Received: 18 March 2023 | Revised: 17 April 2023 | Accepted: 27 April 2023 | Published online: 6 May 2023

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 2(3) 153–160
DOI: 10.47852/bonviewJDSIS3202870

BON VIEW PUBLISHING

Intra-Annual National Statistical Accounts Based on Machine Learning Algorithm

Sandip Garai[1,*], **Ranjit Kumar Paul**[1], **Mohit Kumar**[2] **and Anish Choudhury**[3]

[1]*Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute, India*

[2]*Discipline of Computer Application, ICAR-Indian Agricultural Statistics Research Institute, India*

[3]*Department of Seed Science and Technology, Bidhan Chandra Krishi Viswavidyalaya, India*

Abstract: The methods used for forecasting financial series are based on the concept that a pattern can be identified in the data and distinguished from randomness by smoothing past values. This smoothing process eliminates randomness from the data, enabling the inherent pattern to be used for forecasting. However, acquiring high-frequency national accounts data can be challenging, and complicated methods are required to achieve disaggregated series that are compatible with annual totals. Therefore, there is a need for simpler techniques to obtain high-frequency data from low-frequency equivalents. Machine learning algorithms are rapidly evolving, and feedforward artificial neural networks (ANNs) with appropriate training mechanisms have been proposed to temporally disaggregate economic series without considering related indicators. This study proposes using the ANNs algorithm to disaggregate national statistical accounts. An application of disaggregating annual Australia gross domestic product data into quarterly data has also been presented. The higher frequency data generated have been compared with the observed quarterly data to assess its accuracy. Comparative study suggests that the ANN-based model outperforms over benchmark methods such as Chow and Lin method (CL1) and Fernandez method (f).

Keywords: ANN, Australia GDP, ML, temporal disaggregation

1. Introduction

ANNs are a promising alternative method for temporal disaggregation of economic time series that can potentially overcome some of the issues associated with traditional mathematical methods. For instance, they can avoid heavy mathematical calculations required by some of the existing methods, making them more accessible and user-friendly. Additionally, unlike some of the traditional methods, they do not require the use of related indicators for the disaggregation, which can be advantageous when such indicators are not available or their use is deemed inappropriate. Moreover, they can potentially capture complex nonlinear relationships between variables and provide flexible models that can be trained to adapt to changing data patterns.

The world economy has been undergoing significant transformations in recent years. In this era of globalization, the financial sector has become increasingly important in driving economic growth. With the rapid expansion of financial markets, the demand for accurate forecasting of financial series has increased significantly. Accurate forecasting helps in predicting future market trends and making informed decisions.

Forecasting methods for financial series are based on identifying patterns in the data that can be distinguished from randomness. This process involves smoothing past values to eliminate randomness, enabling the inherent pattern to be used for forecasting.

However, obtaining high-frequency national accounts data can be challenging, and complicated methods are required to achieve disaggregated series that are compatible with annual totals. Therefore, there is a need for simpler techniques to obtain high-frequency data from low-frequency equivalents.

Machine learning (ML) algorithms are rapidly evolving, and feedforward artificial neural networks (ANNs) with appropriate training mechanisms have been proposed to temporally disaggregate economic series without considering related indicators. The ANNs algorithm is a promising method to disaggregate national statistical accounts, and this study proposes using it for this purpose.

In this study, an application of disaggregating annual Australia gross domestic product (GDP) data into quarterly one has been presented. The higher frequency data generated have been compared with the observed quarterly data to assess its accuracy. The comparative study suggests that the ANN-based model outperforms benchmark methods such as the Chow and Lin method and the Fernandez method.

The Australian economy is one of the largest in the world, and GDP data are closely monitored by economists and policymakers. The accuracy of GDP forecasting is crucial in determining economic policies and planning for the future. The study demonstrates that the ANNs algorithm can be an efficient tool in disaggregating economic series and generating high-frequency national accounts data.

However, the ANNs algorithm is not limited to the Australian economy alone. The efficient tool can be used for deriving high-frequency national accounts from available low-frequency ones for many other countries where national statistical institutes are

*Corresponding author: Sandip Garai, Division of Statistical Genetics, ICAR-Indian Agricultural Statistics Research Institute, India. Email: sandip.garai@icar.gov.in.

not advanced enough. This can significantly improve economic planning and decision-making, particularly in developing economies where the accuracy of economic data is crucial.

2. Literature Review

Econometrists require data at different time frequencies to make informed decisions and policies. However, data may not always be available at equal time intervals, and some data may be missing. Therefore, it is essential for national statistical institutes to have efficient tools to obtain well-distributed time series at higher frequencies from available lower frequency data.

Disintegration of time series variables involves estimating intra-period realizations with constraints, such as correlation structures and summing to original values. Two approaches have been taken: (a) using related indicators to estimate infra-annual data compatible with annual total and (b) using methods that do not involve related series.

