POLITECNICO
MILANO 1863

Mathematical Engineering A.Y. 2022-23
Nonparametric Statistics

Non Parametric Analysis of US Agricultural Productivity

Final Report

S. Moroni

September 7, 2023

Contents

1 Introduction

United States are one of the largest consumers and producers in world agricultural commodity markets. Productivity growth in the U.S. farm sector has implications for both U.S. and global food markets. Slowing productivity growth, that fails to keep pace with increasing food demand, may lead to rising food prices. While in the short term, transitory events—such as energy shocks or supply shortages due to bad weather—may cause agricultural commodity prices to rise, the long-term growth trend in U.S. agricultural productivity has enhanced food security and benefited consumers by reducing the real (inflation-adjusted) price of agricultural outputs over time.

Technological developments in agriculture have been influential in driving long-term growth in U.S. agricultural productivity. Innovations in animal and crop genetics, chemicals, equipment, and farm organization have enabled continuing output growth while using much less labor and farmland. As a result, total agricultural output nearly tripled between 1948 and 2019—even as the amount of labor and land (two major inputs) used in farming declined by about 75% and 24%, respectively. Indeed the composition of total inputs has shifted considerably toward more use of farm machinery (part of capital goods) and intermediate inputs and less use of labor and land.

The following study about the agricultural productivity growth is conducted in order to build a clear picture of how the agricultural economy is performing and to identify the primary drivers of the agricultural sector's growth. This allows policymakers and their economic advisers to assess how much emphasis should be placed on the mobilization and composition of production components against rises in productivity . In the following sections an analysis will be conducted into how the main agricultural inputs impact on the agricultural productivity level since 1948 to 2019 (Sections 3.1-3.5). Then a study to identify possible different trends in productivity growth across United States, also related to the demand of commodities in US counties, will be carried out(Sections 3.1, 3.6, 3.7).

2 Data description

The data consists of yearly estimates of productivity growth and agricultural inputs growth at national and states level in the U.S. farm sector. The variables are quantity indexes that describe the changes in production or consumption of the farm sector with respect to a fixed year (i.e. 2015), that has the indexes set equal to 1.

There is a wide range of available variables that can be summarize in four different groups:

- **Output**: gross production leaving the farm
- **Capital Input**: wide range of physical assets used in agricultural activities
- **Labor Input**: total human effort, in terms of both quantity and skill, in agricultural production
- **Itermediate Input**: materials, resources, and products that are used within the agricultural production process but are not directly transformed into the final agricultural output.

Capital Input	
Variables	**Description**
Durable Equipment	Long-lasting assets used in production processes
Service Buildings	Infrastructures that support farming operations
Inventories	Stocks of goods and commodities
Land	Quantity of land used in production processes

Labor Input	
Variables	**Description**
Hired Labor	workforce employed by US farms
Self Employed and Unpaid Family	individuals who are engaged in activities on their own farms

Intermediate Input	
Variables	**Description**
Feed and Seed	
Energy	
Fertilizer and Lime	
Pesticides	
Purchased Services	

Table 1: Variables of the analysis

As mentioned above, also State-level estimates of quantity indexes of Total output, Total Capital Input, Total Labor Input and Total Intermediate Input for the period 1960-2004 have been used in the analysis (Section 3.1, 3.6). For the spatial analysis section, the data provided by [4] allowed to have spatial data regarding the amount of commodities and the population of most of the USA counties in 2007 [2]. Due to lack of data, it was decided to include just continental USA, without taking into consideration Alaska and Hawaii. Finally, in order to avoid any scalability concerns, the variables were logged before starting with the study.

3 Analysis

3.1 Analysis of the Inputs over years

One aim of the study is to observe the trends of the input variables over time and how do they impact the agricultural production index.

The `Labor Input` index stands out with significant change over time. The substantial reduction during the first period could be attributed to technological advancements or changes in agricultural practices that allowed for the same or even greater production with reduced labor input. After 1990, the index stabilized around 1, indicating a relative consistency in the use of labor for agricultural activities. This suggests that the sector has achieved a balance in labor allocation, likely after maximizing efficiency gains from previous innovations.

The behavior of `Capital Input` index presents an intriguing dynamic. Prior to 1990, the index remains relatively stable around 1.3, indicating a possible saturation in capital resource utilization or an optimization of equipment and infrastructure. This may reflect a phase where further increases in capital did not lead to proportional production increments, suggesting efficient resource allocation. Post-1990, a second plateau emerges with slightly lower values that suggest a stable yet more refined utilization of capital, possibly in response to changing technological or market dynamics.

