Comparison of Arima and Holt-Winters forecasting models for time series of cereal production in Peru

Humberto Sernaqué[1], *Moly Meca*[1], *Eduardo Zapata*[1], *Berenise Marchan*[1], *Junior Medina*[1], *Denis Nole*[1], *Cristhian Aldana*[1], *Yesenia Saavedra*[1], *Luis Trelles*[1], *Nelson Chuquihuanca*[1], *Gustavo Mendoza* [1]

[1] Universidad Nacional de Frontera - UNF
Av. San Hilarión 101, Sullana, Piura, PERÚ

ABSTRACT

Agricultural commodities present remarkable volatility in their production levels, which severely affects farmers. The variational dynamics in the prices of the inputs used and the constant variations in weather conditions have a significant influence on the cereal production chain in Peru; therefore, compared to the ARIMA model, the Additive Holt-Winters forecasting model presented a better fit according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), forecasting the production of *Oryza sativa*, *Zea mays L. var. Indurata and Amaranthus caudatus*; however, due to the high seasonality, volatility of production, and the greater amount of outliers due to production in certain periods and geographical areas, the Holt-Winters Multiplicative model predicted the national production of *Zea mays L. ssp amiláceo and Chenopodium quinoa*, in Peru in the period 2000-2021.

Keywords: ARIMA, Holt-Winters, AIC, BIC, forecasting, cereals, production

INTRODUCTION

Cereal production is one of the most developed economic activities worldwide and its crops are among the most extensive, all by their ability to adapt to different environmental conditions, ease of conservation, and their fickleness (Zambrano et al. 2019). In the 2018/2019 period, world cereal production fell to 2,595 million tons, representing a decline of 62.5 million tons compared to the previous period (INEI, 2021). In Spain, cereal production represents an important economic activity, with an estimated €3 602.6 million concerning the value of its production, this figure represents 7.1% and 12.3% of the production of the agricultural branch and the production of the vegetable branch, respectively. Latin America and the Caribbean experienced 2017, a growth in their cereal production levels, exceeding by 20% the production level obtained during the 2016 period, this is thanks to the record figures recorded by Argentina and Brazil in terms of their corn production volumes (Spain Government, 2021).

The present research was carried out to adjust an econometric ARIMA and Holt-Winters forecasting model to estimate a forecast for 2021 of the production volumes of cereals in Peru: *Oryza sativa, Chenopodium quinoa, Amaranthus caudatus, Zea mays L. var. Indurata and Zea mays L. ssp amiláceo*; taking as a historical time series the production data in Peru, period 2000-2021 (Correa, 2020). Among the most important cereals in Peru, we can find rice (*Oryza sativa*), the most produced agricultural product at the national level; likewise, there are also important Andean cereals such as Quinoa (*Chenopodium quinoa*) and Kiwicha (*Amaranthus caudatus*), which are highly valued in international markets for their high nutritional value, such as the positioning at international level that Peru has managed to consolidate itself as the main exporter of quinoa, exporting an average annual value of approximately 125 million dollars (INEI, 2021). Concerning hard yellow corn (*Zea mays L. var. Indurata*) and starchy corn (*Zea mays L. ssp starchy*), it should be noted that their production ranks first worldwide, followed by wheat and rice. For the 2019-2020 crop year, planting intentions for hard yellow corn amounted to 273.0 thousand hectares, an increase of 3.7% for the 2018-2019 crop year; it should be noted that its production is mainly used as an input for the feed industry for poultry, swine and fattening animals, among others (INEI, 2021).

Finally, it is feasible to carry out the present research since it is intended to adjust econometric prediction models (ARIMA and HOLT-WINTERS), to make a forecast of the production of these agricultural goods, generating a futuristic scene that in turn would allow having tools that provide information to farmers and contribute to the implementation of plans and/or strategies to increase their benefits (Aurna et al. 2021).

MATERIALS AND METHODS

For the development of this research and the adjustment of the econometric models, the Box-Jenkins and Triple Exponential Smoothing methodology were used; this implied the use of the formulas corresponding to the ARIMA and Holt-Winters models (Additive and Multiplicative), to obtain a forecast following the five historical datasets obtained for its analysis (Hamzah et al. 2020).

