APPROVAL SHEET

This

is submitted in partial fulfillment of the requirements
for the degree of

Author:

Advisor:

Advisor:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

A

Presented to
the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

by

Application of Multivariate Fuzzy Time Series Models to Consumer Purchasing Decisions

Kevin Byrne[a], Matthew Bolton[b]

[a] University of Virginia Engineering Systems and Environment, 151 Engineer's Way, Charlottesville, Virginia, 22904, United States of America
[b] University of Virginia Engineering Systems and Environment, 151 Engineer's Way, Charlottesville, Virginia, 22904, United States of America

Abstract

Economists have long acknowledged that the population's imprecise and subjective perceptions of the economy follow true economic indicators. Researchers have primarily studied this phenomenon through the application of consumer confidence surveys to traditional econometric models. Fuzzy time series models are an alternative modeling paradigm that have been shown to accurately forecast financial and economic movements by leveraging qualitative and pattern-based reasoning inherent to human decision making. Despite this, nobody has assessed if simulating consumers' qualitative economic perceptions with fuzzy time series is a viable approach to forecasting their purchasing decisions. This paper addresses the gap in the literature by applying multivariate economic fuzzy time series to forecast vehicle purchases in the United States. We evaluated the utility of our approach by comparing the fuzzy time series models to Long-Short-Term-Memory (LSTM) and Vector Autoregressive (VAR) time series models. Results show that the fuzzy time series models perform comparatively or significantly better than LSTM and VAR models in out-of-sample forecasts of vehicle sales. These results suggest that the fuzzy time series approach could have significant future utility for forecasting and interpreting aggregate consumer purchasing trends by imitating their rationality.

Keywords: Fuzzy time series, econometrics, consumer confidence, fuzzy logic, LSTM

1. Introduction

Research has shown that most Americans lack knowledge of precise economic figures. In everyday conversations and news coverage, economic phenomenon are frequently described in qualitative terms (e.g. "the housing market is *tight*"; "stocks are *falling*"; "unemployment is *low*"). However, it is incorrect to assume that Americans are unaware of changing economic situations; their knowledge base, rather, is synthesized in non-quantitative terms. The need for anything more than vague descriptions of national statistics is seemingly unimportant to the purchasing decisions of average Americans.

Curtin (2019) documented the ignorance of specific economic knowledge during the Great Financial Crisis. In both 2007 and 2009, survey respondents were asked to provide the current rate of three readily published economic statistics: unemployment, Consumer Price Index (CPI), and Gross Domestic Product (GDP). Curtin's hypothesis was that attentiveness to economic statistics would increase during a highly publicized recession. Instead, only knowledge about the unemployment rate significantly increased (47% to 58%), while GDP remained roughly unchanged (23% to 25%), and CPI decreased (27% to 22%). Additionally, Curtin found, news media often report economic news in qualitative terms, "summarizing the latest statistic by using subjective phrases, such as economic growth had improved or worsened". Boydstun et al. (2018) established that there is a strong relationship between media tone (categorized as "positive", "neutral", or "negative") and consumer sentiment, even when controlled for by variations in leading economic indicators.

Despite widespread unfamiliarity with economic statistics, research has shown that Americans are keenly aware of economic changes. The strongest evidence of this phenomenon is documented in the work on consumer confidence. Consumer confidence surveys were originally designed in the late 1940s by Katona as a "means to directly incorporate empirical measures of consumer expectations into models of [discretionary] spending and saving behavior" (Curtin, 2007). Consumer confidence is largely measured by surveys that use verbal, qualitative, descriptions of economics statistics. In the literature, the University of Michigan Index of Consumer Sentiment (ICS) is the most studied metric of consumer confidence. The survey underlying the ICS asks participants to verbally describe their personal financial situation and forecast future economic changes, which are binned under terms like "positive", "negative", or "uncertain", and synthesized into a single summary measure (*Index Calculations*, University of Michigan).

Formal statistical analysis has consistently shown that consumer confidence is an accurate reflection of the economy. While seemingly imprecise, the three-point verbal scale utilized by the ICS is highly correlated with real economic statistics (Curtin, 2007). Barnes and Olivei (2017) showed that principal components of the ICS have statistically significant correlation to fundamental economic variables. They also asserted that confidence measures contain information in economic variance outside of traditional fundamental indicators. Carroll et al. (1994) concluded that consumer confidence can be used to accurately forecast spending using time series models. However, they show, consumer confidence surveys contributed marginal information to forecasts when controlled for by leading economic indicators.

To date, the research utilizing qualitative economic perception for spending forecasts has relied heavily on consumer confidence surveys like the ICS, despite documented methodological flaws in the surveys themselves. Dominitz and Manski (2004) argue the synthesis of a single consumer confidence index value results in substantial information loss. Pickering et al. (1973) point out the high degree of multicollinearity in consumer confidence survey questions, a problem which is compounded by the summation of the questions into a single index value. They additionally express concerns about the potential for information loss in only 3-4 response variables for each survey question. Their work points to a requirement for up to 11 response variables being "necessary to optimize information transmission".

