Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.31020
Page: 694-705

A Comparative study of Average-Based FTSMC, GBM, and LSTM for JCI Forecasting

[1]Syifa Aulia
Applied Mathematics, IPB University, Bogor, Indonesia

[2]I Wayan Mangku
Mathematics, IPB University, Bogor, Indonesia

[3]I Gusti Putu Purnaba
Mathematics, IPB University, Bogor, Indonesia

Article Info

Article history:

Accepted: 10 July 2026

Keywords:

Composite Stock Price Index;
Fuzzy Time Series;
GBM;
LSTM;
Prediction.

ABSTRACT

Stocks are high-risk investment instruments, so a model is needed to assist investors in decision-making. The Jakarta Composite Index (JCI) is a key indicator that measures the performance of all stocks on the Indonesian stock exchange. This study compares the performance of the Average-Based FTSMC, GBM, and LSTM models in predicting JCI movements. The research data consists of daily stock index values from June 1, 2024, to June 29, 2026. The models were compared using the evaluation metrics MAPE, RMSE, and MAE. The results show that the Average-Based FTSMC model provides the highest accuracy (MAPE: 0.925%, RMSE: 86.753, and MAE: 65.162). These findings indicate that the fuzzy time series approach is more adaptive to JCI fluctuations than both classical stochastic models and deep learning models. This study is limited to historical data using a single input variable; therefore, generalizing the results to market conditions requires further investigation.

Corresponding Author:

Syifa Aulia,
Applied Mathematics
IPB University
Email: syiifaaulia@apps.ipb.ac.id

1. INTRODUCTION

The capital market plays a crucial role in a country's economy because it serves as a means of connecting investors with issuers [1]. The existence of the capital market can attract interest from both foreign and domestic investors. The capital market offers several financial instruments, one of which is popular and attracts investor attention, namely stocks [2]. Investing in stocks offers the potential for high returns and high liquidity [3]. As a dynamic capital market instrument, stock price movements reflect the market's collective expectations regarding macroeconomic conditions and corporate performance. Investors need to pay attention to stock price movements as a means of portfolio optimization, risk management, and other decision-making.

Stock price fluctuations can be seen from volatility; the higher the volatility, the faster the stock price fluctuations [4]. Stock price fluctuations are influenced by several internal and external factors. Internal factors include company performance, dividend policies, and corporate actions. External factors, on the other hand, include inflation rates, interest rates, currency exchange rates, and geopolitical sentiment. The high market sensitivity to factors that influence stock price fluctuations gives rise to highly volatile, non-linear, and non-stationary stock price characteristics [5]. These characteristics make predicting stock price movements a complex problem.

The main indicator that records the performance of all stocks listed on the capital market (Indonesia Stock Exchange) is the Jakarta Composite Index [6]. The Jakarta Composite Index (JCI) monitors the price movements of stocks traded on the capital market. Investors use the Jakarta Composite Index as a benchmark to assess aggregate capital market conditions [7]. The market capitalization of all listed companies is integrated so that its movements reflect overall national economic sentiment. A positive Jakarta Composite Index movement indicates market optimism, while a sharp decline indicates economic uncertainty that needs to be anticipated [8]. Thus, the accuracy of the Jakarta Composite Index prediction model has implications not only for individual investment decisions but also for systematic risks in policy analysis.

Various modeling approaches have been developed to predict Jakarta Composite Index movements, which can be grouped into three groups: stochastic, fuzzy theory-based, and machine learning. Within the stochastic group, several studies have been conducted, such as those by Bratian et al. [9], Sinha [10] and Sinha [11], which used the GBM model to predict stock indexes. However, based on these studies, none have yet been applied to stock indices in emerging markets such as Indonesia. Furthermore, determining the drift and volatility parameters is crucial to ensure the model aligns with the characteristics of emerging markets. Other approaches that combine stochastic models with artificial intelligence such as ARIMA and LSTM by Nensi et al. [12] and GARCH-LSTM and GARCH-GRU by Mutinda and Langat [13] accommodate nonlinear patterns that cannot be captured by classical stochastic models. In the fuzzy theory group, research Guan [14], Aulia [15], Ramadani [16] discussed predictions using the Fuzzy Time Series Markov Chain model. Research using these fuzzy theory-based models has proven effective in modeling the uncertainty of time series data without specific assumptions. However, the accuracy of fuzzy time series models is still influenced by the determination of the length and number of intervals; therefore, it is necessary to determine interval divisions that are appropriate for the data characteristics. In the machine learning group, research by Wen and Li [17], Sabzipour et al. [18], Yunita et al. [19] employs deep learning models such as LSTM, which excel at capturing both long and short-term temporal dependencies in non-stationary data.

The differing assumptions underlying each approach have implications for the performance of each model in addressing the volatile and non-stationary characteristics of the Jakarta Composite Index. This difference is most pronounced between GBM and LSTM: GBM is a parametric stochastic model that assumes returns follow a normal distribution with drift and volatility, while LSTM is a non-parametric machine learning model that learns nonlinear patterns without assuming any specific distribution. Given these paradigmatic differences, it is relevant to examine the GBM and LSTM models to determine whether stochastic-based and machine-learning approaches can compete in predicting volatile, non-stationary stock price indices. Previous studies have generally compared models within the same category or combined two approaches simultaneously. Therefore, there is currently no empirical evidence indicating which approach stochastic modeling, fuzzy theory, or machine learning is the most reliable when considered together. This gap underscores the importance of evaluating all three approaches to effectively estimate risk in decision-making. Furthermore, addressing this research gap can provide methodological insights into selecting the appropriate model to capture the dynamics of the JCI's movements. Based on this gap, this study aims to compare the performance of the Average-Based FTSMC, Geometric Brownian Motion (GBM), and Long Short-Term Memory (LSTM) models in predicting movements in the Jakarta Composite Index to determine the model with the highest accuracy.

