 doi:10.20944/preprints202110.0049.v1

Article

Minimum Message Length in Hybrid ARMA and LSTM Model Forecasting

Zheng Fang [1,†], David L. Dowe [2*], Shelton Peiris [3], and Dedi Rosadi [4]

[1] Department of Artificial Intelligence and Data Science, Monash University, Clayton, Victoria 3168, Australia. Email: zfan51@student.monash.edu;
[2] Department of Artificial Intelligence and Data Science, Monash University, Clayton, Victoria 3168, Australia.
[3] School of Mathematics and Statistics, University of Sydney, Camperdown, NSW 2006, Australia.
[3] Department of Statistics, Gadjah Mada University, Sleman, Yogyakarta 55500, Indonesia.
* Correspondence: david.dowe@monash.edu
† These authors contributed equally to this work.

Abstract: We investigate the power of time series analysis based on a variety of information-theoretic approaches from statistics (AIC, BIC) and machine learning (Minimum Message Length) - and we then compare their efficacy with traditional time series model and with hybrids involving deep learning. More specifically, we develop AIC, BIC and Minimum Message Length (MML) ARMA (autoregressive moving average) time series models - with this Bayesian information-theoretic MML ARMA modelling already being new work. We then study deep learning based algorithms in time series forecasting, using Long Short Term Memory (LSTM), and we then combine this with the ARMA modelling to produce a hybrid ARMA-LSTM prediction. Part of the purpose of the use of LSTM is to seek capture any hidden information in the residuals left from the traditional ARMA model. We show that MML not only outperforms earlier statistical approaches to ARMA modelling, but we further show that the hybrid MML ARMA-LSTM models outperform both ARMA models and LSTM models.

Keywords: long short-term memory, minimum message length, time series, neural network, deep learning, Bayesian statistics, probabilistic modeling

1. Introduction

Forecasting in time series is difficult in practice due to the presence of trends and/or seasonal components. For example, economic time series data are highly impacted by seasonal factors and often show long run cycles. Such trends and seasonality are difficult to capture by the traditional ARIMA (autoregressive integrated moving average) model [1]. The ARIMA model generally use for the stationary time series or differencing with integer order in order to have the stationery. Using more than 1 integer of differencing order may distort the seasonal or trend factor [2]. This is the motivation for the more general ARFIMA and Seasonal ARIMA (SARIMA) models, respectively built to include fractional differencing and an explicit seasonal factor. But the deep learning LSTM (long short-term memory) technique might be more suitable to capture the information that is less obvious in the time series, as it allows for a much more general model class. The time series science community takes much effort to discover the appropriate model in order to identify time dependency in time series data [3]. Historically, the ARMA model was introduced by Box and Jenkins in 1976, and is popular and widely use in the time series science community and provides accurate forecasts in both in-sample and out-sample data when it correctly selects the relevant parameters [4]. It is a hybrid of the autoregressive (AR) and moving average (MA) processes but the ARMA model can only be used in the stationary time series - so the ARIMA (autoregressive integrated moving average) model was introduced to allow the integer differencing, aiming to make the time series data stationary from the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

publication Time Series Analysis: Forecasting and Control by Box and Jenkins (1970)[1] [5]. In parallel, machine learning has seen the development of neural network models in computer science, ultimately influencing statistics. Similar to the families of the ARMA model, the deep learning also including serval of several variants such as DNN, CNN, and RNN. This report is investigating the particular form of RNN which is called Long Short Term Memory (LSTM), which are typically used for time series [6]. In recent years, LSTM has been shown to work well in the forecast for serial data with complex time dependency, with the LSTM-based model being widely used in analyzing (e.g.) stock market and energy consumption prediction [7]. On the other hand, the Bayesian Minimum Message Length (MML) principle [8] and the Akaike Information Criterion (AIC), Bayesian information criterion (BIC) are based in information theory [9,11,19].

We use the information-theoretic Minimum Message Length (MML), AIC and BIC to select the model orders of ARMA(p, q), and we then train the LSTM model by the residuals left from the ARMA model [12,13]. Most other papers using this kind of hybrid model use information-theoretic model selection techniques such as AIC and BIC. However, our results show that MML compares favorably with these approaches when doing ARMA alone - and this is a new contribution of our paper. We also compare the ARMA-LSTM selected by MML with the ARMA model alone, and our results also show that MML performs well in this kind of time series model. Our results further show that the Bayesian information-theoretic MML principle provides more reliable and high accurate results in the model selection of hybrid model ARMA-LSTM.

Section 2 introduces the Box and Jenkins theory for the ARIMA model and discusses its limitations. Section 3 introduces the information-theoretic Minimum Message Length criterion in model selection, and Section 4 introduces the deep learning model LSTM. Section 5 provides the algorithm of the hybrid ARMA-LSTM model, and Section 6 provides the experimental results with comparison.

2. ARIMA Modelling

This section reviews the theory of Autoregressive Integrated Moving Average Model (ARIMA) modelling due to Box and Jenkins (1970) [5,14,18].

Let $\{Y_t\}$ be a homogeneous nonstationary time series and suppose that the d^{th}, $d = 1, 2 \ldots$ differencing of the series is stationary and is given by $X_t = (1-B)^d Y_t$, where B is the backshift operator. Then a stationary ARMA(p,q) model can be fitted for $\{X_t\}$ satisfying

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{1}$$

where $\{\epsilon_t\} \sim WN(0, \sigma^2)$.

Let $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$; $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$, are two polynomials of degree p and q respectively such that the zeros of $\phi(B)$ and $\theta(B)$ are outside the unit circle. Then the ARMA(p,q) in (1) can be written in a compact form as

$$\phi(B) X_t = c + \theta(B)\epsilon_t. \tag{2}$$

Now the corresponding ARIMA(p,d,q) model for the original series $\{Y_t\}$ is given by

$$\phi(B)(1-B)^d Y_t = c + \theta(B)\epsilon_t. \tag{3}$$

It is known that ARIMA is a form of a linear regression model with the lag order of time series data and corresponding residuals. In an environment where the ARIMA model fits well for the given data, then the corresponding residuals through the model should form a random scatter plot with a constant mean and variance over the time, see, for example, [18]. If the ARIMA model is not well fitted for the data or an incorrect model has

[1] The book by Box, Jenkins, and Reinsel (1994) is an updated version of Box Jenkins (1970) by adding the outlier detection and unit roots test

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

been fitted, then the residuals will not show a random scatter plot and instead indicate autocorrelations within the residuals. This concludes that the information hidden in the data has not been completely captured by the fitted ARIMA model and consider refitting an alternative ARIMA model [26].

