INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Estimation of Farmgate Rice Price Using Temporal Causal Modeling and Kalman Filtering

Frederick F Patacsil[1]*, Paulo V Cenas[1], Andhee M Jacobe[1], Monica M Brosas[1], Pauline Joy V Roa[1]

1 College of Computing, Pangasinan State University, Urdaneta City Campus, Philippines

OPEN ACCESS

Received: 23-03-2022

Accepted: 04-08-2022

Published: 27-08-2022

Citation: Patacsil FF, Cenas PV, Jacobe AM, Brosas MM, V Roa PJ (2022) Estimation of Farmgate Rice Price Using Temporal Causal Modeling and Kalman Filtering. Indian Journal of Science and Technology 15(34): 1642-1647. https://doi.org/10.17485/IJST/v15i34.666

* **Corresponding author**.

frederick_patacsil@yahoo.co.uk

Funding: Pangasinan State University (PSU), for providing financial assistance

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To develop a forecasting model for the farmgate prices of rice crop in the Philippines and to improve the derived model forecasts by applying Kalman filters. **Methods:** The researcher's utilized monthly rice farmgate price and inflation rate from 1990 to 2015 as training information in building the temporal-causal model. On the other hand, the dataset for rice farmgate price from 2016 to 2020 acted as the testing set, allowing the researchers to determine model accuracy using mean square error, mean absolute error, and mean absolute percentage error. **Findings:** Results indicate that applying Kalman filters to the derived temporal-causal model indeed improves prediction performance, as evidenced by the lower MSE values. In particular, applying Kalman filter to the derived Temporal-Causal model 15% (without inflation as control input) and 3% (with inflation as control input) decrease in the MSE. In terms of the Temporal-Causal-Kalman filter with no control input, a decrease of 1.8% is observed for the MAE as well as a decrease of 3% in the MAPE, indicating a substantial improvement in the accuracy of the base model. Interestingly though, adding a control-input variable in the Kalman filter generated gave an increase of 4.4% and 3.8% in the MAE and MAPE respectively. This might be due to the not-so-strong correlation between the farmgate price and control input (inflation). Seeking conditions when will external inputs be helpful in enhancing Kalman filters as well as combining the models with other data analytics techniques may be valuable in future research. As for the comparison with nonlinear setups, results for unscaled, partially scaled, and fully scaled artificial neural networks show that Kalman filtering can attain almost on-par prediction performance with such methods. **Novelty:** This research presented a new scheme of predicting farmgate price. Compared to typical time series models, the derived Temporal-Causal model included inflation as a factor. Moreover, the combination of Temporal-Causal and Kalman filter is a new method for improving forecasts.

Keywords: Kalman Filter; Temporal Causal Model Modeling; Price Estimation; Rice Price

