STAT230 Work Team Project

Work Team 7: Evan, Becca, Laith

Introduction & Motivation

We are interested in how we can attempt to predict the profitability of farms by different farming practices. Specifically we are curious about sustainable agriculture decisions and decisions about crop types. From a conservationist perspective, it is important that sustainable choices are economically viable for farmers.

Our data is sourced from the United States Census of Agriculture, between the years 2009 and 2017, along with county demographic information from Wikipedia, sourced from the U.S Census. The data is collected by county, often summing the a variable across all operations in a given county. This limits our specificity, but we are still able to estimate what is typical for farms across the county by looking at averages for each county. https://www.nass.usda.gov/AgCensus/

Our response variable in reported farm income, and we are planning to predict it with a number of variables (population density, farm acres, bee colonies, grazing rotation, fertilizer use).

Grazing rotation refers to the number of farming operations in a county that reported using either crop cycling or alternating patterns of grazing to minimize impact. Farm acres and bee colonies are counts that are summed across all farms within a state. Fertilizer use is the total dollar amount spent on fertilizers in that year by all farming operations in a given county. Population density was estimated based on county sizes and populations based on the U.S census via Wikipedia. We feel that this combination of predictors is appropriate because it examines many dimensions of farming practices. It includes information about context of the farming (population density, farm size) as well as the types of methods of sustainable or unsustainable practices carried out in each of these counties. It is valuable to understand how economically viable sustainable farming is in the short term in order to access if further supportive measures should be taken in order to ensure both sufficient food production, as well as producing that food in ways that will limit harm to the Earth.

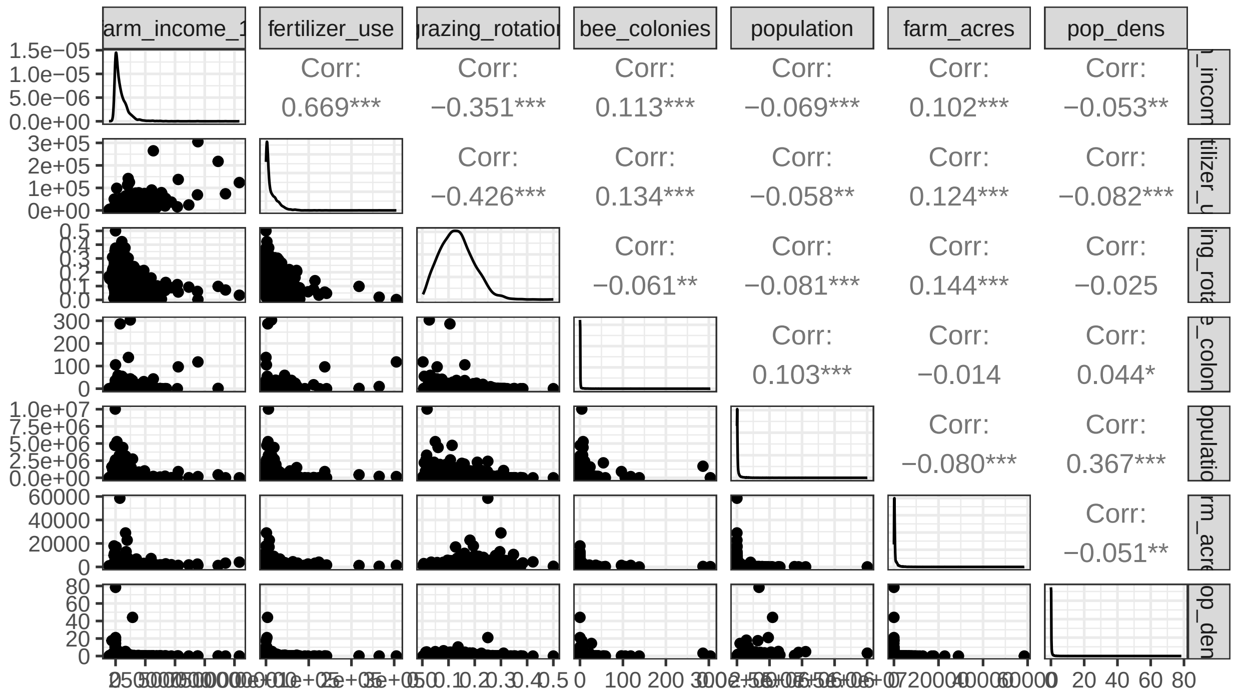
Data set & Wrangling

The initial combining, and wrangling of this large data set can be found at https://github.com/evanmacarthur/Stat230Proj1 The data wrangling was then perfected as shown in appendix 1

Exploratory Data Analysis (EDA)

There are significant correlations with all variables, but lots of these associations do not look linear or normal. Transforming the data may help improve these relationships.

Starting with the response variable, we see that the mean is higher than the median suggesting a right skewed relationship. We used a log transformation. The scatter plots for all transformations are shown in appendix 2. The statistical summaries are shown in appendix 3.



1. We transposed+log-transformed farm income, which strongly improved normality and linearity. It went from right skewed to less right skewed. (Appendix 2, Appendix 3)The mean, median, and IQR for both: $46857, 27177, 61419 \rightarrow 11.4, 11.319, 0.699$

2. We square-root transformed fertilizer use, slightly which improved normality and linearity. It went from initially right skewed to less right skewed. (Appendix 2, Appendix 3) The mean, median, and IQR for both: $12600, 6182, 16406 \rightarrow 91.652, 76.297, 86.104$

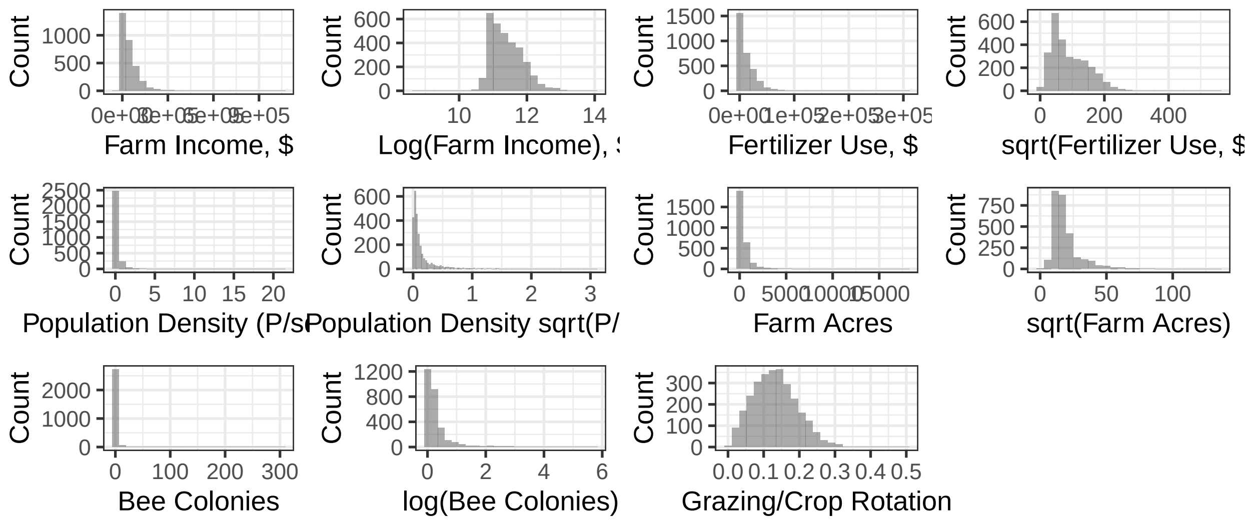
3. We log-transformed bee colonies, which slightly improved normality and linearity. It went from initially heavily right skewed, to slightly less right skewed. (Appendix 2, Appendix 3) The mean, median, and IQR for both: $1.1202, 0.15986, 0.34899 \rightarrow 0.31796, 0.1483, 0.28748$