Rather than use consumer confidence surveys, the contribution of this work is to mirror the approximate economic perceptions of consumers by directly modeling imprecision into quantitative economic variables themselves. Imprecision in the time series can be computed using Fuzzy Time Series models (FTSM). It was our hypothesis that, by virtue of using fuzzy logic (a logic designed to capture the imprecision of human reasoning), FTSMs could simulate consumers' subjective economic perceptions without the need for a confidence survey.

This is accomplished by transforming the time series of fundamental economic variables into interval-based variables representing imprecise linguistic descriptions. The quantitative economic time series are mapped to the new "fuzzy" variables, creating a "fuzzy time series" representing a non-quantitative perception of economic changes (see Figure 1). Time series fits and forecasts are determined by identifying patterns in the relational changes in the fuzzy variables.

This work attempts to see if mimicking consumers' imprecise economic rationality is an accurate modeling alternative for forecasting their purchasing decisions. Not only can this method offer insight into how qualitative perception of the economy impacts consumer purchasing decisions, it removes the issues inherent in using a survey-based consumer confidence medium.

The primary research question of this work is as follows: when forecasting consumer purchasing decisions that are influenced by an imprecise perception of the economy, does the qualitative treatment of a time series explicit to FTSMs result in a more accurate forecast than other time series models? As a barometer of modeling accuracy, Long-Short-Term-Memory (LSTM) and Vector Autoregressive (VAR) models are additionally fit for comparison.

2. Background

This section introduces the model paradigms tested (FTSM, LSTM, and VAR), and why they are selected for forecasting discretionary consumer purchases.

2.1. Fuzzy Logic and Fuzzy Time Series

Fuzzy logic has emerged as one of the preeminent methods of integrating human decision making and qualitative rationalization into complex systems modeling. Introduced by Zadeh in the 1960s, fuzzy logic is "concerned with the formal principles of approximate reasoning"; more specifically, "modeling the imprecise modes of reasoning that play an essential role in the remarkable human ability to make rational decisions in an environment of uncertainty and imprecision" (Zadeh, 1988). The power of fuzzy logic lies within its ability to describe quantitative variables in qualitative terms with varying degrees of precision. Linguistic predicates of a quantitative variable are established with distributions. These formalize the relationship between the verbal descriptions and the quantitative values of the variable.

Fuzzy time series models utilize the concepts of fuzzy set theory to describe temporal changes in non-numerical terms. Song and Chissom (1993) introduced the concept of fuzzy time series. Fuzzy time series map the values of a quantitative time series into intervals that represent linguistic descriptions. The two primary methods of establishing the intervals are to use equal length and uniform sets, or through utilizing a c-means algorithm (Duru and Yoshida, 2012). While increasing the number of intervals represents increased precision, it increases both the complexity and noise of the model. This work adopts the simplicity of uniform set to test the sensitivity of the number of intervals on the model accuracy.

Definition 1: Let F(t) be an univariate fuzzy time series. The fuzzy logical relationship is represented by:

$$F(t\text{–}n), \ldots, F(t\text{–}2), F(t\text{–}1) \rightarrow F(t) \tag{1}$$

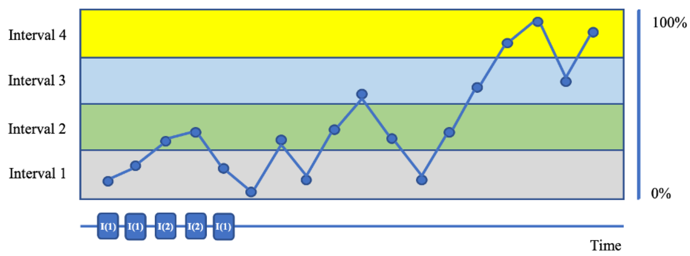

Figure 1: Representation of an univariate fuzzy time series

Arguably, the most important step in the construction of FTSMs is establishing the fuzzy relationships: the driver of the temporal changes in the time series. The complexity of the relationships is a function of the number of intervals and the number of variables in the model. Song and Chissom's (1993) method of computing fuzzy relations uses a min-max operation of matrix transpositions that does not translate easily to complex FTSMs. This work alternatively adopts an approach introduced by Egrioglu et al. (2009) that uses a feedforward neural network to establish the fuzzy relationships. Their method is especially efficient for high order and multivariate time series.

Definition 2 Let F and G1, G2, ..., Gk=1 be k-multivariate fuzzy time series. The fuzzy logical relationship is represented by:

$$\begin{aligned} &(F(t\text{–}1), G1(t\text{–}1)G2(t\text{–}1)...Gk\text{–}1(t\text{–}1)), ..., \\ &(F(t\text{–}n), G1(t\text{–}n)G2(t\text{–}n)...Gk = 1(t\text{–}n)) \rightarrow F(t) \end{aligned} \tag{2}$$

At each time-step, a fuzzy interval can be converted back into a quantitative value, a process called defuzzification. The method of defuzzification in this work adopts the mid-point approach used by Egrioglu et al. (2009), which is discussed in detail in the Methods section.