For example, the additive Holt-Winters (AHW) model used, as shown, worked with the equations (1), (2), (3) y (4):

$$L_t = a\left(Y_t - S_{t-p}\right) + (1-\alpha)[L_{t-1} + T_{t-1}] \quad (1)$$

$$T_t = y[L_t - L_{t-1}] + (1-y)T_{t-1} \quad (2)$$

$$S_t = \delta(Y_t - L_t) + (1-\delta)\, S_{t-p} \quad (3)$$

$$\hat{Y}_t = L_{t-1} + T_{t-1} + S_{t-p} \quad (4)$$

Where (L_t) represents the level over time (t_1); (α) weighting for the level; (T_t) trend over time (t_1); (y) weighting for the trend; (S_t) seasonal component at time (t); (δ) weighting for the seasonal component; (p) seasonal period; (Y_t) value of the data over time; $(\hat{Y}_t)$ fitted value, or the forecast one period ahead, at time (t) (Putri et al. 2021). The methodology of these models used historical data, thus generating confidence in the fit of the model (Wang et al. 2019). It should be noted that, for the validation stage of the models that best fit the series analyzed, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used (Taylor, 2008), which are closely related and proved to be a fundamental criterion when selecting this type of model (Purwanto et al. 2019).

The target population consisted of information from the digitized database extracted from the National Institute of Statistics and Informatics - INEI (https://systems.inei.gob.pe/SIRTOD/app/consulta). From the registered population, information was obtained on the monthly tons produced throughout the national territory, the bulk of national production, and the distribution of the production of these cereals throughout Peru. In addition, different computer codes were implemented in the Python software for the development of the Holt-Winters and ARIMA models, as well as for the processing of the time series in terms of their identification, elimination, validation, and forecasting, some of which are detailed below:

Program code for Holt-Winters model

```
results["Additive"] = [fit1.params[p] for p in params] + [fit1.sse]
results["Multiplicative"] = [fit2.params[p] for p in params] + [fit2.sse]
ax = dataarroz["PROD"].plot(figsize=(12, 8),
   color="black",
ax.set_ylabel("Toneladas de arroz)")
ax.set_xlabel("Year")
fit1.fittedvalues.plot(ax=ax,style="--", color="red")
fit2.fittedvalues.plot(ax=ax,style="--", color="green")
plt.show()
print("Figure: Forecast production of rice")
print(fit1.aic, fit1.bic)
print(fit2.aic, fit2.bic)
print(fit1.forecast(8))
print(fit2.forecast(8))
results
```

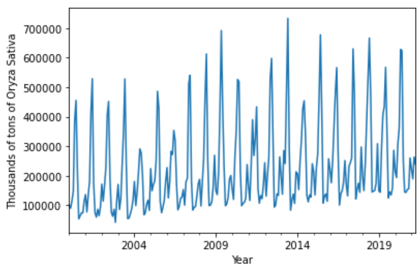
Program code for ARIMA model

```
arma_mod20=ARIMA(datakiwicha,order=(1,1,1),
seasonal_order=(1,0,1,12)).fit()
print(arma_mod20.params)
print(arma_mod20.aic,arma_mod20.bic, arma_mod20.hqic)
arma_mod30=ARIMA(datakiwicha,order=(0,1,1),
seasonal_order=(1,1,1,12)).fit()
print(arma_mod30.params)
print(arma_mod30.aic,arma_mod30.bic
```

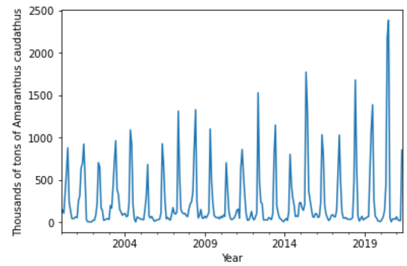
Finally, the ARIMA and Holt Winters models were compared for modeling the forecast of the national production of the following cereals in Peru, 2021: *Oryza sativa, Chenopodium quinoa, Amaranthus caudatus, Zea mays L. var. Indurata and Zea mays L. ssp amiláceo*.