4. We square-root transformed population density, which improved normality and linearity. It went from initially strongly right skewed to somewhat less right skewed, although there are still many 0 values, which are real, but problematic normality. (Appendix 2, Appendix 3) The mean, median, and IQR for both: $0.2402, 0.07206, 0.15057 \rightarrow 0.16445, 0.06958, 0.13644$

5. We square-root transformed farm acres, which strongly improved normality and linearity (Appendix 2, Appendix 3) The mean, median, and IQR for both: $516.31, 251.13, 305.07 \rightarrow 19.312, 15.847, 9.021$

6. We chose not to transform grazing/rotation since normality and linearity already looked good (Appendix 2, Appendix 3) The mean, median and IQR are: $0.13179, 0.12854, 0.08457$

After these transformations, correlations between our predictor variables and farm income look much more linear (Appendix 5). The significance of the correlations are fairly preserved, except for with bee colonies (Appendix 5).



Model selection

	adjr2	cp	bic	rss	farm_acres_2	pop_dens_2	grazing_rotation	fertilizer_use_2	bee_colonies_2
1 (1)	0.521	100.154	-2053.1	283.48				*	
2 (1)	0.530	44.510	-2101.5	277.86		*		*	
3 (1)	0.536	5.694	-2134.1	273.89		*	*	*	
4 (1)	0.537	4.283	-2129.6	273.56	*	*	*	*	
5 (1)	0.537	6.000	-2121.9	273.53	*	*	*	*	*

We initially selected a model using best subsets and stepwise regression procedures. We also performed backwards elimination to check the AIC value of the full model. The final model suggested by best subsets regression was the 4-predictor model with pop_dens_2, grazing_rotation, fertilizer_use_2, and farm_acres_2 without bee colonies 2. This model had the lowest C_p score of 4.2828. This model was also recommended by stepwise regression and backwards elimination, with the lowest AIC score of -6545.7 (Appendix 7). However, since the 5-predictor model including bee_colonies_2 had a C_p score < 6, and had a very similar AIC score of -6544.0, we decided to manually compare the 5 term and 6 term models.

Model comparison - conditions

As shown in Appendix 8, the residuals vs fitted plot, QQ plot, and histogram of the residuals of both models are practically identical and will not serve as a point of difference. The linearity condition is mostly met, as there is mostly symmetry over the line residual=0 on the residuals vs fitted values plot, and there does not seem to be major bending or curving of points. However, there is a noticeable lack of symmetry around a fitted value of 11, and there overall seems to be more variance for positive residual values, which is concerning. The equal variance condition is questionable as there is lower residual variance for higher fitted values. The normality condition is also questionable. Although the QQ-plot is mostly a straight line with mostly no bending, there does seem to be a very large and bent tail at the high end. Furthermore, the histogram of the residuals, while it is unimodal and centered close to 0, is right skewed. We assumed that data values were collected independently. Finally, there appear to be several unusual points in both models.

We next confirmed that our variable transformations helped improve model conditions. As displayed in Appendix 8, there are many more extreme outliers in the untransformed model, with many very high fitted values that would exert undue influence over our model. As seen in the histogram, the data is highly right skewed. Based on these results, we felt confident that our transformations improved model conditions, even though they did not fully remove condition concerns, and we proceeded with the transformed model.

Model comparison - outliers and influential points

We next did a more thorough comparison of outliers and influential points between the two models. Standardized residual, Cook's distance, standardized residual vs leverage, and Cook's distance vs leverage plots looked fairly similar between the two models (Appendix 9). For fmA, the high leverage cutoff is 2(k+1)/n = 2(4+1)/2813 = 0.0035549 and the problematic leverage cutoff is 3(k+1)/n = 0.0053324 There are 236 points with high leverages, 136 of which have problematic leverages. There are 91 points with the absolute value of their standardized residuals above 2 but not 3, making them outliers. There are 62 points with an absolute value of their standardized residuals above 3, making them big outliers. There do not appear to be many influential points, as there are no points with Cook's distances above or near 0.5.

For fmB, the high leverage cutoff is 2(k+1)/n = 2(5+1)/2813 = 0.0035549 and the problematic leverage cutoff is 3(k+1)/n = 0.0053324. There are 337 points with high leverages, 207 of which have problematic leverages. There are 91 points with the absolute value of their standardized residuals above 2 but not 3, making them outliers. There are 62 points with an absolute value of their standardized residuals above 3, making them big outliers. There do not appear to be many influential points, as there are no points with Cook's distances above or near 0.5. So, fmA seems to have 71 less points with problematic leverages. They have the same number of big outliers, and neither have influential points.

Model comparison - multicollinearity

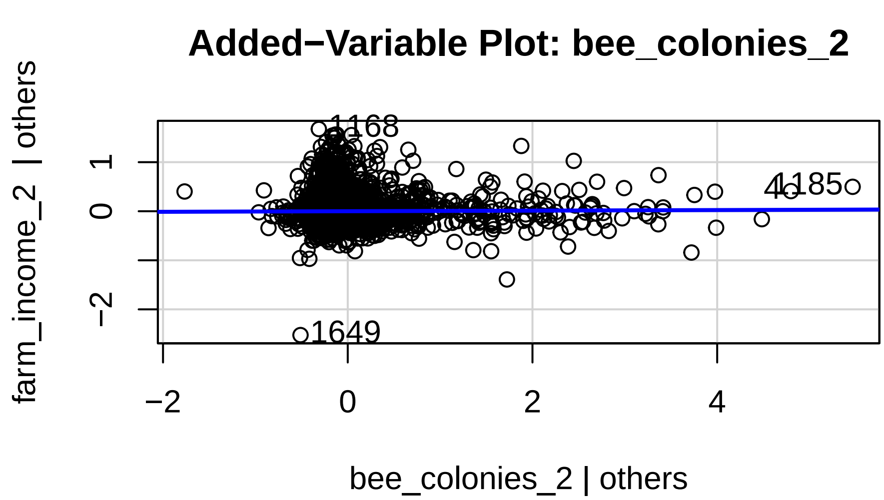
VIF tables

```
##       pop_dens_2 grazing_rotation fertilizer_use_2     farm_acres_2
##           1.1428           1.4723           1.6138           1.2880

##     farm_acres_2       pop_dens_2 grazing_rotation fertilizer_use_2
##           1.2984           1.1975           1.4745           1.6378
##   bee_colonies_2
##           1.0786
```

We next checked for multicollinearity in both models by analyzing correlation plots and VIF scores. All VIF scores for both models were below 2, indicating that multicollinearity is not a concern for either model. The inclusion of bee_colonies_2 in fmB has little impact on VIF scores. The correlation matrix for the bee_colonies_2 pairs isn't too concerning with correlation coefficients of 0.23 or lower (Appendix 10). This further indicates that including bee_colonies_2 does not increase multicollinearity.