Unobserved accounts can be derived using pure mathematical criteria or statistical models. Interpolation is used to estimate missing values within a series, typically for stock data. Distribution, on the other hand, is used to obtain high-frequency values from available lower frequency data, particularly for flow data such as GDP. Extrapolation involves obtaining values outside of the given period of a series. However, available related series can be useful in all three cases.

Friedman [1] noted that traditional interpolation methods tend to be less accurate when a single related economic series is used to estimate a particular value of another economic series. These methods typically rely on data from three time periods: the pre-, post-, and required dates of the related series, as well as two periods (pre- and post-dates) of the given series. To address this issue, Friedman proposed a new method that involves a single bivariate regression, taking into account the correlation structure between the two series in the regression equation. Years later, economists Boot and Feibes [2] proposed a direct approach to obtaining monthly accounts from yearly figures with fewer assumptions than other methods. They noted that this approach was more reliable and accurate than previous methods. Meanwhile, Boot et al. [3] proposed another method for predicting quarterly values from annual totals. Their approach was different from the procedure proposed by Lisman and Sandee [4], which relied on a more naïve mathematical approach.

In the meantime, Chow and Lin [5] proposed a unified best linear unbiased estimator to solve the problem of interpolation, extrapolation, and distribution of financial time series to estimate monthly series by using related indicators. However, this method may not work well in the presence of nonlinear trends or seasonal patterns. Silva and Cardoso [6] extended this work to a more flexible and dynamic approach to deal with stationary time series for disaggregation. For the adjustment of monthly or quarterly, i.e., infra-annual values obtained from any source to accord with the annual totals or averages acquired from another, Denton [7] formulated a quadratic constraint minimization problem. Fernández [8] extended Denton's approach to minimize the quadratic loss of obtained high-frequency series and linear combination of existing values. However, this method may not be suitable for handling nonlinear trends or seasonal patterns. In Fernandez's procedure, one need to find optimal filter to remove all serial correlation from quarterly residuals and often pre-filtering of the data is necessary. Therefore, Litterman [9] found Markov model was useful to distribute a time series. Several works were done to produce disaggregated data at higher frequency for multivariate level by polynomial method with related series.

Univariate polynomial method was proposed firstly by Almon in the book "The Craft of Economic Modeling" in 1988 for temporal disaggregation of univariate time series using relates variable [10]. However, this method also has some limitations, such as its inability to handle nonlinear trends and seasonal patterns. Later, Zaier and Trabelsi [11] modified and extended this work to multivariate case to deduce more than one high-frequency series using low-frequency counterparts without using a related series. Thereafter, Zaier and Trabelsi [12] introduced a two-stage method for temporal disaggregation of economic series. In the first stage, they used regression model to have preliminary estimates for disaggregated series using related indicator of same frequency. Second, they adjusted the newly obtained series using statistical benchmarking method [13] and autoregressive integrated moving average (ARIMA) method [14].

Stram and Wei [15] showed the use of generalized least square method to estimate disaggregated values from aggregated series instead of using ordinary least square approach to save lots of computations. Al-Osh [16] introduced dynamic linear modeling technique using Kalman filtering approach to produce basic series at higher frequency from available aggregates which were non-overlapping totals of disaggregates series. Wei and Stram [17] shed light on how a disaggregate model could be derived using a ARIMA aggregate model by exploring the estimate of plausible autocovariance structure of unobserved unaggregated values from autocovariance structure of observed aggregated series. Shen [18] took the help of variance autoregressive (VAR), Bayesian VAR approach, and ARIMA model to forecast quarterly macroeconomic series using monthly frequency data. Similarly, Miller and Chin [19] used quarterly and monthly series in combination to improve the forecast of quarterly data using VAR modeling approach. State-space method was utilized in multivariate model for temporal disaggregation by Moauro and Savio [20] and Proietti [21]. Schumacher and Breitung [22] proposed factor model for forecasting short-term growth of GDP using monthly and quarterly series. Mixed data sampling (MIDAS) regression is particularly helpful to relate high-frequency counterparts with future low-frequency series. Guay and Maurin [23] proposed a flexible approach to accommodate calendar effects in temporal disaggregation method using MIDAS approach.

2.1. Research gap

Temporal disaggregation and estimation of economic series are complex tasks that require careful consideration of various methods and potential issues. While these methods can be helpful, they also have limitations and potential problems that need to be taken into account to ensure accurate results.