The above family of ARIMA models are also capable of modelling a wide range of seasonal data with slight modifications. A seasonal extension of the model (3) can be written for a set of time series data with seasonality m. Incorporating both the seasonal and nonseasonal components with additional polynomials, the new model is

$$\phi(B)\Phi(B^m)(1-B)^d(1-B^m)^D Y_t = c + \theta(B)\Theta(B^m)\epsilon_t, \tag{4}$$

$\Phi(B^m) = 1 - \Phi_1 B^m - \ldots - \Phi_P B^{mP}$, $\Theta(B^m) = 1 + \Theta_1 B^m + \ldots + \Theta_Q B^{mQ}$ and D is the degree of seasonal differencing. For simplicity, this is written as

$$Y_t \sim SARIMA(p,d,q)(P,D,Q)_m \tag{5}$$

The corresponding model in (4) is known as the Seasonal ARIMA or SARIMA model.

To estimate the parameters of the SARIMA model in (4), it is important to identify changes of variance in the autocorrelation plot (ACF) of data. This ACF provides an indication of linear dependencies between the observation of time series value, which is related to the order of the model. In addition, the corresponding partial autocorrelation function (PACF) can be used to confirm the approximate order required in the model.

In this study we use only non-seasonal ARIMA modelling. As the non-seasonal degree of differencing d can be predetermined in practice, we simulate the stationary time series data.

Assuming the data is generated from a mean zero stationary ARMA(p,q) process with Gaussian errors, we use the facxt that the distribution of data is multivariate Gaussian distribution with mean $\mu = 0$.

Suppose that we have N observations $y = (y_1, ..., y_N)$ generated through the model in (2) with $c = 0$ and let **fi** is the vector of all the parameters. Then the corresponding unconditional log-likelihood function, $L(y|\beta)$ can be given as:

$$L(y|\beta) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Sigma| - \frac{1}{2\sigma^2}y^T\Sigma^{-1}y, \tag{6}$$

where $|\Sigma|$ is the determinant of Σ and $\sigma^2\Sigma$ is an $n \times n$ theoretical autocovariance matrix of y.

3. Minimum Message Length

The Bayesian information-theoretic Minimum Message Length (MML) principle [8,9, 13,15] is based on coding theory, and can be thought of in several equivalent ways. It can be thought of in terms of a transmitter encoding a two-part message and transmitting it to a receiver, where the first part of the message contains information encoding the model and the second part of the message encodes the data given the model. The length of the first part of the message can be thought of as the complexity of the model, and the length of the second part of the message (effectively, the statistical negative log-likelihood) is a measure of goodness of fit to the observed data. For example, with $X = \{A, B, C, D\}$, possible encodings would be (e.g.) $A = 00, B = 01, C = 10, D = 11$ or instead (e.g.) $A = 1, B = 01, C = 001, D = 0001$, with the length of code represented as $I()$ - e.g., with $A = 00$, $I(A) = 2$. The code length is typically (close to) the negative logarithm of the probability.

MML thus gives a quantitative information-theoretic trade-off between model complexity (length of first part of message) and goodness of fit (length of second part of message) [20]. A smaller MML value (or, equivalently, a shorter message length) indicates the model is less complex and highly fitted to the data [19]. In practice, minimizing the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

message length can be expressed as:

$$\arg\min_{\theta\in\Theta} \{\mathrm{I}(\theta) + \mathrm{I}(y^n|\theta)\}, \tag{7}$$

where $I(\theta)$ encodes the **assertion** (or model) and $I(y^n|\theta)$ encodes the **detail** (or data given the model). In MML, there is prior knowledge of π over the parameter space so the MML is part of the Bayesian approach. Following Wallace and Freeman [15], MML87 is extended version of MML, it has been shown to work well in time series model such as autoregressive model (AR) and moving average model (MA) [12,16]. Assuming the ARMA model parameters are given by $\beta = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \sigma^2$, following MML87, we seek the parameter values in order to minimize the message length is:

$$MessLen(\beta, y) = -\log(\frac{h(\beta)f(y_1, ..., y_n|\beta)\epsilon^n}{\sqrt{|F(\beta)|}}) + \frac{k}{2}(1 + \log\kappa_k) - \log h(k), \tag{8}$$

where ϵ is measuring the accuracy of data, $h(k)$ is the prior on the number of parameters, $h(\beta)$ is the Bayesian prior distribution over the parameter set, $f(y_1, ..., y_n|\beta)$ is standard statistical likelihood function, $F(\beta)$ is the expected Fisher Information matrix of the parameter set β, κ_k is the lattice constant (which accounts for the expected error in the log-likelihood function from ARMA model equation 6 due to quantization of the n-dimensional space, it is bounded above by $\frac{1}{12}$ and bounded below by $\frac{1}{2\pi e}$. For example, $\kappa_1 = \frac{1}{12}$, $\kappa_2 = \frac{5}{36\sqrt{3}}$, $\kappa_3 = \frac{19}{192*2^{1/3}}$, and $\kappa_n \to \frac{1}{2\pi e}$ when $n \to \infty$).

The MML87 message length for ARMA model parameter β can also be represented as:

$$I(y, \beta) = -\log\pi(\beta) + \frac{1}{2}\log|F(\beta)| + \frac{k}{2}\log\kappa_k + \frac{k}{2} - \log f(y|\beta) \tag{9}$$

The MML87 is model invariant and avoid explicitly constructing the quantized parameter space [11]. MML87 is used for model selection and parameter estimation by choosing the model with minimize the message length.

4. Long short-term memory (LSTM)

With the development of computational power in electronic equipment, powerful computers provide many learning algorithms and approaches in time series forecasting [21–23]. The deep learning is one of the popular approaches in recent years, it provides a complex model that is able to capture the information from the predictors than the traditional model. Long Short-Term Memory is special kind of Recurrent Neural Network introduced by Hochreiter and Schmidhuber in 1997 [17] . LSTM manages the two state vectors, the short term state h_t and long term state c_t and using the gating mechanism, by adding linear component from the previous layer in order to provide the long memory. LSTM has been widely use in the time series forecasting, because it able to capture more information in the time series data, particularly for the financial econometrics area where the price of financial assets are depends on varies of different factors that are difficult to represented by linear model [23,24]. Each LSTM layer including the cells of forgot gate, input gate, and output gate shown in the Figure 1 [2].