The dynamics of both `Intermediate Input` and `Total Output` indexes show a consistent and constant increase.

The significance of the difference of the means of the Output Variable and of the Input

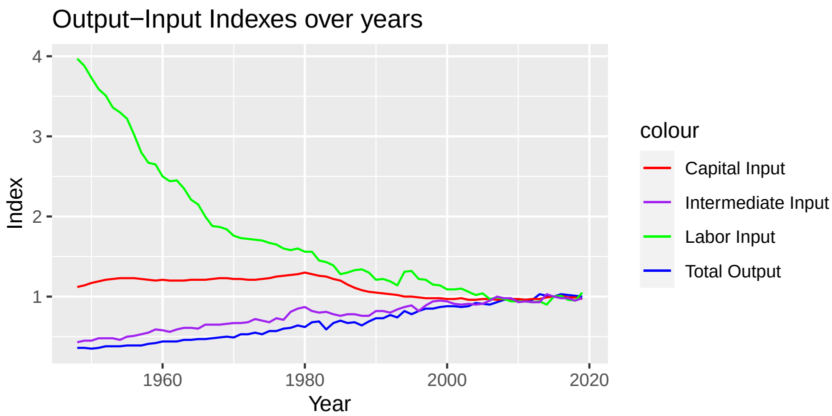


Figure 1: Inputs and Output trends over years

Variables between the two historical periods has been assessed using two paired samples non-parametric permutational tests with $alpha = 0.05$.

Input	Total Output 60-90	Total Output 90-15
Capital Input	0.1656040	-0.3183969
Labor Input	-0.9876408	-0.8513482
Intermediate Input	0.9568786	0.7858585

Table 2: Spearman correlation between Input-Output

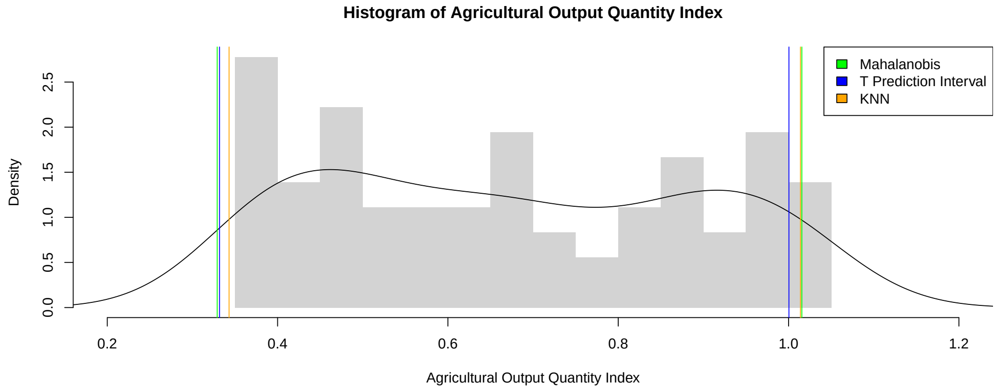
Also the Spearman Correlation index was calculated, in order to assess the type and strength of the correlation between Input factors growth and Output growth. The results suggest that `Intermediate Input` plays a significant role in increasing agricultural output, while excessive use of Labor Input can have negative effects. `Capital Input` shifts its correlation from being weakly positive to negative. This variation highlights the need to consider broader factors such as sustainability, innovation, and market demands that played a significant role in shaping the correlation between Capital Input and Total Output over time.

3.2 Comparison of conformity measures for agricultural production index prediction

In this section the aim is to forecast the agricultural output quantity index for the next years. A first attempt to have an estimate of a quantity index range involves employing a conformal approach with different measures. I considered the following options:

- using Mahalanobis distance
- using T Prediction Interval
- using KNN distance

The confidence level was set to $\alpha = 0.05$. We can dismiss these results since we obtained wide intervals that lack significant information.



	LOWER	UPPER
Mahalanobis	0.3299623	1.015766
T Prediction Interval	0.3377823	1.002827
KNN	0.3429020	1.016266

3.3 Agricultural Output quantity Index Prediction using GAMs

A Generalized Additive Models (GAMs) has been used to improve the results obtained in the preceding section by accounting for the input variables and learning which are the main inputs that impact the agricultural production growth and how they influence it.
Starting from using all the available regressors, the model has been reduced and smooth terms using cubic B-splines have been added. Finally, the significance of all regressors was assessed using permutation tests with $\alpha = 0.10$.

