I. Module 2 Homework 1 [practice]: A Little Model of Time-Varying Expected Returns

Due No due date **Points** 18 **Questions** 10 **Time Limit** None
Allowed Attempts 2

Instructions

Expected returns vary over time. Here is a nice structure we use to represent this idea:

$$x_t = \phi x_{t-1} + \varepsilon_t$$
$$r_{t+1} = x_t + \delta_{t+1}$$

x_t denotes the expected return, and r_{t+1} denotes the actual log return. (We usually run these in logs, I showed levels in class because it's easier.) Actual returns are expected returns x_t plus unpredictable noise δ_{t+1}. ε_{t+1} and δ_{t+1} can be correlated -- good returns can be positively or negatively associated with good news about expected returns. In fact, ε_{t+1} and δ_{t+1} are negatively correlated -- when prices go up we have a good actual return $\delta_{t+1} > 0$ but it's bad news for subsequent expected returns $\varepsilon_{t+1} < 0$. In the lecture, I used $x_t = a + b \times dp_t$, but we more generally think of expected returns as following a latent (we can't observe it directly) state variable of this form, and then prices reveal x_t to us.

We use this sort of time series model widely in finance -- for example, all the term structure models are built this way. It's worth getting familiar with it.

When the problem calls for numerical values, use $\sigma_\varepsilon = 0.018$, $\phi = 0.94$, $\sigma_\delta = 0.18$ and the correlation between ε_t and δ_t shocks $\rho = -\frac{\phi}{1-\phi^2}\frac{\sigma_\varepsilon}{\sigma_\delta} = -0.80756$. These numbers are close to what we see in dividend-yield regressions, and the latter number reverse-engineers a very nice special case.

Attempt History

	Attempt	Time	Score
KEPT	Attempt 2	39 minutes	13 out of 18
LATEST	Attempt 2	39 minutes	13 out of 18
	Attempt 1	138 minutes	5 out of 18

Correct answers are hidden.

Submitted Jun 12 at 8:04pm

Incorrect

Question 1

0 / 1.5 pts

Suppose you observe x_t directly and perfectly. If you were to run a regression of returns on x_t,

$$r_{t+1} = a + bx_t + e_{t+1}$$

in large samples, calculate the values you expect to see of the regression coefficient b, the standard deviation of expected returns $\sigma(a + bx)$, the standard deviation of actual returns $\sigma(r)$ and the R^2 of this regression. (I used notation e_{t+1} because this is not the "true" δ_{t+1} defined in the model, but a regression residual.)

Find the formula for each of these quantities in terms of the given parameters. Then compute them, and report them in order, as numbers separated by spaces. Report standard deviations in net (0.05) not percent (5.00) units. You should input a total of four numbers.

Note, the next question asks you to check these answers by simulation. If you're having trouble with finding the numbers by formulas, you might want to do that simulation first.

Here report the numerical value of the regression coefficient b. You will report the other three numbers in the following three questions.

Note: Questions 1-4 are "Numerical Answer" questions. Your answers will be marked as correct, if they are within a margin of error.

1.0127

Incorrect

Question 2

0 / 1.5 pts

Here report the numerical value of the standard deviation of expected returns $\sigma(a + bx)$ of the regression described in Question 1.

0.5077

Question 3

1.5 / 1.5 pts

Here report the numerical value of the standard deviation of actual returns $\sigma(r)$ of the regression described in Question 1.

0.1876

Question 4

1.5 / 1.5 pts

Here report the numerical value of the R^2 of the regression described in Question 1. Report the numer as a fraction of one.

0.08

Question 5

3 / 3 pts

Simulate a long time series -- at least 100,000 points -- of this model. Use the simulation to check your answers to the last question -- run a regression in simulated data, and see if you get the b, $\sigma(a + bx)$ and R^2 that you expected. (You can use standard errors to gauge the uncertainty due to a finite sample.)

Hint: you can create a vector of correlated variables from a random number generator by

$$\begin{bmatrix} \varepsilon_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon & 0 \\ \rho\sigma_\delta & \sigma_\delta\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} z_t \\ w_t \end{bmatrix}$$

where $\begin{bmatrix} z_t & w_t \end{bmatrix}' \sim N(0, I)$. The 2×2 matrix is the Choleski decomposition of a covariance matrix. If you want to produce a vector of random normals with covariance matrix Σ, the fact that chol(Σ)chol$(\Sigma)'$= Σ means that chol$(\Sigma)\times$ a vector of $N(0, I)$ variables will have covariance matrix Σ.