- Forget gate: $f_t = \sigma(U^f x_t + W^f h_{t-1} + b^f)$
- Input gate: $i_t = \sigma(U^i x_t + W^i h_{t-1} + b^i)$
- Output gate: $o_t = \sigma(U^o x_t + W^o h_{t-1} + b^o)$

The forget gate is using a sigmoid function $\sigma(x)$ from equation 11 which having the value between 0 and 1, it determine how much information should be forgot. If the result from

[2] Source: Dinh Phung, Department of Artificial Intelligence and Data Science, Monash University, Clayton, Victoria 3168, Australia

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

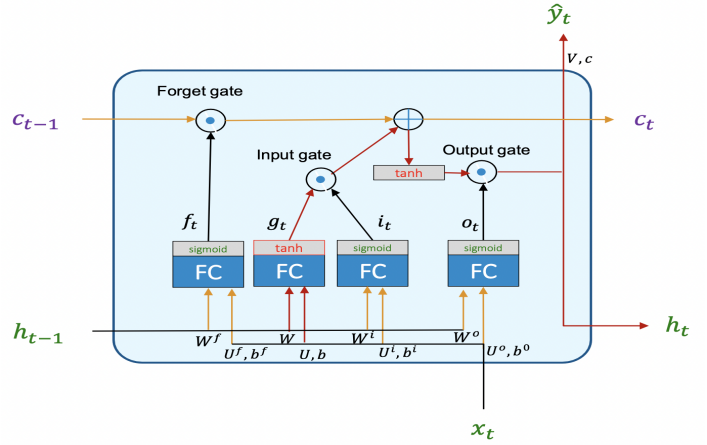

Figure 1. LSTM Structure

sigmoid function close to 0, it means the more information should be forgot, versa vice.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{10}$$

The input gate also uses the sigmoid function, the input gate control the value input from the input function of $g_t = tanh(Wh_{t-1} + Ux_t + b)$ with using $tanh(x)$ function:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{11}$$

The input gate control how much information should be remember. LSTM long-term state is element wise operation with $c_t = f_t \odot c_{t-1} + g_t \odot i_t$. Output gate o_t is control how much

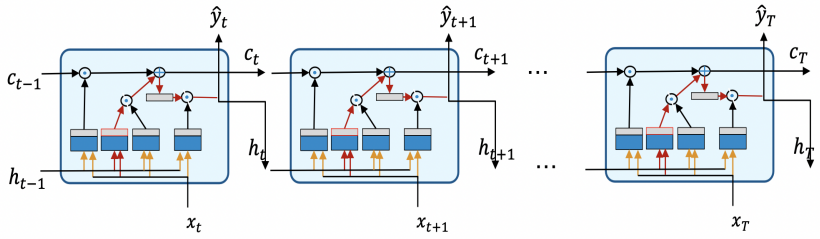

Figure 2. LSTM Overlapping

long term information c_t should be carried forward to the next layer and it also contribute to the short term state of h_t. The result from output gate function is also between 0 and 1, and LSTM short term state also element wise operation with $h_t = o_t \odot tanh(c_t)$. Overlapping more than one layer of LSTM shown in the Figure 2 [3], and makes LSTM provide time series model to capture long and short term information in order to forecast. As usual, the LSTM same with other neural network which trained by the back propagation. LSTM requires $N >= 1$ time steps sequence data to train the model, its timing information will be modeled and characterized to deep representation.

5. Hybrid ARIMA-LSTM model

In recent year, LSTM and its variation along with some hybrid models dominate the financial time series forecasting domain [23]. The LSTM is able to capture the dependency of residuals across the time, and the LSTM is trained by the time step [25]. In this paper, we are using Moving Average lag order q from ARMA parameters selected MML87, AIC, and BIC, if $q = 0$ then only use ARMA to forecast the time series data without LSTM. Our

[3] Source: Dinh Phung, Department of Artificial Intelligence and Data Science, Monash University, Clayton, Victoria 3168, Australia

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

LSTM model is composed of a single input layer with input share of MA order and the sequence learning features. The following LSTM layer also contains the sequence learning features, and the third LSTM layer with the same unit following by the fourth dense layer with one unit.

Algorithm 1 Algorithm with LSTM Model

Require: number of epoches $= 10$
  while MA(q) order in order set selected by MML, AIC **do**
    model.add(LSTM(30, return_sequences=True, input_shape=$(q, 1)$))
    model.add(LSTM(30, return_sequences=True))
    model.add(LSTM(30))
    model.add(Dense(1))
  end while

The hybrid model ARMA-LSTM is training the LSTM model by the residuals from the ARMA model. In this paper, MML87, AIC, and BIC have been used to select the model parameter orders from the ARMA, so this paper does not only compare the errors by hybrid model and single ARMA model but also compare the hybrid model by the selection of MML87, AIC, and BIC. The forecast from the ARMA model is the fitted mean μ_{t+1}, and also because the information hidden in residual from ARMA model, so the forecast of hybrid model will be

$$\hat{Y}_{t+1} = \mu_{t+1} + \mathrm{E}_{t+1} \tag{12}$$

where μ_{t+1} represent the linearity modeling of data from ARMA model selected according to the information-theoretic MML87 and AIC. The ϵ_t is the residual left by the ARMA model $Y_t - \hat{Y}_t$, and $\mathrm{E}_{t+1} = f(\epsilon_t) = f(Y_t - \hat{Y}_t)$, which is forecasted by the LSTM based on the pass residuals value $\epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-q}$, where the parameter q selected by the MML87, AIC, and BIC. The hybrid ARMA-LSTM model combine both linear and non-linear tendencies in time series data [30].

The algorithm of the hybrid model is shown below:

Algorithm 2 Algorithm with Hybrid Model ARMA-LSTM

Require: number of data $n \geq 0$
  while $N \leq$ number of different simulations **do**
    while $n \leq$ number of dataset in simulation **do**
      while $i \in$ MA orders selected from MML, AIC **do**
        if $i \neq 0$ **then**
          Train LSTM model by the residuals of ARMA model
          Rolling forecast the residual by LSTM
          Calculate root mean squared error by Y_{t+1}
        else if $i = 0$ **then**
          Calculate root mean squared error by forecast from ARMA only
        end if
      end while
    end while
  end while

6. Experiments

The purpose of this experiment part is designed to compare the results of the ARMA model itself with the hybrid ARMA-LSTM model and also compare with the hybrid model that the parameters selected by the MML87, AIC, and BIC. In order to analyze the accuracy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

of forecasting, we are using the Root Mean Squared Error, RMSE $= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_i)^2}$ to compare the different results, where T stands for the forecast window size.

6.1. Simulated Dataset

In this section, we use uniform distribution from minimum -0.9 to maximum 0.9 to randomize the parameters ARMA(p, q) for the data simulation. There are 10 different parameter sets from $p_1, ..., p_5$ and $q_1, ..., q_2$. The values in the table are the average of RMSE in particular parameters simulated dataset. The dataset including N = 50, 100, 200, 300, 500 time series data point in one data set and also include forecast windows T = 3 , 10, 30 and 50. Each value in Table 1 is the average of RMSE of forecast errors over the datasets. The bold texts indicate the smallest forecast errors from the different kinds of models. Table 1 to Table 3 provides the comparison for different sizes of forecast window with T = 3, 10, and 30.

