The Impact of Genetically Engineered Upland Cotton on Land Utilization

Tara Chakkithara

Abstract

This paper examines the environmental impact of genetically modified upland cotton production in the United States. Linear regression models were employed to analyze the disparities in land usage between genetically engineered cotton and non-engineered cotton cultivation. The results revealed a significant difference in land usage, indicating the efficiency of genetically engineered cotton cultivation. This study aims to contribute to the growing discourse on sustainability within the agricultural domain.

Contents

Introduction

Cotton is one of the largest agricultural industries worldwide, commonly used in fabric production and a wide variety of other applications. The United States is one of the world's leading exporters of cotton, dominating around a third of the global cotton market. However, cotton cultivation faces significant challenges as its prone to a multitude of pests and disease. To address these issues, genetically engineered cotton was commercially introduced in the United States in 1995.

Over the years, genetically modified cotton has been enhanced to increase yield capacity as well.

This paper aims to investigate the long term impact of genetically modified cotton on land usage and overall cotton cultivation trends within the United States, given that it has been over two decades since its introduction. Existing research mainly investigates genetically modified crops as a monolith, but this paper places its focus on Upland cotton.

Linear models were constructed to compare land usage discrepancies between GMO cotton and non modified cotton. The findings indicate that genetically engineered cotton generates higher yield and requires considerably less land for cultivation. Amidst growing sustainability concerns rising population and income, analysis of cotton production and land utilization are becoming more significant.

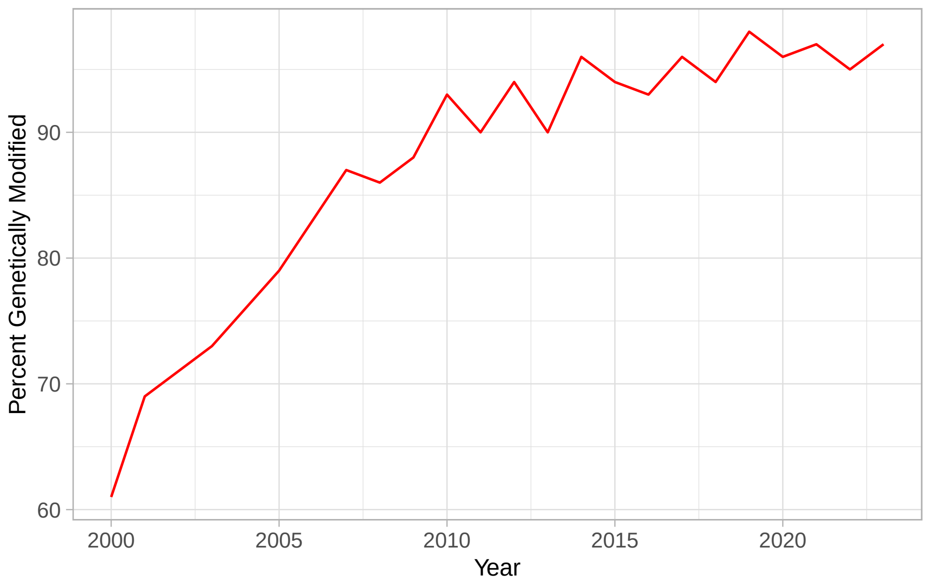
This paper is divided into 4 main sections: data, models, results, and the discussion. The data section examines the data used to construct the models presented in the models section. The results section uses the constructed models to make inferences and predictions. Finally, the last section discusses any other important details pertaining to the topic at hand.

Data

To analyze the effects of genetic engineering on upland cotton cultivation, raw data was obtained from the United States Department of Agriculture. These datasets were cleaned and transformed into a clean dataset using R and the tidyverse package. The resulting dataset consists of the variables "Year", "Percent GM", "Total Acres", "Acres Harvested", "Percent Abandoned", and "Yield".

The variable "Year" encompasses observations spanning from 2000 till 2023. "Percent GM" denotes the percentage of cotton planted that was genetically modified. "Total Acres" represents the total land area in acres dedicated to farming upland cotton in the United States, while "Acres Harvested" is the number of acres that were actually harvested. "Percent Abandoned" is the percentage of cotton that was not harvested and was abandoned. The variable "Yield" measures cotton harvested in pounds per acre.

This dataset was generated over other potential datsets since the original source has high credibility, and the cleaned variables "Year", "Percent GM", "Total Acres", "Acres Harvested", "Percent Abandoned", and "Yield" help investigate trends in upland cotton cultivation that have occurred in recent American history.



