Class 2 Homework - Regression and Regularization

Scenario:

You've been hired by HM Inc., who represents a range of manufacturing technology products. They're trying to plan inventory, and need a model to help them set levels.

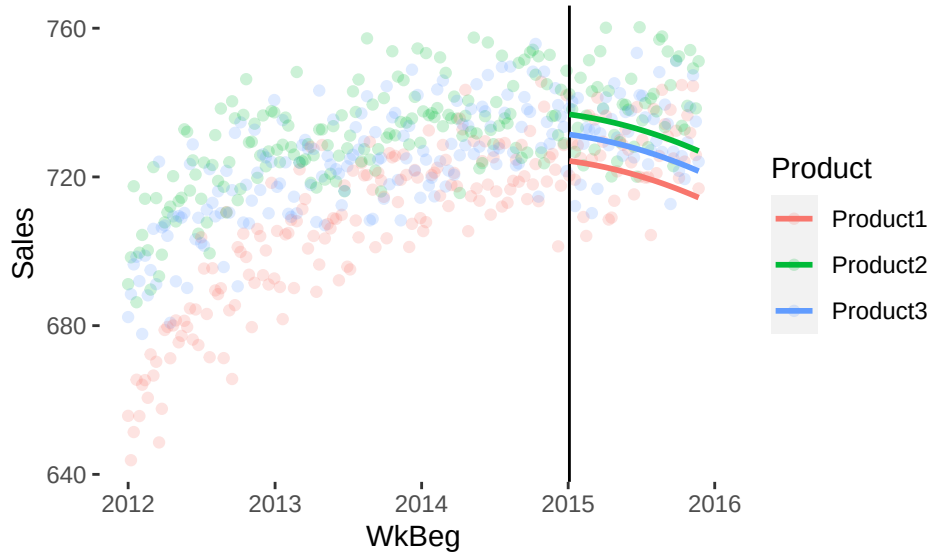
Their Product Life Cycle *(PLC)* team explains that new products often follow a PLC curve with rapid sales growth during the introductory period, but leveling out in the third year, and then slowing in the fourth year as competitive technologies emerge. They have selected 3 products representative of the overall industry PLC. They explain that the model should be useful for planning, and also for auditors to validate revenues.

They give you a **ProductSalesv2.csv** with this data. Use WkBeg < "2015-01-01" for training data, and WkBeg >= "2015-01-01" for **test data** *(cutline shown below for reference)*. The Sales data is scaled in thousands *(so, TOTAL sales in week 204, beg 11/22/2015 for all 3 products was $2.192M)*, so model RMSE is a concern *(big $'s, especially when you consider they rep many products)*

Deliverables

Create the following deliverables:

1. Explanatory model for forecasting sales with RMSE < 12.85k *note: to be clear, RMSE is determined by the* **test data:** $\sqrt{\frac{\sum(y-\hat{y})^2}{n}}$
2. Plot of data and model predictions



3. Determine the Wk where the PLC peaks *(use analytic approach).*
4. Respond to the following client questions:
5. Explain your reasoning and approach with regard to model selection and generalization.
6. If HM introduces a new product most similar to Product 1, what would be the expected sales in Year 1? Year 3?
7. Assume that new product development takes 20 weeks, and they want to introduce the product at the point where sales begin to decrease *(PLC peak).* if they're monitoring slope ($\frac{\Delta Sales}{\Delta Wk}$), at what Wk should we begin product development?