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	**0.982** (0.646)	1.108 (0.529)	1.033 (0.469)	**1.204** (0.308)	1.217 (0.502)	1.234 (0.7)
p_1, q_2	1.133 (0.592)	**1.053** (0.669)	1.166 (0.635)	1.301 (0.922)	**1.289** (0.95)	1.384 (0.838)
p_2, q_1	1.027 (0.423)	1.024 (0.421)	**1.023** (0.418)	1.025 (0.408)	1.012 (0.376)	**1.005** (0.48)
p_2, q_2	1.333 (0.793)	1.278 (0.841)	**1.271** (0.848)	1.241 (0.745)	**1.182** (0.711)	1.194 (0.674)
p_3, q_1	0.955 (0.377)	0.956 (0.377)	**0.944** (0.37)	**0.965** (0.341)	0.975 (0.35)	0.986 (0.426)
p_3, q_2	1.293 (0.331)	1.241 (0.296)	**1.238** (0.296)	1.114 (0.284)	1.211 (0.266)	**1.105** (0.259)
p_4, q_1	0.901 (0.483)	0.916 (0.448)	**0.871** (0.398)	0.948 (0.397)	0.944 (0.41)	**0.932** (0.442)
p_4, q_2	1.207 (0.539)	1.226 (0.515)	**1.206** (0.513)	1.252 (0.777)	1.261 (0.778)	**1.251** (0.772)
p_5, q_1	1.006 (0.54)	0.907 (0.626)	**0.903** (0.578)	1.122 (0.538)	1.117 (0.553)	**1.018** (0.467)
p_5, q_2	**1.026** (0.583)	1.052 (0.553)	1.061 (0.559)	1.042 (0.559)	**1.021** (0.592)	1.046 (0.53)

Table 1: Dataset for N = 100 & T = 3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.234 (0.178)	**1.208** (0.165)	1.221 (0.293)	**1.201** (0.404)	1.132 (0.184)	1.138 (0.317)
p_1, q_2	1.571 (0.375)	1.553 (0.386)	**1.398** (0.304)	1.555 (1.109)	1.549 (1.125)	**1.494** (0.834)
p_2, q_1	**1.025** (0.194)	1.041 (0.203)	1.043 (0.193)	**1.013** (0.174)	1.02 (0.182)	1.037 (0.265)
p_2, q_2	1.353 (0.438)	1.327 (0.373)	**1.325** (0.368)	1.274 (0.213)	1.257 (0.206)	**1.255** (0.205)
p_3, q_1	0.947 (0.194)	**0.895** (0.129)	0.901 (0.134)	1.018 (0.135)	**0.946** (0.116)	0.989 (0.154)
p_3, q_2	**0.978** (0.266)	1.06 (0.239)	1.048 (0.226)	1.149 (0.328)	1.137 (0.338)	**1.135** (0.328)
p_4, q_1	1.083 (0.206)	**1.059** (0.2)	1.075 (0.179)	1.081 (0.261)	**1.029** (0.179)	1.061 (0.128)
p_4, q_2	1.121 (0.192)	1.112 (0.17)	**1.104** (0.174)	1.093 (0.212)	**1.088** (0.191)	1.096 (0.181)
p_5, q_1	1.279 (0.322)	1.244 (0.296)	**1.242** (0.29)	1.169 (0.335)	1.167 (0.327)	**1.166** (0.306)
p_5, q_2	0.903 (0.078)	**0.867** (0.067)	0.877 (0.074)	1.053 (0.231)	1.033 (0.192)	**0.972** (0.126)

Table 2: Dataset for N = 100 & T = 10

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.263 (0.167)	**1.252** (0.156)	1.256 (0.159)	1.217 (0.295)	**1.118** (0.119)	1.192 (0.247)
p_1, q_2	2.641 (0.905)	**2.554** (0.838)	2.694 (0.961)	**1.771** (1.135)	1.848 (0.739)	1.803 (1.373)
p_2, q_1	1.221 (0.139)	1.186 (0.096)	**1.184** (0.102)	1.102 (0.084)	**1.088** (0.083)	1.124 (0.101)
p_2, q_2	1.044 (0.091)	1.145 (0.108)	**1.041** (0.088)	1.138 (0.255)	1.153 (0.211)	**1.136** (0.256)
p_3, q_1	1.086 (0.181)	1.066 (0.19)	**1.061** (0.182)	1.038 (0.172)	1.036 (0.171)	**1.035** (0.145)
p_3, q_2	1.112 (0.295)	**1.096** (0.309)	1.101 (0.306)	1.202 (0.38)	1.153 (0.328)	**1.099** (0.264)
p_4, q_1	1.053 (0.22)	1.038 (0.189)	**1.035** (0.185)	1.058 (0.14)	**1.051** (0.124)	1.063 (0.152)
p_4, q_2	1.263 (0.2)	1.247 (0.194)	**1.238** (0.21)	1.204 (0.133)	1.191 (0.114)	**1.183** (0.152)
p_5, q_1	1.613 (0.27)	1.679 (0.301)	**1.599** (0.342)	1.541 (0.884)	1.531 (0.609)	**1.521** (0.848)
p_5, q_2	1.092 (0.132)	1.047 (0.234)	**1.047** (0.114)	1.074 (0.144)	1.041 (0.117)	**1.041** (0.115)

Table 3: Dataset for N = 100 & T = 30

The Table 2 show the result for the average of RMSE in the datasets for different simulated ARMA parameter set, with the forecast window of 10. Table 3 provides the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

comparison of root mean squared error results of those datasets in different criteria, also compare different simulated dataset with the forecast window of 30.

The large forecast window usually decreases the accuracy for the time series model, T = 50 [4] window size is 50% for the size of in sample set, the MML87 hybrid model still outperforms its peer. It indicates that the MML information criteria is efficient in the model selection and the algorithm of the hybrid model is also efficient in time series analysis, the result of T = 50 shown in the Table 8.

Table 4 show the average of the ten different parameters simulated dataset in the forecast window sizes of T = 3, 10, 30 and 50 [4] with the in sample size of N = 100.

