Investigating Causal Relationships Between Inflation News Among Other News Topics In Philippine News Media Using Granger Causality

Erwin Antepuesto   
Department of Computer, Information Sciences, and Mathematics   
University of San Carlos   
Cebu City, Philippines 19103939@usc.edu.ph   
  
Stan Kiefer Gallego   
Department of Computer, Information Sciences, and Mathematics   
University of San Carlos   
Cebu City, Philippines 21103445@usc.edu.ph  
  
Gerard D. Ompad  
Department of Computer, Information Sciences, and Mathematics  
University of San Carlos   
Cebu City, Philippinesgdompad@usc.edu.ph  
  
Angie Ceniza  
Department of Computer, Information Sciences, and Mathematics  
University of San CarlosCebu City, Philippines amceniza@usc.edu.ph

*Abstract*—A major economic problem for Philippine society is inflation, which is defined as a decline in the purchasing power of the country's currency and has an effect on job, welfare, and health care, among other things. This issue has prompted extensive coverage on social media, reaching Filipino consumers widely. It has led to a trend where numerous news outlets have published numerous articles focused on inflation. However, inflation's effects have also spilled over into other news topics. The purpose of this study is to look into how inflation affects the Philippines by developing a Time Series model employing Granger Causality to discern causal relationships between inflation news articles and other news topics. News articles were collected using Meltwater, an online media monitoring platform from certified Philippine news sources, focusing on inflation-related news articles and selected news topics dating January to April 2024. Data preprocessing was applied to ensure cleanliness and suitability for analysis. Preprocessed data was then converted into time series followed by Granger causality analysis to determine causal relationships.

Keywords—inflation, multivariate time series, granger causality, causality testing, media analysis

# INTRODUCTION

Inflation continuously erodes the purchasing power of the Philippine Peso, often at rates higher than the global norm. Between 2010 and 2020, the Philippines experienced an average inflation rate of 4.2%, within the central bank's target of 2-4% [1]. However, instances of inflation exceeding this range have created economic challenges, particularly for the low- and middle-income populations, which constitute 58.4% and 40%, respectively [2].

Understanding the effects of inflation is a complex process, with several methods already being utilized to determine its effects on various aspects on a country. While traditional methods like CPI and GDP analysis are commonly used to understand inflation, this study introduces Granger causality (GC) to uncover causal relationships between inflation and media coverage—a method not previously explored in the Philippine context. This method allows the researchers to understand and determine any causal relationships between inflation and other news topics bi-directionally by using online news articles gathered from verified news agencies in the Philippines.

Time series analysis (TSA) with its many applications in different fields is essentially a method of investigating the characteristics of the response variable with time as the independent variable [3]. Time series models help in comprehending the underlying forces and structure behind the observed data in this case are the gathered articles on various news topics. These models can also be utilized for forecasting, monitoring, as well as feedback and feedforward control [4]. Taking advantage of this, TSA can help understand how news reacts to inflation by identifying their relationship with Granger causality. TSA and Granger causality is not unfamiliar in the field of economics. This technique was applied to explore connections between two or more variables, as demonstrated in the research conducted investigating the link between the Human Development Index and economic growth using Granger Causality [5]. Thus, analysing how inflation and news coverage interact provides valuable insights into the effects of inflation.

The objectives of this study are to: (1) collect news articles on nine topics, including inflation, from thirty-one verified Philippine news sources; (2) compile these into a time series; (3) determine the appropriate lag length using the Vector Autoregression model; (4) perform time series analysis and Granger causality tests; and (5) validate the results with the Johansen Cointegration Test and the Vector Error Correction Model. The findings of this study offer valuable insights into the dynamics between interrelated issues in the Philippines, enhancing the understanding of factors that affect inflation and vice versa.

Due to the nature of the data used in this study, the researchers defined the scope and took into consideration the following limitations: (1) The news articles are only the ones available online, not physical newspapers; (2) the origin of the articles must only in the list of verified Philippine news agencies the researchers compiled; (1) the dates of the research articles must be from January 2024 to April 2024 due to storage limitations in data collection. Due to this scope, the study’s findings may be influenced by the available articles on the specified scope.

