(25) Stock, James H., Watson, Mark W. "Forecasting Output and Inflation: The Role of Asset Prices." Journal of Economic Literature, 2003.

The paper assesses the usefulness of asset prices as predictors of inflation and output growth. After extensively reviewing the relevant literature, they undertake an empirical analysis of the predictive content of asset prices and other leading indicators for output growth and inflation using data for seven OECD countries covering the period 1959-1999.

Testing forecasts of inflation by comparing the pseudo out of sample mean square forecast error of the predictor to that of an AR forecast benchmark, while dividing the data into two periods subsamples, the results show that some variables forecast relatively well in some countries on one or the other subsamples. For example, the housing price inflation and commodity prices works well in the U.S. in the first period only, the short rate works well in France in the first period, monetary aggregates work well in Germany in the first period, and commodity prices works well in Canada in the second period only. The measures of aggregate activity, such as industrial production, unemployment gaps work well in Canada, US, and Germany in both periods, but even for these predictors, the improvement upon the AR forecasts is neither universal nor always stable.

Similar results are obtained testing forecasts for output growth, where the results show that term spread outperforms the AR benchmark in the first period in the US and Germany, and in the second period in Canada and Japan, the short rate works well in Japan in the second period, and in Germany and the US in the first period, while the nominal exchange rates work well in the second period for Canada, Germany, Italy and Japan, and real stock returns work well for Canada, Germany, Japan and the US in the first period only.

Looking at the stability of forecasts made using a given predictor-country-horizon combination, they find that 6% of all these combinations outperform the AR benchmark in both periods, 18% outperform it in the first period only, 29% in the second only and 46% are worse than the benchmark in both periods. The conclusion they derive is that whether an asset price-country-horizon-dependent variable combination outperforms the benchmark in one period is nearly independent of whether it does so in the other period. In addition, it appears that performance in the two periods is nearly unrelated, or if it is related the correlation is negative, such that asset prices that perform well in the first period tend to perform poorly in the second period. They finally conclude that there appear to be no subsets of countries, predictors, horizons, or variables being forecast that are immune to this instability, where forecasting models that outperform the AR in the first period may or may not outperform the AR in the second, but whether they do appears to be random.

They then turn to in-sample measures of predictive content and stability for these asset price based forecasts, and the results of full sample Granger causality tests for predictive content and the Quandt likelihood ratio QLR tests for

instability suggest three conclusions: first, the Granger causality tests rejects the null hypothesis of no predictive content for 35% of asset prices, indicating that a large fraction of these relations have substantial in-sample predictive content. Over all categories of predictors, 40% reject Granger non-causality for inflation, and 44% reject Granger non-causality for output growth. Second, the QLR statistic detects widespread instability in these relations, as it rejects the null hypothesis of stability in 78% of the inflation forecasting relations and in 71% of the output forecasting relations. Third, a statistically significant Granger causality statistic conveys little if any information about whether the forecasting relation is stable, such that a significant Granger causality statistic makes it no more likely that a predictor outperforms the AR in both periods.

They also considered forecasts based on trivariate models, where those for inflation included lags of inflation, industrial production gap, and the candidate predictor, while those for output growth included lags of output growth, the term spread, and the candidate predictor. They find that the main conclusions drawn from the bivariate models also hold for the trivariate models.

They further examine the possibility that combining forecasts based on the individual indicators can improve their performance, as by pooling forecasts based on different data, the combined forecast uses more information and thus should be more efficient than any individual forecast. They find that the trimmed mean of all individual forecasts of inflation outperforms the benchmark AR in most of the country-horizon-period combinations. For forecasting output growth, adding predictors beyond asset prices does not reliably improve upon the combination forecasts based on asset prices. Results for combination forecasts based on the trivariate models show that the trivariate forecasts typically improve upon the benchmark AR forecasts, however the improvements are not as reliable, nor are they as large as for the bivariate forecasts.

From all these results, they reach four general conclusions: first, some assets have been useful predictors of inflation and/or output growth in some countries and in some time periods. Second, there is considerable instability in bivariate and trivariate predictive relations based on asset prices and other predictors, which suggest that asset prices that forecast well in one country or in one period might not do so in another. Third, in-sample Granger causality test provide a poor guide to forecast performance, as they find that rejection of Granger causality statistic provides no information about whether the relation will be useful for forecasting. Finally, simple combination forecasts reliably and stably improve upon the AR benchmark and forecasts based on individual predictors.