Impact of Canadian Crop Prices on Crop Production

Foundations of Data Science - Group Project

Muskaan Brar[1], Jeush Stanly[2], Monika Tomar[3], Jason Wilcox[4], Elmira Sepehri[5], and David Fishman[6]

Email ids: [1]mbrar44@uwo.ca, [2]jeush.stanly@mail.utoronto.ca, [3]25monikatomar25@gmail.com, [4]jason.wqost@slmails, [5]elmira.sphr@gmail, [6]davjfish@gmail.com

December 2023

1 Summary

1.1 Data source

1.1.1 What is the nature of the data you chose?

The dataset we first selected was sourced from the Government of Canada's Open Government Portal. It records the monthly price of various crop, commodity and livestock products produced in Canada. The raw data contains the price of each farm product, the month the price was captured, a product description, and the province of production. Our desire to perform analysis regarding the relationship between price and production led us to seek additional datasets containing production data for selected commodities. This production data was also sourced from the Government of Canada's Open Government Portal. It recorded the monthly production of farm products by province of production over a period of years.

1.1.2 Why is it interesting?

The datasets we eventually selected were interesting as between them they contain a time series of prices for different products and from different provinces, and the production for those products over of a period of over time, from January 1980 to the present. The duration of the dataset provided an opportunity to find patterns over time and the possibility of a relationship between price and the amount of production within the regions of Canada. Analysis of these trends could provide us with insight to current day problems such as inflation and the cost of living crisis.

1.2 Analysis

1.2.1 What were the challenges?

Within our chosen datasets of Canadian farm product prices and production spanning from January 1980 to present, our analysis required us to overcome a series of challenges.

The format of certain features within the data, such as the REF_DATE were imported as a text datatype and as such were not immediately usable, leading us to perform some feature engineering. The structure of the pricing dataset allowed for differing units of measure to be utilized within the data for a singled commodity. Also, the Wheat Production dataset could contain multiple entries for any given year which made merging with the pricing data problematic. The differing ranges of data of each of the farm products in the

dataset posed a challenge in terms of selecting the best data for analysis The publication of Price and Production datasets in distinct files further added layers of complexity. Our analysis required these be merged. The presence of null values in the merged pricing and production datasets added also required additional action prior to our final analysis. Finally, the limited overlap in time series, resulted in a comparatively small sample size, and led to the selection of suitable techniques suitable for smaller datasets.

1.2.2 How did you overcome them?

To overcome the challenges presented in previous section, we took a number of steps to prepare the data and to create and validate a model for analysis. The date feature REF_DATE was imported as a text datatype and feature engineering was leveraged to change its datatype to a date. During the validation of the model, we noted that the Wheat Production data reported production in March, July and December, as opposed once a year like the others. As the March and July values were the same as the previously reported December value, we were able to make the assumption that the December values where the correct ones. Accordingly, we selected the December values and re-indexed the data to match the Price dataset prior to merging. The source of the data noted that the unit of measure many not be consistent. As part of our data cleaning we validated that the unit of measure for our the selected products was indeed consistent. When reviewing a line plot of the time series data for all provinces. It became evident that all the province had a similar patten of price over the time series, with the exception of Newfoundland and Labrador. Therefore, we elected to exclude Newfoundland and Labrador from our subsequent analysis. To analyze to impact of product production on product price it was needed to add production data to our model. Using the Government of Canada's Open Government Portal we searched for and found production data for our selected product products. After similar feature engineering to correct the REF_DATE feature, we merged the production data using the product description as a key. The availability of price and production data led us to select Meat Chickens, Wheat and Hogs as products for final analysis on any relationship between price and production. The availability of price and production data also required us to drop null values resulting in few data points for comparison. To evaluate the relationship between price and production we applied the technique of linear regression on the resulting data model.

1.3 Conclusions

1.3.1 What did you learn about your dataset?

After reviewing the liner regression results for Meat Chickens, Wheat and Hogs we found the following:

For Meat Chickens, the model intercept was found to be 26.4810 and the coefficient price variable is 50.9459.

$$y = 26.4810 + 50.9459x_1$$

- Where y is the number of chickens, measured in millions of individuals

- and x_1 is the price of chicken meat, in dollars per kilogram

The F-statistic for this model was observed to be 12.47; under a normal distribution the chances of observing this value are approximately 0.01%. Setting our p-value to 0.05, we would be forced to reject the null hypothesis and conclude there is indeed a relationship between the price and production of chicken meat in Canada. The observed value for R^2 was 0.675 which means that approximately 67% of the variance of the data can be accounted for by this model.

Similarly, we found there to be a significant, positive relationship between the prices and production levels of wheat. The model for this relationship can be described as follows:

$$y = 1.285e4 + 41.3883x_1$$

- Where y is the production of wheat, in thousands of metric tonnes

- and x_1 is the price of wheat, in dollars per metric tonne

The F-statistic for this model was observed to be 8.065; under a normal distribution the chances of observing this value are approximately 0.01%. With a p-value set to 0.05, we would be forced to reject the null hypothesis and conclude there is indeed a relationship between the price and production of wheat in Canada, albeit not as strong of a relationship as with chicken meat. The observed value for R^2 was 0.287 which means that approximately 29% of the variance of the data can be accounted for by this model. While this finding is significant, it is clear there are more factors involved in understanding the variations of production of wheat in Canada than just price alone.

Finally, we did not find there to be a significant relationship between the production of hogs and prices in Canada.

The major takeaway from this analysis is that the production of different crops will respond to different factors, differently. In the case of crops such as chicken meat, price appears to be a stronger driver behind crop production; when prices of meat is high, producers tend to raise more animals. The general trend behind wheat production was similar although a more significant portion of the variation in production was unaccounted for by the current price of the commodity. On the other hand, current hog prices clearly do not have an influence on animal production on hog farms; This result was unexpected, especially considering the trend detected in the production of chickens. Swine, unlike chicken meat is not under national supply management and thus more subject to fluctuation of prices due to international markets. This fact might help account for the different responses observed in the two types of meat production. Specifically, producers, in anticipation of highly variable prices, might decide to hedge their bets by investing in their farms regardless of the current state of the market.