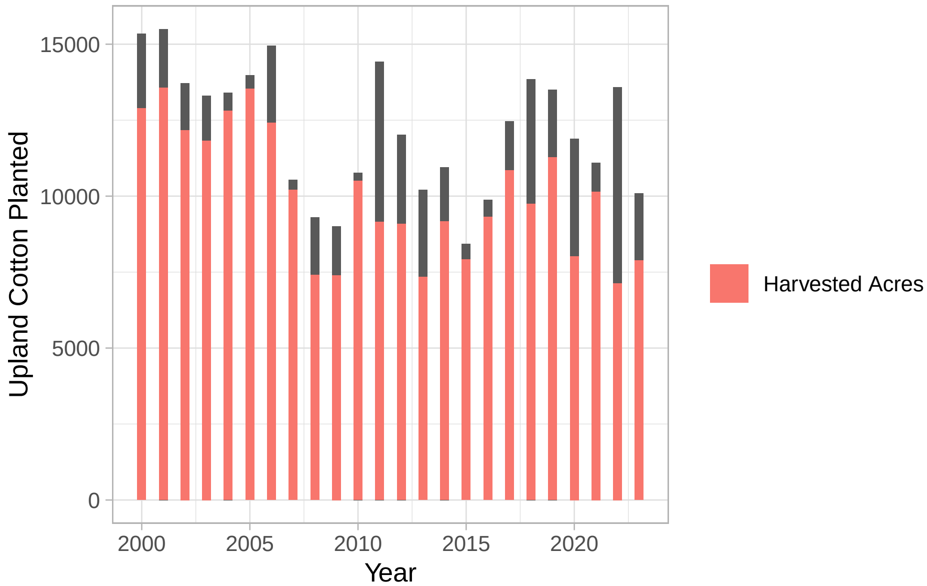
The percent of genetically modified Upland cotton planted in the United States increased drastically from 2000 to 2023. Starting with a minimum value of 61% in 2000, the percentage increased till it reached its peak in 2019 at 98%. The mean percentage of genetically modified upland cotton during this time period was 87%.

Measuring the percentage of genetically modified cotton is a difficult task as cross breeding can occur during cultivation. This type of measurement error is difficult to control as cultivating crops in complete isolation takes a lot of resources. This variable was generated using reports of farms tracked by the USDA.

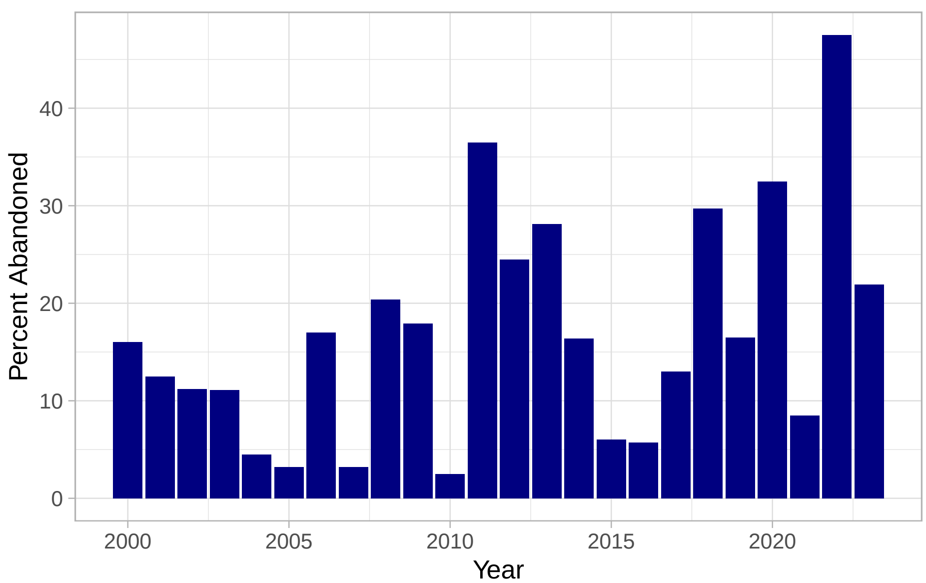
Acreage dedicated to Upland cotton in the United States varied drastically from 2000 to 2023, but notably decreased overtime. It reached a minimum value of 8,422 acres in 2015 and a peak of 15,499 acres in 2001. The mean acreage dedicated to planting upland cotton in the United States was 12,174 acres.