One limitation of traditional interpolation methods is that they may not be accurate when using a single related economic series to estimate a value of another economic series. This can lead to inaccuracies in the temporal disaggregation of the series. Additionally, some methods make certain assumptions that may not always hold true, such as Boot and Feibes' method [2], which assumes that monthly accounts can be obtained from yearly figures with fewer assumptions than other methods. Complexity and computation requirements can also be a potential problem with some methods, such as the unified best linear unbiased estimator proposed by Chow and Lin [5, 24] and Proietti [21]. Other methods, such as Fernández approach [8, 21], require pre-filtering of data to remove serial correlation from quarterly residuals, which can add an additional step and potentially increase complexity. Disaggregating data at a higher frequency for multivariate levels using polynomial methods with related series can result in information loss and introduce biases. Similarly, ARIMA models, commonly used for forecasting in linear domain, can be complex and require careful selection of model parameters. Guay and Maurin's flexible approach may also require accommodation of calendar effects, adding complexity to the method [23]. It is important to note that the accuracy of temporal disaggregation methods can be affected by data availability, quality, frequency, and timing of collection. Therefore, it is crucial to consider these factors and choose appropriate methods to ensure continuity and accuracy in temporal disaggregation and estimation of economic series.

ANNs are a promising alternative method for temporal disaggregation of economic time series that can potentially overcome some of the issues associated with traditional mathematical methods. For instance, they can avoid heavy mathematical calculations required by some of the existing methods, making them more accessible and user-friendly. Additionally, unlike some of the traditional methods, they do not require the use of related indicators for the disaggregation, which can be advantageous when such indicators are not available or their use is deemed inappropriate.

Moreover, they can potentially capture complex nonlinear relationships between variables and provide flexible models that can be trained to adapt to changing data patterns. In this study, ANNs have been used to disaggregate Australia's annual GDP data to quarterly ones and thoroughly compared with the results obtained from Fernandez (f) and Chow and Lin (CL1) methods.

2.2. ANNs and comparative methods

Temporal disaggregation of economic time series has been an area of active research for decades. A plethora of methods have been proposed in the literature to solve this problem. The methods vary in terms of their assumptions, computational complexity, and accuracy. Most of them are not suitable to handle nonlinear trends and seasonality patterns. Recently, ANNs have been proposed as a promising alternative for temporal disaggregation of economic series. ANNs have emerged as a promising alternative for temporal disaggregation of economic series in recent years. Another advantage of ANN-based methods is their ability to handle complex relationships between variables. Traditional time series models often rely on linear relationships, which may not accurately capture the dynamics of the underlying data. ANNs can model nonlinear relationships and can identify patterns and dependencies that other methods may miss. In addition to these advantages, ANN-based methods also have some potential drawbacks. One of the main concerns is the potential for overfitting, where the model becomes too complex and starts to fit noise in the data rather than the underlying patterns. To address this concern, proper regularization techniques can be applied during training.

In summary, ANN-based methods represent a promising approach to temporal disaggregation of economic time series. They offer several advantages over traditional methods, including the ability to handle complex relationships and to learn from data without requiring a priori assumptions about the underlying patterns. While there are potential drawbacks to these methods, ongoing research and development in this area may help to address these concerns and further improve the accuracy and usefulness of ANN-based approaches.

In this article, a comprehensive overview of the existing methods for temporal disaggregation and comparative study of ANN with other methods in terms of their strengths and weaknesses have been provided.

3. Proposed ANN-Based Temporal Disaggregation Algorithm

ANNs are a powerful ML tool that can efficiently perform computational tasks with high accuracy while being cost-effective in terms of time [25–27]. ANNs work based on the principle of the "theory of learning" [28]. However, their "black box" nature, where the underlying functioning is unclear, has led to their neglect in the past. Nevertheless, if constructed properly, they can be a viable alternative to existing optimization tools.

ANNs learn from data and adjust their self-defined parameters within the network accordingly [29–31]. They can capture nonlinear patterns in the data [32–34]. Several learning mechanisms have been proposed for training the network architecture [35–38]. ANNs are different from other statistical or mathematical modeling approaches because the underlying functioning of the architecture and inherent data patterns can be ignored to obtain the desired result.

A method for temporal disaggregation of economic time series has been proposed here, which uses a feedforward neural network with the Levenberg–Marquardt (LM) backpropagation algorithm. In a feedforward network, the output from one neuron is forwarded to the next neuron for processing. The input and output neurons are determined based on the input and required output structure. The number of hidden layers and hidden nodes per layer is selected based on the specific problem during network training. It has been observed that the accuracy and generalization capability of the network decrease as the number of hidden units increases. Moreover, increasing the number of hidden units leads to more parameters and a more specific model for the given input and output combinations, which indicates poor generalization performance of the model for new observations (overfitting). In the current scenario, one input node is used to accommodate each yearly GDP data, Y_t $(t = 1, \ldots, n)$, where n denotes available yearly datapoints. Output layer consists of four output neurons each representing quarterly GDP.