2. RESEARCH METHOD

This study uses daily data from the Jakarta Composite Index (JCI) obtained from the website https://investing.com. The data covers the period from June 1, 2024, to June 29, 2026, with a total of 492 data points. The data was divided into training and testing sets in an 80:20 ratio. A normality test was conducted using the Kolmogorov-Smirnov test as part of fulfilling the assumptions of the GBM model. In the GBM model, the time interval $\Delta t = 1/252$ represents the standard convention for the number of effective trading days in a calendar year in the capital market. The Monte Carlo simulation was run for 1,000 iterations to ensure the convergence of the forecast distribution. In LSTM modeling, data normalization is first performed using min-max normalization to standardize the data scale. For the LSTM model, hyperparameter tuning was performed, covering the number of neurons, the learning rate, and the dropout rate [20]. The selection of this hyperparameter combination uses the grid search method and is validated using the walk-forward validation scheme. 5-fold walk-forward validation was used to preserve the temporal order of the data and prevent data leakage, which commonly occurs in standard random cross-validation. The LSTM architecture defined by the researchers consisted of 1 LSTM layer and 1 dense output layer, with 14-time steps, the ADAM optimizer, and the MSE loss function. The selection of LSTM hyperparameters in this study was based on a combination of literature reviews and the characteristics of JCI's historical data [12], [20], [21]. An early stopping scheme was also applied to determine the required number of epochs. Modeling of the Average-Based FTSMC, GBM, and LSTM was performed using the training data, followed by prediction and validation using the test data. The performance of each model was evaluated using metrics including MAPE, RMSE, and MAE.

2.1 Fuzzy Time Series Markov Chain

Tsaur proposed a new approach called the Fuzzy Time Series Markov Chain (FTSMC), which combines the principles of fuzzy time series with Markov chains to improve the accuracy of the prediction results [22]. The stages in the FTSMC method are described as follows.

a. Define the universe of discourse
Determining the population set begins with identifying the highest (Y_{max}) and lowest (Y_{min}) values in the historical data. Next, two positive integers, d_1 and d_2, are selected as adjustment factors for the upper and lower bounds, so that the data range can be divided into an appropriate number of intervals. The formula for defining the population set is: $U = [Y_{min} - d_1, Y_{max} + d_2]$.

b. Determine the length and number of intervals
The previously obtained universal set range is partitioned into a number of subintervals with uniform widths. The number of subintervals is obtained using the following equation: $n = 1 + 3.32 \log N$.

c. Determining fuzzy sets
Each fuzzy set A_i is formed by assigning a membership degree to each subinterval u_k. The resulting fuzzy set A_i is denoted as follows: $A_i = \left\{\frac{\mu_{ik}}{u_k}\right\},\ \ k = 1,2, \dots, n.$
The fuzzy membership degree is defined as follows: $\mu_{ik} = \begin{cases} 1 & k = i \\ 0.5 & k = i + 1\ or\ k = i - 1 \\ 0 & others. \end{cases}$

d. Fuzzify of historical data
The fuzzification stage aims to map each data value into a relevant fuzzy set. If a data value lies in the subinterval u_k, then the value is represented in the corresponding fuzzy set A_i.

e. Compiling FLR and FLRG
Fuzzy Logical Relationships (FLR) are designed to capture the transitional relationships between fuzzy sets from one time to the next. Formally, these relationships are expressed as: $A_i \to A_k$, where A_i acts as left hand side and A_k as right hand side. FLRs that have identical initial states are then collected into one group called the Fuzzy Logical Relationship Group (FLRG). For example: $A_i \to A_j, A_i \to A_l$, then all these relations are merged into one FLRG, namely $A_i \to A_j, A_l$.

f. Constructing a Markov transition probability matrix
The value of each element of the matrix represents the probability of moving from state A_i to state A_j, which is obtained through the following equation: $P_{ij} = \frac{M_{ij}}{M_i};\ (i,j = 1,2, \dots, n)$, where M_{ij} represents the number of recorded movements from state A_i to state A_j, while M_i represents the total of all movements departing from state A_i.

g. Calculate predictions and adjustment values
The calculation of the predicted value is calculated following the following rules.
Rule 1: If $A_i \to A_k$, m_k is the middle value of u_k, then the prediction result is: $F(t) = m_k P_{ik}$.
Rule 2: If $A_i \to A_1, A_2, A_3, \dots, A_n$, m_k is the middle value of u_k, and data at time $t-1$ (Y_{t-1}) is in state A_i, then the predicted result is: $F(t) = m_1 P_{i1} + \dots + m_{i-1} P_{i(i-1)} + Y_{t-1} P_{ii} + \dots + m_n P_{in}$.
The adjustment value to the initial prediction result is calculated following the following rules.
Rule 1: Suppose A_i is at time $t-1$, $A_i \to A_i$, $A_i \to A_j$ $(i < j)$; the adjustment value is: $D_{t1} = \frac{l}{2}$.
Rule 2: Suppose A_i is at time $t-1$, $A_i \to A_i$, $A_i \to A_j$ $(i > j)$; the adjustment value is: $D_{t1} = -\frac{l}{2}$.
Rule 3: Suppose A_i is at time $t-1$, $A_i \to A_{i+s}$ $(1 \le s \le n - i)$; the adjustment value is: $D_{t2} = s\left(\frac{l}{2}\right)$.
Rule 4: Suppose A_i is at time $t-1$, $A_i \to A_{i-v}$ $(1 \le v \le i)$; the adjustment value is: $D_{t2} = v\left(\frac{l}{2}\right)$.
In general, the general form for the adjusted prediction results is: $F'(t) = F(t) \pm D_{t1} \pm D_{t2}$.

2.2 Average Based Fuzzy Time Series Markov Chain

Average-Based FTSMC is a variant of the Fuzzy Time Series model that uses an average-based algorithm to determine the interval length [23]. The steps in the average-based algorithm are as follows [24].

a. The absolute difference between Y_t and Y_{t+1} is calculated for each sequential data pair, then the difference values are averaged using the following equation: $M = \frac{\sum_{t=1}^{N-1} |Y_{t+1} - Y_t|}{N-1}$

b. The absolute mean value obtained in the first step is divided by two to produce the interval length (l).

c. Based on the previous step, the interval length base is determined by referring to Table 1.