The final model is the following:

$$\begin{aligned}\texttt{Total.Agricultural.Output}_i = \beta_0 &+ \beta_1 * \texttt{Intermediate.Energy.Input}_i \\ &+ \beta_2 * \texttt{Intermediate.Pesticides.Input}_i \\ &+ \beta_3 * \texttt{Capital.Service.Buildings.Input}_i : \texttt{Period}_i \\ &+ \beta_4 * \texttt{Self.Employed.Unpaid.Family}_i : \texttt{Period}_i \\ &+ f_1(\texttt{Capital.Durable.Equipment.Input}_i) \\ &+ f_2(\texttt{Labor.Hired.Labor.Input}_i) + \varepsilon_i \end{aligned} \tag{1}$$

with $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0,\, \sigma^2)$. The variable `Period` has been introduced in order to distinguish the two periods identified in Section 3.1.

$$\beta_1, \beta_2, \beta_{3,2} > 0 \tag{2}$$
$$\beta_{3,1}\, \beta_{4,1}\,, \beta_{4,2} < 0 \tag{3}$$

3.4 Interpretation of the model

The model coefficients are the following:

	Coefficient
(Intercept)	0.92356426
`Intermediate.Energy.Input`	0.04835645
`Intermediate.Pesticides.Input`	0.18684457
`periodfirst:Capital.Service.buildings.Input`	-0.18769120
`periodsecond:Capital.Service.buildings.Input`	0.04182428
`periodfirst:Labor.Self.employed.and.unpaid.family.Input`	-0.08790924
`periodsecond:Labor.Self.employed.and.unpaid.family.Input`	-0.29701090

`Intermediate.Energy.Input` measures how the quantity of energy used in agricultural activities has changed with respect to 2015. A positive coefficient associated with this variable indicates that an increase in the use of energy is associated with an increase in overall agricultural production. In other words, as expected, when more energy is employed in agricultural activities, the total quantity of agricultural production tends to grow.
`Intermediate.Pesticides.Input` has a positive coefficient. This suggests that increased pesticide usage as an intermediate input is associated with higher agricultural production. However, this does not imply causal causation, since other variables, like characteristics of the land, weather conditions or government policies, may be at work. The negative coefficient for `periodfirst:Capital.Service.buildings.Input` suggests that during 1948-1990, increased input of capital services and buildings was associated with reduced agricultural production. This phenomenon may seem counterintuitive, but can be associated with overcapitalization, misallocation of resources and high fixed costs. Conversely, the positive coefficient for `periodsecond:Capital.Service.buildings.Input` indicates that in 1990-2015, greater input of capital services and buildings are correlated with higher agricultural production. The changing impact of this variable over time on agricultural output may reflect changes in management practices, technologies, or economic factors.
The negative coefficient for `period:Self.employed.and.unpaid.family.Input` indicates that the relationship between the labor input of self-employed and unpaid family members and the agricultural output varies across the two time periods (1950-1990 and 1990-2015). Specifically, it suggests that during the second period, the negative influence of self-employed and unpaid family labor on agricultural output is stronger compared to the first period.