3.1. Feedforward mechanism

z_t^0 denotes input to the network (Equation (1)). Chosen number of hidden units is generally which gives minimum training error.

$$z_t^0 = Y_t \ (t = 1, 2 \ldots, n). \tag{1}$$

The weighted sum of input, final weights being decided in adaptive manner till the end of training process, is fed to hidden neurons. Weight at connection from input to i^{th} hidden unit is W_i^H. Added bias with the sum is θ_i^H, for making the sum suitable for the activation function. Activation function (f_H) transforms this into desired output (z_i^H) (Equation (2)) and is supplied to output node. z_i^H is emitted from i^{th} hidden unit.

$$z_i^H = f_H(z^o W_i^H + \theta_i^H); \ i = 1, 2, \ldots. \tag{2}$$

Weighted sum of outputs from all hidden units is added with bias, θ_j^O (Equation (3)) and is transformed again through another activation function in output node. The final output from j^{th} output node is denoted by z_j^O (Equation (4)). Scaling factor has been utilized to each of the output so that they add to yearly GDP value of the corresponding year.

$$sum_j^O = \sum_i z_i^H W_{ji}^O + \theta_j^O. \tag{3}$$

$$z_j^O = f_O\left(sum_j^O\right)SF. \quad (4)$$

Weight between i^{th} hidden node and j^{th} output node is W_{ji}^O. Quarterly GDP data are denoted as $y_{t,j}$ ($t = 1, \ldots, n$ and $j = 1, \ldots, 4$), used in four neurons in output layer of the network. The hidden layer activation function is unipolar sigmoidal activation function (Equation (5)) which can take input in the range $(-\infty, +\infty)$. Function gives output in a range of $(0, +1)$.

$$f_H(x) = \frac{1}{1+e^{-x}}. \quad (5)$$

Activation function in output layer (f_o) is linear transfer function where output is proportional to the input.

3.2. Backpropagation

Backpropagation helps in fitting a network multiple number of times to get fine-tuned. Weight adjustments are the actual learning step completely dependent on this backpropagation algorithm. LM backpropagation algorithm [38] has been employed here for fine tuning temporal disaggregation model. Training error is generally calculated as the sum of squares of differences between actual and estimated values of the network. First derivatives of this error function with respect to weights and biases are used to calculate the Jacobian matrix, J. Hessian matrix, H, is a symmetric matrix of order $T*T$, where T indicates the number of weights and biases in the network (Equation (6)).

$$H = J^TJ = \begin{bmatrix} \frac{\partial^2 E_p}{\partial^2 w_{lk}} & \cdots & \frac{\partial^2 E_p}{\partial w_{lk} w_{m'l'}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 E_p}{\partial w_{m'l'} w_{lk}} & \cdots & \frac{\partial^2 E_p}{\partial^2 w_{m'l'}} \end{bmatrix}. \quad (6)$$

Gradients, g, of LM algorithm are calculated by Equation (7), where vector, e, contains errors of the network.

$$g = J^Te. \quad (7)$$

Training error at each iteration is depicted as E_p (Equation (9)). Total error (E) at training phase is the sum of errors at all training phase (Equation (8)).

$$E = \sum_p E_p. \quad (8)$$

$$E_p = \sum_j (e_j)^2 = \sum_j \left(z_j^O - y_{t,j}\right)^2. \quad (9)$$

Adjustment of weights and biases at k^{th} iteration in Gauss–Newton algorithm happens through Equations (10) and (11), respectively. Learning rate of the network, α, needs to be determined at the earliest stage. If the net input is a_i^k, sensitivity index, which is the change in performance of the network with respect to net input, is represented as σ_i^k at k^{th} iteration (Equation (12)). Sensitivities are propagated backward, and weights and biases are updated concurrently.

$$\nabla w_{ij}^k = -\alpha \frac{\partial E_p}{\partial w_{ij}^k} = -\alpha\sigma_i^k \frac{\partial a_i^k}{\partial w_{ij}^k}. \quad (10)$$

$$\nabla \theta_i^k = -\alpha \frac{\partial E_p}{\partial \theta_i^k} = -\alpha\sigma_i^k \frac{\partial a_i^k}{\partial \theta_i^k}. \quad (11)$$

$$\sigma_i^k = \frac{\partial E_p}{\partial a_i^k}. \quad (12)$$

Alternatively, weights and biases may be represented together using the vector x. Change of x is ∇x in LM algorithm (Equations (13) or (14)). The constant, λ, used in this algorithm is inversely proportional to step size. Null value of λ indicates Gauss–Newton's backpropagation. If error is increased in any step, this is multiplied with another factor (say, β) or is divided by the same in the opposite scenario.

$$\nabla \mathbf{x} = -\left(\nabla^2 \mathbf{E}_p\right)^{-1} \nabla \mathbf{E}_p = (H)^{-1}\mathbf{g}. \quad (13)$$

$$\nabla \mathbf{x} = -(\mathbf{J}^T\mathbf{J} + \lambda\, \mathbf{I})^{-1}\, \mathbf{J}^T e. \quad (14)$$

To train a moderate-sized network, LM algorithm is found very fast and efficient. It tries to shift toward Gauss–Newton's method as early as possible, i.e., it tries to minimize the error sooner. The added constant is reduced after each successful reduction of error and nullified for minimum possible error as Gauss–Newton's method performs well in a minimum error environment.