1 Introduction

The Philippines is an archipelago with more than 7,000 islands and a roughly 30-million-hectare land area. One-third of the total land area was categorized as an agricultural land where 9.7 million Filipino workers and their families entirely depend on it. Being agriculture-dependent means the Gross Domestic Product (GDP) might be adversely affected by sudden fluctuations in crop prices. Hence, a forecast of future crop prices will be of great help to administration officials, traders, and farmers especially in deciding when to sell, buy, as well as what crops to plant. However, crop price estimation is not as easy as it sounds and is one of the numerous agricultural problems in many countries. Various research has been made to come up with models that will help generate accurate price predictions. Related works typically fall into two types: univariate time series analysis of historical crop price data and predictor-searching analysis**.** The former focuses on the analysis of the trend, seasonal and cyclical variations, and random components of the historical data. Traditional models like exponential smoothing, error-trend-seasonal (ETS), moving average (MA), autoregressive (AR), autoregressive moving average (ARMA), and the autoregressive integrated moving average (ARIMA) have been extensively used in literature. However, these models are founded on the belief that the errors are linearly correlated across time and the time series data are Gaussian by nature. Kurumatani [1] tried to exploit auto-regressive integrated moving average (ARIMA or the Box Jenkins method) and its features in forecasting the prices of agricultural products which show seasonality in their time series, and conventional methods. They found out that recurrent neural networks, representing the latest machine learning technology, can forecast future time series better than conventional methods. Researchers also tried to integrate machine learning methods to improve the prediction performance of traditional time series techniques. Integrating artificial neural networks (ANN), long short-term memory (LSTM), and support vector machine (SVM) to ETS and ARIMA have yielded superior results than the traditional models alone, even for non-Gaussian data and data having nonlinearly correlated errors [2,3]. However, the problem with these traditional and hybrid univariate time series methods is that they only consider previous values as bases for the modeling, that is, explanatory/predictor variables were not considered. On the other hand, predictor-searching analysis deals with the seeking of patterns of interrelations among dependent and independent variables. Recent applications of such methods to crop price estimation include the usage of regression-based decision trees and random forests [4], multiple linear regression [5], and cross-sectional and panel regression modeling [6]. However, regression-based models assume that errors are not correlated, a characteristic not exhibited by time-series data. Furthermore, prediction-improving algorithms have found popularity in the field of forecasting. One such algorithm is the Kalman filter. Kalman filtration has found recent applications in time-varying brain networks [7], root tracking [8], chaotic oscillators [9], rice price estimation [10], ultrasonic signals [11], biological systems [12], rice production [13], coffee price [14] and among many others. Results of the studies show that Kalman filter-embedded models provide superior estimates than the original model. [15] propose the SLSTM method, a system for predicting the sales of agricultural products shows a low forecasting error rate and NMAE in performance comparison experiments with auto_arima, Prophet, and standard LSTM models. The experimental results show that the error rate of the proposed SLSTM model is signifificantly lower than those of other classical methods.

In this study, the researchers aimed to test the feasibility of implementing a time-series model which incorporates the effects of explanatory variables, a Temporal Causal model. Moreover, this study also tested the feasibility of improving the predictions of a Temporal Causal Model by embedding Kalman filters. In particular, this study sought to derive a Temporal Causal model for the farmgate price of rice with inflation as a factor and to improve the model by applying Kalman filters.

2 Material and methods

2.1 Research Design

This study utilized a quantitative research design called longitudinal - exploratory design. The longitudinal aspect pertained to the time-based data under study while the exploratory part dealt with the modeling aspect of the research.

2.2 Source of Data

This research used secondary data, taken from the Philippine Statistics Authority, specifically from the OpenSTAT database (openstat.psa.gov.ph). The data includes the monthly farmgate prices of rice and the inflation rate from 1990 to 2020.

2.3 Treatment of Data

To attain the objectives of this research, the following tools were implemented.

2.3.1 Temporal–Causal Modeling

Temporal – Causal modeling was used to determine the initial model for the time series data (farmgate price of rice). A temporal-causal model is an autoregressive time series model which aims to predict future values of an underlying quantity based on its previous values and previous values of dependent time series variables, in this case, the inflation rate.

Temporal–Causal Modeling is deemed as the appropriate time-series modeling tool since rice price is believed to be affected by the previous rice prices as well as the previous values of inflation. In particular, the current rice price (P_i) is believed to follow dynamical equations:

$$P_i = \sum_{k=1}^{12} \alpha_k P_{i-k} + \sum_{k=1}^{12} \beta_k I_{i-k}, \quad I_i = \sum_{k=1}^{12} \gamma_k I_{i-k}$$

where $I_i$represents the current inflation rate. The terminal index 12 represents the monthly period of data and α_k, $\beta_k$and $\gamma_k$are the parameters being estimated.

2.3.2 Kalman Filtering

As for the enhancement of the temporal–causal model, two multidimensional Kalman filters were employed. One relies solely on the farmgate price as the input variable, while the other integrates the inflation rate as a control input. After the determination of the appropriate time-series model, it was enhanced by applying Kalman Filtering. Kalman filtering, also known as Linear Quadratic Estimation, is a recursive algorithm which estimates the internal state of a linear dynamic system by applying appropriate weights to previous estimates and previous measurements. In our case, the input measures, as well as the initialization values, come from the historical data (crop price and inflation rate). The first Temporal-Causal-Kalman (TC-KF1) integration makes use of the forecast result from the time-series model of the previous section as input to the Kalman filter. The discrete Kalman filter algorithm consists of the time-update and measurement-update equations, with no control input variable and zero mean multivariate normal distribution and zero-mean Gaussian white noise.