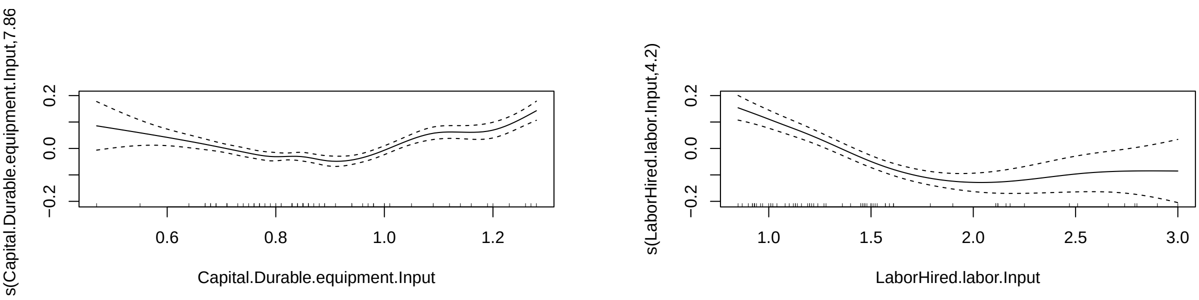


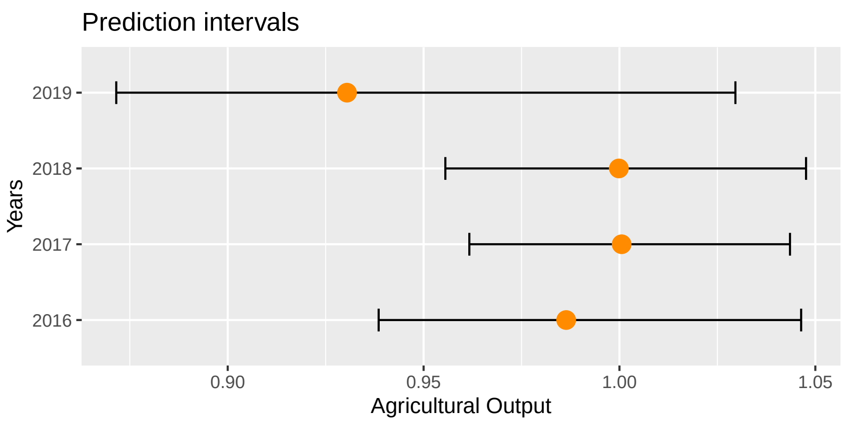
Figure 2: Non-linear effects of covariates

The initial decrease followed by an increase in the curve of `Durable.Equipment` (Figure 2) could indicate a threshold effect. This suggests that up to a certain level of durable equipment growth(approximately 0.95), the impact on agricultural production might be negative or limited. Beyond this threshold, increasing equipment input could have a more significant positive impact on production. It's important to note that identifying such threshold can have significant implications for decision-making, resource allocation, and optimizing agricultural practices.

The pattern of `Labor.Hired.Labor.Input` implies a non-linear relationship with the response variable. The initial decrease from positive to negative values followed by the plateau-like region could signify a point where increasing the input of hired labor has decreasing returns on the agricultural output. This suggests that beyond a certain level (around 1.3), additional increases in hired labor might not result in significant gains in productivity. It could indicate efficient resource utilization, where further labor additions yield diminishing proportional benefits. The impact of new technology, as well as farm size and workplace dynamics, might all have an impact on this behavior.

3.5 Reverse percentile intervals using bootstrap approach with GAM

GAM is used to gain insights into the dynamics of agricultural productivity. Thus the model is employed to predict the agricultural output quantity index of 2016-2019 using available input quantity indexes. By comparing the model's predictions with actual historical data, we can get knowledge of its predictive accuracy over time. Furthermore, an interval is computed using bootstrap approach at level $\alpha = 0.10$ based on the aforementioned GAM model (1).



	2016	2017	2018	2019
True Value	1.03	1.02	1.01	1.00
Prediction for Agricultural Output				
Lower bound	0.9385257	0.9616798	0.9555540	0.8715392
Pointwise	0.9863989	1.0005593	0.9998518	0.9304462
Upper bound	1.0463802	1.0435294	1.0476331	1.0296020

Table 3: Comparison between true values and predictions.

The model's pointwise predictions demonstrate its ability to closely estimate the agricultural output quantity index for the years 2016-2018. The greater deviation in 2019 underscores the challenge of accurately predicting outcomes when considering dynamic, external factors. Overall, this analysis showcases both the strengths and limitations of the model.

3.6 Analysis across US States

This session will investigate possible differences in the agricultural productivity across U.S. First a sequential clustering analysis on the states based on the 'Total Output Quantity Index' within the period of 1960-2004 has been conducted. This procedure aimed to identify groups of states exhibiting similar patterns in the agricultural output trends over the years. To calculate the similarity between the temporal sequences, I employed the Dynamic Time

Warping (DTW) function, which accounts for variations in timing and speed within the sequences.

