Lab: Recursive Regression Prediction of Equity Returns

There have been numerous studies demonstrating that a substantial proportion of variation in excess equity returns is predictable. In this lab we replicate some of the regression models applied in Pesaran and Timmerman (1994) to low frequency equity returns. Regression models are used to generate recursive predictions of excess returns on the S&P500 index using only *ex-ante* dated variables, using monthly data for the period March 1954 to December 1992.

The model we focus on is:

$$Y_t = b_0 + b_1 YSP_{t-1} + b_2 PI12_{t-2} + b_3 DI11_{t-1} + b_4 DIP12_{t-2} + e_t$$

Where:

- $\mathbf{Y_t}$ is the excess return on the S&P500 over the 1-month T-Bill rate.
- **YSP** is the dividend yield, defined as:

  12-month average dividend / month-end S&P500 Index value

- **PI12** is the rate of change of the 12-month moving average of the producer price index:

  PI12 = Ln{PPI12 / PPI12(-12)}

- **DI11** is the change in the 1-month T-Bill rate
- **DIP12** is the rate of change of the 12-month moving average of the index of industrial production

  DIPI12 = Ln{IP12 / IP12(-12)}

- $\mathbf{e_t}$ is white noise

All of the explanatory information should be publicly available at time t-1 , when excess returns for time t are being forecast. Data on Industrial Production and the Producer Price Index are released by the US Government with a delay, and hence these variables are included with lags of length 2.

Under the EMH, it should not be possible to use such publicly available data to predict excess returns.

1. Using the EXCEL LINEST function to compute recursive estimates of the five model coefficients and a one-period ahead forecast of the excess return Y_t and its standard error for the period Jan 1960 through to Dec 1992.

Estimates and forecasts for the first period (Jan 1960) should be produced using the data for the preceding period Mar 54 – Dec 59. For each successive period we re-estimate the regression using one addition period of data.

Examine how the estimates of the one of the parameters fluctuates as the recursion proceeds.

2. Using the LINEST function, produce final estimates of the parameter values, their standard error and t-test statistics using the entire data set.

3. Use the LINEST function to compute the F statistic and the coefficient of determination for the final regression model. What is the strength of the relationship between the forecast and actual excess return series?

4. Perform the following diagnostic tests on the model:

    - Examine the residual plot for evidence of bias or heteroscedasticity.
    - Examine the ACF and PACF of the residuals for evidence of serial autocorrelation.
    - Perform Box-Pierce and Durbin-Watson tests for serial autocorrelation (use the user-defined functions Box-Pierce and DW).
    - Estimate the skewness and kurtosis of the error distribution and test for non-Normality using the Jarque-Bera test. The J-B test statistic, which is asymptotically distributed as $\chi(2)$, is:

$$J\text{-}B = n[\text{Skewness} / 6 + (\text{Kurtosis} - 3)^2 / 24]$$

What do you conclude about the validity of the regression model?

5. Consider a simple trading system in which, at the start of each period, we go long or short the index depending on whether 1-period ahead prediction of the excess return is positive or negative. Calculate the excess return generated in each period, and examine the performance of the model using the following criteria:

    - Total excess return vs. Buy and Hold
    - Average annual excess return
    - Sharpe ratio
    - % Correct sign predictions
    - SIGN / Pesaran-Timmerman test

Solution: Recursive Regression Prediction of Equity Returns

1. We begin by using the LINEST function to compute OLS parameter estimates using the data for the initial period to Dec 1959. The results, shown in row 72 of the solution sheet are generated using the Excel formula:

   I72 =INDEX(LINEST($B$2:$B71,$C$2:$F71),1,i)

