting macro variables like inflation and output. See Croushore (1997) for an overview.

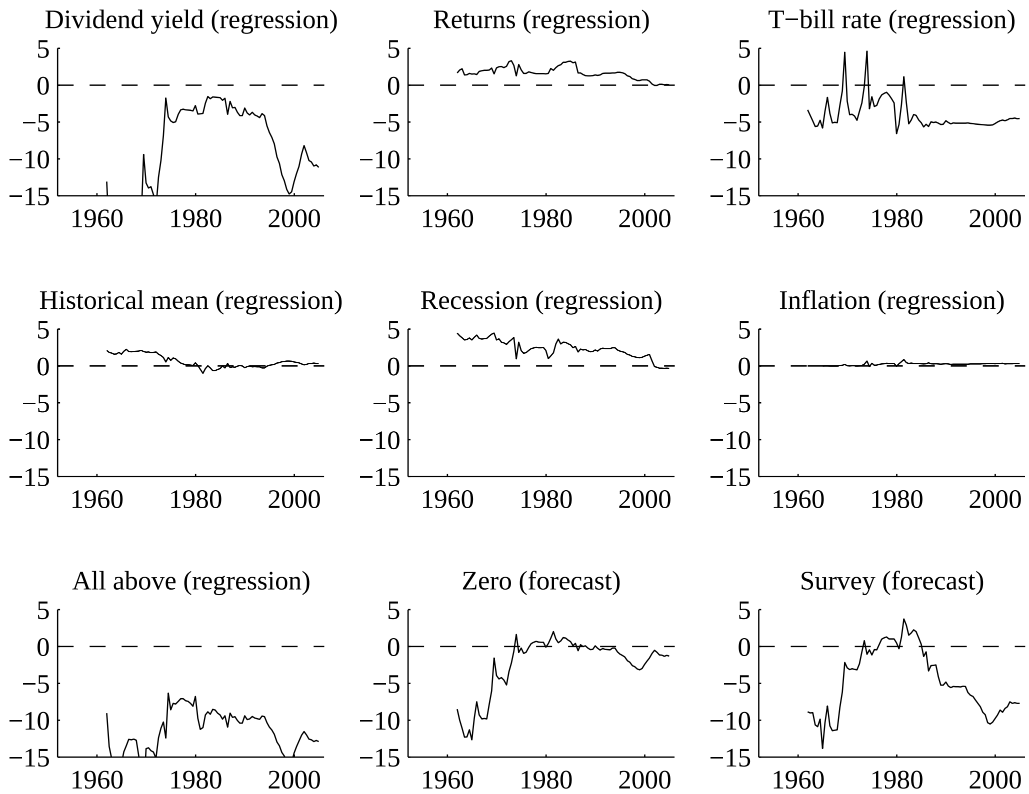

Figure 2: Recursive out-of-sample R^2 from forecasting 6-month excess capital gains with different models, %. This figure shows the out-of-sample R^2 for samples that start in 1952 and end in the period marked on the horisontal axis. The first sample is for 1952-1961.12 and the last sample is for 1952–2005.12.

4 Characterising the Survey Forecasts

This section studies the relation between the survey forecasts and traditional predictors.

Figure 3 shows the survey predictions and the NBER recessions (marked by shaded areas). The Livingston forecasters clearly did not believe in the random walk hypothesis: there are distinct movements in the expected capital gains. In particular, the expectations have local maxima in almost all recessions, which suggests a belief in a medium term mean reversion.[10] (Recall that the forecasts are for the 6 months starting 6 months ahead in time.)

[10]The NBER recessions are declared after a long period (for instance, the November 2001 trough was declared only in July 2003), but they still serve as reasonable proxy for the perceived (in real-time) state of the business cycle.

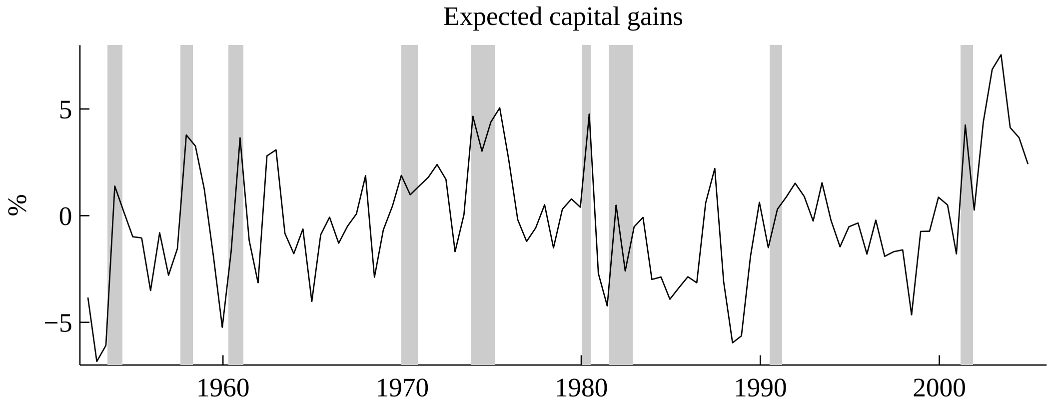

Figure 3: Expected excess capital gains on the combined S&P index, %. This figure shows the survey predictions of the capital gains on the S&P index (in excess of a riskfree rate) for the 6-month horizon starting 6 months ahead in time and the NBER recessions (shaded areas).