Time update and measurement update equations are as follows:

$$x_{k+1} = F_k x_{k+1} + w_k$$

$$z_k = H_k x_k + v_k$$

The second Temporal-Causal-Kalman (TC-KF2) integration follows the preceding model, with the addition of the inflation rate as a control input variable.

2.3.3 Training and Testing

In order to avoid overfitting, the dataset was divided into two parts for both modeling setups. The dataset from 1990 to 2015 served as the training set. On the other hand, the existing information from 2016 to 2020 served as the testing set, assessing the accuracy of the three models.

2.3.4 Accuracy Determination and Model Comparison

To determine the accuracy of the models the study utilized mean square error, mean absolute error, and mean absolute percentage error. The forecasting measurements are expressed as follows, where Y_i and $\acute{Y}_i$are the actual and forecasted values, respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^{N} {}_i - \acute{Y}_i \big|$$

$$MAPE = \frac{100}{n} \sum_{i=1}^{N} \frac{\left(Y_i - \hat{Y}_i\right)}{Y_i}$$

The lower the values of the mean square error, mean absolute error, and mean absolute percentage error means better model. Moreover, the model which acquired the least MSE, MAE, and MAPE was deemed as the best forecasting model[16].

3 Results and Discussion

Several forecasting models were proposed to determine the farmgate price of rice in the study. The models are the single Temporal-Causal model, and the integrated Temporal-Causal-KF1 and Temporal-Causal-KF2. Table 1 present the results of the forecasting.

The temporal-causal (TC) analysis model alone is used as a baseline in comparing the other hybrid models. From the results, MSE = 5.2912, MAE = 1.7631, and MAPE = 10.5311% for the model.

Adding to the temporal-causal model, Kalman filtering was added in the TC model to produce TC-KF1. The use of the hybrid model aims to determine the results of using a linear quadratic estimation in crop price forecasts. The results show that with TC-KF1, MSE = 4.4954, MAE = 1.7322, and MAPE = 10.2181%.

With TC-KF2, the integration of the temporal-causal model and the Kalman filtering is combined with an added control input variable in the Kalman filter. The results show that with the added control input, MSE = 5.1272, MAE = 1.8414, and MAPE = 10.9313%

From the results of the MSE, the TC -KF1 outperforms both TC -KF2 and TC model, with TC -KF2 having lower MSE than the single TC. Integrating a Kalman filter, whether TC -KF1 or TC -KF2, demonstrated better performance than the TC alone. The TC model has difficulty in capturing nonlinear features of the farmgate price data, which might lead to its low performance. Integrating a Kalman filter effectively improved the forecasting performance instead of the single TC model. This confirmed the study of[13,14] that price that is non-stationary can be estimated and forecasted utilizing the Kalman filtering algorithm on a single linear state space model to estimate and forecast the optimal value of coffee price.

Table 1. Comparison of theActual Farmgate Price to the TC, TC-KF1, and TC -KF2 Price Forecast from 2016-2020