Model comparison - Added Variable Plot



The AV plot of the additional predictor of fmB, bee_colonies_2, shows a horizontal band that is curvilinear. This indicates that bee_colonies_2 contains very little additional information about farm_income_2, beyond that from the other predictors. This indicates that bee_colonies_2 is probably not a strong predictor of farm_income_2.

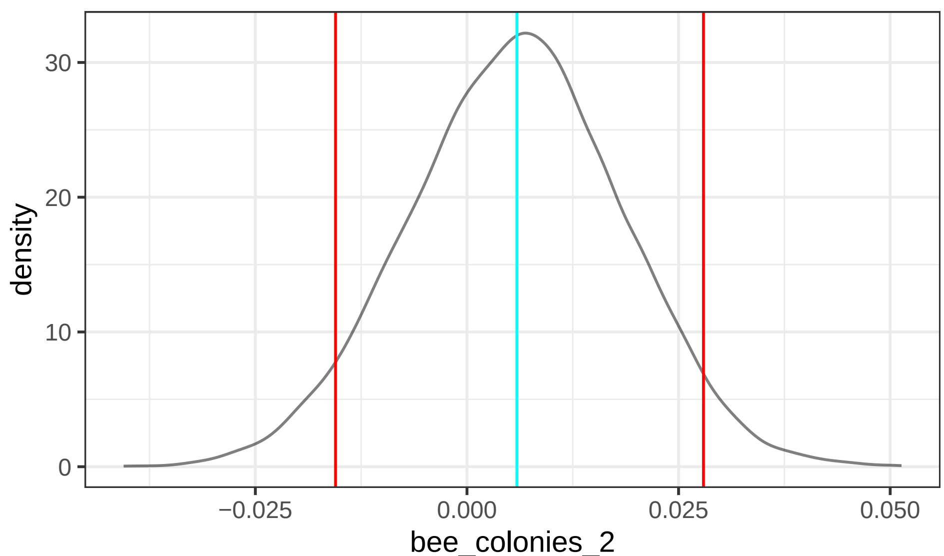
Model comparison - Nested F-Test

$H_0 : \beta_{bee-colonies} = 0$ $H_A : \beta_{bee-colonies} \neq 0$

We performed a nested F test on the two models, with the above null and alternative hypotheses (Appendix 11). The test resulted in a p-value of 0.59, so we fail to reject the null hypothesis and do not have enough evidence to conclude the that the predictor bee_colonies_2 is significant after controlling for other predictor effects.

Bootstrapping

Since the normality condition was not met for either model, we will test the significance of bee_colonies_2 using bootstrapping. We assumed that our bootstrapped distribution was nonnormal and skewed. The resulting 95% confidence interval for the coefficient of bee_colonies_2 is (-0.016319, 0.027324). Since this interval includes 0 (see figure below),we are not 95% confident that the bee_colonies_2 is not 0, and conclude that this predictor is not significant.



Final evaluation of 4 and 5 term models

We conclude that fmA, the four term model is preferable to fmB, the five term model that includes bee colonies as a predictor. As described above, fmA and fmB have very similar deviations from model conditions. However, fmB has about 100 more points with high leverage than fmA, indicating that fmB has more potential to make inaccurate predictions than fmA.

After controlling for the effects of the other four predictors, bee colonies has an insignificant effect on farm income, as indicated by a nested f test and individual coefficient t stat with p values of 0.59. We double checked this result with a bootstrap test since our model did not fit conditions, but still found that the effects of bee colonies on farm income were insignificant.

Furthermore, as we saw in our initial model selection process, fmA is a stronger model than fmB. Both models have an adjusted R^2 of 0.537 (Appendix 12), and fmA has lower AIC and Cp scores than fmB. Overall, we found no reasons why fmB is preferable and many reasons why fmA is preferable. We will therefore proceed with analyzing fmA.

Outlier removal

One outlier, at index 1649, stands out in our equal variance plot and has the largest standard residual value in fmA. We tried removing it and reassessing conditions and model fit. See the appendix for our specific model results for this new model, fmA3. Removing this outlier decreased residual variance at the outlier's fitted value and taking out such an extreme value improved data normality. However, both equal variance and normality are still very questionable.

As seen in Appendix 13, removing this outlier did slightly decrease the significance of pop_dens_2 (p = 1.1e-13 compared to 5.5e-14 with the outlier) and farm_acres_2 (p = 0.072 compared to 0.065 with the outlier). However, removing the outlier increased R^2 by 0.5%, from 0.537 to 0.542. Given that taking out this outlier improved model conditions and increased model strength slightly, and since the impacts on model significance were small and less important than improvement to model conditions, we decided that it made sense to leave the outlier out of our model.

Interactions

Finally, we decided to test some interactions in our model. We specifically tested all interactions involving population density, since we thought that the population density of a county could modulate the effects of farm size and farming practices on farm income. For example, more populous counties might have more access to resources that would help them profitably employ fertilizer and grazing rotation practices.

See Appendix 14 for specific model results for this model, fmC. We first checked the conditions of our model with these three interaction terms. Equal variance, normality, and linearity looked very similar to our model without interactions. Looking at model significance, we found that all three interactions were significant or close to significant: p=0.0561 for pop_dens_2:grazing_rotation; p=0.0274 for pop_dens_2:fertilizer_use_2; p=0.0199 for pop_dens_2:farm_acres_2. However, adding these interaction terms barely increased the overall strength of our model: R^2 adjusted changed from 0.542 to 0.543. Although these interactions are interesting and fairly significant, they are not strong predictors of farm income and do not improve model conditions. We decided not to include them in our final model in order to simplify our model. This allowed us to interpret the effects of individual predictors on farm income.

Final Conclusion:

Overall, our final model was fmA3, which has outlier 1649 removed and has no interaction terms. This model indicates that 54.2% of the variability in farm_income_2 can be explained by the multiple linear relationship between farm_income_2 and pop_dens_2, grazing_rotation, fertilizer_use_2, and farm_acres_2. Additionally, this model indicates that average farm income in a county increases with increasing population density, proportion of farms that use grazing rotations, average fertilizer use, and average farm size. We reverse transformed our model coefficients (Appendix 15), and found that average farm income increases by $0.31 for every 1 person/acre density increase; farm income increases by $0.05 for every 10% increase in the percent of farms that use grazing rotations; farm income increases by $1.00 for every $1 increase in average farm fertilizer expenditure; and farm income increases by $1.00 for every 1 acre increase in average farm size. That is, counties with bigger farms that use more fertilizer, counties with more farms that use grazing rotations, and counties with large population densities are likely to contain more profitable farms. > Fitted model equation:

$$\widehat{farmincome} = 61262 + 0.309(popdensity) + 0.466(grazingrotation) + 1.005(fertilizeruse) + 1.001(farmacres)$$