# RELATED LITERATURE

## Time Series

A time series is a sequence of observations arranged in a specific order, typically over time, though it can also follow other dimensions, like spatial arrangements. Time series can be continuous, such as when recording an electrical signal like voltage. When only a single variable is tracked, it is known as a univariate time series. In cases where multiple variables are recorded simultaneously such as the mode of this study, it is referred to as a multivariate time series [6]. Data validity is crucial before applying any data preprocessing when dealing with time series analysis [7]. A feature refers to a measurable characteristic or attribute of the data that can be used to identify patterns, relationships, or predictive information within the time series [8]. Features have to be identified and engineered to be extracted from the data to understand which features significantly influence predictions [9]. Features can be identified as correlated features which exhibit statistical dependence or association with each other. Correlated features can lead to confounded inferences if not properly accounted for by dropping these features before performing Granger causality [10]. Time series are also categorized into two types: stationary and non-stationary. A stationary time series is characterized by constant statistical properties over time. This means that the mean, variance, and autocorrelation structure remain stable across different periods of the series. A non-stationary time series exhibits changing statistical properties over time, these can disrupt the causality outcomes by leading to incorrect conclusions due to the irregularity of the data [11].

## Time Series Causality

Time series causality explores cause-and-effect relationships between variables in a time series, crucial for informed decision-making in fields like economics, medicine, and environmental sciences [12]. It has two main objectives: treatment effect estimation, which measures the impact of an event on a variable, and causal discovery, which seeks to identify causal relationships between variables in the time series, aligning with the goals of this study [13].

## Granger Causality

Granger causality assesses whether one time series can predict another by analysing if past values of one improve the forecasting of future values. Introduced by C.W.J. Granger [14], it compares the forecasting performance of two series. Granger causality has been applied in various fields, notably in finance, where it identified key systemic events like the October 1997 mini-crash, the 1998 Russian default, and the 2002 stock market downturn [15].

## Existing Works

Related works pertaining to news causality utilized much different methodologies to identify causal relationships between two or more topics. Some include using causal Bayesian networks, combination of causality models such as necessary, sufficient-component, and probabilistic causality. One comparative study examined time-series causality techniques but on only one news source, and didn’t infer topic causality [16]. Other methods include lexico-syntactic causal patterns on mined text [17], supervised learning via language models [18], knowledge bases [19], and machine learning [20]. Beyond news, Granger causality has been used to show that indices like the RCI, A-COVID Index, and uncertainty index significantly predict stock market volatility in Latin America [21], and that geopolitical tensions influence oil prices and forecast accuracy [22].

# METHODOLOGY

## Sources of Data

The data for this study is sourced from verified URLs of Philippine news agencies, all collected through Meltwater, a comprehensive media intelligence platform. The collected articles span nine general news topics: inflation, economy, politics, technology, environment, health, business, welfare, and foreign affairs. Only articles published between January 2024 and April 2024 were included. Each article was analysed by Meltwater to ensure it fell under one of the specified news topics and met the date range criteria before being collected.

## Data Collection and Preprocessing

In collecting news article data from Meltwater, the researchers used the keyword feature to find articles that pertain to the selected news topics. The keyword feature also has filters, the filters used were: only Philippines made articles, to exclude non-Philippines based news agencies, articles written in Tagalog, English and Cebuano, and fall within the specified time frame for data collection.

The dataset for this research comprises 118,266 articles categorized into various news topics. The business category had the highest count with 32,505 articles, followed by technology with 20,286 articles. Health and environment-related topics were nearly equal, with 14,686 and 14,683 articles, respectively. The economy and politics categories had similar counts of 11,618 and 11,616 articles. Additionally, welfare news contributed 5,081 articles, and inflation-focused articles summed 5,734. Foreign affairs had the least representation with only 2,057 articles. In terms of article origins, the largest contribution came from The Manila Times, which provided 64,198 articles, making up a significant portion of the dataset. Other notable contributors included Philstar.com with 9,580 articles, Daily Tribune with 7,636, Business Mirror with 7,039, and Manila Standard with 6,497 articles. Additional sources like Business World, SunStar Philippines, and ABS-CBN News contributed 5,446, 4,498, and 2,781 articles, respectively. Other outlets such as GMA News Online and Rappler added to the dataset with contributions of 2,762 and 2,236 articles, while regional outlets like Cebu Daily News and Visayan Daily Star contributed smaller amounts.