Table 1. Interval Base

Interval Length	Base
0.1-1.0	0.1
1.1-10	1
11-100	10
101-1000	100
1001-10000	1000

d. The interval length is then rounded up to match the interval base value obtained from the table.

2.3 Kolmogorov Smirnov

One test that can be used to determine whether a dataset is normally distributed is the Kolmogorov-Smirnov test. The hypothesis testing procedure for the Kolmogorov-Smirnov test is as follows [25]:
H_0: Data are normally distributed.
H_1: Data are not normally distributed.
The Kolmogorov-Smirnov test statistic is expressed in the following equation: $D = max|F_t - F_s|$
where D denotes the maximum deviation, F_t denotes the theoretical distribution function (as hypothesized), and F_s denotes the empirical distribution function (based on the sample). The test criterion is to accept H_0 if the value of $D < D_\alpha$ or the p-value $> \alpha$, where α = 0.05, so it can be concluded that the data are normally distributed.

2.4 Lemma Itô

Suppose X_t is a continuous-time stochastic process that satisfies the following equation:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \quad (1)$$

where W_t is the Wiener process. If there is a function $G(X_t, t)$ which is a differentiable function of X_t and t, then [26]

$$dG = \left[\frac{\partial G}{\partial X_t}\mu(X_t, t) + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial X_t^2}\sigma^2(X_t, t)\right]dt + \frac{\partial G}{\partial X_t}\sigma(X_t, t)dW_t. \quad (2)$$

2.5 Wiener Process

A Wiener process is a special type of Markov stochastic process with a drift rate of 0 and a variance rate of 1. A stochastic process is said to follow a Wiener process if it satisfies the following condition:

a. The change in $\Delta W(t)$ over a given time interval Δt is $\Delta W(t) = \epsilon\sqrt{\Delta t}$, where ϵ is a random variable with a standard normal distribution $N(0,1)$ and Δt is the difference between two distinct time intervals.
b. The value of $\Delta(t)$ for each different time interval is independent.

The change in the value of $W(t)$ over a time period T can be considered as the sum of the changes in $W(t)$ over N time intervals of length t, that is: $N = \frac{T}{\Delta t}$. Therefore, the change in the value of $W(t)$ over the time period T can be written as: $W(T) - W(0) = \sum_{i=1}^{N} \epsilon_i\sqrt{\Delta t}$, where $\epsilon_i (i = 1,2, \dots, N)$ is a random variable with a standard normal distribution [27].

2.6 Brownian Motion

A stochastic process $\{W(t), t \geq 0\}$ is called Brownian motion if it satisfies the following conditions [9]:

a. $W(0) = 0$,
b. $\{W(t), t \geq 0\}$ has a stationary and independent increment,
c. $W(s) - W(t) \sim N(0, s - t), \forall s < t$.

2.7 Geometric Brownian Motion

Geometric Brownian Motion (GBM) is a stochastic process used in stock price modeling. The GBM model assumes that stock price returns are independent and normally distributed. The stochastic differential equation of the GBM model is written as follows [28].

$$dP_t = \mu P_t dt + \sigma P_t dW_t \quad (3)$$

with P_t representing the stock price at time t and W_t representing the Wiener process. Equation 3 can also be written as equation 4 as follows.

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t. \quad (4)$$

Suppose a function G of the variables P_t and t is $G(P_t, t)$ with P_t satisfying equation 3. Suppose the function $G(P_t, t) = \ln P_t$, then the following equation will be obtained: $\frac{dG}{P_t} = \frac{1}{P_t}$, $\frac{\partial^2 G}{\partial P_t^2} = -\frac{1}{P_t^2}$, and $\frac{dG}{t} = 0$. Based on Lemma It$\hat{o}$, the function G will follow the following equation:

$$
\begin{aligned}
dG = \left[\frac{\partial G}{\partial P_t}P_t + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial P_t^2}(\sigma P_t)^2\right]dt + \frac{\partial G}{\partial P_t}\sigma P_t dW_t &= \left[\frac{1}{P_t}\mu P_t - \frac{1}{2}\frac{1}{P_t^2}\sigma^2 P_t^2\right]dt + \frac{1}{P_t}\sigma P_t dW_t \\
&= \left[\frac{1}{P_t}\mu P_t - \frac{1}{2}\frac{1}{P_t^2}\sigma^2 P_t^2\right]dt + \frac{1}{P_t}\sigma P_t dW_t \\
&= \left[\mu - \frac{\sigma^2}{2}\right]dt + \sigma dW_t \\
\ln P_t - P_{t-1} &= \left[\mu - \frac{\sigma^2}{2}\right]dt + \sigma dW_t \\
\ln\left(\frac{P_t}{P_{t-1}}\right) &= \left[\mu - \frac{\sigma^2}{2}\right]dt + \sigma dW_t \\
\frac{P_t}{P_{t-1}} &= exp\left[\left[\mu - \frac{\sigma^2}{2}\right]dt + \sigma dW_t\right] \\
P_t &= P_{t-1}\, exp\left[\left[\mu - \frac{\sigma^2}{2}\right]dt + \sigma dW_t\right].
\end{aligned} \tag{5}
$$

where P_t represents the stock price at time t, P_{t-1} represents the stock price at time $t-1$, μ represents the drift value, σ represents volatility, and W_t represents the Wiener process. Since $dW_t \sim dt.Z$ where $Z \sim N(0,1)$, then in discrete simulations, the Wiener component is approximated as $\epsilon\sqrt{\Delta t}$ where $\epsilon \sim N(0,1)$, so equation 5 can be written in discrete form as equation 6 as follows [27].

$$P_t = P_{t-1}\, exp\left[\left[\mu - \frac{\sigma^2}{2}\right]\Delta t + \sigma\epsilon\sqrt{\Delta t}\right]. \tag{6}$$

2.8 Long Short-Term Memory (LSTM)

LSTM is a deep learning model specifically designed to address the vanishing gradient problem in RNNs. The fundamental difference between the LSTM model and RNN lies in the introduction of memory cells and the gate mechanism [29]. Cell states function as memory cells that carry and retain relevant information in the long term. The gate mechanism in the LSTM model consists of three parts: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). The forget gate removes irrelevant information, the input gate controls the information entering the cell state, and the output gate regulates the information used as the hidden state (h_t) for the next step [30]. The structure of the LSTM neural network is visualized in Figure 1 as follows [31].

