26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Forecasting Australian pulse export trade flows and uncovering key trade determinants using machine learning

Deenuka Niroshini Perera[a], Sobhan (Sean) Arisian[a], Yuquan Du[a], Shalinka Jayatilleke[a], Su Nguyen[b]

[a]*La Trobe Business School, La Trobe University, Bundoora, Melbourne, Victoria, 3086, Australia.*
[b]*Department of Accounting, Info Sys & Supply Chain, RMIT University, City Campus, Australia.*
Emails: D.Perera@latrobe.edu.au, S.Asian@latrobe.edu.au, Bill.Du@latrobe.edu.au, S.Jayatilleke@latrobe.edu.au and su.nguyen@rmit.edu.au

Abstract: Australia is a major lentil exporter, yet policy volatility and external shocks complicate trade. Using a dual framework of machine learning (ML) and a PPML gravity model, this study forecasts exports at global, national, and bilateral levels with 81 trade-related variables. ML models explained ~65% of global and ~60% of Australia's export variance, with accuracy strongest in stable years. Lagged exports and rolling averages dominate feature importance, highlighting path dependence, while production and macroeconomic indicators provide additional signals. Results inform strategies to enhance resilience in pulse trade.

Keywords: *Machine Learning, Gravity Model, Trade Forecasting, Lentil Exports, Trade Determinants, Australia*

1. INTRODUCTION

Plant-based proteins, such as pulses, are recognised for their affordability and high nutritional value compared to red meat, making them a crucial dietary component. With its favourable climate and advanced agricultural techniques, Australia has long been a leading producer and exporter of pulses, with India being one of the primary markets (ABARES, 2024). Australia has a strategic opportunity to strengthen its position as a leading pulse exporter. However, the pulse trade in Australia faces significant challenges from the producers' and exporters' perspectives, primarily due to frequent and unpredictable fluctuations in agricultural trade policies, particularly tariff rates and quotas (DAFF, 2023). In addition, factors such as crude oil prices, pulse production levels in destination markets, limitations in logistics infrastructure, and evolving geopolitical dynamics also present significant obstacles to the stability and growth of Australia's pulse export sector (Kingwell et al., 2020; White et al., 2018). Hence, forecasting pulse trade flows and identifying the most influential key trade determinants is crucial for maintaining Australia's competitive edge in the agricultural sector and enhancing resilience, as it can inform evidence-based policymaking, guide export strategies, support strategic investments in trade facilitation, and help producers and exporters to make informed decisions. Although the broader research context focuses on the global pulse trade, lentils were selected as the case study for this analysis due to their substantial contribution to Australia's export volumes and their representative supply chain characteristics (DAFF, 2023; FAOSTAT, 2024). While Australia is recognised as one of the world's leading lentil exporters, no known studies have applied advanced predictive modelling techniques to forecast lentil export flows or to systematically identify the key determinants influencing these exports. Addressing this gap is critical to inform targeted policy interventions, optimise export strategies, and enhance resilience in the face of global trade volatility.

2. METHODOLOGY AND RESULTS

To fill the above research gap, the study employs a dual method modelling framework, with supervised machine learning (ML) serving as the primary predictive approach. The ML component utilises a diverse range of algorithms, tree-based models (e.g., Random Forests and Extra Trees), gradient boosting frameworks (e.g., LightGBM, XGBoost and CatBoost), and neural network architectures. For this purpose, AutoGluon, an AutoML toolkit for tabular data (Erickson et al., 2020), is employed to train and evaluate multiple ML algorithms on a time series regression task aimed at forecasting lentil export quantities. Complementing this, a classical econometric trade modelling framework, which is the Gravity model (PPML approach), is formulated to provide a theoretically grounded benchmark against the predictive performance and explanatory capacity of the ML models. In addition, its well-established theoretical foundation provides the underlying structure necessary for applying ML methods in predictive contexts (Athey & Imbens, 2019; Silva & Tenreyro, 2006). Moreover, this integrated framework is operationalised at three analytical levels: (i) Global level - forecasting total lentil export quantities worldwide; (ii) Country level - forecasting lentil exports of individual countries, with a specific focus on Australia; and (iii) Country-pair level - forecasting bilateral lentil export flows, particularly between Australia and its principal trading partners. This study incorporates a comprehensive set of trade-related variables, such as lentil export and production quantities, rainfall data, lentil tariff data, crude oil prices, exchange rates, uncertainty indices, derived features, export lags and gravity variables (81 in total, including more than 40 gravity variables, for example, GDP, geographic distance, population, and common language). These variables were sourced from reputable and widely used databases, such as FAOSTAT, the CEPII Gravity Database, ERA5-Land, and the World Bank.

The ML forecasting results explained approximately 65% of the variance in global lentil export quantities over the evaluation period ($R^2 = 0.6497$), with an overall RMSE of 10,098, indicating moderate prediction accuracy. Yearly results highlight fluctuations in performance, with the model achieving strong accuracy in 2021 and 2022 ($R^2 > 0.79$, RMSE < 6,700) but weaker results in 2020 and 2023, likely reflecting market volatility and external disruptions. The ML forecasting model for Australia's lentil exports explained about 60% of the variance across all years ($R^2 = 0.5952$) with an overall RMSE of 43,201 tons, indicating sizable prediction errors on average. Its performance was strongest in 2020 and 2022 ($R^2 > 0.92$, RMSE < 16,000) but weakened in 2019 and 2023, highlighting sensitivity to market shocks and external disruptions.

3. CONCLUSIONS

The feature importance analysis reveals that lagged export volumes and rolling averages are the strongest predictors of Australia's lentil trade, underscoring its path-dependent nature and the strategic value of maintaining consistent trade relationships. In addition, production quantities in both origin and destination countries, along with macroeconomic indicators, competitor supply dynamics, and geopolitical factors, collectively shape export performance, highlighting the need for exporters to integrate crop monitoring, economic analysis, and diplomatic risk management into their trade strategies.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all my supervisors for their valuable feedback, continuous support, and insightful guidance throughout this research.

REFERENCES

ABARES. (2024). *National overview*. Australian Bureau of Agricultural and Resource Economics and Sciences. https://www.agriculture.gov.au/abares/research-topics/agricultural-outlook/australian-crop-report/overview#above-average-national-winter-crop-productions

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685-725.

DAFF, A. G. (2023). *Insight – Australia is a key global exporter of pulses*. Department of Agriculture, Fisheries and Forestry: Australia. Retrieved 9 from https://www.agriculture.gov.au/about/news/insight-australia-key-global-exporter-pulses#daff-page-main

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., & Smola, A. (2020). Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*.

FAOSTAT. (2024). *FAOSTAT*. Food and Agriculture Organization of the United Nations. https://www.fao.org/faostat/en/#data

Kingwell, R., Loxton, R., & Mardaneh, E. (2020). Factors and scenarios affecting a farmer's grain harvest logistics. *The Australian journal of agricultural and resource economics*, *64*(2), 244-265. https://doi.org/10.1111/1467-8489.12355

Silva, J. M. C. S., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4), 641-658. https://doi.org/10.1162/rest.88.4.641

White, P., Carter, C., & Kingwell, R. (2018). *Australia's grain supply chains: costs, risks and opportunities*. https://aegic.org.au/wp-content/uploads/2021/03/FULL-REPORT-Australias-grain-supply-chains-DIGITAL__.pdf