4. Performance Measure

Four indicators are used to evaluate the performance of the proposed tool to be used to disaggregate economic series. Here, actual and predicted quarter on quarter changes in GDP has been represented as a_i and p_i, respectively. Total number of observations is mentioned as N.

4.1. Theil's U-Statistic

Theil's U-Statistic (U) [39] is described as in Equation (15).

$$U = \frac{\sqrt{N^{-1}\sum_{i=1}^N (p_i - a_i)^2}}{\sqrt{N^{-1}\sum_{i=1}^N {p_i}^2} + \sqrt{N^{-1}\sum_{i=1}^N {a_i}^2}}. \quad (15)$$

Range of U-Statistic is $0 \le U \le 1$. It is desired to be smaller.

4.2. Mean absolute error

Mean absolute error (MAE) is the mean of the absolute differences between actual and disaggregated values (Equation (16))

$$MAE = N^{-1}\sum_{i=1}^N |p_i - a_i|. \quad (16)$$

4.3. Root mean square error

Root Mean Square Error is represented ($RMSE$) as the squared root of mean of the squared differences between actual and disaggregated values (Equation (17))

$$RMSE = \sqrt{N^{-1}\sum_{i=1}^N (p_i - a_i)^2}. \quad (17)$$

RMSE and MAE values are expected to be lesser.

Table 1
Descriptive statistics of GDP data

Statistic	Annual GDP	Quarterly GDP
Minimum	−1.332	−2.017
1st quartile	2.492	0.305
Median	3.195	0.775
Mean	3.446	0.851
3rd quartile	4.434	1.382
Maximum	7.192	4.423
SD	1.780	1.033
CV	51.654	121.387

4.4. Signal-to-noise ratio

The signal-to-noise (SN) ratio is measured as the ratio of actual values (signal) to the sum of squared differences between the predicted and the observed values (noise) (Equation (18))

$$SN\ Ratio = \frac{\sum_{i=1}^{N} {a_i}^2}{\sum_{i=1}^{N} (p_i - a_i)^2}. \tag{18}$$

Higher the SN ratio, better the model is.

5. Application and Results

5.1. Data description

The Australian Bureau of Statistics provides data on the country's GDP, and 58 years of seasonally adjusted (SA) GDP data from 1961 to 2018 has been obtained. Table 1 provides a summary of the annual and quarterly SA GDP data, which are expressed as percentages (%). Figures 1 and 2 show the annual and quarterly GDP data, respectively. The data reveal that the minimum annual growth rate in GDP is −1.332%, while the maximum growth rate is 7.192%. Similarly, the minimum quarter-on-quarter growth rate is −2.017%, while the maximum growth rate is 4.423%. Furthermore, it is evident that quarterly growth rates are more erratic than annual growth rates.

5.2. Model building and validation

The success of training the ANN for the annual and quarterly GDP data relies heavily on the optimal architecture setup. To achieve this, the number of hidden neurons and iterations was adjusted to achieve minimal error and network convergence during training. The findings showed that using four hidden neurons with a single hidden layer produced a simpler and more efficient model. Therefore, the ANN architecture with four hidden neurons and one hidden layer was deemed the best fit for the current scenario. The network was trained 500 times with a learning rate of 0.05, using the first 50 years of data for training and the last 8 years for validation purposes (more the number of data points, more accurate results are expected). After training, the optimal model was used for predicting test data.

A comparative study was conducted, including different methods such as "CL1" and "f" method. The U, MAE, RMSE, and SN ratio were compared as described in the previous section. Figure 3 depicts line charts of prediction performance of three models at the testing phase. The results show that the ANN-based model outperformed the other two models in temporal disaggregation efficiency.