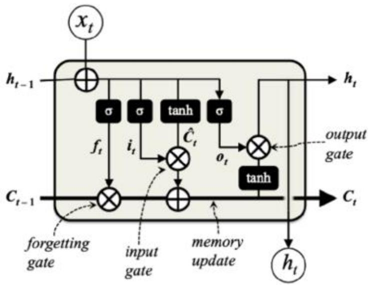

Figure 1. LSTM neural network structure

There are activation functions used in each gate, namely the sigmoid and tanh functions. The sigmoid function (σ) gives a value ranging from 0 to 1, while the tanh function gives a value between -1 and 1. The sigmoid and tanh activation functions, respectively, are: $\sigma(x) = \frac{1}{1+e^{-x}}$ and $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. The equations for the forget gate (f_t), input gate (i_t), output gate (o_t), and hidden state (h_t) are presented as follows: $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$, $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$, $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$, and $h_t = o_t \otimes \tanh(C_t)$. The $\tilde{C}_t$ and C_t values are calculated as follows: $\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$ and $C_t = C_{t-1} \otimes f_t + \tilde{C}_t \otimes i_t$.

2.9 Evaluation Model

Model evaluation is carried out using several evaluation metrics to measure the accuracy of the prediction model, including the following [32].

a. Root Mean Square Error (RMSE): calculates the square root of the average prediction error; large prediction errors will be penalized. The formula for RMSE is: $RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_t - A_t)^2}$.

b. Mean Absolute Error (MAE): calculates the average of the absolute differences between predicted and actual values; does not overly penalize outliers. The formula for MAE is: $MAE = \frac{1}{n}\sum_{t=1}^{n}|F_t - A_t|$.
c. Mean Absolute Percentage Error: calculates the average absolute percentage error between the predicted value and the actual value; has a very intuitive interpretation in terms of relative error. The formula for MAPE is: $MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\%$.

According to Lewis [33] as cited in Chen *et al.* [34], the lower the MAPE error percentage, the more reliable the prediction results. A MAPE value of less than 10% is categorized as "very accurate", 10%–20% as "good", 21%–50% as "reasonable", and more than 50% as "inaccurate".

3. RESULTS AND ANALYSIS

3.1 Descriptive Data Analysis

A descriptive analysis of the data was conducted to identify the characteristics and patterns of the historical data. The results of the descriptive analysis are presented in Table 2 below.

Table 2. Descriptive analysis

Minimum	Maximum	Data Range	Mean	Median	Standard Deviation
5342.140	9134.700	3792.560	7414.364	7312.615	711.732

Based on Table 2, the results of the descriptive analysis show that the JCI data used in this study have a minimum value of 5,342.140 and a maximum value of 9,134.700. The mean is recorded at 7,414.364, while the median is 7,312.615. The relatively small difference between the mean and the median indicates that the data distribution tends to be symmetrical. The standard deviation of 7 11.732 describes the degree of deviation from the mean. When compared to the mean, this standard deviation indicates that the data variability is moderate, meaning it is not too extreme but still reflects the dynamics of price movements.

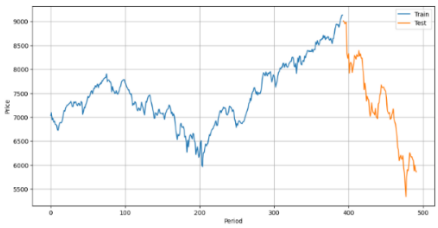

Figure 2. Plot the daily price of JCI

The data in this study were divided into training data and test data, with a split of 80% training data and 20% test data; the distribution of training and test data is shown in Figure 2. As shown in Figure 2, the training data exhibits an upward and downward trend pattern during certain periods, whereas the testing data shows a downward trend until the end of the period. It can also be observed that during the testing period, there was an extreme decline over a relatively short timeframe. This indicates the presence of significant volatility, signaling that the market is in an unstable condition.

3.2 Average Based FTSMC Modeling

The steps for modeling the Average-Based FTSMC are presented as follows:

a. Define the universe of discourse
From the training data, the minimum (Y_{min}) and maximum (Y_{max}) values were determined in advance; the minimum and maximum values obtained were 5967.99 and 9134.7, respectively. In this study, d_1 = 625.85 and d_2 = 17.44 were selected, resulting in the following universal set: $U = [Y_{min} - d_1, Y_{max} + d_2] = [5967.99 - 17.44\,, 9134.7 + 625.85] = [5342.14, 9152.14]$.

b. Determine the length and number of intervals
The first step in determining the interval length using an average-based algorithm is to calculate the average absolute value between two observations. The average absolute value is as follows:

$$M = \frac{\sum_{t=1}^{393}|Y_{t+1} - Y_t|}{393} = 57.323.$$

The length of the provisional interval is obtained by dividing the previously calculated absolute mean by two, resulting in l = 28.662. Next, based on the base table, the length of the interval is rounded to l = 30. Based on the length of the interval, a total of 127 subintervals were formed. The subintervals formed include the following: u_1 =[5342.14, 5372.14], … , u_{127} =[9122.14, 9152.14].

c. Determining fuzzy sets

The fuzzy sets A_i are defined based on the intervals obtained from the partitioning of the universal set, where each interval u_i represents a fuzzy set A_i. The fuzzy sets in this study are defined as follows:

$$A_1 = \left\{\frac{1}{u_1}, \frac{0.5}{u_2}, \frac{0}{u_3}, \dots, \frac{0}{u_{125}}, \frac{0}{u_{126}}, \frac{0}{u_{127}}\right\}, \dots, A_{127} = \left\{\frac{0}{u_1}, \frac{0}{u_2}, \frac{0}{u_3}, \dots, \frac{0}{u_{125}}, \frac{0.5}{u_{126}}, \frac{1}{u_{127}}\right\}.$$

d. Fuzzify of historical data

Each historical data point is fuzzified into the corresponding fuzzy set. Data points falling within subinterval u_i are mapped to fuzzy set A_i. The results of the data fuzzification are shown in Table 3.