Appendices

Appendix 1: Data Wrangling

```r
df0 <- read_csv("gooddata.csv")
df0 <- df0 |>
  clean_names()
df0 <- df0 |>
  rename("num_operations" = farm_operations_number_of_operations)

df <- df0 |>
  dplyr::select(c(farm_income_1, num_operations,
                  fertilizer_use, grazing_rotation, bee_colonies,
                  population, county_acreage_2000, farm_acres))

df1 <- df |>
  mutate(pop_dens = population/county_acreage_2000)

df1 <- df1 |>
  mutate(fertilizer_use = fertilizer_use/num_operations) |>
  mutate(grazing_rotation = grazing_rotation/num_operations) |>
  mutate(bee_colonies = bee_colonies/num_operations) |>
  mutate(farm_acres = farm_acres/num_operations)

df1 <- df1 |>
  dplyr::select(!c(num_operations, county_acreage_2000))

df1 <- df1 |>
  filter(!is.infinite(pop_dens))
df2 <- df1 |>
  mutate(farm_income_2 = log(farm_income_1+55152))
df2 <- df2 |>
  filter(farm_income_2 > 5)
df3 <- df2 |>
  mutate(fertilizer_use_2 = sqrt(fertilizer_use))
df3 <- df3 |>
  mutate(pop_dens_2 = log(pop_dens+1))
df3 <- df3 |>
  filter(!is.infinite(pop_dens_2))
df3 <- df3 |>
  mutate(farm_acres_2 = sqrt(farm_acres))
df3 <- df3 |>
  mutate(bee_colonies_2 = log(bee_colonies+1))
df4 <- df3 |>
  dplyr::select(c(farm_income_2, farm_acres_2,
                  pop_dens_2, grazing_rotation,
                  fertilizer_use_2, bee_colonies_2))
df4 <- na.omit(df4)
df3<-na.omit(df3)
```

Appendix 2: variable transformation scatter plots

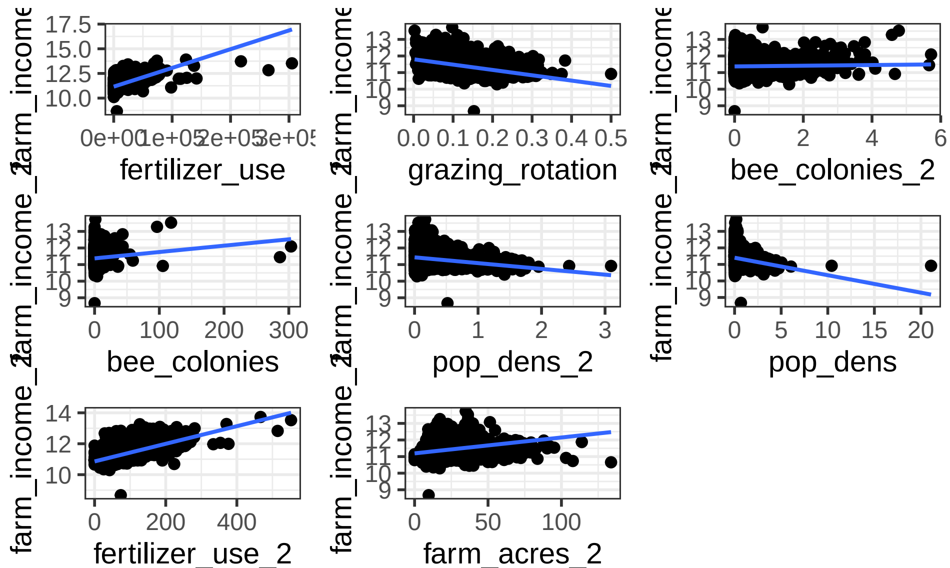


Table 1: Summary of transformed Bee Colonies

min	Q1	median	Q3	max	mean	sd	n	missing
0	0.04669	0.1483	0.33417	5.7187	0.31796	0.55139	2813	0

Table 2: Summary of Bee Colonies

min	Q1	median	Q3	max	mean	sd	n	missing
0	0.04779	0.15986	0.39678	303.5	1.1202	9.1848	2813	0

Table 3: Summary of Transformed Farm Acres

min	Q1	median	Q3	max	mean	sd	n	missing
0	12.398	15.847	21.419	133.78	19.312	11.976	2813	0

Table 4: Summary of Farm Acres

min	Q1	median	Q3	max	mean	sd	n	missing
0	153.72	251.13	458.79	17896	516.31	932.61	2813	0

Table 5: Summary of Transformed Population Density

min	Q1	median	Q3	max	mean	sd	n	missing
0.00038	0.02955	0.06958	0.16599	3.0948	0.16445	0.26464	2813	0

Table 6: Summary of Population Density

min	Q1	median	Q3	max	mean	sd	n	missing
0.00038	0.02999	0.07206	0.18056	21.082	0.2402	0.66928	2813	0

Table 7: Summary of Transformed Fertilizer Use

min	Q1	median	Q3	max	mean	sd	n	missing
0	45.296	76.297	131.4	552.37	91.652	58.097	2813	0

Table 8: Summary of Fertilizer Use

min	Q1	median	Q3	max	mean	sd	n	missing
0	2105.3	6182	18511	305114	12600	16789	3069	6

Table 9: Summary of Transformed Farm Income

min	Q1	median	Q3	max	mean	sd	n	missing
8.6733	11.013	11.319	11.712	13.908	11.4	0.48011	3075	0

Table 10: Summary of Farm Income

min	Q1	median	Q3	max	mean	sd	n	missing
-55150	5494.5	27177	66913	1041275	46857	67968	3076	0

Table 11: Summary of Grazing/Crop Rotation

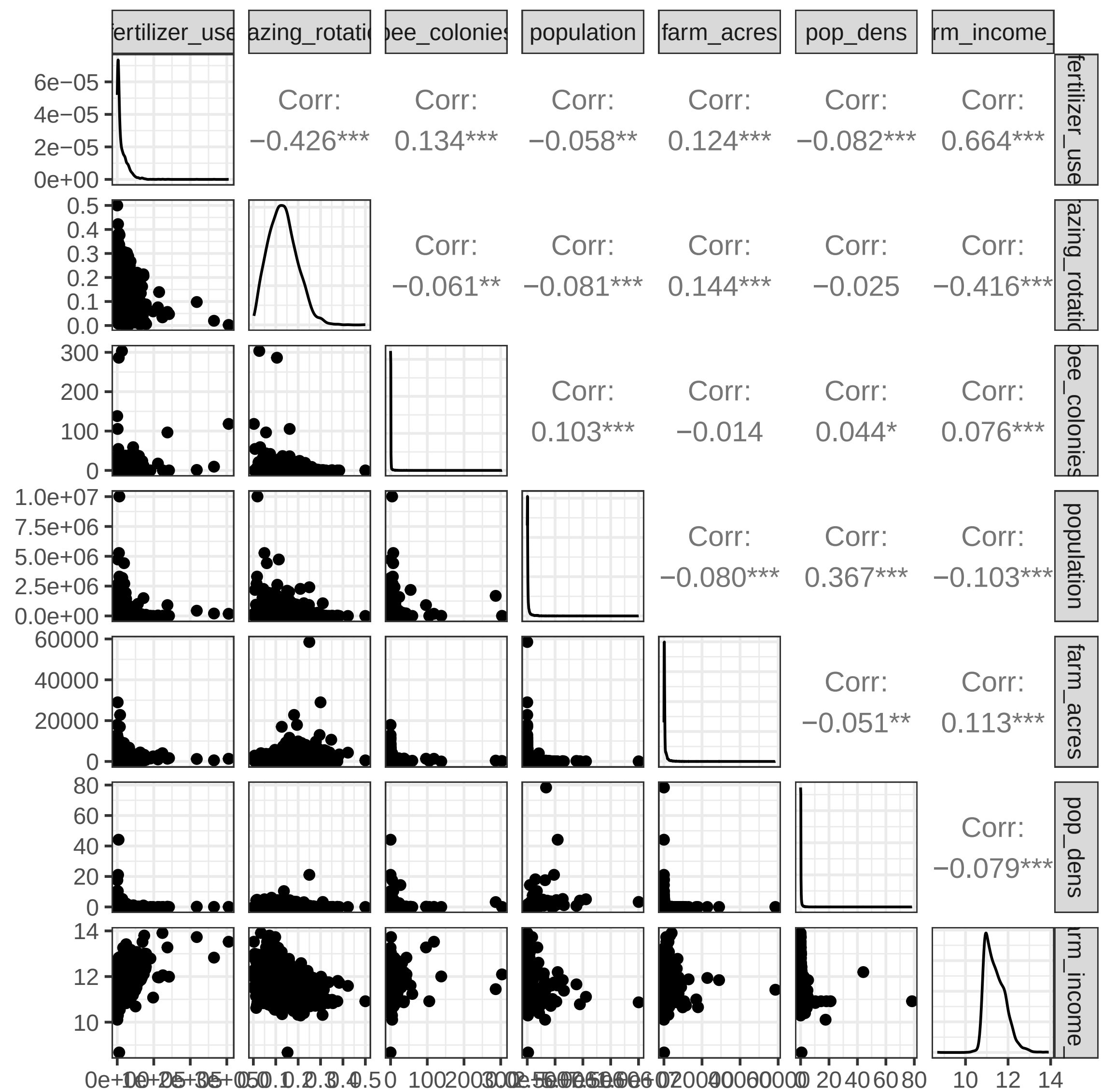
min	Q1	median	Q3	max	mean	sd	n	missing
0.00253	0.08698	0.12854	0.17155	0.5	0.13179	0.06138	2813	0