RESULTS

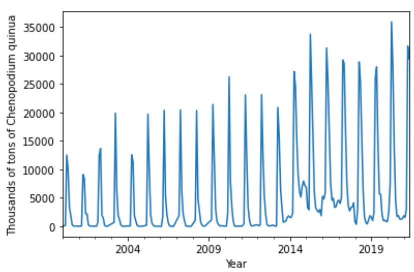
The results show the production behavior of: *Oryza sativa, Chenopodium quinoa, Amaranthus caudatus, Zea mays L. var. Indurata and Zea mays L. ssp amylaceous*; in which it is observed that said behavior of the series studied in Peru, period 2000-2021, is notably seasonal (see Figure 1). After applying the econometric models to obtain future estimates of the production levels that could be obtained for the period 2021, it can be determined that, of the models applied, the Holt-Winters Model presents more favorable measures of goodness of fit concerning the ARIMA Model, where the use of the Holt-Winters seasonal model can more accurately detail the hidden patterns among the data (Hussain et al. 2016). For the validation and acceptance of the best model adjusted to the time series, the values of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were used as a reference, taking as the best model that presented the lowest values concerning these indicators, since, according to the theory, this ensures a better capture of the behavior of the series studied (see Table 1).



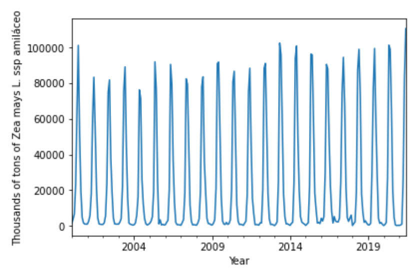
(a) *Oryza sativa*



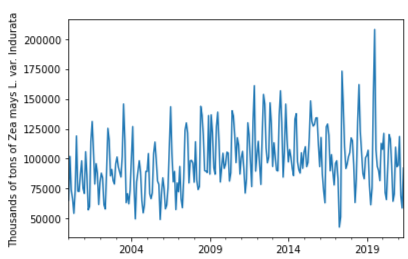
(b) *Amaranthus caudatus*



(c) *Chenopodium quinua*



(d) *Zea mays L. ssp amiláceo*



(e) *Zea mays L. var. Indurata*

Figure 1. Production of main cereals in Peru, 2000-2020 period

Table 1. Validation of ARIMA and Holt-Winters models, *Oryza sativa*, *Zea mays L. var. Indurata, Zea mays L. ssp amylaceous, Amaranthus caudatus and Chenopodium quinua* production.

Production		**ARIMA**	**Holt-Winters**	
			Additive	**Multiplicative**
Oryza sativa	**BIC**	6559.58	5751.18	5754.20
	AIC	6573.75	5694.46	5697.48
Zea mays L. var. Indurata	**BIC**	5490.97	5057.24	5060.43
	AIC	5473.46	5000.46	5003.64
Zea mays L. ssp amylaceous	**BIC**	5472.26	4952.63	4944.28
	AIC	5454.51	4895.79	4887.43
Amaranthus caudatus	**BIC**	3346.38	3031.74	3182.14
	AIC	3332.39	2974.95	3125.35
Chenopodium quinua	**BIC**	4375.99	4579.35	4346.56
	AIC	4355.61	4522.56	4289.78

For the case of *Chenopodium quinoa and Zea mays L. ssp amylaceous*, the best fitting model, according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), is the multiplicative Holt-Winters, since the production of these cereals in Peru is highly seasonal, obtaining production in certain periods and in a certain area of the national territory; In other words, there is no constant production, which leads to many more outliers, generating a highly volatile behavior that can be better captured by the aforementioned model. The validation of the ARIMA models was used for each series as follows: *Oryza sativa* (1,0,1)(1,1,1,12), *Zea mays L. var. Indurata* (1,0,1)(1,1,1,12), *Zea mays L. ssp amylaceous* (1,1,1)(1,0,1,12), *Amaranthus caudatus* (0,1,1)(1,1,1,12) *and Chenopodium quinua* (0,0,0)(2,3,3,12).

Finally, after the validation of the different econometric models with the proposed methodology, we proceeded to make the forecasts with the model that presented the best fit, obtaining the following forecasts for the remainder of the 2021 period; which reflected a cyclical trend that characterizes each of the series studied (see Table 2).

Table 2. Forecasting with the Holt-Winters Additive model, of the production of *Oryza sativa, Zea mays L. var. Indurata and Amaranthus caudatus*, Peru, 2021.