Table 3. Data fuzzification results

No	Data	Fuzzification	No	Data	Fuzzification	No	Data	Fuzzification
1	7036.19	A_{57}	5	6897.95	A_{52}	391	9075.41	A_{125}
2	7099.31	A_{59}	6	6921.55	A_{53}	392	9133.87	A_{127}
3	6947.67	A_{54}	⋮	⋮	⋮	393	9134.70	A_{127}
4	6974.90	A_{55}	390	9032.58	A_{124}	394	9010.33	A_{123}

e. Compiling FLR and FLRG

The FLR is constructed by linking the fuzzification results for time period t with those for the subsequent time period t+1. The details of the FLR generated from this process are presented in Table 4 below.

Table 4. Fuzzy Logical Relationship results

No	Data	FLR	No	Data	FLR
1	7036.19	-	⋮	⋮	⋮
2	7099.31	$A_{57} \rightarrow A_{59}$	392	9133.87	$A_{125} \rightarrow A_{127}$
3	6947.67	$A_{59} \rightarrow A_{54}$	393	9134.70	$A_{127} \rightarrow A_{127}$
4	9010.33	$A_{127} \rightarrow A_{123}$	394	9010.33	$A_{127} \rightarrow A_{123}$

FLRs with the same left-hand side are grouped into the same group. The grouping of FLR results into a group is referred to as a FLRG. The results of the FLRG formation are shown in Table 5.

Table 5. Fuzzy Logical Relationship Group results

Group	FLRG	Group	FLRG
1	$A_1 \rightarrow \emptyset$	31	$A_{31} \rightarrow (1)A_{28}, (1)A_{31}, (1)A_{35}, (1)A_{40}$
2	$A_2 \rightarrow \emptyset$	⋮	⋮
⋮	⋮	126	$A_{126} \rightarrow \emptyset$
30	$A_{30} \rightarrow (1)A_{33}, (1)A_{38}$	127	$A_{127} \rightarrow (1)A_{123}, (1)A_{127}$

f. Constructing a Markov transition probability matrix

The results of the FLRG formation were then converted into a 127 × 127 Markov transition probability matrix.

$$P = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

g. Calculate predictions and adjustment values

Predictions and adjustment values were calculated in accordance with existing rules. The initial prediction results and the adjusted prediction results are shown in Table 6.

Table 6. Preliminary and final prediction results

No	Data	Initial Prediction	Adjustment Value	Final Prediction
1	7036.19	-	-	-
2	7099.31	7070.890	30	7100.890
⋮	⋮	⋮	⋮	⋮
394	9010.33	9075.920	-75	9000.920

3.3 Geometric Brownian Motion Modeling

Modeling Geometric Brownian Motion begins by calculating the return using the following log-return formula: $r_t = log\left(\frac{S_t}{S_{t-1}}\right)$, where S_t represents the price at time t and S_{t-1} represents the price at time $t-1$. The log return data were then tested for normality using the Kolmogorov-Smirnov test to verify that they met the assumptions of the Geometric Brownian Motion model. Visually, the distribution of the log returns can be seen in Figure 3.

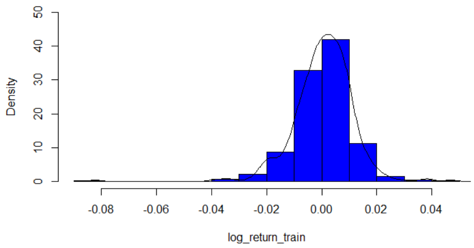

Figure 3. Plot of the log-return histogram for the JCI

Figure 3 shows that the histogram of the log returns has a bell-shaped distribution, indicating that the log returns follow a normal distribution. To confirm that the log return data follow a normal distribution, a Kolmogorov-Smirnov test was conducted. The Kolmogorov-Smirnov test yielded values of D_α = 0.0685, D = 0.0676, and a p-value of 0.0552. Based on these results, it is evident that D = 0.0676 $<$ D_α = 0.0685. In addition, the p-value is 0.0552, which is greater than the significance level α = 0.05. Thus, since $D < D_\alpha$ and the p-value $> \alpha$, the log returns are said to be normally distributed. Next, to estimate the drift and volatility parameters used in the GBM model, the mean and standard deviation of the returns are first calculated as follows.

$$\bar{r} = \frac{1}{393}\sum_{t=1}^{393} r_t = 0.000629, \text{ and } s = \sqrt{\frac{1}{392}\sum_{t=1}^{393}(r_t - \bar{r})^2} = 0.011.$$

The values of the volatility (σ) and drift (μ) parameters are calculated as follows, respectively:

$$\sigma = \frac{s}{\sqrt{\Delta t}} = 0.2188, \text{ and } \mu = \frac{\bar{r}}{\Delta t} + \frac{\sigma^2}{2} = 0.2719.$$

Therefore, the Geometric Brownian Motion model in this study is expressed as follows:

$$P_t = P_{t-1}\, e^{\left(\left[0.2719 - \frac{(0.2188)^2}{2}\right]\Delta t + \sigma\epsilon\sqrt{\Delta t}\right)} = P_{t-1} e^{(0.247963\Delta t + 0.2188\epsilon\sqrt{\Delta t})}.$$

Based on the GBM model that has been developed, a Monte Carlo simulation is then performed by generating 100 random numbers drawn from a standard normal distribution. This simulation will produce 100 different prediction trajectories for the GBM model. Therefore, the median of all the resulting trajectories will be calculated with a 90% confidence interval. The results of the Monte Carlo simulation for the GBM model are shown in Figure 4.