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
T = 3	1.086	1.076	**1.072**	1.121	1.123	**1.115**
T = 10	1.149	1.136	**1.121**	1.159	1.136	**1.134**
T = 30	1.338	1.331	**1.325**	1.234	1.221	**1.220**
T = 50	1.308	1.296	**1.295**	1.225	**1.195**	1.221

Table 4: Average of Table 1 to Table 3

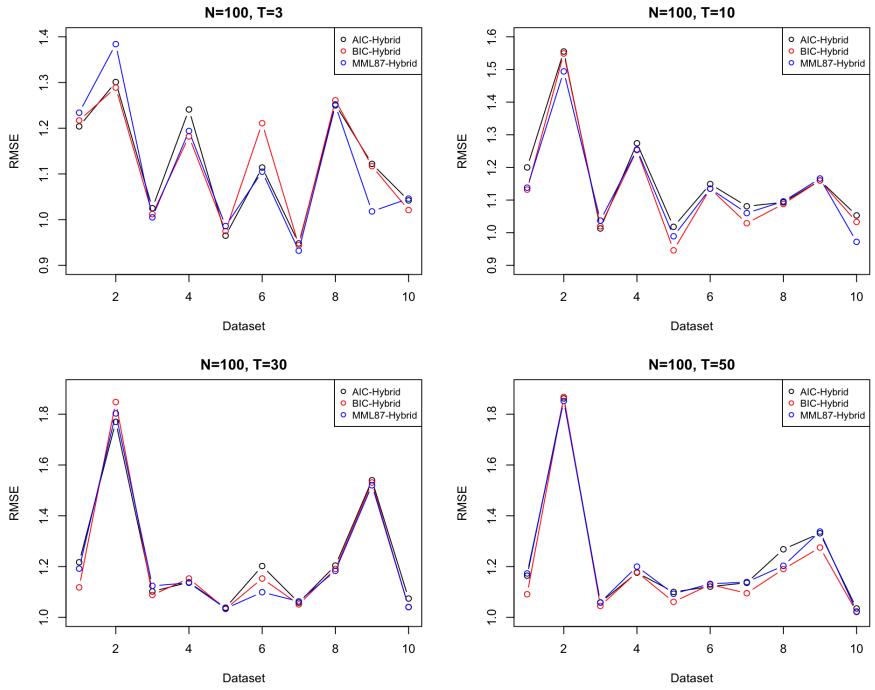

Figure 3. Comparison in different forecast windows

MML87 is outperformed in the in-sample size of N = 100 in all the T = 3, 10, 30, and 50, the results show that the MML87 not only considers the goodness of fit of data, also considers the model complexity. That's why the Figure 3 show MML87 have lower Root mean squared error in majority of case. The hybrid model selected by MML87 have the lowest error rate in the T = 3, 10 and 30, it makes sense that the MML87 would be widely used in the time series model selection. The results of N = 100 with T = 50 also show that the large size of the forecast window, the complex hybrid model ARMA-LSTM is

[4] Data provided by Table 8 in Appendix

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

performed better than the simple time series model because the model will become more accurate when forecasting in a long time after trained by larger amount of parameters.

Table 9 to Table 12 compare six different models or model selection techniques in the RMSE of the dataset in N = 50, 200, 300, and 500 with the forecast window size T = 10. The AIC tends to overfit in the small amount of data set such as N = 50 [5]. Through an increase in the amount for the in-sample dataset, the RMSE decreases in the hybrid model ARMA-LSTM because the larger size of data helps to LSTM to train and fit an accurate model. So the results show the RMSE for model with MML87 is lower than the other models than others in the N = 100, 200, and 300 because of the efficiency in controlling the model complexity in MML87, the model can avoid the overfitting problem in the small size of the dataset.

The hybrid model with LSTM is overfit the small in-sample size, because there are a larger amount of parameters that need to be estimated than the ARMA model, the hybrid model tends to perform well in the large in-sample size because deep learning model better off in large in-sample size, such as N = 200 [6], 300 [7], and 500 [8].

For the small in-sample size, such as N = 50, BIC performance is good on hybrid ARMA-LSTM, because BIC is able to select the model without overfitting. The MML87-Hybrid have smallest average RMSE in the N = 100, 200 and 300 for the different randomized dataset, it states that the MML87 works well in the time series model selection, and the model selected able to provide the lower forecasting errors. The hybrid models work efficiently when there is enough or a large size of in-sample data, otherwise, it also overfits the small dataset. In the meantime, by comparing the RMSE from MML87-ARMA and AIC-ARMA and BIC-ARMA, the results favor the MML87 rather than the AIC and BIC, it shows that the MML87 having a good performance in the time series model selection and is able to select the ARMA model with lower forecasting errors.

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
N = 50 [5]	1.301	1.291	**1.280**	1.224	**1.202**	1.244
N = 100	1.149	1.136	**1.121**	1.159	1.136	**1.134**
N = 200 [6]	1.177	1.187	**1.183**	1.159	1.161	**1.154**
N = 300 [7]	1.163	1.152	**1.147**	1.131	1.125	**1.123**
N = 500 [8]	**1.194**	1.197	1.196	1.180	**1.173**	1.181

Table 5: Average of Table 1 to Table 3

[5] Data provided by Table 9 in Appendix
[6] Data provided by Table 10 in Appendix
[7] Data provided by Table 11 in Appendix
[8] Data provided by Table 12 in Appendix

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

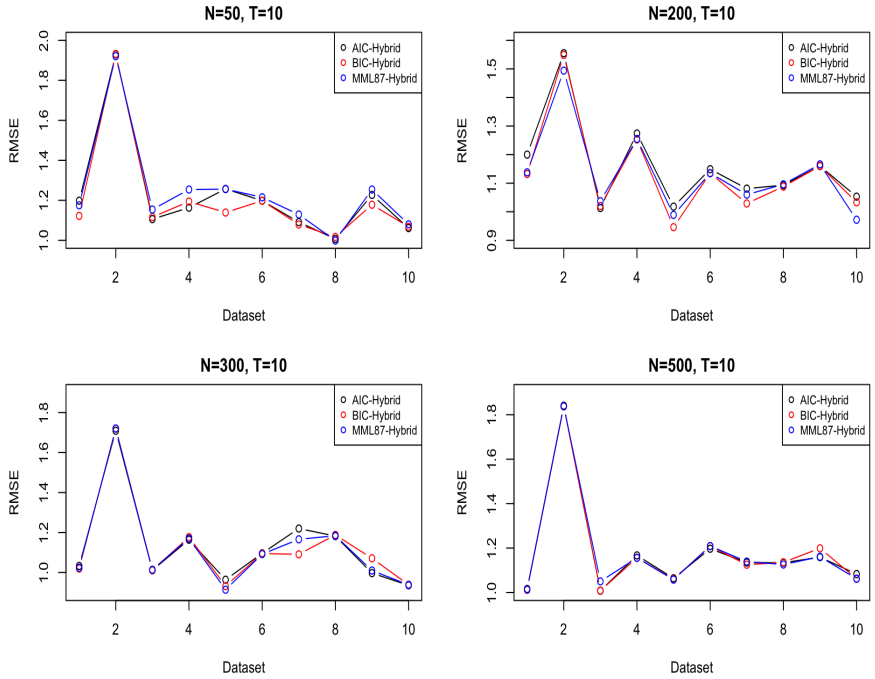

Figure 4. Comparison in different forecast windows

6.2. Financial data

The stock return prediction is one of most popular research topic in financial application [27,31]. This section studies the performance of the hybrid model by MML87, and hybrid model by AIC, BIC, and the ARIMA model selected by MML8, AIC, and BIC. The stock prices were selected from the components of the Dow Jones Industrial Average, including Apple, Boeing, Cisco System, Goldman Sachs, IBM, Intel, Johnson & Johnson, JPMorgan Chase, Coca-Cola, and 3M. The time horizontal of the data selected is adjusted closed price for three years or 1,258 days in each trading day from 2016-09-23 to 2021-09-22. This experimental studies the different performances in forecast window sizes T = 3, 5, 10, 30, 50, 70, 100, 130, 150, and 200.

