

You bring up an interesting point about the idea of the error being multiplicative. As you say, in general, for a linear model, it seems that the model itself should have an additive error term. I'm curious of how a multiplicative error term would be understood contextually, since this is the first time that I've encountered it. When I thought about it initially, I was thinking that perhaps there was some nonlinear model that produced an error relative to the estimated true line that was modeled as some multiplicative constant, rather than some additive constant. However, I couldn't think of this in any real sense beyond the basic model.

I see how you mention also transforming the model to give it the new additive term. I am curious in what situations it the new error term would follow the general assumptions of an error term such as normality, etc. For example, does it usually map to a normal error term, or is that sort of an edge case?

In the second part about linearization, you mention some important topics from the book such as steepest descent and fractional incrementation. I think it's interesting how nonlinear models are unable to find an optimized solution as simply as the linear model. This seems to be a huge problem in machine learning, where it takes quite a while to train certain models, in particular neural networks. I think it's worth mentioning that neural networks themselves are a nonlinear model.

With the last section you make a good point how in the additive model there is no iterative process and so there is an exact global minimum. This seems to be a good point to compare to the multiplicative models. However, I think this applies more for the normal least squares and not some other strange case where for example we have some nonlinear problem that also has an additive error term. I also agree that looking at the physical science of the model is a helpful in determining suitable values. Also, plotting seems to be an effective method to me when there are limited terms.