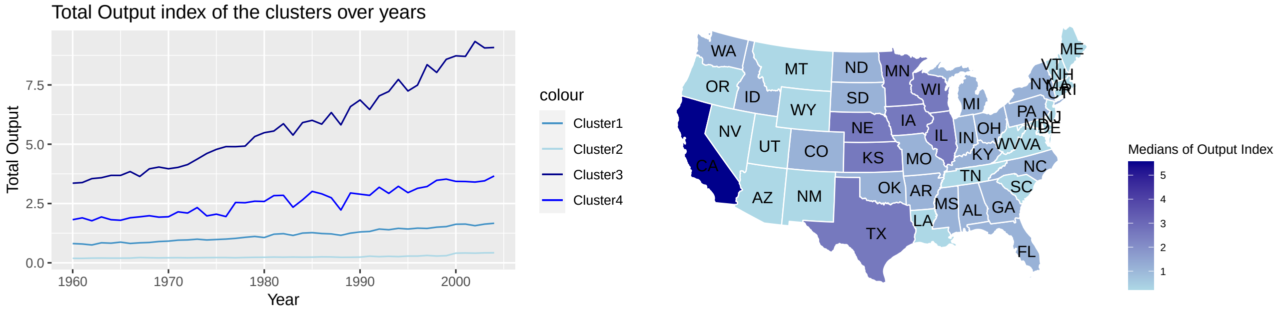
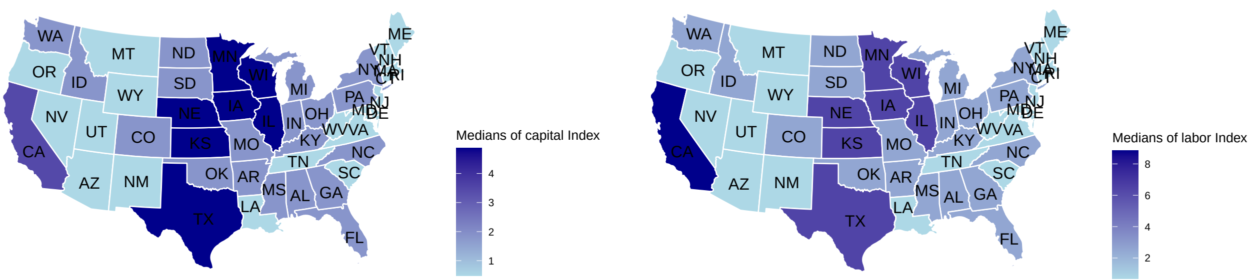


Figure 3: Trends of the clusters in time and space.

The clustering process revealed a partitioning of states into 4 distinct groups.The presence of diversified clusters underscores the importance of accounting for regional variations in agricultural analysis and emphasizes the heterogeneity of agricultural practices and resources within the United States.
Particularly noteworthy is a cluster consisting of only the California. This state exhibits 'Total Output Quantity Index' values significantly higher than those of other states, pointing out high agricultural productivity in the region. The nature of this singleton suggests that California might possess agricultural characteristics or government policies that have led to higher outcomes during the examined period.

In order to gain a more comprehensive understanding of the underlying agricultural dynamics, a Permutational Manova Analysis is conducted on the Agricultural Inputs. The Manova analysis on Capital Input, Labor Input and Intermediate Input shows that the clustering not only captures output distinctions but also reflects statistically significant variations in the use of inputs. This means that the differences in output trends among the identified clusters are not random but potentially supported by distinct input usage patterns.



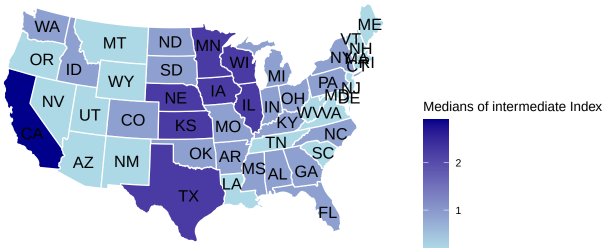


Figure 4: Trends of cluster in space: Capital Input at the top left, Labor Input at the top right and Intermediate Input at the botton

3.7 Nonparametric Spatial Analysis using GAMs

In order to provide valuable insights for business decisions and agricultural development strategies, in this section the analysis focuses on identify regions with the highest sales of agricultural commodities through a spatial GAMs. Areas with high sales could indicate regions with a strong demand for agricultural commodities and a higher necessity for agricultural productivity. The model is the following:

$$
\begin{aligned}
\log(\texttt{Total\_Commodities\_Sales}) = {} & f_1(\texttt{Latitude}, \texttt{Longitude}) \\
& + f_2(\log(\texttt{Population}))
\end{aligned}
\tag{4}
$$

The model explains 32.8% of variability. The effect of the covariates on the sales is shown in Figure 5.