Indeed, *Table 2* shows that the correlation of expected capital gains (over month 7-12) with current ex post returns (over the last 6 months) is negative, and the correlation with a dummy variable for the NBER recessions is strongly positive. Interestingly, the forecasts have a negative correlation with dividend yields and a positive correlation with inflation—in contrast to the ex post capital gains.

The main point, however, is that the survey forecasts are strongly related to the predictors. For instance, in Table 2 most of the correlations with the survey forecasts are significant at the 5% level (none of the correlations with the ex post capital gains are), and regressing the survey forecasts on all the predictors gives an R^2 of almost 50%. This strengthens the impression from the analysis of the forecast performance: it seems as if the survey forecasts are similar to a "too large" forecasting model. This means that the forecasters use "models" that suffer from overfitting and recency bias.[11]

5 Conclusion

This paper analyses the forecasts of equity price changes in the Livingston survey for the sample from 1952 to 2005. The results indicate that the survey forecasts perform clearly worse than the historical mean (estimated on a recursive sample), and that they share many properties with "too large" prediction models.

[11] See Bondt and Thaler (1985) and Bondt and Thaler (1990) for studies of recency bias.

	Ex post	Survey forecasts
Dividend yield	14.8	−20.7
Returns	−4.9	−17.1
T-bill rate	−8.6	−21.2
Historical mean	−7.6	−7.1
Recession	12.6	30.2
Inflation	−11.2	21.4

Table 2: Correlations of ex post and predicted 6-month excess capital gains with different predictors, %. The table shows results for the 6-month horizon starting 6 months ahead in time. See Figure 1 for details on the data.

There are two implications of this finding. First, if this group of forecasters cannot predict stock price changes, should we then really pay any attention to the small set of (carefully selected) regressions that can? Maybe they are just type 1 errors. If so, portfolio recommendations should not rely on predictability. Second, it still seems as if these forecasters thought they could predict price changes: the expectations vary markedly—and are strongly correlated with traditional predictors. Maybe studies (as opposed to recommendations) of portfolio choice and asset pricing should incorporate beliefs on time-varying expected stock price movements.

A Data Appendix

The data of the Livingston Survey is from the Federal Reserve Bank of Philadelphia (http://www.phil.frb.org/).

The monthly S&P 500 and S&P Industrials price indices are from S&P's Trade and Securities Statistics (1976) and Compustat. The dividends on the S&P 500 and the inflation rate are from Shiller's homepage (http://www.econ.yale.edu/%7Eshiller/data.htm), see Shiller (2000). For a discussion of the historical S&P series, see Jones and Wilson (2002).

The 3-month T-bill rate is from FRED II (http://research.stlouisfed.org/fred2/).

References

Batchelor, R., and P. Dua, 1995, "Forecaster Diversity and the Benefits of Combining Forecasts," *Management Science*, 41, 68–75.

Bates, J. M., and C. W. J. Granger, 1969, "The Combination of Forecasts," *Operations Research Quarterly*, 20, 451–468.

Bondt, W. F. M. D., 1991, "What Do Economists Know about the Stock Market?," *Journal of Portfolio Management*, 17, 84–91.

Bondt, W. F. M. D., and R. H. Thaler, 1985, "Does the Stock Market Overreact?," *Journal of Finance*, 40, 793–805.

Bondt, W. F. M. D., and R. H. Thaler, 1990, "Do Security Analysts Overreact?," *American Economic Review*, 80, 52–57.

Campbell, J. Y., and S. B. Thompson, 2005, "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average," NBER WP 11468.

Croushore, D., 1997, "The Livingston Survey: Still Useful After All These Years," *Federal Reserve Bank of Philadelphia Business Review*, pp. 15–26.

Dokko, Y., and R. H. Edelstein, 1989, "How Well Do Economists Forecast Stock Market Prices? A Study of the Livingston Surveys," *The American Economic Review*, 79, 865–871.

Federal Reserve Bank of Philadelphia, 2004, "Documentation for the Discontinued Stock Price Forecasts in the Federal Reserve Bank of Philadelphia's Livingston Survey," http://www.philadelphiafed.org/econ/liv/S&P.html.

Federal Reserve Bank of Philadelphia, 2005, "Livingston Survey Documentation," http://www.philadelphiafed.org/econ/liv/index.html.

Goyal, A., and I. Welch, 2004, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," NBER WP 10483.

Jones, C. P., and J. Wilson, 2002, "An Analysis of the S&P 500 Index and Cowles's Extensions: Price Indexes and Stock Returns 1870–1999," *Journal of Business*, 75, 505–533.

Pearce, D. K., 1984, "An Empirical Analysis of Expected Stock Price Movements," *Journal of Money, Credit and Banking*, 16, 317–327.

Shiller, R. J., 2000, *Irrational Exuberance*, Princeton University Press, data available at http://www.econ.yale.edu/ shiller/data.htm.

Winkler, R. L., and S. Makridakis, 1983, "The Combination of Forecasts," *Journal of the Royal Statistical Society A*, 146, 150–157.