Figure 1
Australia annual GDP data

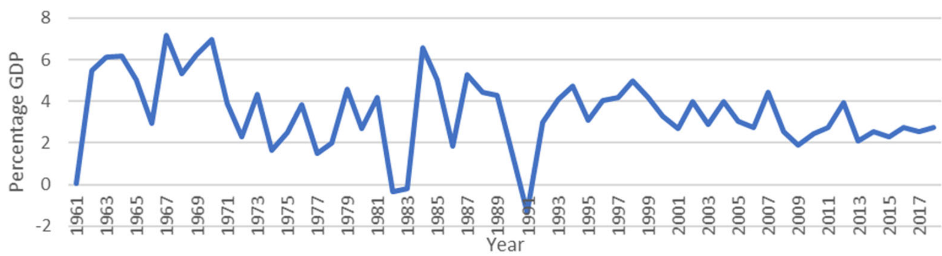


Figure 2
Australia quarterly GDP data

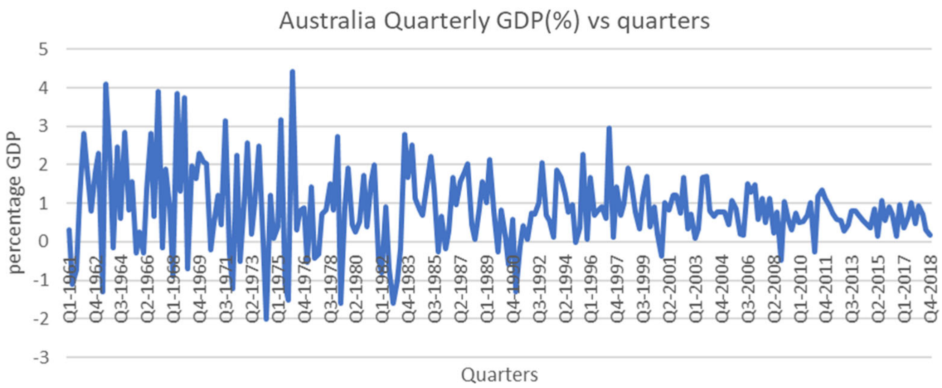
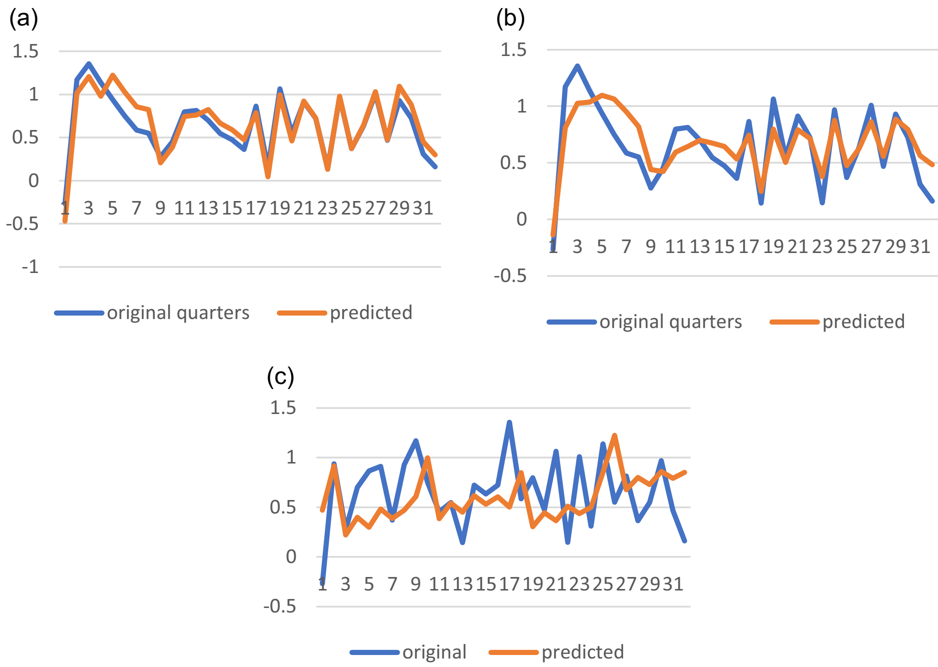


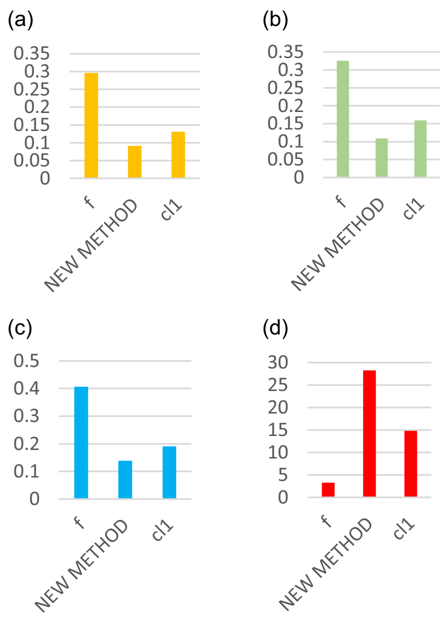
Figure 3
Original vs predicted quarterly GDP values of validation set by (a) ANN disaggregation method, (b) f method, and (c) CL1method



In Figure 4, different comparative measures are presented. The "NEW METHOD" in Figure 4 depicts the ANN-based disaggregation model. The proposed method was found to be the best approach to disaggregating Australia's GDP data, as evidenced by the results. SN ratio is the highest and other metrics are the lowest for the "NEW METHOD" which is desired for a model to be considered as a good predictor.

