Farm Appraisal: A Statistical Analysis

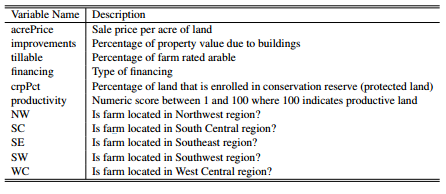
November 22nd, 2016

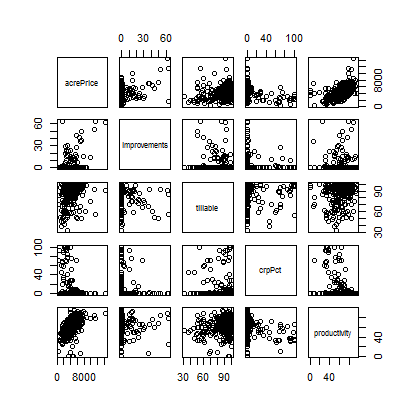
**Introduction and Problem Background**

Farm appraisers are looking for more consistent ways to value a farm prior to sale. They want to ensure that a fair price is reached between the buyer and the seller. The goals of this study are to (1) determine which factors affect the price of the farm and (2) use those factors to help appraise the value of a farm about to go on sale.

**The Data**

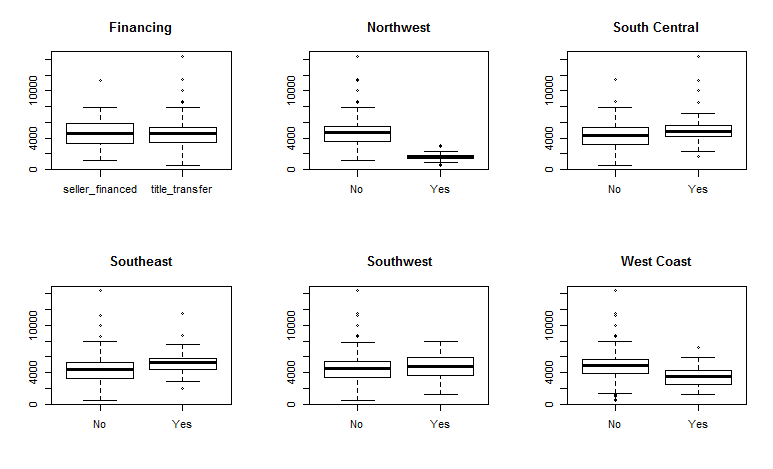
In order to achieve these goals, we have data for 420 farms. Each farm contains information on the variables shown in Figure 1.



Figure 1

The Acre Price is the variable at question here. All of the other variables are predictor variables, or variables we will use to do the predicting of Acre Price. Figure 2 contains the basic relationship between all of the numeric variables with relation to acre price. From a glance, we can see that the strongest relationship seems to be between the acre price and the productivity of the land. There also doesn't seem to be any strong signs of collinearity.

Figure 2

Figure 3 also gives us an opportunity to assess some the relationships between acre price and the location of the farm, as well as the type of financing. According to the box plot, you can see a major drop in acre price when the farm is located in the Northwest. The appraiser also thinks that the effect of productivity on price in the Northwest is different than other areas, so we will also take a look at that interaction.

Figure

Based on these exploratory results, a multiple linear regression model is appropriate for this data and will allow us to achieve the goals mentioned earlier.

**Statistical Modeling**

The best way to approach modeling this data is not necessarily just to use every variable in the model. We used a best subset selection method to see which combination of variables best explain the data, and will be best for prediction. We chose to compare models using the BIC because it has a large variable penalty and we want a small model for interpretation. This technique helped us arrive at a final model that includes the following variables: improvements, tillable, crpPct, productivity, NW, and WC. This combination will do the best job at explaining the data with a simple model.

**Regression Model**

After settling on this model, we looked at some of the assumptions for multiple linear regression (we will go over these more), and we realized that the assumption of equal variance was not being met. We are going to use a log transformation in the data to help meet this assumption. After transformation, we will have the following model:

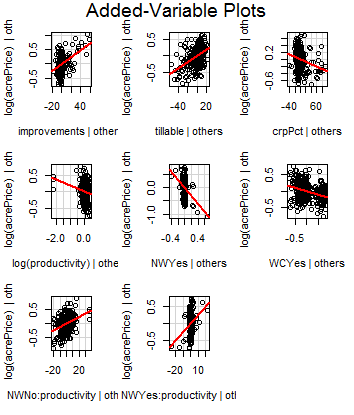
log(acrePrice­­i)=0 + ­1(improvementsi)+ 2(tillablei) + 3­­(crpPcti)+ β4­(log(productivityi)) + β5­I(NWi=Yes) + β6­I(WCi=Yes) + 7I(NWi=No)(productivityi) + 8I(NWi=Yes)(productivityi) + i

where εi ~ N(0,σ2)