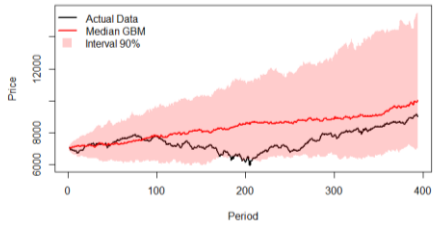

Figure 4. Results of the GBM model simulation

As shown in Figure 4, the red line represents the median plot of all trajectories. This plot does not closely follow the pattern of the actual data. It exhibits an upward trend throughout the period, which is consistent with a positive drift. The median plot from this GBM Monte Carlo simulation has not yet been able to capture the pattern of the actual data well. Nevertheless, the 90% confidence interval generated from the simulation still successfully covers nearly the entire range of the actual data's movement.

3.4 LSTM Modeling

Before the model training phase began, all data were first standardized using the min-max normalization technique so that their values were distributed within the range of 0 to 1. The normalized data were then transformed into an input format compatible with the LSTM architecture. Hyperparameter optimization was performed using a grid search approach and evaluated using a walk-forward validation scheme to ensure the model's generalization to previously unseen data. Forward validation was performed using a 5-fold expanding window scheme on the training data. The size of the training data in each fold increased progressively, while the validation data in that fold always consisted of observations following the training data period, thereby preserving the data sequence. The random seed value, set to 123 to ensure the reproducibility of the results, was applied during both the hyperparameter selection and model training processes. The hyperparameters optimized in this study include the number of neurons, the learning rate, and the dropout value. The complete hyperparameter configuration and LSTM model architecture include: 1 LSTM layer, 1 Dense layer, 128 neurons, a learning rate of 0.01, a dropout rate of 0.01, a batch size of 32, 14 time steps, the ADAM optimizer, and the MSE loss function.

During the model training process, an early-stopping mechanism is implemented as a strategy to automatically halt training iterations if there is no improvement in performance on the validation data before the specified number of epochs is reached. The implementation of this mechanism aims to prevent both overfitting and underfitting while reducing computational time. The maximum number of epochs was set to 100, but the early stopping mechanism terminated the training process at the 61st epoch. The dynamics of the loss values on the training and validation data during the LSTM model training process are shown in Figure 5.

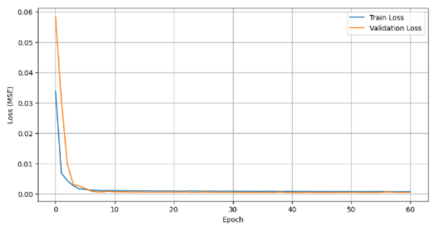

Figure 5. Plot train loss and validation loss

Figure 5 shows that in the early epochs; there is a very sharp decline in the loss values for both curves. This rapid decline indicates that the model successfully learned the underlying patterns of the JCI time series data efficiently during the early stages of training. Furthermore, between epochs 10 and 60, both curves show a stable (convergent) trend with very small and largely constant loss values. The training loss and validation loss curves also appear to overlap without any significant gaps throughout the training process.

3.5 Evaluation Model

The predictive models built during the training phase were then evaluated using test data. Visually, a comparison of the accuracy between the prediction results and the actual data for each model can be seen in Figure 6. Quantitatively, the measures of prediction error for each model can be seen in Table 7.

Table 7. Evaluation metric results

Model	MAPE	MAE	RMSE
Average Based FTSMC	0.925%	65.162	86.753
Geometric Brownian Motion	33.976%	2267.962	2541.952
LSTM	1.775%	123.797	166.885
Naïve (Persistence)	1.569%	110.161	152.734

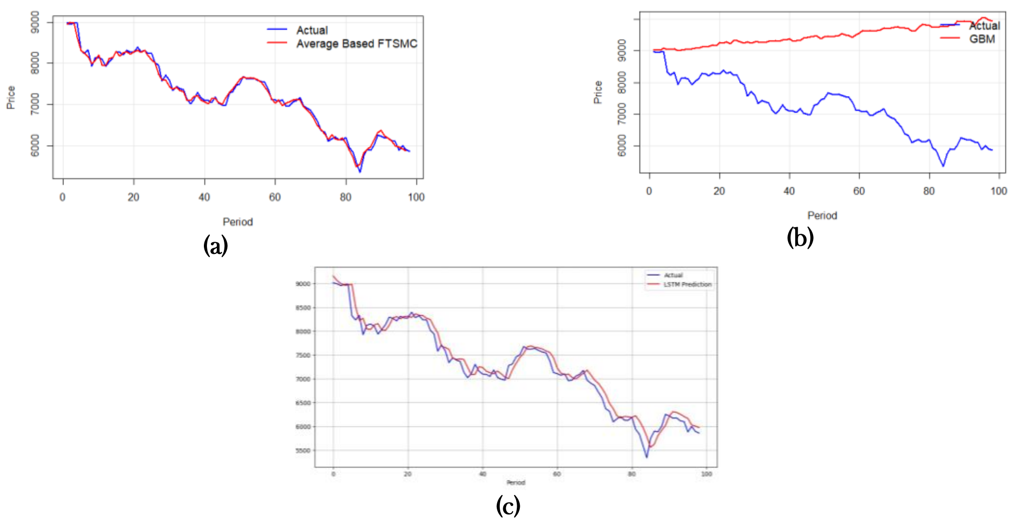

Figure 6. Prediction results: (a) Average-based FTSMC, (b) Geometric Brownian Motion, (c) LSTM

The evaluation results show that the Average-Based FTSMC model achieved the highest accuracy (MAPE 0.925%), followed by LSTM (MAPE 1.755%), and GBM (MAPE 33.976%). A similar pattern was consistently observed for MAE and RMSE, where Average-Based FTSMC and LSTM outperformed GBM. Based on the MAPE criterion, Average-Based FTSMC and LSTM fell into the "highly accurate" category, while GBM fell into the "reasonable" category. For comparison, a naive model (persistence, $\hat{y}_t = y_{t-1}$) was also evaluated over the same test period, yielding a MAPE of 1.569%, an MAE of 110.161, and an RMSE of 152.734. Compared to the naive model, Average-Based FTSMC demonstrated a significant reduction in MAPE. Average-Based FTSMC excels at capturing fluctuation patterns with high precision (Figure 6(a)), as the formation of fuzzy intervals and probability matrices can represent the movements of the JCI. LSTM (Figure 6(c)) is capable of learning nonlinear dependencies and long-term temporal patterns. In contrast, GBM (Figure 6(b)) fails to closely follow actual data patterns due to the assumptions of constant drift and volatility, and thus does not adequately represent actual market conditions.