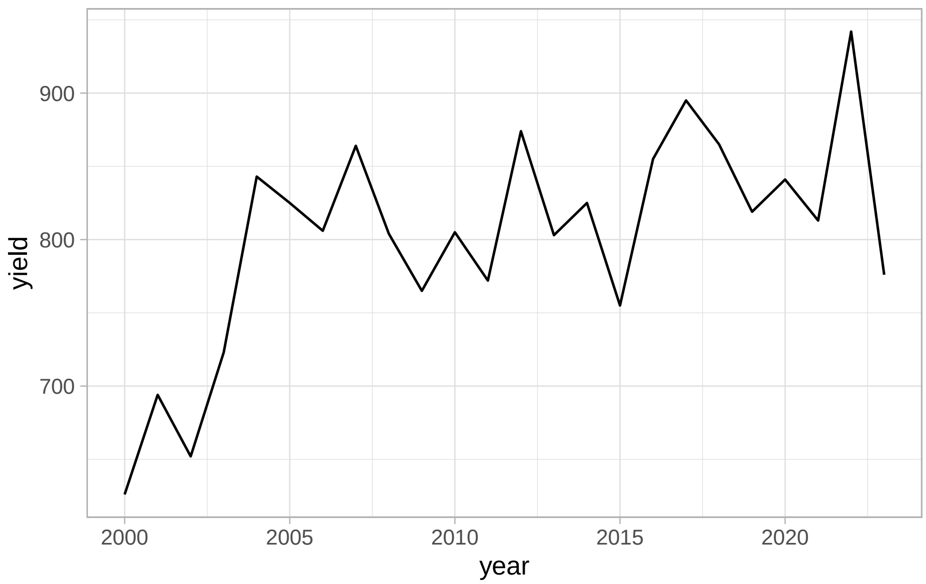
The total acreage of Upland cotton harvested follows a similar trend. The number of harvested acres has the overarching trend of decreasing over time. Harvested acres had a peak value of 13,560 in 2001 and a minimum value of 7,132 in 2022. From 2000 to 2023 the mean value of harvested acres was 10,072.

Measuring acreage is easier and more accurate than measuring the percent of genetically modified cotton. Active farm lands and plantations are tracked by the USDA and since most farming in the United States is commercial this is a fairly accurate measurement.



The percentage of abandoned cotton has a trend of increasing over time. In the year 2010 only 2.5% of the planted Upland cotton was abandoned compared to 2022 when around 48% of Upland cotton planted was abandoned. The mean percentage of cotton abandoned was 16%.





Upland cotton yield in the United States has been steadily increasing over time from 2000 to 2023. Total yield reached a maximum amount in 2022 with 942 pounds per acre compared to the minimum yield of 626 pounds per acre in 2000. The mean yield during this time period was 801 pounds per acre.

Models

To study the long term trends in cotton cultivation from 2000 to 2023, four distinct linear models were constructed. Two of these models were created to analyze overarching trends in cotton production, while the other two were designed to investigate the disparities between genetically engineered cotton and non-GMO cotton.

All models utilized in this study were generated using the tidymodels package. Due to the limited amount of data, bootstrapping using the boot package was also employed for certain models to enhance its accuracy.

Inferencing Long Term Trends

The first model constructed investigates the influence of genetically engineered cotton on the percentage of planted cotton that was abandoned, using bootstrapping. Thus, in this model GMO cotton acts at the independent variable X and percentage of abandoned cotton represents the dependent variable Y.

$$Y = 0.33X - 11.88$$

The second model was also constructed using bootstrapping and looks into the impact of genetically engineered cotton on total yield. In this model, GMO cotton is designated as the independent variable X and total yield acts as the dependent variable Y.

$$Y = 5.07X + 359.24$$

Simulating Non-GMO Cotton Usage

The third model is a multiple linear regression model that uses two predictors: the percentage of GMO cotton planted and the total acres of cotton planted, denoted X_1 and X_2 respectively, to predict the total yield Y.

The error term, denoted as ε represents the deviation between the true values and model predictions. The standard deviation of the error term can be estimated using the residual standard error (RSE) of 51.5, implying an approximate probability distribution of $N(0, 51.5^2)$ for the error term.

The goal of this model is to simulate the effect of land area on total yield while the percentage of GMO cotton is set to zero. This simulation aims to show the discrepancy in yield between genetically engineered cotton and non modified cotton.

$$Y = 5.98X_1 + 0.01X_2 + 178.70 + \varepsilon$$

The last model is also a multiple linear regression model. It uses the two predictors: the percentage of GMO cotton planted and total yield, denoted as X_1 and X_2 respectively. This model is used to predict the total land in acres used to plant cotton which is represented by the dependent variable Y.