Table 12: Summary Bee Colonies

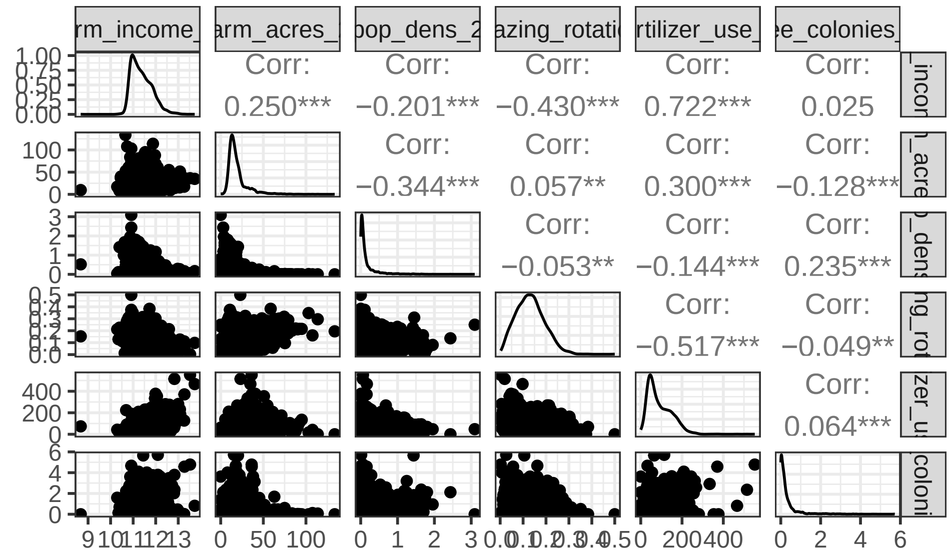
min	Q1	median	Q3	max	mean	sd	n	missing
0	0.04779	0.15986	0.39678	303.5	1.1202	9.1848	2813	0

Appendix 3: Numeric summaries of transformed variables

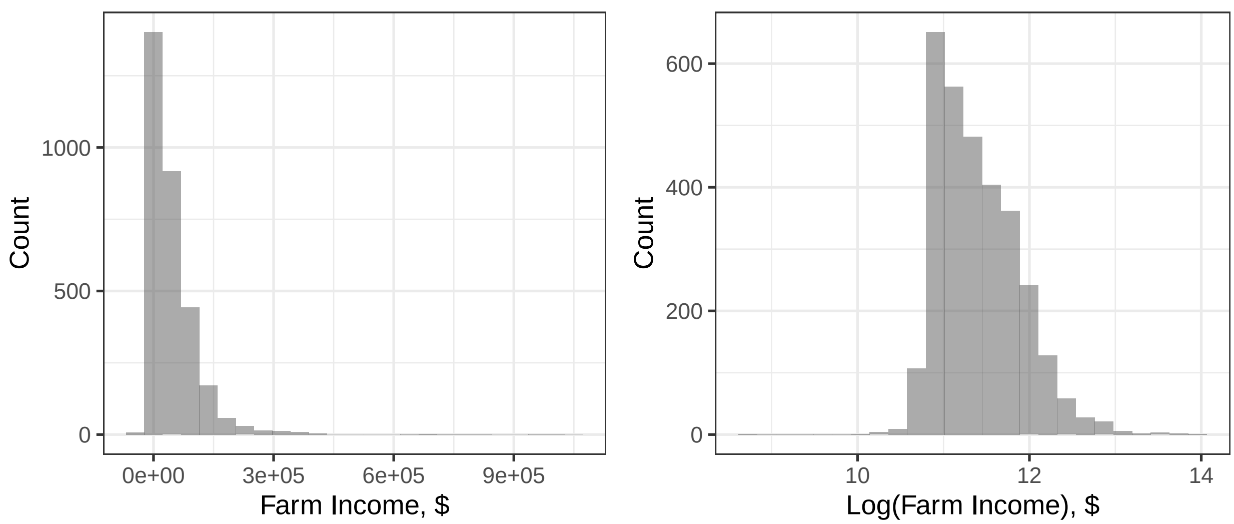
Appendix 4: ggpairs w/ just Farm Income Transformed



Appendix 5: ggpairs with all transformed Variables



Appendix 6: Comparison of Farm Income transformation



Appendix 7: Model selection

Stepwise regression:

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## farm_income_2 ~ 1
##
## Final Model:
```

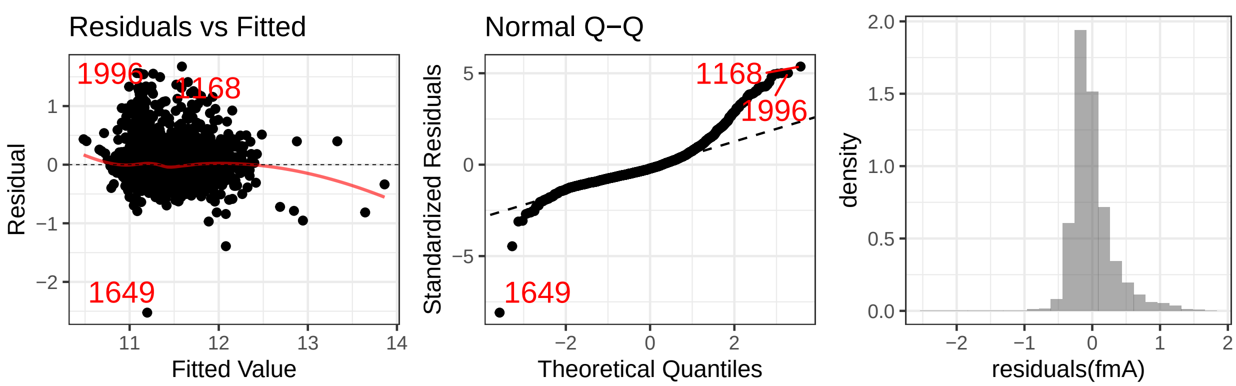
```
## farm_income_2 ~ fertilizer_use_2 + pop_dens_2 + grazing_rotation +
##     farm_acres_2
##
##
##                 Step Df  Deviance Resid. Df Resid. Dev     AIC
## 1                                     2812     591.50 -4384.4
## 2 + fertilizer_use_2  1 308.01801      2811     283.48 -6451.4
## 3       + pop_dens_2  1   5.61712      2810     277.86 -6505.7
## 4 + grazing_rotation  1   3.97730      2809     273.89 -6544.3
## 5     + farm_acres_2  1   0.33237      2808     273.55 -6545.7
```

Backwards elimination:

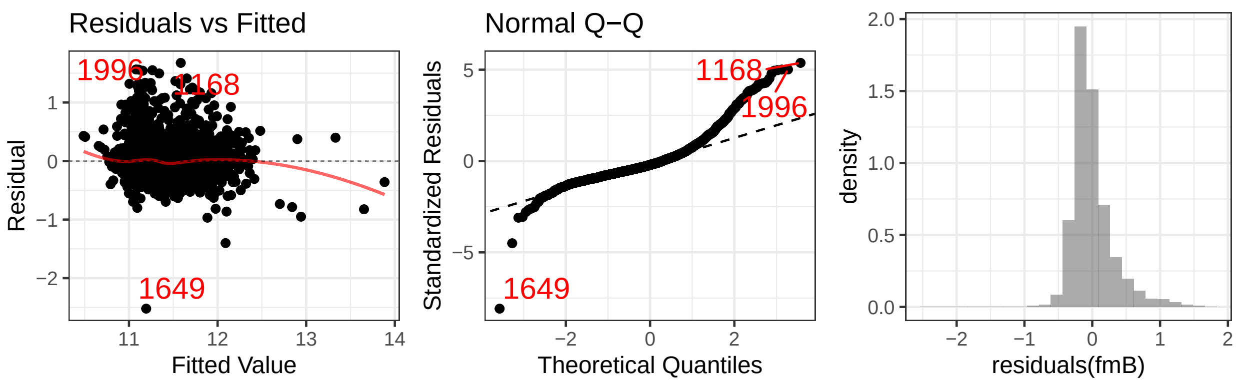
```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##     fertilizer_use_2 + bee_colonies_2
##
## Final Model:
## farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##     fertilizer_use_2
##
##
##               Step Df Deviance Resid. Df Resid. Dev     AIC
## 1                                   2807     273.53 -6544.0
## 2 - bee_colonies_2  1 0.027553      2808     273.55 -6545.7
```