Date	Farmgate Price	TC	TC-KF1	TC-KF2
Jan-16	17.04	17.54	-	-
Feb-16	17.23	17.73	-	-
Mar-16	17.32	17.86	14.7221	14.7753
Apr-16	16.81	18.13	15.6001	15.6744
May-16	17.22	18.37	16.1387	16.2342
Jun-16	17.45	18.38	17.7705	17.8810
Jul-16	19.15	18.39	19.3552	19.4977
Aug-16	18.63	18.29	19.0715	19.2332
Sep-16	17.62	18.19	16.4600	16.6289
Oct-16	16.08	18.13	18.4328	18.6241
Nov-16	16.72	18.2	19.8359	20.0424
Dec-16	17.86	18.35	19.2325	19.4579
Jan-17	17.89	18.47	17.9009	18.1399
Feb-17	17.91	18.55	19.0282	19.2717
Mar-17	17.97	18.67	17.8717	18.1504
Apr-17	18.33	18.77	17.8965	18.1951
May-17	18.16	18.84	18.0897	18.3884
Jun-17	18.3	18.86	19.1383	19.4291
Jul-17	18.94	18.84	20.6148	20.8601
Aug-17	18.76	18.82	18.0137	18.2401
Sep-17	18.13	18.8	17.6140	17.8394
Oct-17	17.53	18.81	18.9395	19.1986
Nov-17	18.09	18.85	19.9250	20.1904
Dec-17	18.53	18.9	17.6443	17.8979
Jan-18	18.89	18.96	19.5188	19.7534
Feb-18	19.91	19.02	18.9174	19.1718
Mar-18	20.5	19.08	17.1790	17.4686
Apr-18	20.16	19.12	18.8298	19.1606
May-18	20.14	19.15	19.5282	19.8817
Jun-18	20.42	19.17	19.3323	19.7008
Jul-18	21.02	19.17	19.8420	20.2640
Aug-18	22.7	19.18	20.2514	20.7074
Sep-18	22.04	19.19	19.1706	19.6595
Oct-18	19.98	19.21	19.6448	20.1384
Nov-18	19.4	19.24	19.3865	19.8943
Dec-18	19.64	19.27	18.4397	18.9654
Jan-19	18.91	19.3	19.7684	20.2721
Feb-19	18.4	19.34	19.1942	19.6672
Mar-19	17.51	19.37	17.7675	18.1925
Apr-19	16.8	19.4	19.4530	19.8306
May-19	16.91	19.42	18.8850	19.2366
Jun-19	16.53	19.43	18.9102	19.2583
Jul-19	17.32	19.44	19.0058	19.3390
Aug-19	16.2	19.46	19.2955	19.6161

Continued on next page

Table 1 continued

Sep-19	14.75	19.47	17.2886	17.5837
Oct-19	14.4	19.49	17.9601	18.2318
Nov-19	14.58	19.51	19.8994	20.1554
Dec-19	15.32	19.53	18.4999	18.7503
Jan-20	15.53	19.55	19.0570	19.3405
Feb-20	15.9	19.58	18.7485	19.0596
Mar-20	16.96	19.6	19.0207	19.3363
Apr-20	18.17	19.61	20.3719	20.6755
May-20	18.86	19.63	19.9542	20.2463
Jun-20	18.35	19.64	19.2804	19.5681
Jul-20	17.75	19.66	20.5403	20.8334
Aug-20	16.93	19.67	19.5744	19.8907
Sep-20	14.9	19.68	19.7812	20.0930
Oct-20	15.05	19.7	18.7461	19.0631
Nov-20	15.74	19.71	18.9519	19.2571
Dec-20	16.54	19.73	19.3109	19.6254
	MSE	5.2912	4.4954	5.1272
	MAE	1.7631	1.7322	1.8414
	MAPE	10.5311%	10.2181%	10.9313%

Comparing the integrated models, the TC -KF1 performs better than the TC -KF2 with the addition of a control input. The addition of an external control input, in the form of the inflation rate, did not improve the TC -KF1 model. Since most standard time-series models are concerned with the internal system dynamics, adding an external input did not improve the model. The farmgate price and inflation rate also has low inverse correlation, which may result into lowering the performance of the model. Comparing the MAE and MAPE results of all models, TC -KF1 has the lowest MAE and MAPE results, TC coming in second, and TC -KF2 having the largest MAE and MAPE among the three. While this shows that the TC works better than the ARIMA-KF2, it is noted that both MAE and MAPE are more robust to data with outliers; the data used in the study have no significant outliers.This shows that with the available data, the combination of the ARIMA model and Kalman filter has the best performance based on the MSE, MAE, and MAPE.