The standard deviation for the error term in this model can be estimated using an RSE of 1774.72, so the probability distribution of the error term ε can be estimated with $N(0, 1774.72^2)$.

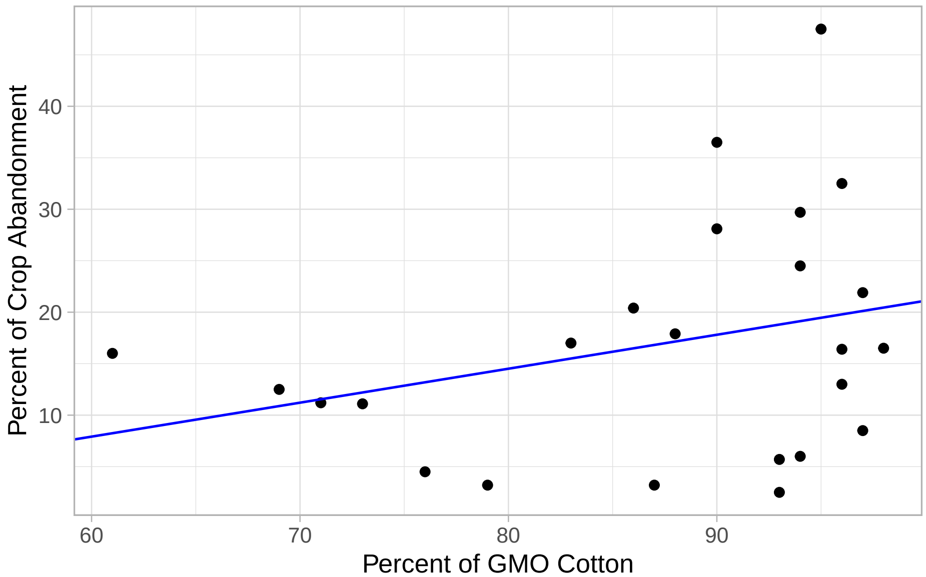
The goal of this model is to simulate the effect of yield on total land usage while the percentage of GMO cotton is set to zero. This simulation aims to show the discrepancy in land usage between genetically engineered cotton and non modified cotton.

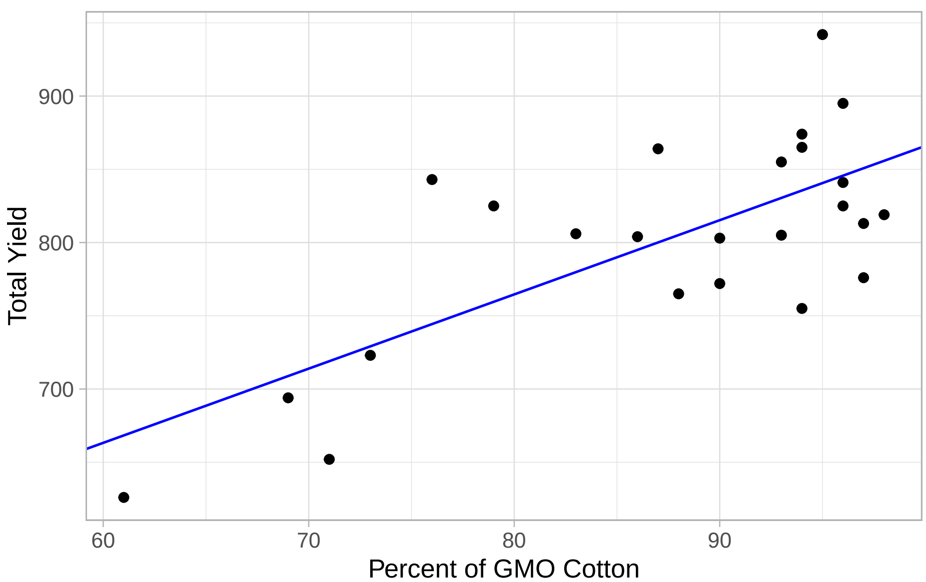
$$Y = -159.52X_1 + 9.86X_2 + 18199.24 + \varepsilon$$

Results

The first model $Y = 0.33X - 11.88$, attempts to investigate the impact of GMO cotton on crop abandonment. The coefficient of 0.33 for GMO cotton is not statistically significant, with a p-value of 0.15. Thus, there is insignificant evidence to conclude that GMO cotton impacts crop abandonment.