After downloading the data across all topics, a secondary filtering process was initiated. This began with centralizing datasets per topic into individual CSV files due to download size limitations of 20 kilobytes. This resulted in nine CSV datasets representing all topics. Following this centralization, additional filtering was applied based on source and date. Nine empty datasets were created to store the fully filtered data per topic. The final data format consisted of daily mappings for each news source within the date range from January 1 to April 30, 2024.

The datasets for selected news outlets were converted into data frames and merged into a single dataset. Since some dates appeared in certain datasets but not in others, null values were checked by counting them in each column to ensure data quality. Rows containing null values were dropped to maintain consistent dates across the dataset. Unnecessary columns and those filled with zeroes over 50% were also discarded to reduce dimensionality and enhance manageability. To ensure analytical integrity, correlated features between data frames were identified by finding pairs with an absolute correlation coefficient greater than 0.8. Columns exhibiting perfect correlation scores of 1.0 were dropped to avoid numerical instability in future analyses. This preprocessing step not only eliminated redundant information but also improved overall performance of the dataset for subsequent analysis.

## Data Analysis

The Augmented Dickey-Fuller (ADF) test served as a tool to determine the stationarity of time series data, a critical step for subsequent modelling techniques like Granger causality. It worked by transforming non-stationary time series into stationary ones through first difference transformation. Additionally, the selection of the Vector Autoregression (VAR) model aided in identifying the optimal lag order necessary for modelling multivariate time series data using Vector Autoregression.

In the analysis of causal relationships among different variables, the Granger causality test was utilized. This test examined all possible combinations of non-stationary time series data, extracting p-values for different lag lengths. The minimum p-value across all lag lengths was selected and stored in the resulting data frame. Additionally, the Johansen cointegration test was conducted to determine whether any cointegrated vectors existed among the non-stationary time series variables. The presence of cointegration implied the existence of a long-run relationship among the variables, indicating meaningful and stable associations. Conversely, the absence of cointegrated vectors suggested that there was no long-run relationship among the variables, potentially leading to spurious correlations identified through Granger causality analysis.

## Validation

* **Vector Autoregression (VAR) Model Selection:** The VAR model selection is commonly used to model multivariate time series data, serving as a basis for selecting the best-performing candidate model. In this approach, Information-Theoretic Criteria such as the Akaike Information Criterion (AIC) (1) and the Bayesian Information Criterion (BIC) (2) are applied to compare candidate models.



2

The model with the lowest criterion value is generally considered the best estimate of the unknown true model. When conducting Granger Causality analysis, model selection typically relies on the AIC, where the model with the lowest AIC value is chosen as the optimal model for testing causal relationships [23].

* **Johansen Cointegration Test:** Johansen Cointegration analyzes long-term relationships between multiple non-stationary time series. Unlike the Granger Causality test, which requires differencing for stationarity, Johansen's test checks for a long-term equilibrium without differencing. If cointegration exists, it suggests the variables move together over time. This method is often paired with Vector Error Correction Models to capture both short-term and long-term dynamics [24].
* **Vector Error Correction Model (VECM):** The VECM (3) is derived from the VAR model and is used to model systems of integrated time series with cointegrating relationships. The VECM captures both short-term deviations and long-term equilibrium adjustments [25].

3

* **Trace Statistic**: Trace statistic (4) is a number calculated from a statistical test of a hypothesis. It shows how closely your observed data match the distribution expected under the null hypothesis of that statistical test [26].

4

* **Max Eigenvalue Statistic (MES):** MES (5) tests the null hypothesis of having exactly r cointegrating vectors against the alternative of [27].

5

## Conceptual Framework

A black background with a black square

Description automatically generated with medium confidence

1. Conceptual Framework

# RESULTS AND ANALYSIS

## Data Preprocessing

## Lag Length Selection

With the data stabilized through preprocessing, the VAR model selection was conducted. The model determines the optimal lag length by evaluating four key criteria: AIC, BIC, Final Prediction Error (FPE), and the Hannan-Quinn Information Criterion (HQIC).