Now, find the coefficient this model predicts for the 5 year return,

$$(r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4} + r_{t+5}) = a_5 + b_5 x_t + e_{t+5}^{(5)}$$

(I added the superscript "$^{(5)}$" to clarify these are different residuals than before.) You should calculate the formula for b_5 from the model, (i.e. in terms of $\phi, \sigma_\epsilon, \sigma_\delta, \rho$ etc.) and then input numbers. Calculating the R^2 in this way is possible but a lot of unpleasant algebra. So instead, run the regression in your simulated data, check the coefficient b_5 and just report the R^2 that you get from the regression in simulated data. (Big picture: This is a nice way to avoid doing a lot of algebra!)

Report the numerical value of the coefficient b_5.

Note: Questions 5-6 are "Numerical Answer" questions. Your answers will be marked as correct, if they are within a margin of error.

4.4783

Question 6

3 / 3 pts

Report the numerical value of the R^2 of the regression described in the previous question. Report the numer as a fraction of one.

0.311

Incorrect

Question 7

0 / 1 pts

Plot 100 points of x_t and r_{t+1} -- line up the data so that the forecast x_t and the actual r_{t+1} are on the same date. If you had only the return data, and could not see the x_t, could you see the variation in average returns buried in the noise of actual returns? This plot may give you a sense of why people missed return predictability for so long.

Now plot 100 years of the 5 year return $(r_{t+1} + r_{t+2} + r_{t+3} + r_{t+4} + r_{t+5})$ along with the predicted value from your model, $a_5 + b_5 x_t$, again lining up $a_5 + b_5 x_t$ with r_{t+5}. This should look quite similar to the graph reported in class using dividend-yield data. The correlation of expected returns with actual returns should be much clearer, as the R^2 is much higher.

To verify your graphs for the problem set, check the correct statements and leave others unchecked.

- [ ] The one-year return plot shows larger returns than the 5year return plot.
- [x] The expected return lines seem to move before actual returns move.
- [ ] The expected return lines are more correlated with actual returns in the one year plot than the 5 year plot.
- [x] The one-year return plot is more jumpy less serially correlated than the 5-year return plot.

Question 8

2 / 2 pts

At this point you should have a sense that this model produces artificial data that looks pretty much like our forecasts of returns from dividend yields, and it integrates the long run and short run observations.

But how is that consistent with our forecasts of returns from past returns? If returns are truly predictable from dividend yields, why do we not see that when we forecast using past returns? After all, the main way a dividend yield gets high is for the price to decline. Does the forecastability of returns from dividend yields mean that returns really are "safer" for long run investors who can afford to wait out the "temporary declines" in stock prices?

To answer all these questions, we need to find what our little model predicts for the forecast of returns based on past returns. It is possible to find algebraically what the univariate return process that our little system implies. It turns out that returns follow an ARMA(1,1), $r_{t+1} = \phi r_t + v_{t+1} - \theta v_t$ where v_t is the regression shock. But rather than do a lot of algebra at this point, let's attack the question by simulation:

In your long artificial data set, run the regression $r_{t+1} = a + br_t + v_{t+1}$, and report b.

Hint: I cooked the problem so that the right answer with infinite data is a round number. Feel free to input the round number rather than the exact value from your simulation.

Note: Question 8-9 are "Numerical Answer" questions. Your answers will be marked as correct, if they are within a margin of error.

-0.0005

Question 9

2 / 2 pts

Report the numerical value of the R^2 of the regression described in the previous question. Report the numer as a fraction of one.

0

Incorrect

Question 10

0 / 1 pts

So, does the fact that expected returns vary over time, "high" prices mean low subsequent returns, and "low" prices will in fact "mean revert" -- price fluctuations are in fact "temporary" -- necessarily imply that stocks are safer at long horizons?

(This is the point of the problem! Link your result in the previous problem to our findings about the variance of long run returns in previous problems -- do k-year returns have a standard deviation more or less than $\sqrt{k}$ times that of one-year returns? You can infer the answer from your last regression, or attack it by simulation if you forgot.)

Answer the correct statements with a check, the incorrect or unsure questions with a blank.

- [ ] Even with this predictability model, the unconditional variance of returns $var(r_{t+1} + r_{t+2} + \ldots r_{t+k})$ is still linear in horizon k, so stocks are no safer in the long run.
- [x] The predictability of returns -- the fact that there really is such a thing as "temporarily" low prices -- means that long-horizon investors face a lot less risk than short horizon investors.