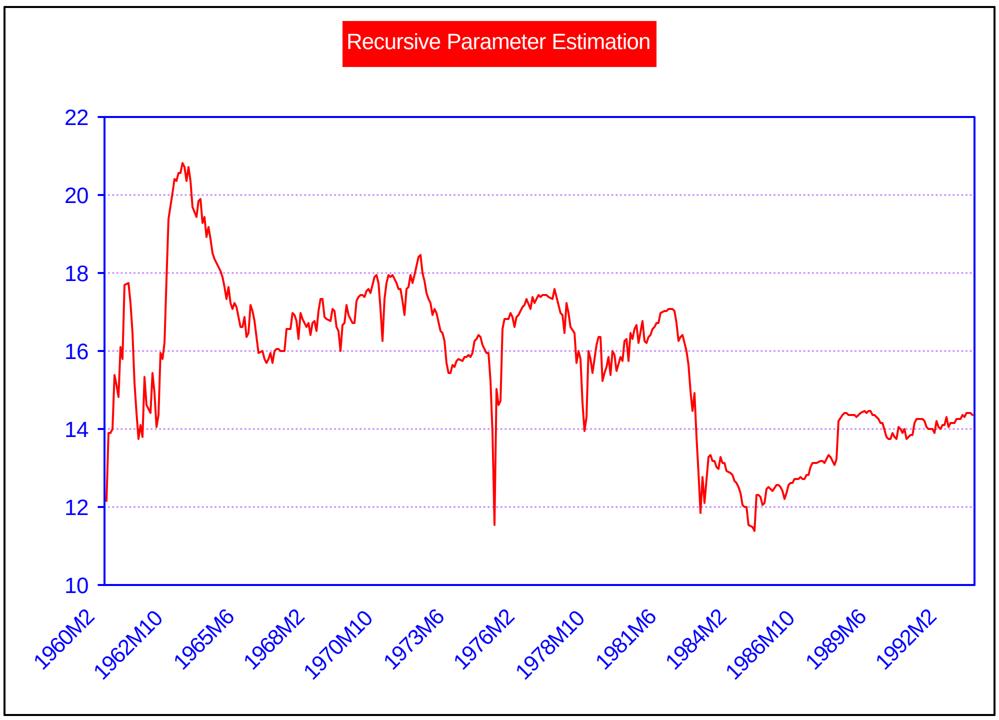
   Where I is 5 for the constant, coefficient, 4 for the first coefficient, etc.

   The 1-period ahead forecast of excess returns, F_{t+1}, is produced by applying the current parameter estimates in the regression equation. The Excel formula is:

   N72 =I72+SUMPRODUCT(C72:F72,J72:M72)

   The error terms $e_t = Y_t - F_t$ are calculated in column O of the worksheet.

   We can examine how the recursion proceeds by looking at the fluctuating estimate of the parameter β_1. This would appear to suggest the possibility of a major structural change, or regime switch, towards the end of 1975 and again at the end of 1983 in which the size of the impact of dividend yields on future excess returns appears to have declined substantially.



2. The final results of the recursive regression, in which all of the data is included for estimation purposes, are summarized in the table below.

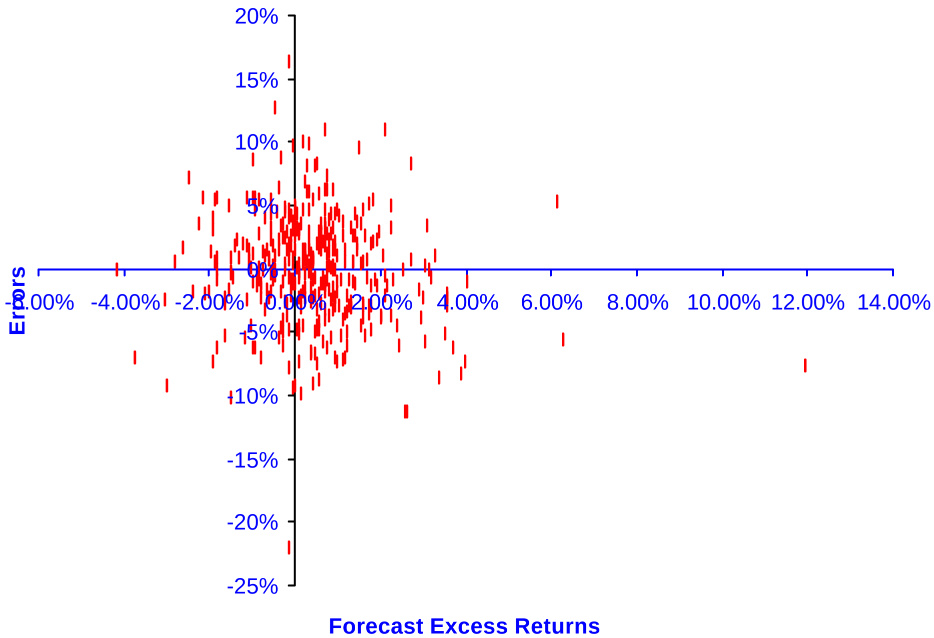
   Subject to the usual regression assumptions with regard to Normality, etc, all of the regression coefficients would appear to be significant to at least the 1.5% level.

