Panel data analysis of a Chocolate Factory

Luis Penteado 20230441 | Pedro Adonis 20230994 | Peter Falterbaum 20230956 | Stepan Kuznetsov 20231002

Introduction

In the last few years the global business of chocolate has witnessed dynamic shifts following economic trends, industry developments and changing on customer behaviour. According to the International Cocoa Organization (ICCO), which has been tracking the production trends, the cocoa-based products increased their production 3% compared 2019 to 2021. This raise was influenced by several factors including demand in emerging markets, technological advancements in the chocolate manufacturing process and sustainable practices adopted by key players in the industry.

In addition to economic trends there are also industrial factors that can cause instability, such as production costs. As stated by AGĂNENCEI in the period 2018 to 2021 there was an increase on energy, raw materials and ingredients, especially cocoa, that affected the whole chocolate market.

Within this context, we created a panel dataset to help a fictitious chocolate company to improve their production by finding improvement opportunities in the production line. The dataset contains information regarding the reference year and the production. Additionally, information on employees, machines and operations and energy consumption is considered.

Hypothesis

What are the factors that most influence chocolate production?

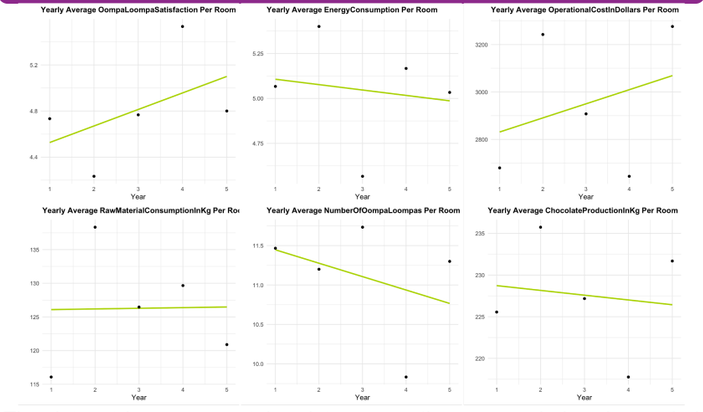
Methodology

To answer our research question, we created a panel dataset about a fictitious chocolate factory inspired by the Fantastic Chocolate Factory movie. Our dependent variable was Chocolate Production In Kilograms (Kg) and the independent variables, which can influence chocolate production, were the following: Number of Oompa Loompas* (number of employees), the quantity of materials in Kg (such as cocoa, sugar, milk…), the operational cost in dollars, the energy consumption rate of the industry, and the Oompa Loompa satisfaction (which is the employees satisfaction). The dataset contains specific data to each rooms of the company, covering a period of five years.

With that in mind, we created three different models for our dataset: a Fixed Effect model, a Random Effect model and a Pooled Ordinary Least Squares (OLS) model. To assess the presence of heteroscedasticity, the Breusch-Pagan and the Special White test were applied to all models. The Hausman test was used to determine if the differences in parameter estimates between fixed effects and random effects models were statistically significant. Both tests contributed to the selection of the appropriate model which can answer the research question.

* Oompa Loompas were the employees that worked in the chocolate factory from the movie the Fantastic Chocolate Factory Film.

Data Analysis



The dataset in scope contains data over a five-year period and revealed several trends which might impact chocolate production. Employee satisfaction among the Oompa Loompas showed a positive trend, suggesting that improvements in workplace conditions may have contributed to a happier workforce. On the opposite, the total number of Oompa Loompas declined and also slightly the total chocolate production.

Operational costs increased over the years on average, which could be due to inflation, increased labor costs, or investment in more advanced production technology. This rise in costs doesn't appear to adversely impact chocolate production, which remained relatively constant despite an observed increase in expenses. Correspondingly, energy consumption declined marginally, mirroring the trend observed in chocolate production.

The upward trends in employee satisfaction and operational costs will require ongoing attention to ensure the sustainable growth of the chocolate factory. Additionally, it's relevant to find out which factors impact the production of the chocolate factory to ensure a stable chocolate production in the future.

Results

Pooled OLS, fixed effects, and random effects models each offer distinct approaches to analyzing panel data, with the choice between them based on the presence of individual-specific variations and their correlation with explanatory variables. To test whether the unique errors in the RE model are correlated with the regressors, the hausman test was conducted. As both, breusch pagan and the special white test, didn't provide evidence that any of the models have heteroskedasticity, we employed the normal hausman test and not the robust hausman test.

Test	P Value	H0	Conclusion
Hausman test	0.05328	Random effect model is more efficient	Reject the H0
Breusch-Pagan (OLS)	0.07269	Homosk.	No sign of significant heteroskedasticity
Breusch-Pagan (FE)	0.07269	Homosk.	
Breusch-Pagan (RE)	0.07269	Homosk.	
Special White-test (OLS)	0.1975	Homosk.	No sign of significant heteroskedasticity
Special White-test (FE)	0.1408	Homosk.	
Special White-test (RE)	0.1975	Homosk.	

The special white test, conducted along with the Breusch-Pagan test, assesses both linear and non-linear heteroskedasticity, providing a comprehensive analysis of potential heteroskedastic patterns in the regression model. The obtained p-value in the Hausman test is slightly above the 0.05 threshold, using a 95% confidence interval. This leads to the conclusion to not reject the null hypothesis. As the p value is marginal to the threshold, its interpretation has been discussed. The results show that both, RE and FE are valid estimators, but the RE is more efficient. Given the p-value we suggest to use the RE model. Considering the p value is close to the threshold, further analysis might be beneficial to confirm our findings.

Model	Number of OompaLoompas	Raw Materials Consuption	Operational Costs	Energy Consumption Rate	Oompa Loompa Satisfaction	R-squared
Pooled OLS	<2e-16	<2e-16	1.980e-13	<2e-16	5.183e-11	93.512
Fixed Effects	<2e-16	<2e-16	3.067e-10	<2e-16	2.059e-09	94.04
Random Effects	<2e-16	<2e-16	4.619e-16	<2e-16	1.222e-12	93.512

Our analysis showed that all regressors are significant factors in the production process, given a confidence interval of 95%. The high R-squared values indicate that over 93% of the variance in chocolate production is explained by the models.

Conclusions

In this study on the chocolate industry, we employed panel data models to analyze a fictitious dataset. Several tests such as Breusch-Pagan, White, and Hausman, highlighted the superiority of the Random Effects model and that it does not have heteroskedasticity. This model explained over 93% of the variance in chocolate production, emphasizing the importance of the number of employees (Oompa Loompas) and their satisfaction in the production process. Contrary to initial data analysis, employee satisfaction shows a positive correlation with production, suggesting that improvements in working conditions could enhance productivity. We also observed that despite an increase in operational costs over the years, this factor did not negatively impact chocolate production. To conclude, given the small difference to the threshold of the Hausman test, we suggest further analyses to validate the robustness of our conclusions, for example data set expansion.

References

1. International Cocoa Organization (ICCO). https://www.icco.org/
2. International Cocoa Organization. (2021). Quarterly Bulletin of Cocoa Statistics, November 2021.
3. Mordor Intelligence. (n.d.). Chocolate Market Trends & Segmentation - Industry Analysis. Retrieved from Mordor Intelligence Chocolate Market Report.
4. AGĂNENCEI, MARIA ADRIANA. STUDY ON CHOCOLATE COSTS IN THE PERIOD 2018-2021. Faculty of Management and Rural Tourism, Romania