In comparison to the study made by[10], which estimates future prices of crop prices in the Philippines using a simple ARIMA and ARIMA-Kalman plus additional input of exogenous variables whose influence in prices were integrated in this study which provided better prediction accuracy than the aforementioned predictive models. These variables reduced all the error measures examined in the study. To add, the model used in this study uses a greater sample period of monthly prices from 1990-2020. A commonality among research on crop price forecasting is the comparison of hybrid models which include a variety of different algorithms in the data treatment. From the previous studies, hybrid models combining multiple forecasting tools provide more accurate predictions for crop prices. From Cenas[10], the combination of the ARIMA and Kalman filter has better predictive performance for Philippine crop prices compared to the standalone ARIMA. According to Purohit et al.[3] and Celma & Oliveira[15,16], hybrid methods do provide better results in crop price forecasting than individual models. The result of this study also follows the same conclusion.

To have a baseline comparison with nonlinear methods, three artificial neural networks (ANN) augmentations were also tested by the researchers. An unscaled ANN acquired an MSE of 5.715, MAE of 1.507, and MAPE of 11.18%. An ANN with scaled inflation variable gave an MSE of 5.355, MAE of 1.507, and MAPE of 10.86%. A fully scaled ANN provided an MSE of 0.057, MAE of 0.0821, and MAPE of 0.487%. In conclusion, Kalman filtering beat the unscaled and partially scaled ANN enhancement of the Temporal-Causal model in terms of MSE, indicating less variations in the prediction, though a narrower range will be provided by the ANN-augmented models. Nonetheless, Kalman filtering falls short when compared to a fully scaled ANN-augmented. However, these results show that Kalman filtering can be on-par with nonlinear models, especially for non-scaled ones.

4 Conclusion

This study utilized farmgate price and inflation data from 1990 to 2020. The information from 1990 to 2015 were used to build the Temporal-Causal model. The model accuracy was determined by comparing the forecasts for 2016 to 2020 with the actual values (testing set). The model was then improved by applying two Kalman filters – one without inflation as control input and another with inflation as control input. The popular ARIMA statistical method is compared by combining ARIMA and the Kalman filter in terms of predicting future rice prices. The result of the study reveals integrated Time-Series Temporal-Causal Model and Kalman filter model consistently got lower results among the three models based on the utilized forecasting measurements. Focusing on the MSE indicator, the integration of a Kalman Filter (MSE = 4.4954) and the addition of an input variable (MSE = 5.1272) is lower than the Time-Series ARIMA (MSE = 5.2912). From this, a hybrid model that combines Kalman filtering and the ARIMA statistical analysis has proven to be robust in predicting crop prices based on the study. In addition, the addition of a Kalman filter in the Time-Series model produced better results. Improvements on the Kalman filter may be done as well, such as using other variables as control input, or implementing extensions of the Kalman filter. Results reveal that the embedding of Kalman filters to the derived Temporal-Causal model enhances model performance. Kalman filtration without control input yielded the best performance – a decrease of 15%, 1.8%, and 3% were observed for the MSE, MAE, and MAPE respectively. On the other hand, using inflation as control input gives light to some interesting results. Despite a decrease of 3% in the MSE, an increase of 4.4% and 3.8% in the MAE and MAPE respectively, signifying that there are certain conditions in order to make a Kalman filter with control input work better. Also, comparison with some ANN-augmentation revealed that Kalman filtration can have similar results with nonlinear methods. It is recommended that the proposed hybrid models be used to predict other food crops. Using other input variables aside from the inflation rate are also suggested as intervening variables in price prediction. Additionally, the use of hybrid models can also be applied in forecasting other time-series data. Importantly, the findings are helpful for government decision-makers, farmers, and end users to develop strategies to stock/release supply of rice in the market.

5 Acknowledgement

The authors sincerely thank the Pangasinan State University (PSU), for providing financial assistance for this research and OpenSTAT database for providing the research the needed data.