PARAMETERS	-0.024	14.338	-0.280	-0.007	-0.159
SE	0.010	3.424	0.065	0.003	0.040
t-statistic	-2.442	4.188	-4.321	-2.763	-3.941
Prob	1.497%	0.003%	0.002%	0.595%	0.009%

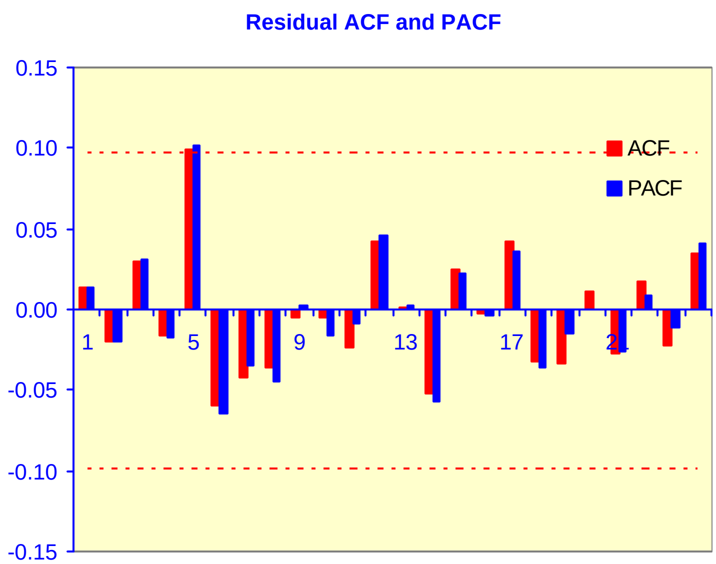
3. The overall ANOVA is shown in the table below. The regression model appears to have statistically significant explanatory power, indicating the excess returns are indeed predictable from ex-ante variables.

ANOVA					
R^2	8.6%				
Correl	20.7%				
F	10.82	DF	461.00	Prob	0.000%

4. The residual plot (shown below) displays no obvious signs that the regression assumptions are invalid. There does, however, appear to be at least one outlier (the forecast excess return of 11.93% in May 1980).



The correlogram of the residuals ACF and PACF displays no evidence of significant (partial) autocorrelations, and this is confirmed by both the D-W statistic (1.97) and the Box-Pierce statistic (12.2).



On the other hand the error distribution has a negative skew and kurtosis well below the standard Normal value of 3. The value of the Jarque-Bera test statistic (13.67) is statistically significant at the 0.1% level, indicating a departure from the Normality assumption (in the third and fourth moments). This result calls into question the validity of parameter significance tests and forecast confidence intervals.

5. The results of the trading simulation exercise are shown in the chart and table following. The simple trading rule appears to be capable of exploiting the predictability of excess returns to produce substantial excess returns in comparison with a simple buy-and-hold strategy. This results chiefly from the remarkable ability of the model to correctly predict the direction of the market: close to 74% correct sign predictions, compared with the expected 50% under EMH. As the Sign test shows, the result is highly significant.

   The conclusion is the low frequency excess returns are indeed predictable using ex-ante variables and, potentially, a successful trading strategy could be developed on the basis of such a simple regression model. In their 1994 study, Pesaran and Timmermann went on to demonstrate the potential for generating abnormal returns using this approach, even in the presence of sizeable transaction costs.

PERFORMANCE	Excess Return	# Trades	396
Total	291.07%	# Correct Sign	293
Over Buy & Hold	120.26%	% Correct Sign	74.0%
Av Annual	8.82%	Sign Statistic	9.55
St Dev	14.86%	Prob	0.000%
Sharpe Ratio	0.59	Pesaran-Timmerman	9.55
		Prob	0.000%