FORECAST (Holt-Winters Additive)								
SERIE	**may-21**	**jun-21**	**jul-21**	**aug-21**	**sep-21**	**oct-21**	**nov-21**	**dic-21**
Oryza sativa	526894.59	675310.68	381730.11	134811.35	145218.82	157897.01	162553.34	284924.52
Zea mays l. Var. Indurata	-	120542.50	132858.67	106440.17	82855.50	90651.79	92710.69	90439.31
Zea mays l.ssp amilaceo	-	-	36921.15	12521.38	3263.84	718.68	713.64	637.97
Amaranthus caudatus	-	395.23	140.72	43.60	14.53	18.02	15.76	16.82
Chenopodium quinoa	-	12098.82	5636.02	1678.94	659.81	469.13	651.98	866.34

CONCLUSIONS

When comparing the ARIMA and Holt-Winters models for modeling the national production of the following cereals in Peru: *Oryza sativa, Chenopodium quinoa, Amaranthus caudatus, Zea mays L. var. Indurata and Zea mays L. ssp amiláceo*; a cyclical behavior was evidenced, which describes a periodic pattern, determining that the investigated series present seasonality. The Additive Holt-Winters model presented the best fit according to the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), forecasting the production of *Oryza sativa, Zea mays L. var. Indurata and Amaranthus caudatus*; however, due to the high seasonality, volatility of production, and the higher amount of outliers due to production in certain periods and geographical areas, the Multiplicative Holt Winters model better predicted the national production of *Zea mays L. ssp amiláceo and Chenopodium quinoa*, in Peru in the period 2021, from a time dataset of 2000-2021. Finally, the productions: *Oryza sativa, Amaranthus caudatus, and Zea mays L. var. Indurata* yielded the values: AIC: 5694.46 and BIC: 5751.18; AIC: 2974.95 and BIC: 3031.7 and AIC: 5000.46 and BIC:5057.24 respectively for the above-mentioned series; while for the production of: *Chenopodium quinoa and Zea mays L. ssp amylaceous* the model took was the Multiplicative Holt-Winters, which gave the following values: AIC: 4289.78 and BIC: 4346.56 and AIC: 4887.43 and BIC: 4944.28 respectively.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Universidad Nacional de Frontera, Sullana, Piura, Perú.

REFERENCES

Aurna, F, Mostafa, R.Siddiqui, A, Karim, T., Saika, S., Arifeen, M. y Kabir, H. (2021). Time series analysis of electric energy consumption using autoregressive integrated moving average model and Holt Winters model, TELKOMNIKA TELECOMMUNICATION, COMPUTING, ELECTRONICS AND CONTROL Volume 19 No. 3. pp. 991-1000

Correa, L. (2020). Proposal of a predictive model adjusted to the milling process and its effects on economic profitability for a rice processing company.

Hamzah, L., Nabilah, S., Russel, E., Usman, M., Virginia, E., y Wamiliana. (2020). Dynamic modelling and forecasting of data export of Agricultural Commodity by Vector Autoregressive Model, JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY Volume 55 No. 3.

Hussain, A., Rahman, M., y Alam Memon, J. (2016). Forecasting electricity consumption in Pakistan: the way forward, ENERGY POLICY Volume 90 C. pp. 73-80.

INEI. (2021). “Regional Information System for decision-making”, in: Regional Information System of decision-making.

Purwanto, Sunardi, Tristanti Julfia, F., y Paramananda, A. (2019). “Hybrid model of ARIMA-linear trend model for tourist arrivals prediction model in Surakarta City, Indonesia”, AIP Conference Proceedings, Surakarta, IN.

Putri, R., Usman, M., Warsono, Widiarti, y Virginia, E. (2021). Modeling Autoregressive Integrated Moving Average (ARIMA) and Forecasting of PT Unilever Indonesia Tbk Share Prices During the COVID-19 Pandemic Period. Journal of Physics: Conference Series, 1751, 012027.

Spain Government. (2021). Food from Spain. Agriculture, fishing and feeding Ministery.

Taylor, J. (2008). A Comparison of Univariate Time Series Methods for Forecasting Intraday Arrivals at a Call Center, MANAGEMENT SCIENCES Volume 54 No. 2. pp. 253-265.

Wang, J., Zhang, L., Zhang, W., y Wang, X. (2019). Reliable Model of Reservoir Water Quality Prediction Based on Improved ARIMA Method, ENVIRONMENTAL ENGINEERING SCIENCE Volume 36 No. 9.

Zambrano, C., Andrade, M. y Carreño, W. (2019). Factors that affect rice productivity in the province of Los Ríos. Quevedo: University and Society.