Next, to test whether the differences in accuracy among the models are statistically significant, a paired test was conducted on the absolute error values of each model. This paired design was used because each model was evaluated using the exact same test dataset, so the absolute error values of the different models on the same data samples are paired with one another. The hypothesis tested for each pair of models is formulated as follows:
$H_0: \mu_d = 0$ (There is no significant difference in the mean error between the two model)
$H_1: \mu_d \neq 0$ (There is a significant difference in the mean error between the two models)

Before conducting the model comparison test, the Shapiro-Wilk normality test (α = 0.05) was applied to the absolute error distributions of each model. The test results indicated that the differences in absolute error between model pairs were not normally distributed (p-value = 0.001 < α); therefore, the Wilcoxon signed-rank test was used as a nonparametric alternative. The analysis was conducted on three model pairs (Average-Based FTSMC–GBM, Average-Based FTSMC–LSTM, and GBM–LSTM) with n = 98 paired observations covering the entire testing period. The test results show that Average-Based FTSMC consistently has a smaller absolute error than GBM (V = 0, p-value < 2.2×10^{-16}) and LSTM (V = 0, p-value = 1.46×10^{-9}) at all observation points. LSTM also consistently outperformed GBM (V = 0, p-value < 2.2×10^{-16}). Since the three tests were conducted simultaneously, a Bonferroni correction was applied with α adjusted to 0.05/3 = 0.0167. After correction, all three comparisons remained significant (p-value < 0.0167), confirming that the performance differences between models are consistent and not merely artifacts of repeated testing. Overall, these results reinforce the main finding of this study that the Average-Based FTSMC is statistically superior to both comparison models, with LSTM ranking second and GBM showing the weakest forecasting performance on the JCI dataset. The consistent superiority of the Average-Based FTSMC model demonstrates that a fuzzy-theory approach combined with a Markov chain is more adaptive in capturing the uncertainty and nonlinearity of JCI data. This superiority can be explained by the differences in the parameter estimation mechanisms of the three models. Unlike the GBM, which assumes constant drift (μ) and volatility (σ), the Average-Based FTSMC forms an FLRG that represents fuzzy state transitions based on the latest conditions. This Markovian property allows the model to adjust its predictions directly to local changes in the data without requiring the assumption of long-term stationarity. Meanwhile, although the LSTM is capable of capturing nonlinear temporal patterns through its gating mechanism, the trained weights remain constant during testing; thus, its adaptability is limited to patterns learned during training, and it is unresponsive to new data outside the training distribution. This mechanistic difference aligns with the empirical findings of the study, in which the Average-Based FTSMC significantly outperformed the GBM at every observation point. Thus, the Average-Based FTSMC demonstrated the best performance in predicting the JCI during the study period.

4. CONCLUSION

This study aims to compare the performance of the Average-Based FTSMC, Geometric Brownian Motion (GBM), and Long Short-Term Memory (LSTM) models in predicting movements in the Jakarta Composite Index from June 1, 2024, to June 29, 2026. Evaluation results using the MAPE, MAE, and RMSE metrics show that, for the specific data characteristics and test period, the Average-Based FTSMC model provides the highest prediction accuracy with a MAPE of 0.925%, an MAE of 65.162, and an RMSE of 86.753, followed by the LSTM model, while the GBM model produces significantly higher error rates. These findings indicate that, in the case of the JCI data studied, approaches based on fuzzy logic and machine learning are better able to capture nonlinear patterns and unstable fluctuations compared to classical stochastic models that assume constant drift and volatility. Therefore, the selection of a prediction model must take into account the specific statistical characteristics and temporal dynamics of the relevant stock market data.

5. REFERENCES

[1] M. Chang and Y. Li, "Impact of capital market volatility on economic growth - An analysis based on stochastic volatility model," *Heliyon*, vol. 10, no. 3, pp. 1–16, 2024, doi: 10.1016/j.heliyon.2024.e25679.

[2] X. Ji, J. Wang, and Z. Yan, "A stock price prediction method based on deep learning technology," *Int. J. Crowd Sci.*, vol. 5, no. 1, pp. 55–72, 2021, doi: 10.1108/IJCS-05-2020-0012.

[3] M. Hamidah and R. Putra, "Factors that encourage investors to decide to invest in shares in the capital market for the 2020-2023 period (Covid-19 pandemic): Systematic Literature Review," *Int. J. Soc. Sci. Hum. Res.*, vol. 7, no. 05, pp. 2604–2611, 2024, doi: 10.47191/ijsshr/v7-i05-08.

[4] R. Bhowmik and S. Wang, "Stock market volatility and return analysis : A systematic literature review," *Entropy*, vol. 22, no. 552, pp. 1–18, 2020, doi: 10.3390/e22050522.

[5] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4741–4753, 2021, doi: 10.1007/s00521-020-05532-z.

[6] Fuad and I. Yuliadi, "Determinants of the composite stock price index (IHSG) on the Indonesia stock exchange," *J. Econ. Res. Soc. Sci.*, vol. 5, no. 1, pp. 27–41, 2021, doi: 10.18196/jers.

[7] P. Sia, C. Puah, C.-M. Leong, K.-J. Yii, and M. M.-J. Tang, "Does inflation or interest rate matter to Indonesian stock prices ? An asymmetric approach," *Emerald*, vol. 27, no. 1, pp. 72–86, 2024, doi: 10.1108/JED-07-2024-0239.

[8] N. N. Marpaung, I. Rini, and D. Pangestuti, "Macroeconomic factors and Jakarta Stock Exchange : A comparative analysis pre- and until the COVID-19 pandemic," *SAGE*, vol. 14, no. 2, pp. 1–15, 2024, doi: 10.1177/21582440241247894.