In summary, the study demonstrated that an ANN architecture with four hidden neurons and one hidden layer could efficiently perform temporal disaggregation for Australia's GDP data. The proposed method outperformed other approaches in terms of U, MAE, RMSE, and SN ratio.

Figure 4
Comparison of (a) U-Statistic, (b) MAE, (c) RMSE, and (d) SN ratio at the validation phase for ANN (NEW METHOD), f, and CL1 methods



6. Conclusion

Traditional interpolation methods for estimating economic series can be inaccurate and require complex computations that are time-consuming. ANNs provide a promising alternative for temporal disaggregation of economic time series as they are more accessible and user-friendly, require less mathematical calculations, and do not depend on related indicators for disaggregation.

This study proposes an ML-based approach for temporal disaggregation of financial series without related indicators, using Australia's GDP data as an example. The results show that the ANN-based model outperforms traditional methods in disaggregating the annual GDP data into quarterly ones. This ML-based approach can be a useful tool for deriving high-frequency national accounts from available low-frequency ones, particularly in developing economies where accurate data and forecasting are crucial for economic planning and decision-making.

The study highlights the importance of accurate data and forecasting in economic planning and the need for simpler techniques to obtain high-frequency data from low-frequency equivalents. Overall, the study shows the potential of ML techniques in economic forecasting and analysis.

Recommendations

Based on the findings of this study, it is recommended that statistical institutes and other organizations involved in economic planning consider the use of ML techniques, particularly ANNs, in temporal disaggregation of financial series. This approach can provide a more efficient and accurate alternative to existing methods, particularly in cases where related indicators are not available or pre-filtering of data is required. It is also recommended to further explore the potential of ML techniques in economic forecasting and analysis, particularly in developing economies where the accuracy of economic data is crucial for decision-making. Finally, the study highlights the need for simpler techniques to obtain high-frequency data from low-frequency equivalents, which can significantly improve economic planning and decision-making.

Acknowledgement

The authors are very thankful to the editorial team for considering this work. The authors are also very thankful to anonymous reviewers for their valuable comments which helped a lot to improve the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in [Australian Bureau of Statistics] at https://www.abs.gov.au/statistics/economy/national-accounts/australian-national-accounts-national-income-expenditure-and-product/latest-release.