The empirical results show that the hybrid model ARIMA-LSTM can substantially outperform the traditional time series ARIMA model, particularly in the size of forecast window T= 5, 30, 100, 130, 150, and 200. Many studies demonstrated that the stock return depends on various factors such as dividend yield, the book to market ratio, and/or interest rate [27–29]. However, traditional linear time series models are difficult to consider the effect of all those factors, it requires a more complex model to capture the information hidden in residual from the ARIMA model. The hybrid model with LSTM is able to capture the relation between publicly available information In order to make the stock price stationary in time series analysis, the ARIMA models are using parameter $d = 1$ to differencing in one order. As the experimental results show, MML87 outperform the other information-theoretic of AIC and BIC in term of lower root mean squared error for out-of-sample forecasting.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

	Mean	S.D	PACF1	PACF2	PACF3
AAPL	66.440217	37.060808	0.996875	0.044454	-0.004848
BA	258.704781	82.478194	0.995870	-0.031231	-0.061804
CSCO	40.585947	8.595774	0.994585	0.073202	-0.016488
GS	227.095242	56.820929	0.993579	0.039741	-0.043412
IBM	124.851224	10.369478	0.982339	0.070195	-0.040622
INTC	46.269478	9.305502	0.992194	0.178757	-0.053398
JNJ	130.715314	18.399352	0.993930	0.050988	-0.031304
JPM	104.046116	24.467471	0.993854	0.067756	-0.049235
KO	44.519034	6.089778	0.993828	0.031639	-0.039178
MMM	173.550240	20.467854	0.991641	0.004475	0.026664

Table 6: Mean, standard deviation, pacf lag 1 to 3 for ten selected stocks

Figure 5. Log prices for ten selected stocks

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
T = 3	**2.987** (3.446)	3.027 (3.555)	3.075 (3.572)	4.414 (4.75)	4.302 (4.608)	**4.289** (4.616)
T = 5	**4.024** (5.091)	4.077 (5.228)	4.163 (5.086)	4.024 (5.45)	3.966 (5.42)	**3.907** (5.449)
T = 10	4.748 (4.707)	**4.747** (4.858)	4.868 (4.347)	5.359 (5.429)	5.261 (5.268)	**5.249** (5.262)
T = 30	5.872 (6.797)	**5.867** (6.6)	5.994 (5.662)	5.754 (4.822)	**5.628** (4.687)	5.643 (4.677)
T = 50	7.834 (7.511)	7.609 (7.298)	**6.659** (6.966)	**7.328** (6.787)	7.411 (6.879)	7.384 (6.898)
T = 70	9.991 (9.491)	9.909 (9.316)	**9.645** (7.99)	10.393 (8.048)	10.221 (7.789)	**10.085** (7.612)
T = 100	14.465 (17.187)	13.991 (15.428)	**9.866** (10.854)	9.304 (9.256)	**9.087** (9.35)	9.253 (9.486)
T = 130	14.482 (9.714)	14.301 (10.571)	**13.551** (10.238)	**13.768** (10.598)	13.811 (11.124)	14.581 (10.972)
T = 150	22.985 (28.173)	22.985 (28.077)	**18.045** (17.856)	17.778 (16.771)	17.526 (16.582)	**17.461** (15.931)
T = 200	31.144 (37.567)	30.502 (38.314)	**30.286** (32.564)	26.831 (31.63)	**26.424** (31.547)	26.507 (31.59)

Table 7: RMSE for different forecast window sizes

The hybrid model tends to outperform for large forecast window size rather than the small forecast window size, because the large ahead size in forecasting has higher uncertainty. For most of the financial security, there is high volatility in the long forecasts. Also because of semi-strong market efficiency, which means the stock price fully and fairly reflect publicly available information in time horizontal in forecast window, and also reflect all past information. So it is more likely that a complex model will provide accurate results in prediction for T greater than 100.

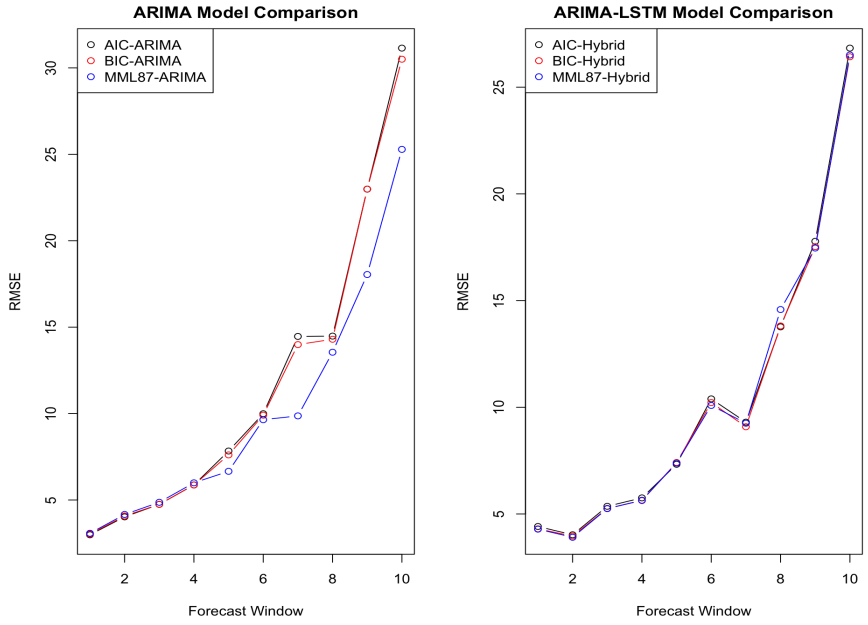

Figure 6. Comparison in different forecast windows

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

7. Conclusions

We have investigated the time series modeling in the Minimum Message Length framework using the Wallace and Freeman (1987) approximation [15]. The hybrid model compares with the traditional time series model based on information-theoretic approaches AIC, BIC and MML87. Two types of data are used in order to study the performances for different models, the hybrid model performs usually performed better than the traditional ARIMA model, and MML87 is able to select the lower forecasting errors than the AIC and BIC, as the experimental results show.

