Eco311: Testing Efficient Market Hypothesis

(Jing Li, Miami University)

1. Big picture: a financial theory called Efficient Market Hypothesis (EMH) states that the change in stock price is unpredictable. Time series regression is commonly used to test whether EMH is valid.

2. Let p denote stock price such as SP500 index. Consider a random walk model for p

$$p_t = p_{t-1} + u_t \quad (random\ walk) \tag{1}$$

where the error term u represents unpredicted news or shock. The random walk model implies that the price would not move $p_t = p_{t-1}$ in the absence of shock. Put differently, the price changes only when unpredicted shock occurs. One example is President Trump announced a new tariff of 125 percent, as opposed to 25 percent expected by the market.

3. According to EMH, market prices fully reflect all available information at any given time. As a result, prices adjust only when unexpected or new information arrives. That is why testing EMH is equivalent to testing the random walk model.

4. Nevertheless, due to a technical issue called unit root (price does not have constant mean value or variance), it is more convenient to focus on the change in log price, denoted by r

$$r_t = \log(p_t) - \log(p_{t-1}) \approx \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{u_t}{p_{t-1}}, \quad (return) \tag{2}$$

We call r return because it approximates the return of a very simple investing strategy: buy at $t-1$ and sell at t (assuming no dividend payment, stock split, etc). The last equality shows that the return r is closely related to shock u.

5. By definition u is unexpected and unpredictable. Therefore, testing the (weak form of) EMH amounts to testing the unpredictability of return. If we consider autoregression, that means we can test whether the lag return helps predict current return

$$r_t = \rho r_{t-1} + e_t \tag{3}$$

EMH fails if we reject the null hypothesis

$$H_0 : \rho = 0, \quad (first\ lag\ does\ not\ matter) \tag{4}$$

A t value of ρ less than 1.96 in absolute value supports EMH.

6. The R codes below download SP500 data from FRED web, compute the return r, and run AR(1) regression (3) for r :

```
> library(tseries)
> library(quantmod)
> symbols = c("SP500")
> getSymbols(symbols, src = "FRED")
> r = diff(log(SP500[,1]))
> rlag1 = c(NA, r[1:(length(r)-1)])
> summary(lm(r~rlag1))$coef
                 Estimate  Std. Error   t value     Pr(>|t|)
(Intercept)  0.0004656203 0.000240757  1.933984 5.323636e-02
rlag1       -0.1340176867 0.020492735 -6.539766 7.551428e-11
```

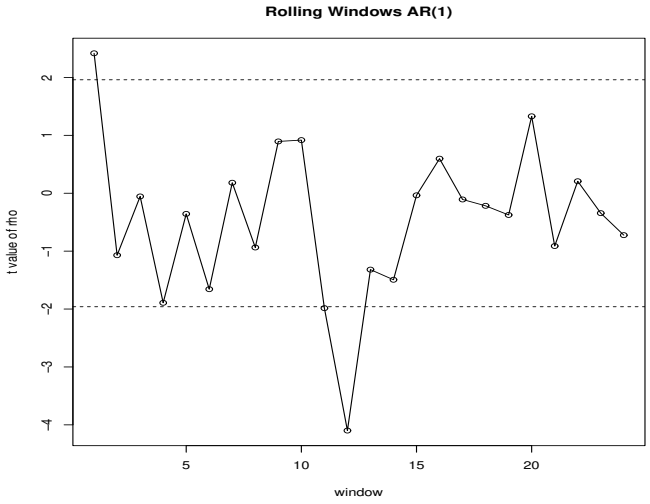
The result does not support EMH as t value -6.539766 exceeds 1.96 in absolute value.

7. In fact, $\rho = -0.1340176867$ being negative implies that we can predict that, in the absence of unexpected news, the market will fall tomorrow $r_{t+1} < 0$ if the market rises today $r_t > 0$. Or, if the market falls today, tomorrow the market will rise. Some people interpret the negative rho as evidence for market overacting (overshooting) at one time and correcting itself subsequently. Another interpretation for negative rho is that the market is "mean-reverting", i.e., the price tends to move toward the mean value or trend once deviating from it. Those self-correcting and mean-reverting arguments rationalize technical analysis that exploits trends, momentum, and other patterns in historical data

8. To summarize, rejecting EMH gives credibility to technical analysis that assumes markets may not be fully efficient, and patterns, anomalies, or opportunity could exist.

9. Be aware that the reported result is based on just ONE sample, and it may change when different sample is used. So next we divide the whole sample into several

rolling windows or subsamples: the first window includes $t = 1, 2, \ldots, 100$; the second window uses $t = 101, \ldots, 200$, and so on. We run the regression (3) repeatedly with the subsample data. The T values of rho are plotted below (where two dash lines denote minus plus 1.96)



We see that among 24 regressions, only three have significant t value of rho, supporting the view that EMH holds most of time. Furthermore, quite a few rhos are positive, casting doubt on the mean-reverting and self-correcting arguments.

10. We can extend the basic analysis by augmenting (3) with additional predictors

$$r_t = \sum_j \beta_j x_j + e_t, \tag{5}$$

where the regressors may include multiple lag values of return, and other factors such as volume. EMH violates if any of β is consistently statistically significant[1].

Takeaway: EMH argues that if new information comes out (like a company's earnings report or macroeconomic news), the market reacts immediately and adjusts prices accordingly. So, by the time you act on that information, it's already "priced in". This makes it very hard or impossible to forecast future movement of price. Whether or not EMH is supported by empirical data is an open question.

[1]Read `http://www.fsb.miamioh.edu/lij14/420n_paper_emh.pdf` for more discussion