Nonparametric Analysis of **US Dairy Production and Consumption**

Teo Bucci
Filippo Cipriani
Gabriele Corbo
Andrea Puricelli

POLITECNICO
MILANO 1863

Nonparametric Statistics
MSc. Mathematical Engineering
Politecnico di Milano

December 13th, 2022
Work in Progress Presentation #2

Dataset presentation

The data consists of **yearly** measurements of production and consumptions of **US dairy** products.

Although there are many variables available, we are **focusing** on:

Production
- **Milk price [$]**
- Milk production [lbs]
- Hay price [$] etc.

Consumption [lbs/person]
- Milk (milk, yogurt, cottage, …)
- Cheese (mozzarella, cheddar, swiss, …)

1980 – 2014

Data provided by tidytuesday

+

We retrieved data to fill the observations up to the past year

2015 – 2021

Goals

We identified our main stakeholders in **cheese factories** with one of the following **needs:**

Improvements to the production chain

Identify which dairy products have seen their **consumption increase** and which **decline** over time, and the reasons behind

The aim is to **understand potential improvements** to production chain and **marketing strategy** of declining goods

Optimal price strategy of a new competitor

A **new potential competitor** wants to decide whether to **join the market** or not, the problem is to identify the **potential of its resources**

It’s therefore important to have a clear idea of the **pricing** of goods and make **predictions** about it

Data Pipeline

Adjust prices for **inflation**:[1]

$$\text{2022 USD value} = \frac{\text{CPI in 2022}}{\text{CPI in 1980}} \cdot \text{1980 USD value}$$

Embed **import/export** data

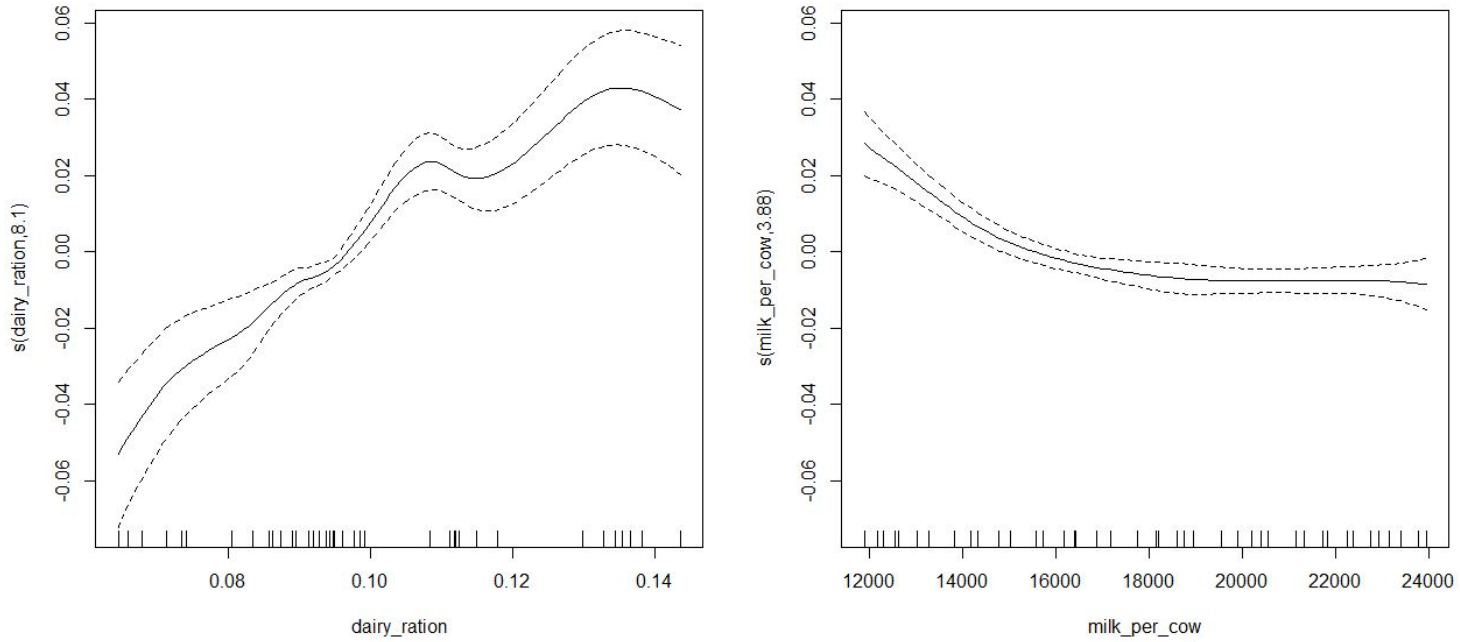
Import volume as price regressor, index of competition with domestic production
Export volume as indicator of drive for production demand

