1) "The book says when correlation is high, it is hard to discriminate the effects of the explanatory variable, which leads to a high variance and standard error. And so, there tends to have erratic estimates."

Just to be clear about what erratic estimates means. In formula 3.11 in the book, you can see that when you have two explanatory variables, there is a term \( 1 - r^2_{X_2,X_3} \) at the denominator of your OLS estimators (\( r^2_{X_2,X_3} \) is the square of the coefficient of correlation between two variables). If there is a high correlation between the two variables \( 1 - r^2_{X_2,X_3} \) will tend to be close to zero. Hence, if you are given two different samples even if there are only slight differences in the values of the variables \( Y, X_2 \) and \( X_3 \) in the two samples, you will have large differences in the point estimates \( (b_2, b_3) \) in the two samples because the denominator close to zero will imply very large differences.

Eg. Suppose that \( 1 - r^2_{X_2,X_3} = 0.001 \) in sample 1 and 0.0009 in sample 2. The numerator of 3.11 is equal to 0.5 in sample 1 and 0.52 in sample 2. Then the point estimates will be: 0.5/0.001 = 500 in sample 1 and 0.52/0.0009 \( \approx \) 577.78 in sample 2. This is why the point estimates are said to be erratic.

From the standard error for the estimator \( b_2 \) and \( b_3 \) in formula 3.36, it is clear that s.e.(\( b_2 \)) will tend to be very large if \( 1 - r^2_{X_2,X_3} \) is very close to zero and warn you that the regression estimates are unreliable.

2) "But according to what i learn from last class, larger standard error MSD(\( X_2 \)) (MSD stands for Mean Square Deviation) or MSD(\( X_3 \)) is a good thing as there will be a wider spread of the data along the fitted line and I thought the data fitted line and estimates should then be more reliable."

This is the third way p139 to reduce the multi-collinearity problem.

3) "So what does standard error really means? Is large standard error good or not?"

Large s.e.(\( b_2 \)) is something informative about the (bad) quality of your estimate, while large MSD(\( X_2 \)) is something interesting because it allows to obtain more precise estimates and may allow us to have precise point estimates even if the explanatory variables \( (X_2 \) and \( X_3 \) are "highly" correlated.