Author Contributions: Conceptualization: Zheng Fang; methodology: Zheng Fang and David L. Dowe; Computation: Zheng Fang and Dedi Rosadi; validation: Zheng Fang, David L. Dowe and Shelton Peiris; investigation: Zheng Fang and Shelton Peiris; writing and preparation: Zheng Fang and David L. Dowe; writing and review: Zheng Fang and David L. Dowe and Shelton Peiris; supervision, David L. Dowe and Shelton Peiris and Dedi Rosadi. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Department of Artificial Intelligence & Data Science, Monash University for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.189 (0.217)	1.191 (0.228)	1.182 (0.222)	1.164 (0.304)	1.091 (0.212)	1.173 (0.241)
p_1, q_2	2.307 (0.458)	2.308 (0.457)	2.298 (0.466)	1.862 (1.169)	1.868 (1.16)	1.852 (1.073)
p_2, q_1	1.113 (0.087)	1.092 (0.103)	1.094 (0.104)	1.058 (0.073)	1.045 (0.096)	1.059 (0.139)
p_2, q_2	1.191 (0.096)	1.189 (0.103)	1.191 (0.1)	1.176 (0.24)	1.178 (0.259)	1.201 (0.289)
p_3, q_1	1.094 (0.159)	1.093 (0.157)	1.097 (0.155)	1.101 (0.192)	1.061 (0.144)	1.093 (0.115)
p_3, q_2	1.127 (0.06)	1.123 (0.055)	1.125 (0.058)	1.121 (0.134)	1.129 (0.155)	1.132 (0.153)
p_4, q_1	1.188 (0.182)	1.189 (0.188)	1.192 (0.186)	1.136 (0.137)	1.095 (0.181)	1.139 (0.113)
p_4, q_2	1.232 (0.165)	1.221 (0.133)	1.212 (0.134)	1.268 (0.457)	1.19 (0.269)	1.203 (0.319)
p_5, q_1	1.593 (0.304)	1.521 (0.199)	1.528 (0.209)	1.331 (0.428)	1.275 (0.234)	1.338 (0.383)
p_5, q_2	1.051 (0.083)	1.033 (0.064)	1.032 (0.063)	1.035 (0.055)	1.021 (0.067)	1.023 (0.069)

Table 8: Dataset for N = 100 & T = 50

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.068 (0.147)	1.071 (0.118)	1.067 (0.115)	1.198 (0.222)	1.122 (0.305)	1.175 (0.259)
p_1, q_2	1.994 (0.655)	1.994 (0.655)	2.04 (0.705)	1.93 (1.553)	1.932 (1.563)	1.921 (1.566)
p_2, q_1	1.242 (0.213)	1.242 (0.213)	1.235 (0.17)	1.106 (0.193)	1.116 (0.196)	1.154 (0.278)
p_2, q_2	1.185 (0.355)	1.183 (0.359)	1.232 (0.476)	1.163 (0.386)	1.194 (0.499)	1.254 (0.601)
p_3, q_1	1.348 (0.557)	1.254 (0.604)	1.304 (0.575)	1.257 (0.499)	1.139 (0.605)	1.256 (0.449)
p_3, q_2	1.283 (0.234)	1.283 (0.234)	1.291 (0.233)	1.198 (0.27)	1.198 (0.27)	1.215 (0.285)
p_4, q_1	1.263 (0.461)	1.251 (0.469)	1.044 (0.172)	1.091 (0.264)	1.079 (0.27)	1.129 (0.243)
p_4, q_2	0.987 (0.132)	0.987 (0.132)	0.989 (0.137)	1.007 (0.137)	1.017 (0.126)	0.999 (0.138)
p_5, q_1	1.533 (0.457)	1.426 (0.535)	1.464 (0.509)	1.227 (0.442)	1.178 (0.445)	1.254 (0.434)
p_5, q_2	1.101 (0.153)	1.098 (0.151)	1.137 (0.185)	1.061 (0.168)	1.068 (0.175)	1.08 (0.117)

Table 9: Dataset for N = 50 & T = 10

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.244 (0.365)	1.277 (0.42)	1.248 (0.404)	1.153 (0.381)	1.13 (0.376)	1.151 (0.353)
p_1, q_2	1.359 (0.445)	1.359 (0.445)	1.359 (0.445)	1.474 (0.813)	1.491 (0.882)	1.474 (0.813)
p_2, q_1	0.927 (0.183)	0.915 (0.172)	0.92 (0.182)	0.939 (0.126)	0.955 (0.15)	0.933 (0.128)
p_2, q_2	1.184 (0.41)	1.191 (0.398)	1.189 (0.402)	1.134 (0.368)	1.114 (0.393)	1.106 (0.37)
p_3, q_1	1.137 (0.347)	1.136 (0.347)	1.117 (0.355)	1.082 (0.314)	1.082 (0.316)	1.085 (0.325)
p_3, q_2	0.915 (0.198)	1.038 (0.08)	0.991 (0.093)	1.088 (0.184)	1.083 (0.172)	1.054 (0.161)
p_4, q_1	1.199 (0.558)	1.166 (0.557)	1.19 (0.562)	1.086 (0.591)	1.109 (0.507)	1.107 (0.732)
p_4, q_2	1.108 (0.196)	1.101 (0.191)	1.129 (0.24)	1.184 (0.358)	1.186 (0.359)	1.184 (0.36)
p_5, q_1	1.581 (0.481)	1.584 (0.475)	1.586 (0.48)	1.383 (0.83)	1.391 (0.802)	1.382 (0.832)
p_5, q_2	1.123 (0.263)	1.101 (0.174)	1.101 (0.174)	1.063 (0.234)	1.069 (0.133)	1.063 (0.128)

Table 10: Dataset for N = 200 & T = 10

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0049.v1

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	1.024 (0.312)	1.028 (0.332)	1.031 (0.322)	1.033 (0.316)	1.02 (0.27)	1.024 (0.32)
p_1, q_2	2.008 (1.123)	1.995 (1.024)	1.988 (1.028)	1.709 (0.918)	1.72 (0.896)	1.72 (0.854)
p_2, q_1	1.022 (0.144)	1.025 (0.138)	1.016 (0.133)	1.011 (0.125)	1.012 (0.121)	1.014 (0.297)
p_2, q_2	1.172 (0.398)	1.168 (0.383)	1.166 (0.384)	1.164 (0.422)	1.177 (0.443)	1.17 (0.413)
p_3, q_1	0.886 (0.198)	0.868 (0.205)	0.865 (0.215)	0.964 (0.261)	0.932 (0.183)	0.914 (0.188)
p_3, q_2	1.07 (0.408)	1.068 (0.412)	1.059 (0.401)	1.096 (0.284)	1.095 (0.289)	1.092 (0.284)
p_4, q_1	1.215 (0.445)	1.191 (0.468)	1.184 (0.464)	1.22 (0.621)	1.091 (0.42)	1.166 (0.453)
p_4, q_2	1.191 (0.338)	1.167 (0.308)	1.162 (0.278)	1.182 (0.427)	1.188 (0.473)	1.184 (0.433)
p_5, q_1	1.169 (0.225)	1.159 (0.216)	1.152 (0.216)	0.997 (0.131)	1.071 (0.213)	1.01 (0.146)
p_5, q_2	0.874 (0.25)	0.846 (0.249)	0.844 (0.247)	0.936 (0.213)	0.939 (0.197)	0.938 (0.196)