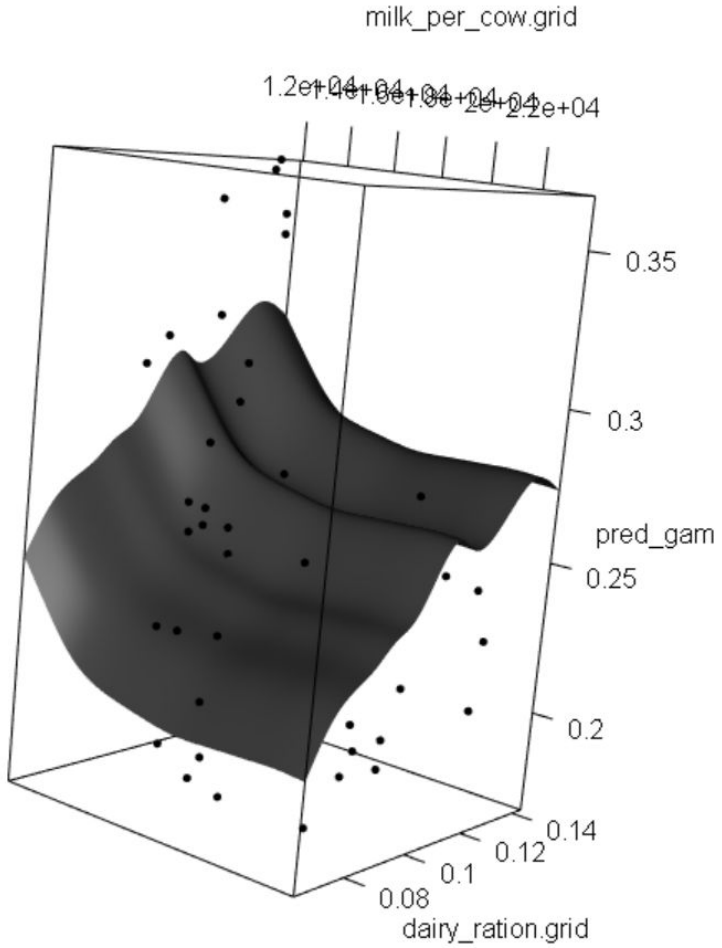
[1]CPI: Consumer Price Index

GAM

$$\begin{aligned}\texttt{avg\_price\_milk} = \beta_0 &+ \beta_1 \texttt{milk\_cow\_cost\_per\_animal}_i \\ &+ \beta_2 \texttt{milk\_volume\_to\_buy\_cow\_in\_lbs}_i \\ &+ \beta_3 \texttt{milk\_feed\_price\_ratio}_i \\ &+ f_1(\texttt{dairy\_ration}_i) + f_2(\texttt{milk\_per\_cow}_i) + \epsilon_i\end{aligned}$$

We performed multiple **Permutation Tests** to assess the significance of each covariate.

