Farmer Demographics and Inflation in U.S. Agriculture

HDSI Agri Datathon 2024
Terabyte Terriers
Oct 6th, 2024

Axusmawe Asmelash
Boston University
1 Commonwealth Ave
Boston, MA 02215
aasmelas@bu.edu

Ria Sonalker
Boston University
1 Commonwealth Ave
Boston, MA 02215
rsonalk@bu.edu

Ananya Agarwal
Boston University
1 Commonwealth Ave
Boston, MA 02215
ananya04@bu.edu

Abstract

This challenge aims to investigate the relationship between farmer age demographics and farm production trends across various regions of the contiguous United States. We analyzed National Agricultural Statistics Service (NASS) data, including demographic and land use totals at county and state levels, to identify patterns in farmer age. Economic data was adjusted for inflation to evaluate trends in crop sales, with special attention to the distinctions between "operators" and "producers". We used Linear Regression to explore how age trends correlate with farm production, incorporating relevant variables and assessing potential impacts on the agricultural economy.

1. Introduction

The National Agriculture Statics Service (NASS) conducts a census every 5 years that provides in depth information on nearly 2 million farms within America. Often, these farms are manned by groups of farmhands who have a direct hand in cultivating, transporting, and ultimately selling their harvest. 1.9 million farms dot America's rural landscape, and 95% are operated by families [Ser22]. It is crucial for any analysis of NASS data to include the myriad of independent variables that ultimately affect the production levels of farms.

Recently, the national economy has experienced relatively high rates of inflation, which can skew data and impact the accuracy of economic analysis. In agricultural production, high inflation rates may distort the perceived trends in farm sales and outputs, making it challenging to draw meaningful conclusions about the relationship between farmer demographics and production. Such inflationary pressures can complicate comparisons over time, leading to misinterpretation of trends that are crucial for understanding the agricultural economy and its future direction.

To address this concern, we have cross-analyzed farm sales data with current inflation rates to isolate the impact of inflation on production levels. This adjustment ensures a more accurate reflection of economic trends at both the county and state levels. By incorporating inflation into our analysis, we strengthen the reliability of our findings and provide a more stable foundation for predictive machine learning models. This approach allows for more precise insights into the relationship between farmer demographics and production, offering a clearer view of potential future trends in the agricultural economy.

2. Data and Methods

In order to aggregate the information taken from over 3 thousand counties in America, we first removed columns with more than 50% NaN values to avoid losing entire rows of data due to missing values in specific columns. We then eliminated rows with NaN or "(D)" values and removed county FIPS codes labeled as 999. To address the issue with operators and producers, we have combined the two fields into one column. Sales and number of operations columns were segregated using pattern filtering, and we calculated the average sales per operation as a measure of productivity for each county. A cleaned dataset was created with key columns, including year, county FIPS, average age, and average sales per operation. Counties were categorized into the top and bottom 25% based on productivity. Data types were adjusted to facilitate accurate calculations, and outliers were retained for analysis.

We applied polynomial regression to model the relationship between years and the average age of farmers in the top 25% and bottom 25% of productive counties. This choice was driven by its capability to capture non-linear trends,

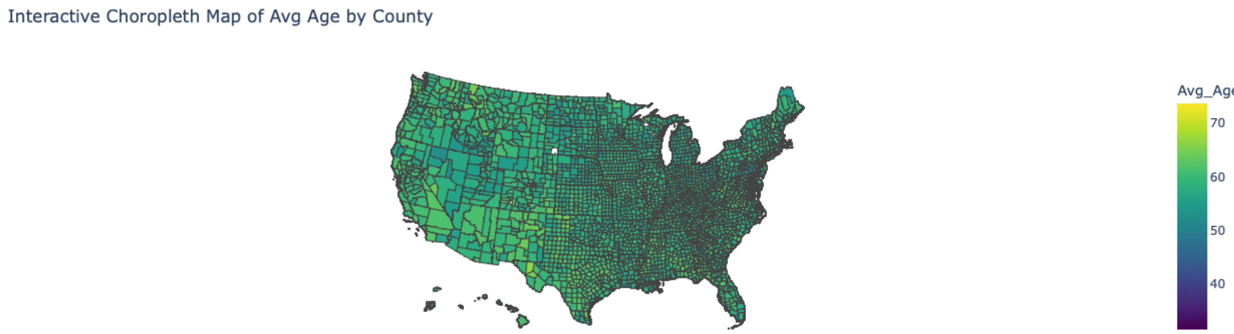


Figure 1: Choropleth map of the average age of producers working in farms in counties across the country. Data set shown is from 2022, the most recent census taken by NASS [Ser22].

which linear models could not accurately represent. Using sci-kit learn, we trained our models on 80% of the data and tested on the remaining 20%, resulting in Mean Squared Error (MSE) values of 5.658 for the top counties and 3.012 for the bottom counties, demonstrating the model's effectiveness. While we considered time series models like ARIMA, the non-continuous nature of our data led us to prefer regression for its trend analysis capabilities. Visualization tools such as matplotlib and seaborn were integral to interpreting our results, highlighting the trends captured by the polynomial regression models. This comprehensive approach ensured our findings were both robust and grounded in the data's inherent patterns.