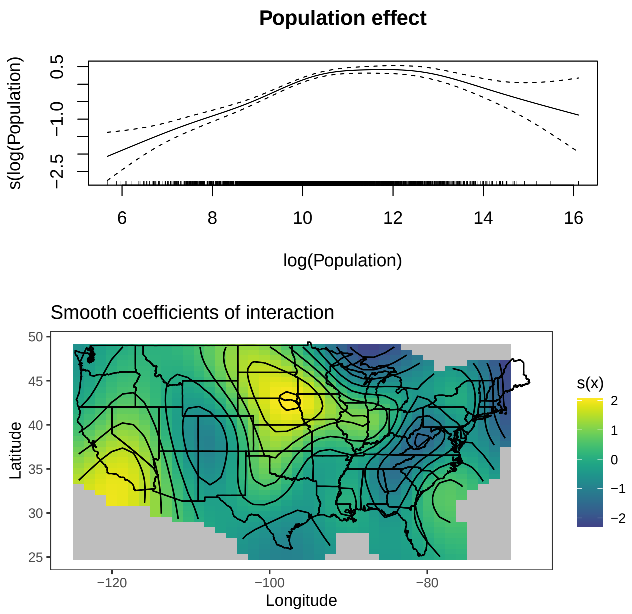


Figure 5: Effect of the coefficients

4 Conclusions

This analysis delved into various aspects of agricultural productivity in the United States, shedding light on the relationships between input factors and output quantities. The study began by observing trends in input variables over time and their impacts on the agricultural production index. The results indicated dynamic shifts in labor, capital, and intermediate input usage, with changes in labor allocation, capital saturation, and evolving relationships with intermediate input.

The use of Generalized Additive Models (GAMs) enriched the understanding of the relationship between input variables and agricultural output. The GAM model revealed significant coefficients for different input factors, including intermediate energy and pesticides input, as well as the role of capital service buildings and self-employed labor input across two distinct periods. The non-linear effects of durable equipment and hired labor input were also captured, highlighting threshold behavior and diminishing returns.

The analysis further extended to spatial and regional considerations. Clustering analysis and Manova revealed clusters of states with distinct agricultural output trends and corresponding input usage patterns, emphasizing the importance of accounting for regional variations. The nonparametric spatial analysis using GAMs identified regions with higher agricultural commodities sales, potentially indicating strong demand and the need for increased agricultural productivity.

In sum, this study provided a multi-dimensional understanding of agricultural productivity dynamics, incorporating historical trends, input-output relationships, predictive approaches, and spatial analysis. The findings underscored the complexity of agricultural systems, where factors ranging from technological advancements to regional variations shape productivity outcomes. The insights from this analysis can guide policy decisions, resource allocation, and strategic planning in the agricultural sector, ultimately contributing to improved efficiency, sustainability, and food security. However, its scope is limited by the exclusion of external factors, such us weather conditions or characteristics of the soil, and the inherent complexities of agricultural systems. Acknowledging these limitations is essential for interpreting the findings and understanding the broader context of agricultural productivity dynamics. Further research that incorporates external variables and employs more sophisticated modeling techniques could provide a more comprehensive understanding of the factors driving agricultural productivity.

References

[1] Ball et al. Is u.s. agricultural productivity growth slowing? *Applied Economic Perspectives and Policy*, 35(3):435–450, 2013.

[2] United States Census Bureau.
https://www2.census.gov/programs-surveys/ popest/datasets/.

[3] Agricultural Productivity in the U.S.
https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-u-s/.

[4] United States Department of Agriculture. The quick stats database.
https://www.nass.usda.gov/Quick$_{S}tats/index.php, Apr.$2019.

[5] Wang et al. Public r&d, private r&d, and us agricultural productivity growth: Dynamic and long-run relationships. *American journal of agricultural economics*, 95(5):1287–1293, 2013.

[6] Sun Ling Wang, Paul Heisey, David Schimmelpfennig, and V Eldon Ball. Agricultural productivity growth in the united states: Measurement, trends, and drivers. *Economic Research Service, Paper No. Err-189*, 2015.

[7] Sun Ling Wang, Richard Nehring, and Roberto Mosheim. Agricultural productivity growth in the united states: 1948-2015. *Amber Waves: The Economics of Food, Farming, Natural Resources, and Rural America*, 2018(2), 2018.

[5] [7] [6] [1] [4] [3] [2]

Code

The code is available at the following https://github.com/SofiaMoroni9/Nonparametric-analysis-agricultural-productivity-US Github page.