Table 11: Dataset for N = 300 & T = 10

Average of RMSE & Standard deviation						
	ARMA			ARMA-LSTM		
	AIC	BIC	MML87	AIC	BIC	MML87
p_1, q_1	0.988 (0.229)	0.966 (0.233)	0.968 (0.232)	1.016 (0.178)	1.012 (0.182)	1.014 (0.179)
p_1, q_2	1.546 (0.728)	1.549 (0.713)	1.562 (0.703)	1.841 (0.915)	1.838 (0.875)	1.838 (0.877)
p_2, q_1	1.002 (0.37)	1.017 (0.349)	1.016 (0.351)	1.008 (0.329)	1.008 (0.325)	1.05 (0.349)
p_2, q_2	1.156 (0.188)	1.165 (0.176)	1.165 (0.176)	1.167 (0.337)	1.156 (0.355)	1.156 (0.355)
p_3, q_1	1.091 (0.175)	1.093 (0.18)	1.09 (0.176)	1.064 (0.225)	1.06 (0.157)	1.058 (0.22)
p_3, q_2	1.23 (0.372)	1.235 (0.364)	1.235 (0.364)	1.197 (0.365)	1.209 (0.393)	1.209 (0.393)
p_4, q_1	1.041 (0.272)	1.07 (0.25)	1.063 (0.257)	1.135 (0.342)	1.053 (0.239)	1.139 (0.343)
p_4, q_2	1.253 (0.265)	1.256 (0.265)	1.255 (0.266)	1.134 (0.218)	1.136 (0.214)	1.126 (0.235)
p_5, q_1	1.559 (0.363)	1.551 (0.385)	1.541 (0.365)	1.159 (0.331)	1.199 (0.421)	1.161 (0.292)
p_5, q_2	1.073 (0.179)	1.068 (0.188)	1.068 (0.188)	1.083 (0.136)	1.062 (0.167)	1.062 (0.167)

Table 12: Dataset for N = 500 & T = 10

References

1. Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1394-1401). IEEE.

 doi:10.20944/preprints202110.0049.v1

2. Hunt, R., Peiris, S., & Weber, N. (2021). A General Frequency Domain Estimation Method for Gegenbauer Processes. Journal of Time Series Econometrics, 13(2), 119-144.
3. Fathi, O. (2019). Time series forecasting using a hybrid ARIMA and LSTM model. Velvet Consulting.
4. De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. International journal of forecasting, 22(3), 443-473.
5. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.
6. Wang, J. Q., Du, Y., & Wang, J. (2020). LSTM based long-term energy consumption prediction with periodicity. Energy, 197, 117197.
7. Chen, K., Zhou, Y., & Dai, F. (2015, October). A LSTM-based method for stock returns prediction: A case study of China stock market. In 2015 IEEE international conference on big data (big data) (pp. 2823-2824). IEEE.
8. Wallace, C. S., & Boulton, D. M. (1968). An information measure for classification. The Computer Journal, 11(2), 185-194.
9. Wallace, C. S., & Dowe, D. L. (1999). Minimum message length and Kolmogorov complexity. The Computer Journal, 42(4), 270-283.
10. Dowe, D. L. (2008). Foreword re CS Wallace. The computer journal, 51(5), 523-560.
11. Wong, C. K., Makalic, E., & Schmidt, D. F. (2018). Minimum message length inference of the Poisson and geometric models using heavy-tailed prior distributions. Journal of Mathematical Psychology, 83, 1-11.
12. Sak, M., Dowe, D. L., & Ray, S. (2005, December). Minimum message length moving average time series data mining. In 2005 ICSC Congress on Computational Intelligence Methods and Applications (pp. 6-pp). IEEE.
13. Wallace, C. S. (2005). Statistical and inductive inference by minimum message length (pp. 93-100). New York: Springer.
14. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time Series Analysis: Forecasting and Control. John Willey and Sons, New Jersey.
15. Wallace, C. S., & Freeman, P. R. (1987). Estimation and inference by compact coding. Journal of the Royal Statistical Society: Series B (Methodological), 49(3), 240-252.
16. Fitzgibbon, L. J., Dowe, D. L., & Vahid, F. (2004, January). Minimum message length autoregressive model order selection. In International Conference on Intelligent Sensing and Information Processing, 2004. Proceedings of (pp. 439-444). IEEE.
17. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
18. Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1976). Time series analysis prediction and control.
19. Dowe, D. L. (2008). Foreword re CS Wallace. The computer journal, 51(5), 523-560.
20. Baxter, R. A., & Dowe, D. L. (1994). Model selection in linear regression using the MML criterion. In Proceedings of the Data Compression Conference. IEEE, Institute of Electrical and Electronics Engineers.
21. Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187-205.
22. Qiu, X., Zhang, L., Ren, Y., Suganthan, P. N., & Amaratunga, G. (2014, December). Ensemble deep learning for regression and time series forecasting. In 2014 IEEE symposium on computational intelligence in ensemble learning (CIEL) (pp. 1-6). IEEE.
23. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing, 90, 106181.
24. Li, J., Bu, H., & Wu, J. (2017, June). Sentiment-aware stock market prediction: A deep learning method. In 2017 international conference on service systems and service management (pp. 1-6). IEEE.
25. Zhang, X., & Tan, Y. (2018, June). Deep stock ranker: A LSTM neural network model for stock selection. In International conference on data mining and big data (pp. 614-623). Springer, Cham.
26. Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of COVID19 per regions using ARIMA models and polynomial functions. Applied soft computing, 96, 106610.
27. Cheng, T., Gao, J., & Linton, O. (2019). Nonparametric Predictive Regressions for Stock Return Prediction.
28. Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. Journal of financial Economics, 17(2), 357-390.
29. Fama, E. F., & French, K. R. (2021). Dividend yields and expected stock returns (pp. 568-595). University of Chicago Press.
30. Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. IEEE Access, 8, 71326-71338.
31. Gao, J. (2004). Modelling long-range-dependent Gaussian processes with application in continuous-time financial models. Journal of applied probability, 41(2), 467-482.