References

1) Kurumatani K. Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*. 2020;2(8):1–7. Available from: https://link.springer.com/article/10.1007/s42452-020-03225-9.
2) Bayona-Oré S, Cerna R, Hinojoza ET. Machine Learning for Price Prediction for Agricultural Products. *WSEAS Transactions on Business and Economics*. 2021;18:969–977. Available from: https://doi.org/10.37394/23207.2021.18.92.
3) Purohit SK, Panigrahi S, Sethy PK, Kumari S, Behera. Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. *Applied Artificial Intelligence*. 2021;35:15–15. Available from: https://doi.org/10.1080/08839514.2021.1981659.
4) Ghutake I, Verma R, Chaudhari R, Amarsinh V. An intelligent Crop Price Prediction using suitable Machine Learning Algorithm ITM Web Conf. 2021;40:3040–3040. Available from: https://doi.org/10.1051/itmconf/2021400304.
5) Ge Y, Wu H. Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications*. 2020;32(22):16843–16855. Available from: https://doi.org/10.1007/s00521-018-03970-4.
6) Paltasingh K, Ranjan G. Phanindra Statistical Modeling of Crop-Weather Relationship in India: A Survey on Evolutionary Trend of. *Methodologies Asian Journal of Agriculture and Development*. 2018;15:43–60. Available from: https://doi.org/10.22004/ag.econ.275688.
7) Pascucci D, Rubega M, Plomp G. Modeling time-varying brain networks with a self-tuning optimized Kalman filter. *PLOS Computational Biology*;16(8):e1007566–e1007566. Available from: https://doi.org/10.1371/journal.pcbi.1007566.
8) Kostoglou K, Lunglmayr M. Root tracking using time-varying autoregressive moving average models and sigma-point Kalman filters. *EURASIP Journal on Advances in Signal Processing*. 2020;2020(1). Available from: https://doi.org/10.1186/s13634-020-00666-7.
9) Forero-Ortiz E, Tirabassi G, Masoller C, Pons AJ. Inferring the connectivity of coupled chaotic oscillators using Kalman filtering. *Scientific Reports*. 2021;11(1). Available from: https://doi.org/10.1038/s41598-021-01444-7.
10) Cenas P. Forecast of Agricultural Crop Price using Time Series and Kalman Filter Method. *Asia Pacific Journal of Multidisciplinary Research*. 2017;5(4). Available from: https://www.semanticscholar.org/paper/Forecast-of-Agricultural-Crop-Price-using-Time-and-Cenas/7eca5c374b0fe4eb95688f6bc1dcdcbf4ebe3825.
11) Kostoglou K, Lunglmayr M. Root tracking using time-varying autoregressive moving average models and sigma-point Kalman filters. *EURASIP Journal on Advances in Signal Processing*. 2020;2020(1):1–16. Available from: https://doi.org/10.1186/s13634-020-00666-7.
12) Pascucci D, Rubega M, Plomp G. Modeling time-varying brain networks with a self-tuning optimized Kalman filter. *PLOS Computational Biology*. 2020;16(8):e1007566–e1007566. Available from: https://doi.org/10.1371/journal.pcbi.1007566.
13) Maminirivo FS, Kyo K. A Bayesian Approach to Evaluating the Dynamics of Rice Production in Madagascar. *International Journal of Agricultural Economics*. 2020;5(2):43–43. Available from: https://doi.org/10.11648/j.ijae.20200502.12.
14) Berhane T, Shibabaw N, Shibabaw A, Adam M, Muhamed AA. Forecasting the Ethiopian Coffee Price Using Kalman Filtering Algorithm. *Journal of Resources and Ecology*. 2018;9(3):302–307. Available from: https://doi.org/10.5814/j.issn.1674-764x.2018.03.010.
15) Yoo TWW, Oh ISS. Time Series Forecasting of Agricultural Products' Sales Volumes Based on Seasonal Long Short-Term Memory. *Applied Sciences*. 2020;10(22):8169–8169. Available from: https://doi.org/10.3390/app10228169.
16) Ribeiro CO, Oliveira SM. A hybrid commodity price-forecasting model applied to the sugar-alcohol sector. *Australian Journal of Agricultural and Resource Economics*. 2011;55(2):180–198. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-8489.2011.00534.x.