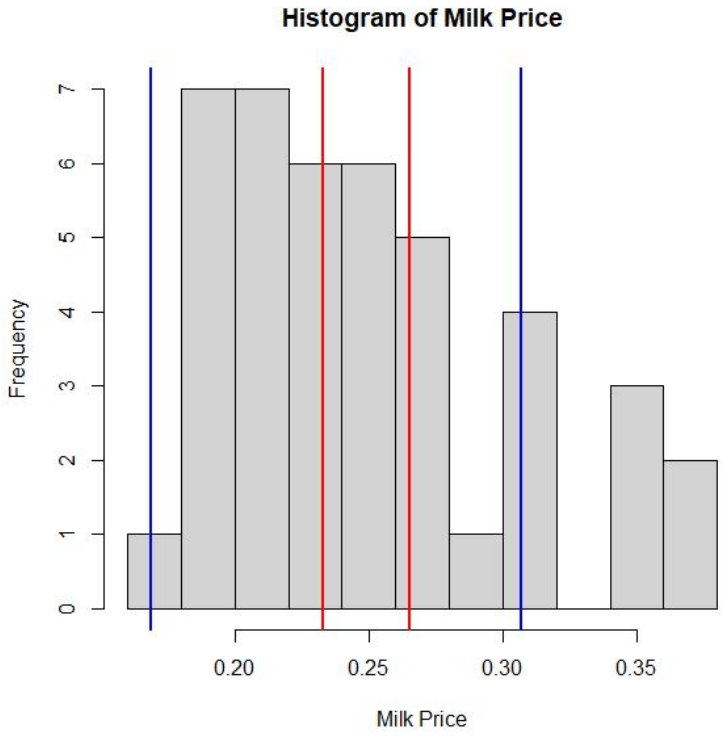




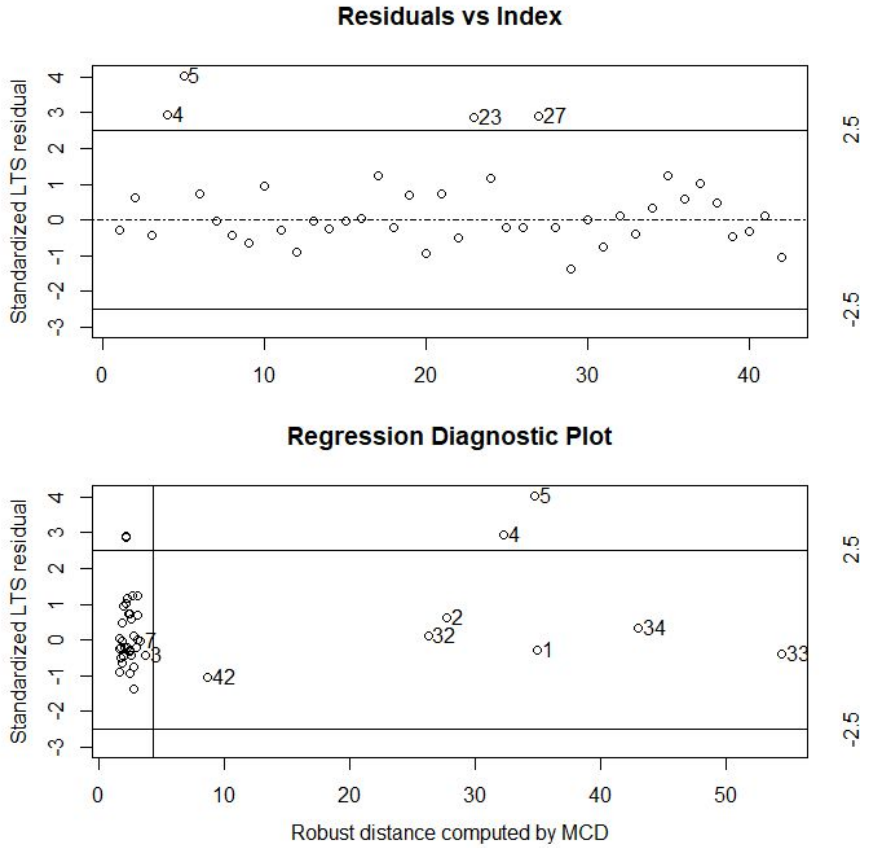
GAM

Conformal Prediction

The two conformal interval are respectively:
Red: Based on the aforementioned **GAM**.
Blue: Created using the **KNN** distance.