[9] V. Bratian, A.-M. Acu, D. M. Mihaiu, and R.-A. Serban, "Geometric brownian motion (GBM) of stock indexes and financial market uncertainty in the context of non-crisis and financial crisis scenarios," *Mathematics*, vol. 10, no. 3, pp. 1–23, 2022, doi: 10.3390/math10030309.

[10] A. K. Sinha, "The reliability of geometric Brownian motion forecasts of S&P500 index values," *J. Forecast.*, vol. 40, no. 8, pp. 1444–1462, 2021, doi: 10.1002/for.2775.

[11] A. Sinha, "Daily and Weekly Geometric Brownian Motion Stock Index Forecasts," *Risk Financ. Manag.*, vol. 17, no. 10, pp. 1–22, 2024, doi: 10.3390/jrfm17100434.

[12] A. Illa *et al.*, "Performance analysis of ARIMA , LSTM , and hybrid ARIMA-LSTM in forecasting the composite stock price index," *Cauchy*, vol. 10, no. 2, pp. 588–604, 2025, doi: 10.18860/cauchy.v10i2.33379.

[13] J. K. Mutinda and A. K. Langat, "Stock price prediction using combined GARCH-AI models," *Sci. African*, vol. 26, pp. 1–14, 2024, doi: 10.1016/j.sciaf.2024.e02374.

[14] H. Guan, H. Jie, S. Guan, and A. Zhao, "A novel fuzzy-Markov forecasting model for stock fluctuation time series," *Evol. Intell.*, vol. 13, no. 2, pp. 133–145, 2020, doi: 10.1007/s12065-019-00328-0.

[15] S. Aulia, I. W. Mangku, and I. G. P. Purnaba, "Perbandingan kinerja model FTS-Markov chain dan geometric brownian motion dalam memprediksi harga saham BBRI," *Milang*, vol. 21, no. 2, pp. 163–177, 2025, doi: 10.29244/milang.21.2.163-177.

[16] K. Ramadani and D. Devianto, "The forecasting model of Bitcoin price with fuzzy time series Markov chain and chen logical method," *AIP Conf. Proc.*, vol. 2296, no. 1, p. 020095, 2023, doi: 10.1063/5.0032178.

[17] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," *IEEE Access*, vol. 11, pp. 48322–48331, 2023, doi: 10.1109/ACCESS.2023.3276628.

[18] B. Sabzipour *et al.*, "Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment," *J. Hydrol.*, vol. 627, 2023, doi: 10.1016/j.jhydrol.2023.130380.

[19] A. Yunita *et al.*, "Performance analysis of neural network architectures for time series forecasting : A comparative study of RNN , LSTM , GRU , and hybrid models," *MethodsX*, vol. 15, p. 103462, 2025, doi: 10.1016/j.mex.2025.103462.

[20] V. Deswal and D. Kumar, "Hyperparameter tuning of the LSTM model for stock price prediction," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 4, pp. 705–712, 2024.

[21] R. A. Ferdiansyah and E. Utami, "Komparasi model transformer dan LSTM dalam prediksi Harga Saham Blue Chip di BEI," *JNANALOKA*, vol. 7, no. 1, pp. 51–59, 2024, doi: 10.36802/jnanaloka.2026.v7-no01-%25p.

[22] R. C. Tsaur, "A fuzzy time series-Markov chain model with an application to forecast the exchange rate between the Taiwan and us Dollar," *Int. J. Innov. Comput. Inf. Control*, vol. 8, no. 7, pp. 4931–4942, 2012.

[23] S. Hariyanto, Y. D. Sumanto, S. Khabibah, and Zaenurrohman, "Average-based fuzzy time series Markov chain based on frequency density partitioning," *J. Appl. Math.*, vol. 1, pp. 1–9, 2023, doi: 10.1155/2023/9319883.

[24] K. Huarng, “Effective lengths of intervals to improve forecasting in fuzzy time series,” *Fuzzy Sets Syst.*, vol. 123, no. 3, pp. 387–394, 2001, doi: 10.1016/S0165-0114(00)00057-9.
[25] A. Zakiah, E. Sulistianingsih, and N. Satyahadewi, “Geometric brownian motion with jump diffusion and value at risk analysis of PT Bank Negara,” *Barekeng*, vol. 19, no. 1, pp. 617–628, 2025, doi: 10.30598/barekengvol19iss1pp617-628.
[26] R. S. Tsay, *Analysis of financial time series*, 2nd ed. Canada: John Wiley & Sons, Inc, 2010. doi: 10.1002/9780470644560.
[27] J. C. Hull, *Options, Futures, and Other Derivatives*, 11th ed. United Kingdom: Pearson Education, 2022.
[28] M. Alhagyan, “The effects of incorporating memory and stochastic volatility into GBM to forecast exchange rates of Euro,” *Alexandria Eng. J.*, vol. 61, no. 12, pp. 9601–9608, 2022, doi: 10.1016/j.aej.2022.03.036.
[29] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, “Applications of long short-term memory (LSTM) networks in polymeric sciences: A review,” *Polymers (Basel).*, vol. 16, no. 18, pp. 1–44, 2024, doi: 10.3390/polym16182607.
[30] M. Krichen and A. Mihoub, “Long short-term memory networks: A comprehensive survey,” *AI*, vol. 6, no. 9, pp. 1–21, 2025, doi: 10.3390/ai6090215.
[31] A. Lawi, H. Mesra, and S. Amir, “Implementation of long short-term memory and gated recurrent units on grouped time-series data to predict stock prices accurately,” *J. Big Data*, vol. 9, pp. 1–19, 2022, doi: 10.1186/s40537-022-00597-0.
[32] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE , MAE , MAPE , MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/peerj-cs.623.
[33] C. D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. London: Butterworth Scientific, 1982.
[34] R. J. C. Chen, P. Bloomfield, and J. S. Fu, “An evaluation of alternative forecasting methods to recreation visitation,” *J. Leis. Res.*, vol. 35, no. 4, pp. 441–454, 2003, doi: 10.1080/00222216.2003.11950005.