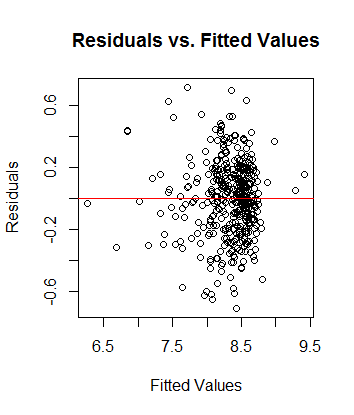
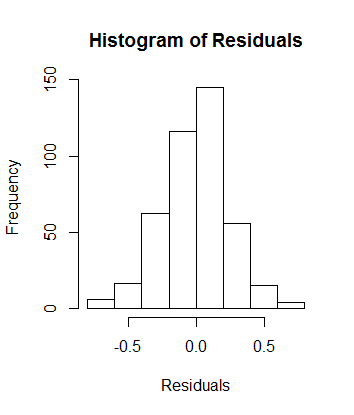
The response variable is the log acre price, which is comprised of an intercept (0) and the predictor variable coefficients (­1-8) and their respective variables. The intercept can be interpreted as the average log acre price of a farm that is not in the Northwest or West Coast with zero improvements, zero percent tillable, zero percent land enrolled in conservation reserve, zero productivity. This represents a "baseline." As any of the variables in this model changes the log acre price will also change with it. For example, ­1 is the average change in log acre price as improvements increases by 1 and everything else is held constant. 7­ and 8 represent coefficients for the interaction between productivity and the Northwest. For example, holding all else constant, 8 represents the additional average change in log acre price as productivity increases by 1 if the farm is located in the Northwest. 7 can be interpreted similarly, but for a farm that is not located in the Northwest.

For this model to hold validity, We are assuming linearity in the data, independent observations, equal variance, and normally distributed errors.

**Model Justification and Verification**



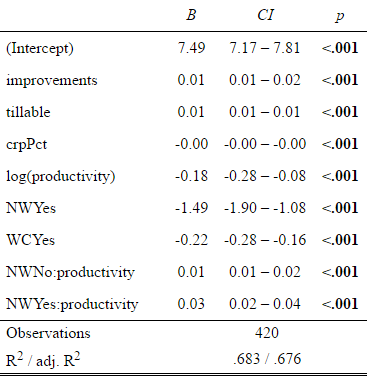
The Added-Variable Plots to the left help us to see the linear relationship between each variable and the log acre price. This verifies our assumption of linearity. Below, the Histogram of Residuals shows normally distributed errors, verifying the assumption of normality. Now, the Residuals vs. Fitted Values plot is not perfect. There is evidence that suggests this assumption may not be satisfied. However, the transformations we performed significantly improved this assumption, and we will move forward assuming equal variance and independence.



**Model Fit**

This model has a multiple R2 measure of .6826. This means that roughly 68% of the variation in acre price is explained by the variables and the interaction included in the model. This is actually fairly high and suggests that our model does fit the data well.

We are also interested in how valuable our model is for predicting the price of a farm. Running a cross-validation study on the data using our model helped us to assess the bias, root predicted mean square error (RPMSE) coverage, and prediction width of our model. The bias, which tells us if we are, on average, too high or too low on our predictions was -118.926. The RPMSE is a measure of how far off your predictions are on average. We had a RPMSE of 1042.853. Considering that the range of acre price's of the farms in our data set was 509 to 14379, these numbers fair well when it comes to predicting. 95% of the prediction intervals that we created using this model actually contained the true acre price of the farm. We can be 95% confident that the prediction interval we come up with contains the true value. Our average prediction interval width was 4376. This isn't the narrowest interval, which means it won't be the most precise, but it is narrow enough to draw valuable insights from our prediction intervals.

**Results of Analysis**

To the right there is a table which comprises the results of the effects of selected variables on acre price. The β column provides the coefficient for each variable, and the following CI column, contains a confidence interval for that coefficient estimate. For example, holding all else constant, as the log of productivity increases by 1, we expect an average decrease of .18 in the log acre price. We are 95% confident that the true decrease in log acre price lies between .28 and .08.

**Interaction between Northwest and Productivity**

Looking at the table above, you can see the last two terms are a combination of NW and productivity. The appraiser had prior intuition that the effect of productivity in the NW is different that in other areas. We found that prior intuition to be correct. We ran an analysis of variance comparing a model that did not include the interaction to a model that did include the interaction. The null hypothesis was that the reduced model (the one without the interaction) was better. The hypothesis test came back with a p-value of essentially 0. We can reject the null hypothesis and conclude that the interaction term is indeed significant.

**Prediction**

Using this model to predict the farm price for a title transfer farm in the NW with zero percentage improvements, a 94% tillable percentage, a conservation reserve percentage of 0, and a productivity score of 96, we obtain an acre price prediction of 5186.799, and we are 95% confident that the true value of the acre price is between 2641.816 and 10183.48.

**Conclusions**

We have identified a reliable model for predicting the acre price of farms using the variables: improvements, tillable percentage, conservation reserve percentage, productivity score, and location (in the NW or WC). While it might seem counter-intuitive that the log(productivity) has a negative effect on acre price, remember that productivity is also counted in the interaction term with location. A majority of the effect of productivity actually lies in the interaction term. So, overall, as productivity increase, the acre price of the farm increases as well. We also learned that tillable percentage and number of improvements have significant effect on the acre price. Location also matters, with farms located in the NW and WC lowering the acre price for a farm.

Moving forward, it would be helpful to consistently gather data on a few more variables that appraisers believe contribute to the acre price of a farm. This would help explain the variation in the data and possibly lead to narrower prediction intervals. I would suggest better understanding the current availability of farmland near the farm site, and maybe the turnover rate of farmland in the area to better approximate the worth of the land in relation to the land's potential for growth and the overall scarcity of land in the area. Current prices for commodities and crops could also play a major factor in the acre price of a farm.