3 predicting asset returns

...is something we all would like to master. In this session, we will try to predict returns of some popular assets.

1. Load `s3_data.txt` into RStudio. This dataset contains the following series:

name	description	source
SP500	S&P 500 index	R. Shiller's website
SPDIV	dividend per "share" of the S&P500 index	
FINANC	S&P Financial Select Sector index	Datastream
VIX	CBOE Volatility Index	St. Louis Fed
eurusd	return in USD of holding 1 Euro	
rf	3-Month Treasury Bill, secondary market rate (de-annualized)	

   Calculate *sp500*: the excess return of *SP500*. Leave out the first observation of the sample.

2. Let's first take a look at the autocorrelograms (ACF plots) of *SP500* and *sp500*. Do you see a unit root? Use 10 lags for your plots.

3. We can also formally test for the presence of a unit root in each series. To this end, conduct two unit root tests (include a constant, but no trend): the ADF test and the KPSS test.

   In what follows we will deal with excess returns only, and whenever a return is mentioned, an excess return is meant (except for *eurusd*).

4. In order to test for the presence of serial correlation in *sp500*, perform a Ljung-Box test with 12 lags.

5. Let's apply a demeaned AR(1) model (using MLE). Such a model assumes that the best prediction of a return today is a fraction of yesterday's return.

$$r_t - \mu = a(r_{t-1} - \mu) + u_t \tag{1}$$

   Using *sp500* data until 2005:05, report a forecast for 2005:06.

6. Perform some model checking using the function `checkresiduals()`. If the model is adequate, the residuals should be uncorrelated and have a zero mean. In addition, constant variance and normal distribution are two useful (but not necessary[1]) properties.

7. Repeat the AR(1) forecasting till the end of the sample, creating a series of recursive (expanding window) forecasts.

[1] It can be shown that MLE behaves asymptotically nice even if we drop the assumption of normally distributed errors. The estimates will tend towards those obtained with normal residuals. This situation is called quasi maximum likelihood (see section 10.4 from the lecture notes).

8. Plot the time-series of the forecasts aligned with the actual values.

9. Calculate the RMSE.

10. Another possibility is that something else can forecast *sp500*:

$$\begin{aligned} r_t &= \alpha + \beta_1 f_{1,t-1} + \beta_2 f_{2,t-1} + \ldots + \beta_K f_{K,t-1} + u_t \\ &= \alpha + F_{t-1} B + u_t, \end{aligned} \tag{2}$$

where F_{t-1} and B_{t-1} are a $1 \times K$ and $K \times 1$ vectors of factors and corresponding loadings respectively. Let us choose yesterday's excess return of *SP500*, last month's appreciation of EUR with respect to USD and last month's value of *VIX* as predictors. Estimate model (2) with OLS and come up with an out-of-sample forecast for 2005:06.

11. Repeat till the end of the sample; report the RMSE.

12. It would be convenient to express the forecasting power of a model relative to some benchmark model. From this perspective, perform a one-sided Diebold-Mariano test[2], with $H_0 : e_t^2 - \epsilon_t^2 \leqslant 0$. You want to determine whether ϵ_t^2 (the squared forecast error from your multi-factor model) is "smaller" than e_t^2 (the squared forecast error of the benchmark AR(1) model), in which case you reject the null.

13. In the same vein, compute the out-of-sample R^2, defined as:

$$R^2_{os} = 1 - \frac{(\text{RMSE})^2}{(\text{RMSE})^2_b} \tag{3}$$

where RMSE_b is the RMSE of the benchmark model. Calculate R^2_{os} of the multi-factor model relative to the AR(1) model.

[2]More precisely, the `forecast` package function `dm.test()` uses a modified version of the test developed by Harvey, Leybourne and Newbold (1997). It introduces finite sample modifications to the DM statistic to improve size control in small samples, with a finite sample bias correction. Student-t critical values are used rather than standard normal ones.

homework

0. Load `s3_data.txt` into RStudio. Calculate the excess returns of *SP500* and *FINANC*. We will try to forecast the latter (*financ*). Calculate also the (logarithm of) price-dividend ratio, a popular variable among predictors:

$$pd_t = \log \frac{P_t}{D_t} = p_t - d_t, \tag{4}$$

Calculate this ratio using *SP500* and *SPDIV*.

0. Get rid of the 1999 data. We will only consider observations from 2000 onwards.

1. Let's first consider an AR model. Using data from 2000:01 to 2009:12, construct a correlogram of *financ* with 12 lags to see which lags are individually significant at the 95% level. Do the same for the PACF (using this time the `ggPacf()` function). Interpret your two plots. **(2 points)**

2. The function `ar.mle()` allows us to select the order of an AR(p) model according to Akaike information criterion (AIC). Report your result and relate it to your previous answer. **(2 points)**

   Hint: Your function could look like this: `ar.mle(as.ts(yourseries))`.

3. Using data from 2000:01 to 2009:12, estimate a demeaned AR(p) where p is the order derived in the previous question. Doing some model checking, comment on the `checkresiduals()` output. **(2 points)**

   For those who did not answer the previous question: use $p = 6$.

4. Based on your AR estimation, make a forecast of *financ* in 2010:01. Report your forecast and compute a 90% confidence interval around it. **(1 point)**

5. Gradually increasing the estimation window length, keep on forecasting till the end of the sample; calculate the RMSE of this model. **(2 points)**

6. Price-dividend ratios are often assumed to be stationary, although in small samples this could be difficult to prove. Conduct two stationarity tests (include a constant but no trend): the ADF test and the KPSS test. Report the test statistics and the critical values at the 1% level. Give your conclusion. **(2 points)**

7. Estimate by OLS the following factor model[3], using the data from 2000:01 to 2009:12:

$$\begin{aligned} r_{FINANC,t} = \alpha + \beta_1 r_{SP500,t-1} + \beta_2 r_{EURUSD,t-1} + \beta_3 (VIX)_{t-1} + \beta_4 (pd)_{t-1} \\ + \beta_5 r_{FINANC,t-1} + \beta_6 r_{FINANC,t-2} + u_t \end{aligned}$$

   Come up with a forecast for 2010:01. Report the forecast and the 95% confidence interval around it. **(2 points)**

[3] Surely it would be better to use the price-dividend ratio for the financial sector only, not for the broad market.

8. Gradually increasing the estimation window length, repeat the forecasting till the end of the sample. Plot predicted and realized values on the same chart and report the RMSE of this model. **(2 points)**

9. Perform a two-sided Diebold-Mariano test, comparing the forecast errors of the two models. Adopt a quadratic loss function. Based on your result, determine whether the two forecasts have the same accuracy. Give your conclusion at a 5% significance level. **(3 points)**

10. Finally, combine your two forecasts, computing a simple mean at each observation (forecast averaging). Derive the RMSE and comment. **(2 points)**