References

[1] Friedman, M. (1962). The interpolation of time series by related series. *Journal of the American Statistical Association*, *57*(300), 729–757.
[2] Boot, J., & Feibes, W. (1967). On Glejser's derivation of monthly figures from yearly data. *Brussels Economic Review*, *36*, 589–596.
[3] Boot, J. C. G., Feibes, W., & Lisman, J. H. C. (1967). Further methods of derivation of quarterly figures from annual data. *Journal of the Royal Statistical Society. Series C* (Applied Statistics), *16*(1), 65–75. https://doi.org/10.2307/2985238
[4] Lisman, J. H. C., & Sandee, J. (1964). Derivation of quarterly figures from annual data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *13*(2), 87–90. https://doi.org/10.2307/2985700
[5] Chow, G. C., & Lin, A. L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, *53*(4), 372–375. https://doi.org/10.2307/1928739
[6] Silva, J. M. C. S., & Cardoso, F. N. (2001). The Chow-Lin method using dynamic models. *Economic Modelling*, *18*(2), 269–280. https://doi.org/10.1016/S0264-9993(00)00039-0
[7] Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, *66*(333), 99–102. https://doi.org/10.1080/01621459.1971.10482227
[8] Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, *63*(3), 471–476. https://doi.org/10.2307/1924371
[9] Litterman, R. B. (1983). A random walk, Markov model for the distribution of time series. *Journal of Business & Economic Statistics*, *1*(2), 169–173. https://doi.org/10.1080/07350015.1983.10509336
[10] Almon, C. (2017). *The craft of economic modeling*. USA: CreateSpace.
[11] Zaier, L. H., & Trabelsi, A. (2007). A polynomial method for temporal disaggregation of multivariate time series. *Communications in Statistics – Simulation and Computation*, *36*(3), 741–759. https://doi.org/10.1080/03610910601096296
[12] Zaier, L. H., & Trabelsi, A. (2009). A benchmarking approach to temporal disaggregation of economic time series by related series. *Communications in Statistics – Theory and Methods*, *38*(19), 3561–3582. https://doi.org/10.1080/03610920802604194
[13] Hillmer, S. C., & Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of the American Statistical Association*, *82*(400), 1064–1071. https://doi.org/10.1080/01621459.1987.10478541
[14] Hillmer, S. C., & Tiao, G. C. (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, *77*(377), 63–70. https://doi.org/10.1080/01621459.1982.10477767
[15] Stram, D. O., & Wei, W. W. S. (1986). A methodological note on the disaggregation of time series totals. *Journal of Time Series Analysis*, *7*(4), 293–302. https://doi.org/10.1111/j.1467-9892.1986.tb00496.x
[16] Al-Osh, M. (1989). A dynamic linear model approach for disaggregating time series data. *Journal of Forecasting*, *8*(2), 85–96. https://doi.org/10.1002/for.3980080203
[17] Wei, W. W. S., & Stram, D. O. (1990). Disaggregation of time series models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 453–467. https://doi.org/10.1111/j.2517-6161.1990.tb01799.x
[18] Shen, C. H. (1996). Forecasting macroeconomic variables using data of different periodicities. *International Journal of Forecasting*, *12*(2), 269–282. https://doi.org/10.1016/0169-2070(95)00659-1
[19] Miller, P. J., & Chin, D. M. (1996). Using monthly data to improve quarterly model forecasts. *Federal Reserve Bank of Minneapolis Quarterly Review*, *20*(2), 16–33.
[20] Moauro, F., & Savio, G. (2005). Temporal disaggregation using multivariate structural time series models. *The Econometrics Journal*, *8*(2), 214–234. https://doi.org/10.1111/j.1368-423X.2005.00161.x
[21] Proietti, T. (2006). Temporal disaggregation by state space methods: Dynamic regression methods revisited. *The Econometrics Journal*, *9*(3), 357–372. https://doi.org/10.1111/j.1368-423X.2006.00189.x
[22] Schumacher, C., & Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, *24*(3), 386–398. https://doi.org/10.1016/j.ijforecast.2008.03.008
[23] Guay, A., & Maurin, A. (2015). Disaggregation methods based on MIDAS regression. *Economic Modelling*, *50*, 123–129. https://doi.org/10.1016/j.econmod.2015.05.013
[24] Chow, G. C., & Lin, A. L. (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association*, *71*(355), 719–721. https://doi.org/10.1080/01621459.1976.10481554
[25] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0
[26] Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD Thesis, Harvard University.
[27] Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, *1*(4), 339–356. https://doi.org/10.1016/0893-6080(88)90007-X
[28] Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*, *27*(11), 1134–1142.

[29] Bonelli, M. G., Ferrini, M., & Manni, A. (2017). Artificial neural networks to evaluate organic and inorganic contamination in agricultural soils. *Chemosphere*, *186*, 124–131. https://doi.org/10.1016/j.chemosphere.2017.07.116

[30] Hannah Jessie Rani, R., & Aruldoss Albert Victoire, T. (2019). A hybrid Elman recurrent neural network, group search optimization, and refined VMD-based framework for multi-step ahead electricity price forecasting. *Soft Computing*, *23*(18), 8413–8434. https://doi.org/10.1007/s00500-019-04161-6

[31] Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

[32] Garai, S., & Paul, R. K. (2023). Development of MCS based-ensemble models using CEEMDAN decomposition and machine intelligence. *Intelligent Systems with Applications*, *18*, 200202. https://doi.org/10.1016/j.iswa.2023.200202

[33] Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Paul, A. K., Roy, H. S., . . . , & Manjunatha, B. (2023). An MRA based MLR model for forecasting Indian annual rainfall using large scale climate indices. *International Journal of Environment and Climate Change*, *13*(5), 137–150.

[34] Paul, R. K., & Garai, S. (2022). Wavelets based artificial neural network technique for forecasting agricultural prices. *Journal of the Indian Society for Probability and Statistics*, *23*(1), 47–61. https://doi.org/10.1007/s41096-022-00128-3

[35] Anderson, J. A. (1995). *An introduction to neural networks*. USA: MIT Press.

[36] Freeman, J. A., & Skapura, D. M. (1991). *Neural networks: Algorithms, applications, and programming techniques (computation and neural systems series)*. USA: Addison-Wesley.

[37] Hagan, M. T., Demuth, H. B., Beale, M. H., & de Jesús, O. (1996). *Neural network design*. UK: PWS Publishing.

[38] Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, *5*(6), 989–993. https://doi.org/10.1109/72.329697

[39] Theil, H. (1958). *Economic forecasts and policy.* Netherlands: North Holland Publishing.

How to Cite: Garai, S., Kumar Paul, R., Kumar, M., & Choudhury, A. (2024). Intra-Annual National Statistical Accounts Based on Machine Learning Algorithm. *Journal of Data Science and Intelligent Systems*, *2*(3), 153–160. https://doi.org/10.47852/bonviewJDSIS3202870