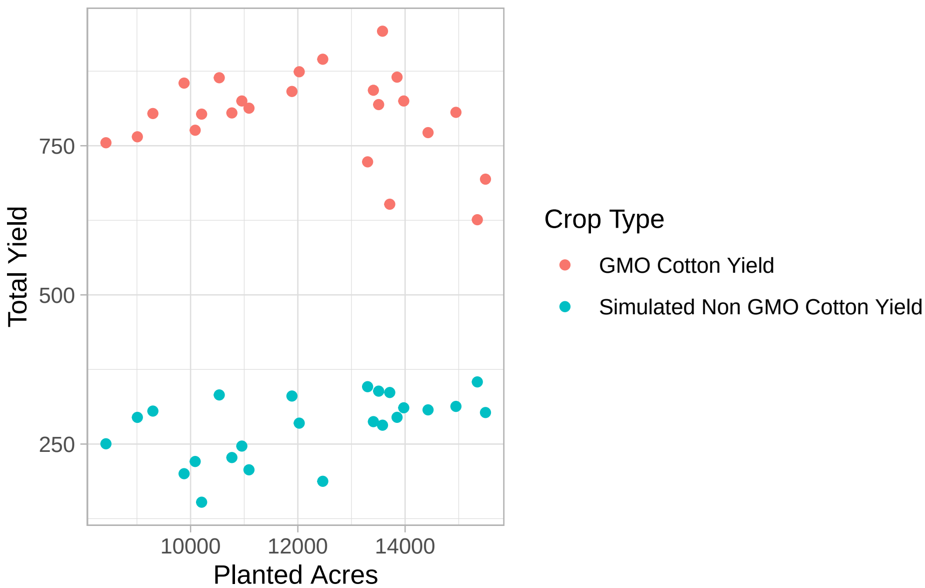
The correlation coefficient between the two variables is close to 0, with a value of 0.30, which indicates a weak linear relationship. The lack of significance may be due to the fact that a linear model may not capture the relationship between the two variables, as seen in the figure below.



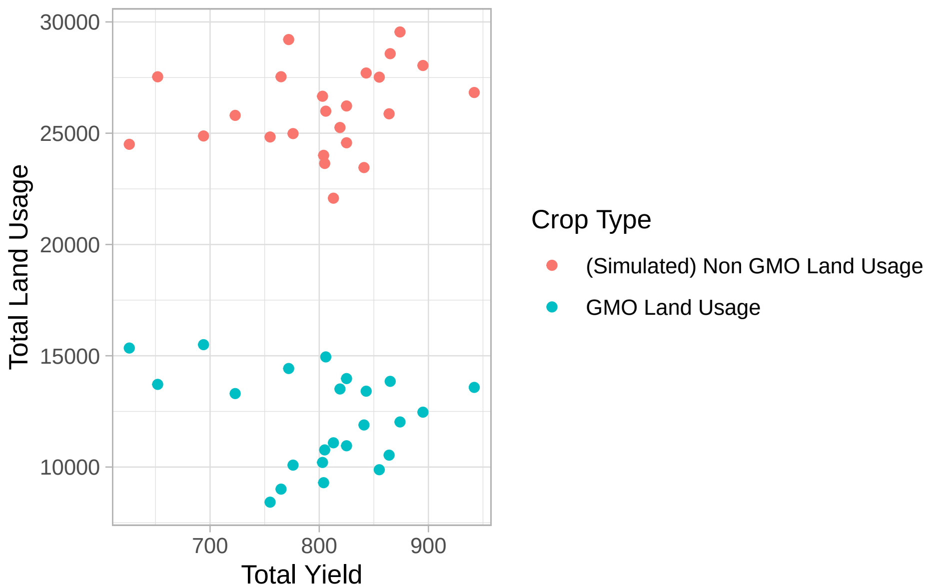


In the second model, represented by the equation $Y = 5.07X + 359.24$, the coefficient 5.07 is statistically significant with a p-value of 7.741e-05. Thus, there is enough evidence of a positive relationship between GMO cotton and yield as seen in the figure below.

The third model was used to simulate hypothetical yield results if GMO cotton had not been employed. As seen in the figure below, the simulated Non-GMO yield is significantly lower than the yield resulting from genetically engineered cotton.



The last model was employed to simulate the hypothetical land usage of Non-GMO cotton. As depicted in the figure below, the simulated land usage is considerably higher than that of genetically engineered cotton.



Discussion

References