Robust Regression



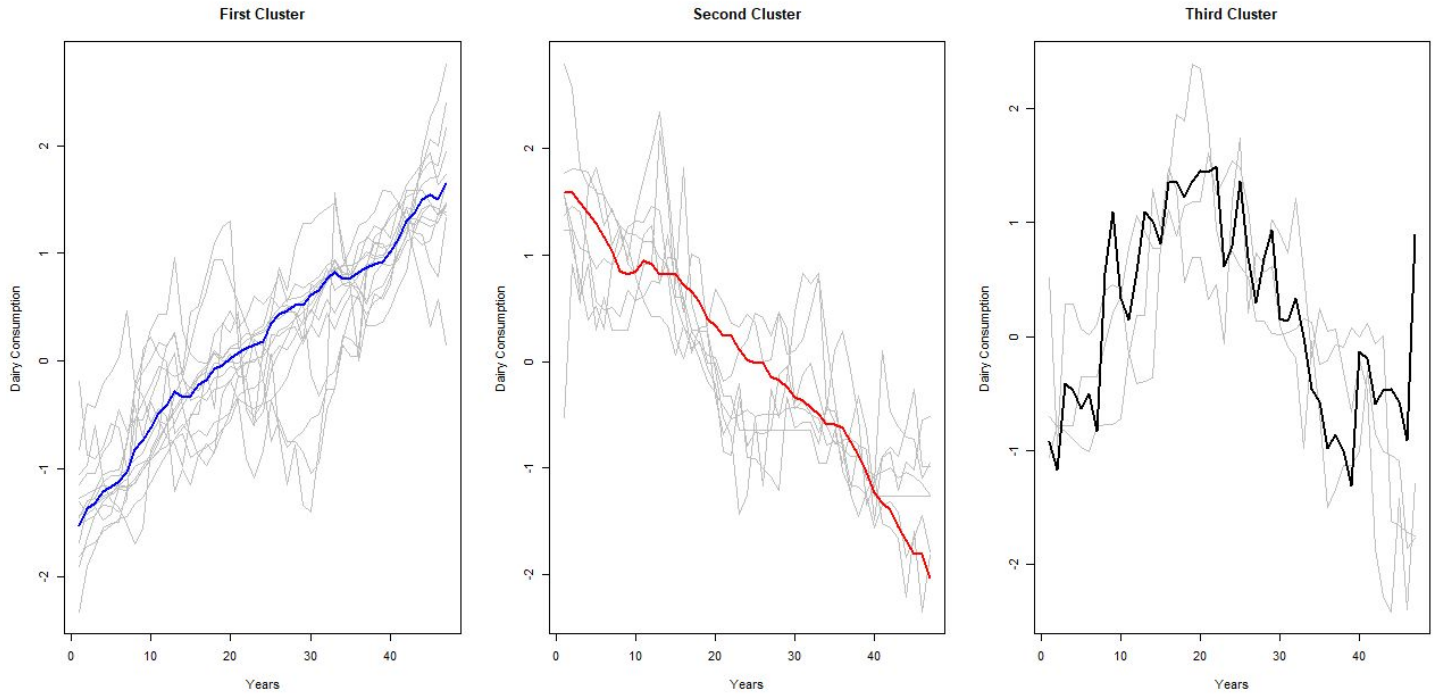
We can note four outliers, corresponding to the years **1983, 1984, 2002,** and **2006**.

Bayesian Clustering

$$\mathbf{y}_i = \mathbf{Z}\boldsymbol{\alpha}_i + \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, n$$

$$\theta_{it} = \rho\theta_{i,t-1} + \nu_{it}$$

$$\boldsymbol{\gamma}_i = (\boldsymbol{\beta}_i, \boldsymbol{\theta}_i) \qquad \boldsymbol{\gamma}_i \mid G \overset{\text{iid}}{\sim} G, \text{ for } i = 1, \ldots, n \text{ with } G \sim \mathcal{PD}(a, b, G_0)$$



Through an **outlier detection**, we verified that there were no outliers in the three clusters.

Future Developments

- Perform a **Stakeholder Analysis** to categorize our stakeholders by using the Power-Interest matrix
- Since the single state's information is available, we may consider **adding a spatial component** by grouping the states on East Coast, Central States, and West Coast and understand their differences
- Consider possible ways of embedding the **import/export** data in our current model
- Explore more **distances** for the conformal prediction
- Perform **prediction** using our models using fictitious (or real? 🧐) cheese factories
- We would like to explain why the years 1983, 1984, 2002, and 2006 have been detected as **outliers** in the Robust Regression and find some actual events related to this anomalies

Thank you for the attention!

References

- rfordatascience/tidytuesday. (2019, January 28). GitHub. Retrieved December 12, 2022, from https://github.com/rfordatascience/tidytuesday
- USDA ERS - Dairy Data. (n.d.). USDA ERS - Dairy Data. Retrieved December 12, 2022, from https://www.ers.usda.gov/data-products/dairy-data/
- Bergmann, D., O'Connor, D., & Thümmel, A. (2015). **Seasonal and cyclical behaviour of farm gate milk prices**. *British Food Journal*.
- Nicholson, C. F., & Stephenson, M. W. (2015). **Milk price cycles in the US dairy supply chain and their management implications**. *Agribusiness*, *31*(4), 507-520.
- Nieto-Barajas, L. E., & Contreras-Cristán, A. (2014). **A Bayesian nonparametric approach for time series clustering**. *Bayesian Analysis*, *9*(1), 147-170.