1. Lag Length Criteria Values

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lag** | **AIC** | **BIC** | **FPE** | **HQIC** |
| **0** | 27.49 | 27.83\* | 8.718e+11 | 27.63 |
| **1** | 28.33 | 33.37 | 2.046e+12 | 30.37 |
| **2** | 29.10 | 38.84 | 5.078e+12 | 33.05 |
| **3** | 29.58 | 44.03 | 1.220e+13 | 35.45 |
| **4** | 29.70 | 48.85 | 3.135e+13 | 37.47 |
| **5** | 28.05 | 51.91 | 3.022e+13 | 37.73 |
| **6** | 23.62 | 52.18 | 7.960e+12 | 35.21 |
| **7** | 10.76\* | 44.03 | 1.766e+10\* | 24.27\* |

This analysis focuses on minimizing the AIC, which balances model fit and complexity by penalizing overly complex models to prevent overfitting. The lowest AIC value, 10.76, was found at lag 7, indicating that the VAR model will use data from the previous 7 days to predict future values based on daily activities.

## Granger Causality

Introductory missing

1. Lag Length Criteria Values

|  |  |  |
| --- | --- | --- |
| **Predictors (X)** | **Response Variables (Y)** | **P\_Values** |
| Business Mirror\_inflation | ABS-CBN News\_business | 0.0002 |
| Business World\_welfare | 0.0002 |
| GMA News Online\_welfare | 0.0003 |
| ABS-CBN News\_environment | 0.0013 |
| InterAksyon\_business | 0.0028 |
| ABS-CBN News\_welfare | 0.0032 |
| GMA News Online\_foreign\_affairs | 0.0038 |
| Philstar.com\_welfare | 0.0074 |
| SunStar Philippines\_welfare | 0.0075 |

1. Lag Length Criteria Values

|  |  |  |
| --- | --- | --- |
| **Predictors (X)** | **Response Variables (Y)** | **P\_Values** |
| Business Mirror\_inflation | ABS-CBN News\_business | 0.0002 |
| Business World\_welfare | 0.0002 |
| GMA News Online\_welfare | 0.0003 |
| ABS-CBN News\_environment | 0.0013 |
| InterAksyon\_business | 0.0028 |
| ABS-CBN News\_welfare | 0.0032 |
| GMA News Online\_foreign\_affairs | 0.0038 |

## Discussion

The Granger causality results shows that inflation news has a significant and far-reaching impact on a wide range of socioeconomic and global issues across multiple news outlets. The analysis shows that inflation-related coverage extends beyond economic topics to include welfare, the environment, foreign affairs, and technology. This suggests that inflation has a significant impact on media narratives beyond financial concerns, influencing how these broader topics are covered throughout the media landscape.

One clear trend is the strong correlation between inflation and welfare reporting, as seen in Business World, ABS-CBN News, SunStar, Philstar.com, and other publications. This demonstrates how economic pressures, such as inflation, have a direct impact on social issues like welfare, prompting the media to emphasize their interdependence. Furthermore, the significant impact of inflation on environmental topics, as seen in outlets such as Rappler, ABS-CBN News, Manila Standard, and Philstar.com, suggests that inflation is being framed in relation to environmental concerns, potentially reflecting public discourse on the socioeconomic effects of environmental degradation or climate change.

Furthermore, inflation news dominates foreign affairs coverage in outlets such as Philstar, GMA News Online, and the Manila Standard. This connection could reflect how inflation, as a global economic phenomenon, interacts with international relations, trade, and global policies. The impact of inflation on technology topics in publications such as Business Mirror and Manila Standard emphasizes inflation's cross-cutting role in shaping how technological advancements and their societal consequences are reported.

## Cointegration Test

To validate the Granger causality results, the Johansen cointegration test was conducted after making all time series stationary through differencing. Although the individual non-stationary variables are not stationary, cointegration helps determine if they maintain a long-term equilibrium relationship.