3. Results

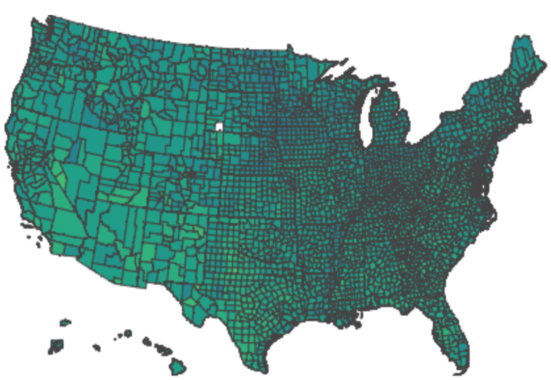


Figure 2: Map of the average age of producers working on farms across counties in the U.S. Data shown is from 1997 [Ser22].

The first question we answer was as follows, "*How has the average age of farmers evolved over time across county, states, and nation in the United States, from 1997 onwards?*". By applying the cleaned FIP, Year, and age data

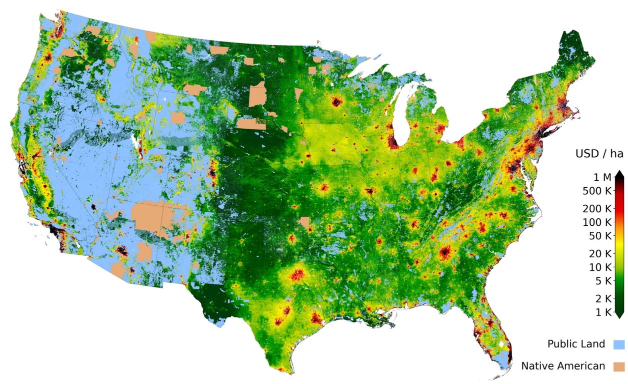


Figure 3: Map of the public vs. private land ownership and the subsequent land valuation. [Ego21].

we accrued we were easily able to create a colormap for age across American counties. As you can see in **Figure 1** and **Figure 2**, the average age of farmers across the nation has noticeably increased in the period of 25 years (1997-2022). However, the national average age for a farmer was 65 years old, in 2022 that number was 58.1 years old. The regions at which the age of farms are more significantly older than recent years is predominately in the western half of the country. A comparison with a map of public versus private land seen in **Figure 3** reveals an interesting pattern: older farming population is most pronounced in regions where public land is more abundant. This analysis highlights a significant shift towards a older farming population in the United States over the past 25 years, with the most pronounced changes occurring in regions abundant with public land, primarily in the western part of the country.

The next question we we address was as follows: "*What are the trends in average age for the most productive counties in the U.S., as measured by the dollar value of agricultural production? What about the lowest productive coun-*

ties in the U.S.?". **Figure 4** displays the results of polynomial regression analysis on the average age of farmers in the top 25% and bottom 25% of productive counties over the years. From the graph, it is evident that counties in the bottom 25% have consistently higher average ages compared to those in the top 25%. The trend for the top 25% of counties shows a slight increase in the average age of farmers over time. In contrast, the bottom 25% counties exhibit an initial increase in average age, followed by a decline in recent years, returning to values close to those seen in 2008. This suggests that younger farmers are more prevalent in the more productive regions, while the less productive areas have a relatively older farming population with fluctuating trends.

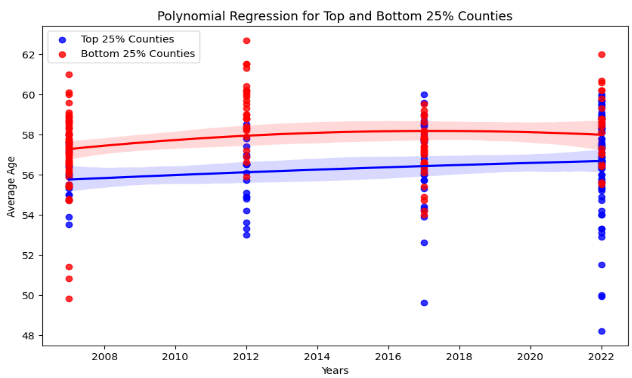


Figure 4: Polynomial Regression for top and bottom 25% counties.