Appendix 8: Model comparison - conditions

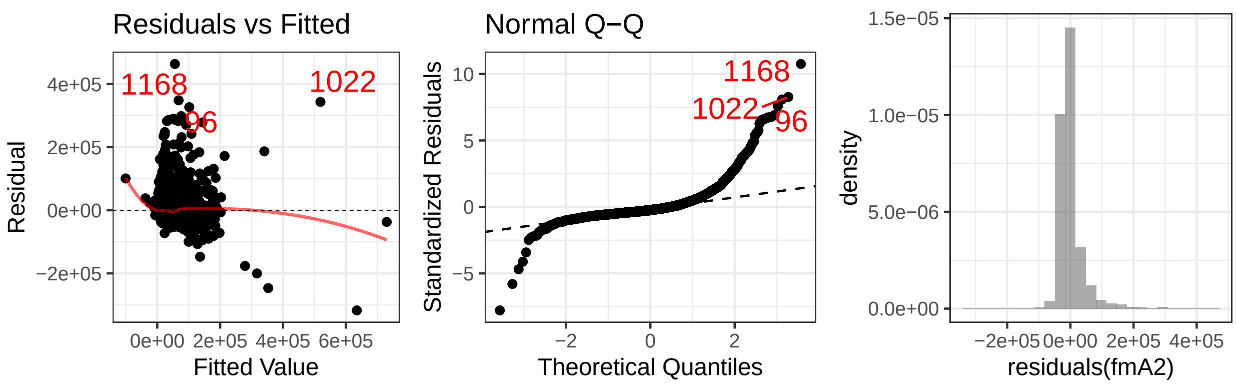
fmA conditions:



fmB conditions:

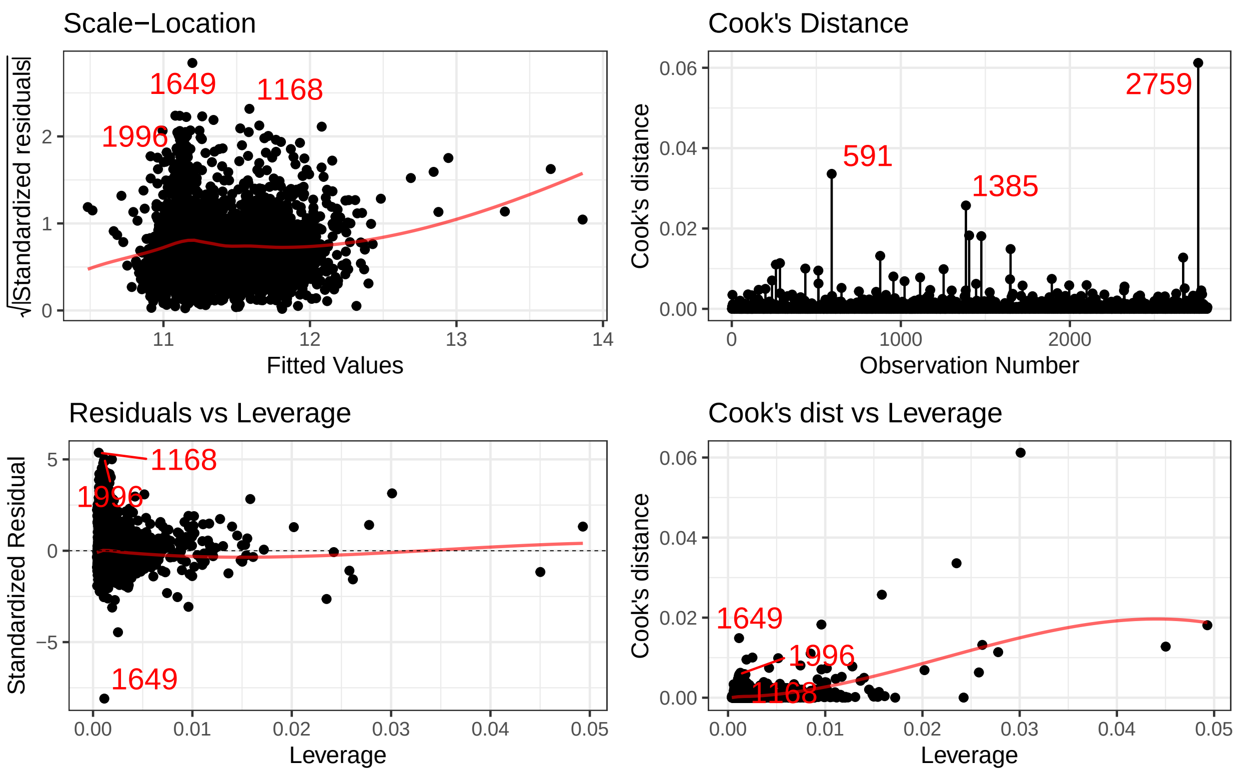


Untransformed fmA conditions:

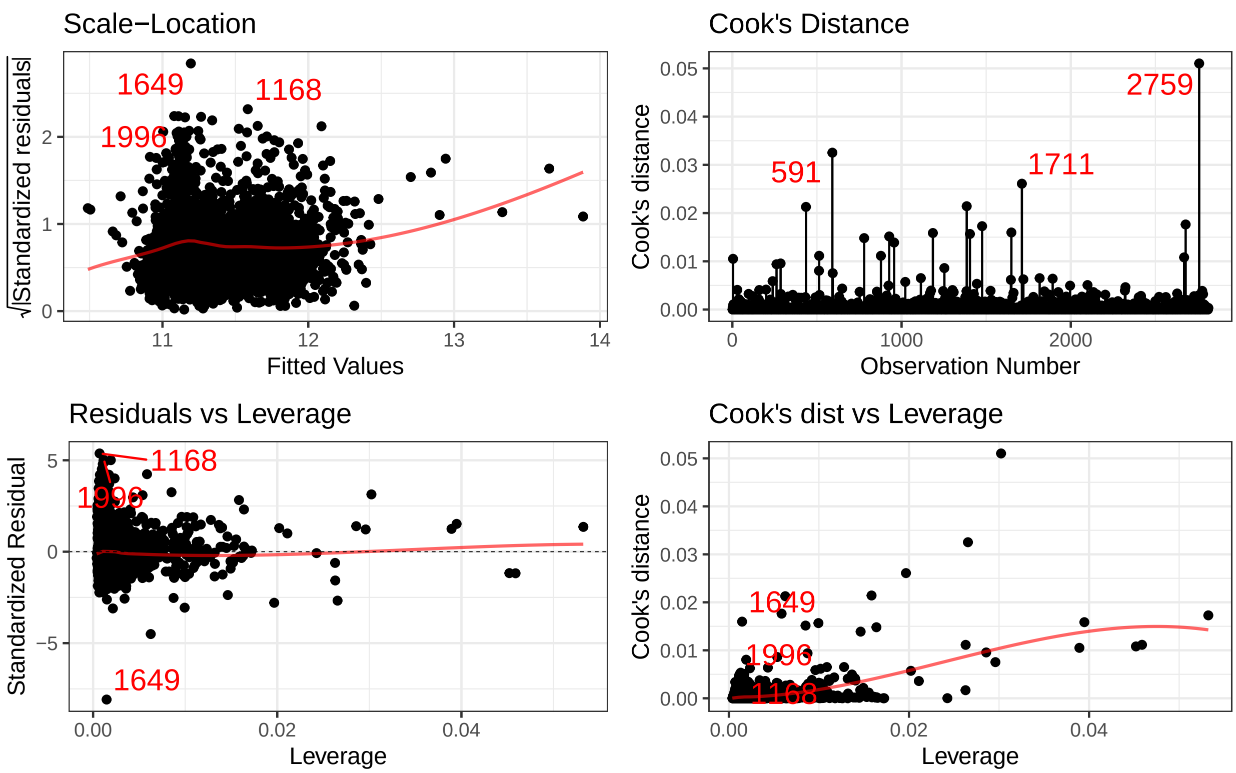


Appendix 9: Model comparison - outliers and influential points

fmA plots:


Scale−Location
√|Standardized residuals|
Fitted Values
1649
1168
1996
Cook's Distance
Cook's distance
Observation Number
2759
591
1385
Residuals vs Leverage
Standardized Residual
Leverage
1168
1996
1649
Cook's dist vs Leverage
Cook's distance
Leverage
1649
1996
1168

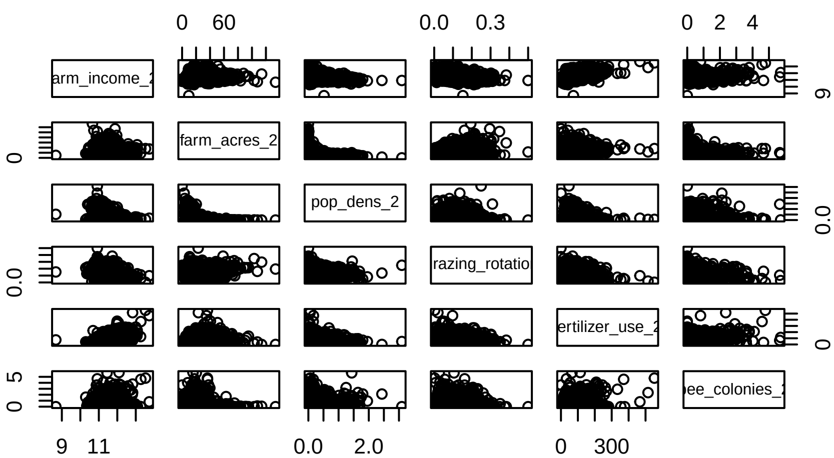
fmB plots:


Scale−Location
√|Standardized residuals|
Fitted Values
1649
1168
1996
Cook's Distance
Cook's distance
Observation Number
2759
591
1711
Residuals vs Leverage
Standardized Residual
Leverage
1168
1996
1649
Cook's dist vs Leverage
Cook's distance
Leverage
1649
1996
1168

Appendix 10: Model comparison - multicollinearity

Correlation matrix and plot matrix for all relevant variables:

```
##                  farm_income_2 farm_acres_2 pop_dens_2 grazing_rotation
## farm_income_2         1.000000     0.250045  -0.200636        -0.430220
## farm_acres_2          0.250045     1.000000  -0.343740         0.056543
## pop_dens_2           -0.200636    -0.343740   1.000000        -0.052888
## grazing_rotation     -0.430220     0.056543  -0.052888         1.000000
## fertilizer_use_2      0.721624     0.299724  -0.144407        -0.517474
## bee_colonies_2        0.024889    -0.128397   0.235198        -0.049328
##                  fertilizer_use_2 bee_colonies_2
## farm_income_2            0.721624       0.024889
## farm_acres_2             0.299724      -0.128397
## pop_dens_2              -0.144407       0.235198
## grazing_rotation        -0.517474      -0.049328
## fertilizer_use_2         1.000000       0.063654
## bee_colonies_2           0.063654       1.000000
```



Appendix 11: Model comparison - nested F test

```
## Analysis of Variance Table
##
## Model 1: farm_income_2 ~ pop_dens_2 + grazing_rotation + fertilizer_use_2 +
##     farm_acres_2
## Model 2: farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##     fertilizer_use_2 + bee_colonies_2
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1   2808 274
## 2   2807 274  1    0.0276 0.28   0.59
```

Appendix 12: Final evaluation of 4 and 5 term models

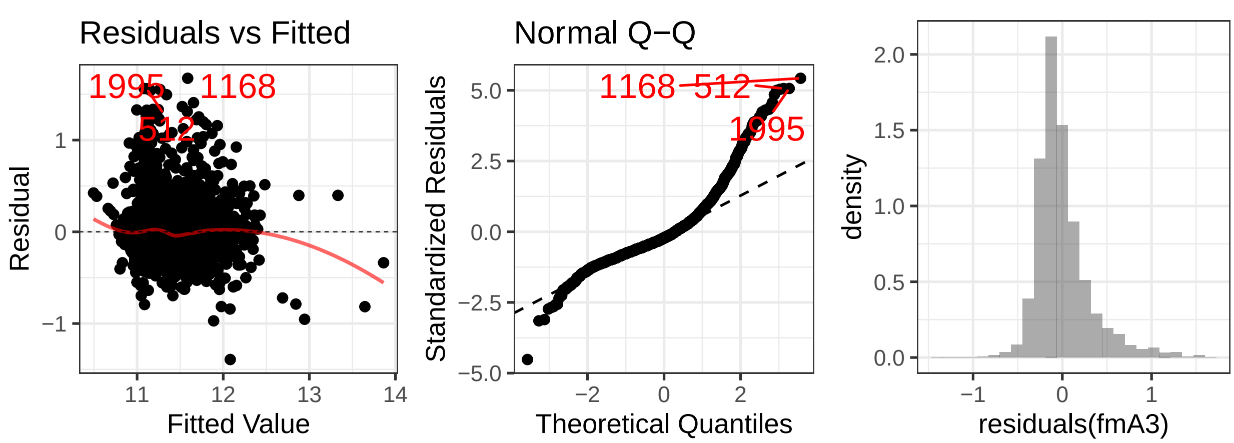
fmA summary:

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       11.023637   0.024966  441.55  < 2e-16 ***
## pop_dens_2        -0.179728   0.023776   -7.56  5.5e-14 ***
## grazing_rotation -0.772752   0.116352   -6.64  3.7e-11 ***
## fertilizer_use_2  0.005092   0.000129   39.57  < 2e-16 ***
## farm_acres_2      0.001030   0.000558    1.85    0.065 .
##
## Residual standard error: 0.312 on 2808 degrees of freedom
## Multiple R-squared:  0.538,  Adjusted R-squared:  0.537
## F-statistic:  816 on 4 and 2808 DF,  p-value: <2e-16
```

fmB summary:

```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       11.02278    0.02502  440.55  < 2e-16 ***
## farm_acres_2       0.00106    0.00056    1.89    0.059 .
## pop_dens_2        -0.18250    0.02434   -7.50  8.7e-14 ***
## grazing_rotation -0.77515    0.11646   -6.66  3.4e-11 ***
## fertilizer_use_2  0.00508    0.00013   39.21  < 2e-16 ***
## bee_colonies_2    0.00590    0.01109    0.53    0.595
##
## Residual standard error: 0.312 on 2807 degrees of freedom
## Multiple R-squared:  0.538,  Adjusted R-squared:  0.537
## F-statistic:  653 on 5 and 2807 DF,  p-value: <2e-16
```

Appendix 13: Evaluation of model without outlier, fmA3

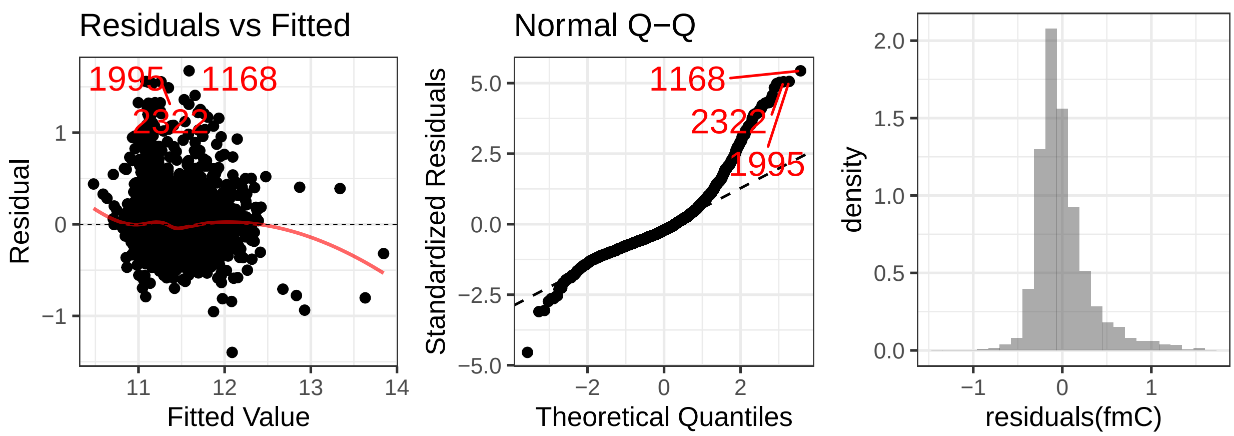


fmA3 summary:

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       11.022920   0.024677  446.68  < 2e-16 ***
## pop_dens_2        -0.175508   0.023507   -7.47  1.1e-13 ***
## grazing_rotation -0.763733   0.115014   -6.64  3.7e-11 ***
## fertilizer_use_2  0.005098   0.000127   40.07  < 2e-16 ***
```

```
## farm_acres_2      0.000991   0.000551    1.80    0.072 .
##
## Residual standard error: 0.309 on 2807 degrees of freedom
## Multiple R-squared:  0.543,  Adjusted R-squared:  0.542
## F-statistic:  833 on 4 and 2807 DF,  p-value: <2e-16
```

Appendix 14: Evaluation of model with interactions, fmC



fmC summary:

```
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 11.049113   0.026524  416.56  < 2e-16 ***
## pop_dens_2                  -0.249320   0.077285   -3.23   0.0013 **
## grazing_rotation            -0.888741   0.129364   -6.87  7.9e-12 ***
## fertilizer_use_2             0.004965   0.000144   34.48  < 2e-16 ***
## farm_acres_2                 0.001302   0.000561    2.32   0.0203 *
## pop_dens_2:grazing_rotation  0.729898   0.381994    1.91   0.0561 .
## pop_dens_2:fertilizer_use_2  0.001493   0.000677    2.21   0.0274 *
## pop_dens_2:farm_acres_2     -0.012375   0.005313   -2.33   0.0199 *
##
## Residual standard error: 0.308 on 2804 degrees of freedom
## Multiple R-squared:  0.544,  Adjusted R-squared:  0.543
## F-statistic:  478 on 7 and 2804 DF,  p-value: <2e-16
```

Appendix 15: Reverse transforming variables

```
coef.intercept<-fmA3$coefficients["(Intercept)"]
exp(coef.intercept)
```

```
## (Intercept)
##       61262
```

```r
#population density
coef.pop<-fmA3$coefficients["pop_dens_2"]
exp(coef.pop)/exp(1)
```

```
## pop_dens_2
##    0.30866
```

```r
#grazing
coef.graze<-fmA3$coefficients["grazing_rotation"]
exp(coef.graze)/1
```

```
## grazing_rotation
##          0.46592
```

```r
#fertilizer
coef.fert<-fmA3$coefficients["fertilizer_use_2"]
exp(coef.fert)/(1^2)
```

```
## fertilizer_use_2
##           1.0051
```

```r
#acreage
coef.acre<-fmA3$coefficients["farm_acres_2"]
exp(coef.acre)/(1^2)
```

```
## farm_acres_2
##        1.001
```