The Johansen Cointegration test was followed by a Vector Error Correction Model, experimenting with time lags from 1 to 6 to find the best fit. However, no cointegrated relationships were found, indicating that the time series variables do not move together in the long run, suggesting that any correlation is likely short-term or spurious. This implies that the variables are either independent or influenced by different factors.

# CONCLUSION

This study examined the causal relationship between inflation and various news topics using time series analysis, particularly the Granger causality test, on articles from selected Philippine news agencies. The Johansen cointegration test was also applied to assess potential long-term relationships. Results indicated that inflation news from outlets like Business Mirror and ABS-CBN News had significant predictive influence on topics such as business, welfare, and technology, showing that inflation news shapes media narratives beyond its economic indicator role. However, the Johansen test revealed no long-term cointegrated relationships, suggesting that while inflation news has short-term effects, these do not persist over time and are influenced by other factors.

In conclusion, the study highlights the complex, short-term dynamics between inflation news and socio-economic topics. The lack of long-term cointegration underscores the importance of distinguishing between temporary and lasting influences to better understand how economic indicators and media narratives interact.

To deepen the understanding of the relationships observed, we recommend extending the analysis timeframe to reveal persistent trends and distinguish between short-term and long-term effects. Including foreign news outlets could provide insights into how global perspectives shape inflation discourse. Future studies should explore alternative methodologies that account for structural breaks or regime changes, improving the accuracy of time-series analysis and capturing shifts in inflation-related narratives. These recommendations aim to enhance the exploration of causal relationships and the factors shaping economic narratives.

##### Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

1. J. Consumer Price Index and Inflation Rate | Philippine Statistics Authority | Republic of the Philippines. (2024, October 4). Retrieved October 4, 2024, from Psa.gov.ph website: https://psa.gov.ph/price-indices/cpi-ir
2. Understanding Social Classes in the Philippines: Which Class Do You Belong to? (2019a). Retrieved April 7, 2024, from Pids.gov.ph website: https://pids.gov.ph/details/news/in-the-news/understanding-social-classes-in-the-philippines-which-class-do-you-belong-to#:~:text=The%20latest%20Family%20Income%20and,in%20the%20high%2Dincome%20class.
3. Pandian, S. (2024, October 23). Time Series Analysis and Forecasting | Data-Driven Insights (Updated 2024). Retrieved April 8, 2024, from Analytics Vidhya website: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/
4. ‌SEMATECH. (2023). Engineering Statistics Handbook. Retrieved April 8, 2024, from Nist.gov website: https://www.itl.nist.gov/div898/handbook/
5. Ozturk, S., & Seher Suluk. (2020). The granger causality relationship between human development and economic growth. International Journal of Research in Business and Social Science, 9(6), 143–153. https://doi.org/10.20525/ijrbs.v9i6.902
6. Mastrangelo, C. M., Simpson, J. R., & Montgomery, D. C. (2013). Time Series Analysis. Springer EBooks, 1546–1552. https://doi.org/10.1007/978-1-4419-1153-7\_1045
7. Su, Y., Gong, Y., & Song, S. (2023). Time series data validity. Proc. ACM Manag. Data, 1. https://doi.org/10.1145/3588939
8. ‌Zou, Z., & Cheng, C. (2023). A review of causal analysis methods in geographic research. Environmental Modelling & Software, 172, 105929–105929. https://doi.org/10.1016/j.envsoft.2023.105929
9. Castro, M., Ribeiro, P., Soriano-Vargas, A., Rafael, Maiara Moreira Gonçalves, Leopoldo Lusquino Filho, … Rocha, A. (2023). Time series causal relationships discovery through feature importance and ensemble models. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-37929-w
10. Shojaie, A., & Fox, E. B. (2021). Granger Causality: A Review and Recent Advances. Annual Review of Statistics and Its Application, 9(1), 289–319. https://doi.org/10.1146/annurev-statistics-040120-010930
11. ‌Sachin Heerah, Molinari, R., Guerrier, S., & Marshall-Colon, A. (2021). Granger-causal testing for irregularly sampled time series with application to nitrogen signalling in Arabidopsis. Bioinformatics, 37(16), 2450–2460. https://doi.org/10.1093/bioinformatics/btab126
12. Moraffah, R., Sheth, P., Karami, M., Bhattacharya, A., Wang, Q., Tahir, A., … Liu, H. (2021). Causal inference for time series analysis: problems, methods and evaluation. Knowl. Inf. Syst., 63, 3041–3085. https://doi.org/10.1007/s10115-021-01621-0
13. Weng Siew Lam, Weng Hoe Lam, Saiful Hafizah Jaaman, & Pei Fun Lee. (2023). Bibliometric Analysis of Granger Causality Studies. Entropy, 25(4), 632–632. https://doi.org/10.3390/e25040632
14. Gebhard Kirchgässner, & Wolters, J. (2007). Granger Causality. Springer EBooks, 93–123. https://doi.org/10.1007/978-3-540-73291-4\_3
15. Karpman, K., & Basu, S. (2018). Learning financial networks using quantile granger causality. Houston, TX, USA: Association for Computing Machinery. https://doi.org/10.1145/3220547.3220548
16. Maisonnave, M., Delbianco, F., Tohme, F., Evangelos Milios, & Maguitman, A. G. (2022). Causal graph extraction from news: a comparative study of time-series causality learning techniques. PeerJ Computer Science, 8, e1066–e1066. https://doi.org/10.7717/peerj-cs.1066
17. Lexico-syntactic causal pattern text mining | Proceedings of the 14th WSEAS international conference on Computers: part of the 14th WSEAS CSCC multiconference - Volume II. (2015). Retrieved October 4, 2024, from Guide Proceedings website: https://dl.acm.org/doi/abs/10.5555/1984366.1984373
18. Vivek Khetan, Roshni Ramnani, Anand, M., Sengupta, S., & Fano, A. E. (2021). Causal BERT: Language Models for Causality Detection Between Events Expressed in Text. Lecture Notes in Networks and Systems, 965–980. https://doi.org/10.1007/978-3-030-80119-9\_64
19. CauseNet | Proceedings of the 29th ACM International Conference on Information & Knowledge Management. (2020). Retrieved October 4, 2024, from ACM Conferences website: https://dl.acm.org/doi/abs/10.1145/3340531.3412763
20. Learning causality for news events prediction | Proceedings of the 21st international conference on World Wide Web. (2024). Retrieved October 4, 2024, from ACM Other conferences website: https://dl.acm.org/doi/abs/10.1145/2187836.2187958
21. Coronado, S., Martinez, J. N., Gualajara, V., & Rojas, O. (2022). Transfer Entropy Granger Causality between News Indices and Stock Markets in U.S. and Latin America during the COVID-19 Pandemic. Entropy, 24(10), 1420–1420. https://doi.org/10.3390/e24101420
22. Fernandois, A., & Medel, C. (2020). Geopolitical tensions, OPEC news, and the oil price: A granger causality analysis. Economic Analysis Review, 35(2), 57–90. Retrieved from https://www.rae-ear.org/index.php/rae/article/view/734
23. Stock, J. H., & Watson, M. W. (2001). Vector Autoregressions. Journal of Economic Perspectives, 15(4), 101–115. https://doi.org/10.1257/jep.15.4.101
24. Danang Indrajaya. (2021). Analysis of Cointegration and VECM of FDI, Labor Force, Government Expenditure and GDP in Indonesia (2005-2019). International Journal of Economics Development Research (IJEDR), 2(1), 65–77. https://doi.org/10.37385/ijedr.v2i1.265
25. Dalina Maria Andrei, & Andrei, L. C. (2015). Vector Error Correction Model in Explaining the Association of Some Macroeconomic Variables in Romania. Procedia Economics and Finance, 22, 568–576. https://doi.org/10.1016/s2212-5671(15)00261-0
26. RIFFENBURGH, R. H. (2006). Using the Reference Guide. Elsevier EBooks, 187–193. https://doi.org/10.1016/b978-012088770-5/50051-4
27. Helmut Lüutkepohl, Pentti Saikkonen, & Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VA R process. Econometrics Journal, 4(2), 287–310. Retrieved from https://www.researchgate.net/publication/312984866\_Maximum\_eigenvalue\_versus\_trace\_tests\_for\_the\_cointegrating\_rank\_of\_a\_VA\_R\_process