The final question we answered was "*When analyzing farm production and economic impact, how does the use of an inflation adjustment measure for the value of agricultural products affect the analysis?*. **Figure 5** illustrates the impact of inflation on agricultural sales data, highlighting the differences between nominal and inflation-adjusted values over time. The blue line, representing nominal sales, shows a consistent and steep upward trend, suggesting significant growth in agricultural sales. In contrast, the orange line, representing adjusted sales, indicates a more moderate increase when inflation is accounted for. This difference suggests that much of the observed rise in nominal sales is due to inflation rather than a genuine increase in agricultural production. By using an inflation adjustment, we can filter out the effects of price increases, providing a clearer view of real economic growth in the agricultural sector. Similarly, the images in **Figure 6** further reinforces the importance of inflation adjustment. The GDP Difference graph underscores this trend by directly showing the growing gap between nominal and real GDP values, emphasizing the inflationary impact on the economy.

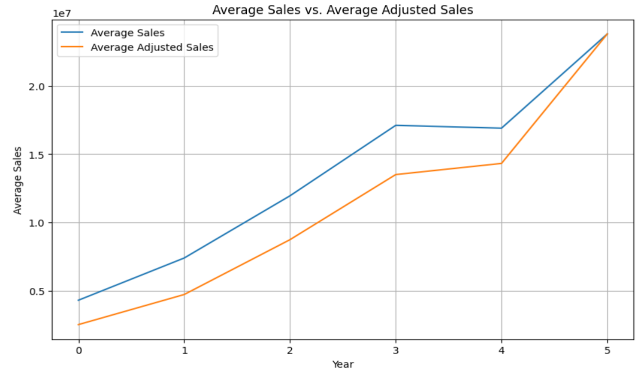
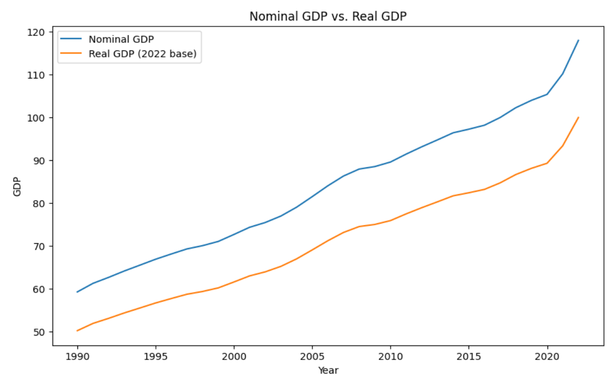
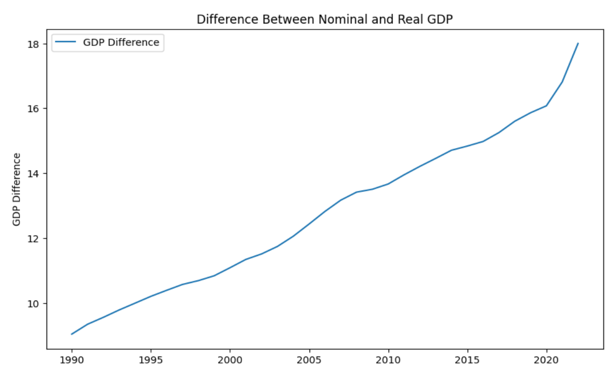


Figure 5: Average sales adjusted to inflation levels over time.



(a) Nominal GDP vs. Real GDP.



(b) Difference in GDP over time.

Figure 6: Comparison of Nominal GDP and Real GDP alongside the Difference in GDP.

We studied animal and crop sales in regions across the America. Across all graphs, a consistent pattern emerges: the adjusted values display a less steep increase compared to nominal data, indicating that inflation significantly influences the perceived growth of agricultural production and economic output. This pattern also highlights the importance of regional analysis, as similar growth trends are observed across different areas but with varying intensities. Incorporating inflation adjustments into the analysis is crucial for accurately assessing true economic trends and avoiding misleading conclusions based on nominal figures alone.

4. Conclusion

In conclusion, this study reveals significant trends in farmer demographics and economic patterns within the U.S. agricultural sector, addressing our key research questions. Our analysis highlighted a shift toward an older farming population, particularly in less productive regions with abundant public land, suggesting a connection between younger farmers and higher productivity levels. The use of polynomial regression models enabled us to effectively capture these trends, providing a deeper understanding of how farmer age relates to agricultural output over time.

Additionally, our findings underscore the importance of adjusting for inflation when analyzing agricultural sales data. Without this adjustment, nominal sales figures can lead to overestimated perceptions of economic growth, masking the true trends in production. By accounting for inflation, we were able to present a clearer picture of real economic progress in the sector. Future research could build on this work by exploring the impact of specific policy interventions on farmer demographics or investigating the role of land ownership patterns in shaping agricultural productivity. Acknowledging that our study was limited by the non-continuous nature of available data, future analyses with more granular, continuous datasets could further enhance the accuracy and insights of these findings.

Please see this link to our Submission Video.
Section 1 Notebook. Section 2 Notebook.
Section 3 Notebook.

References

[Ego21] Alex Egoshin. Value of private land in the u.s. mapped, 2021. 2

[Ser22] USDA National Agricultural Statistics Service. USDA releases 2022 census of agriculture data, 2022. https://www.nass.usda.gov/Publications/AgCensus/2022/. 1, 2