FTSMs have been applied with success to economic and financial forecasting, although the vast majority of the work has been focused on stock indices. Some successful applications to non-stock forecasts include the application of FTSMs to Taiwan's Export Orders Index (EOI) by Huang et al. (2019), and of short-term crude oil forecasting by Zhang et al. (2010). Little in the literature has been published on applications of multivariate FTSMs for non-stock econometrics, and even less with an emphasis on the qualitative rationality of the participants influencing the movements of time series.

2.1.1. LSTM Time Series Models

With the use of a neural-network to identify the fuzzy relationships in FTSMs, it is appropriate to compare the results of this work's FTSMs with a different neural network time series modeling paradigm. Recurrent neural networks (RNN) are supervised machine learning models that differ from other neural network modeling paradigms due to their ability to loop information, effectively enabling memory of past patterns. This allows the network to access past data when solving for node weights. However, exploding and vanishing gradient descent problems have been well documented with RNNs (Hochreiter and Schmidhuber, 1997).

To overcome gradient descent problems, a special type of RNN called the Long-Short-Term-Memory (LSTM) model was created. While LSTM models contain an outer loop indicative of other RNNs, LSTM models contain memory cells which allow gradients to flow through time, rather than be lost or diminished. The key premise behind LSTM models is the use of memory gates to handle long-term and complex dependencies. This makes it possible to store information for a longer period of time and helps to reduce gradient problems.

A high-level explanation of LSTM unit composition follows (for a more detailed description, see (Van Houdt et al., 2020)):

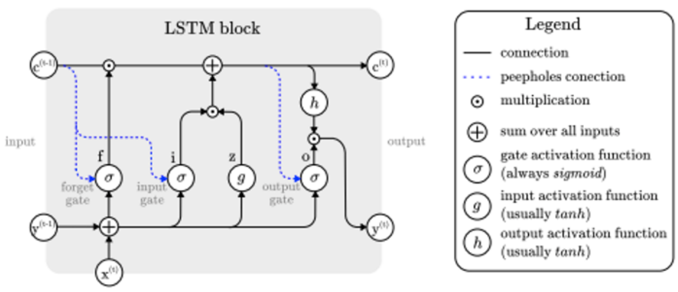

Figure 2: Vanilla LSTM Unit (Van Houdt, 2020)

The traditional LSTM unit is composed of a cell, an input gate, an output gate, and a forget gate (see Figure 2). The input gate and forget gate determine which information from the time-step should be retained in memory and which information should be removed from previous cell states. The parsed information from the input and forget gate are then synthesized together with the previous time-step input and the previous cell value. The updated cell state is combined with the current time-step input and the previous model output. This is passed to the output gate.

LSTM models have the ability to store and recall memory that mimics the sequence-classification of the brain (Bao et al., 2017). This is particularly useful for handling datasets with both obvious patterns – like cyclical business conditions and seasonality – with more sporadic patterns – like the memory of non-cyclical economic occurrences that could influence spending habits. Strong performance of financial and economic LSTM time series models is well documented. Siami-Namini et al. (2018) found that LSTM models consistently outperformed ARIMA models (which are a standard in economic forecasting) for forecasting international stock index returns. Ala'raj et al. (2021) showed that a bidirectional LSTM model was the most accurate amongst other neural network paradigms for forecasting credit card payment delinquencies. There are countless other published applications of LSTM models in the fields of economics and finance.

2.2. Vector Autoregressive Models

In more traditional econometrics literature, reduced-form vector autoregressive models are used to estimate time series utilizing multiple economic variables. VAR models are standard in describing economic movements, and many of the papers cited in this work, including Carroll et al. (1994), Barnes and Olivei (2017), and Bram and Ludvigson (1998) utilize this model paradigm. Reduced-form VAR models express all combinations of the variables involved as a set, or vector, of related, simultaneous regressions. In other words, reduced-form VAR estimates each variable's time series as a linear function of its own past values, as well as the past values of the other variables in the model. The respective regressions are estimated using ordinary least squares. Correlation between the variables is essential, otherwise the inclusion of an uncorrelated variable would result in no additional explanatory information in the estimated time series. As discussed in the Methods section, a Granger Causality test was conducted to check the correlation of the variables selected. For the purpose of this study, only the estimated model parameters and results for vehicle sales were examined.

3. Methods

3.1. Variable Selection and Transformation

The dependent variable chosen for this paper is the monthly percentage change in total vehicle sales in the United States. Vehicle sales represent a large discretionary purchase by consumers; the timing and type of vehicle purchased may be motivated by qualitative perceptions of economic strength. To simulate consumers' imprecise economic perception, three exogenous economic variables were chosen that are commonly digested in both the media and in the marketplace: Gross Domestic Product (GDP), Personal Consumption Expenditure (PCE), and Real Disposable Income (RDI). In colloquial terms, GDP can be thought of as aggregate output of the economy, PCE as the level of inflation, and RDI as excess consumer liquidity.

All data was downloaded from the St. Louis Federal Reserve Economic Data (FRED) website. The frequency of the time series was monthly, spanning from January, 1978 to December, 2019. To account for the changing economic expectations of consumers, models were fit and tested in three periods: 1) January, 1978 – July, 1990; 2) January, 1992 – July, 2002; and 3) January, 2004 – July, 2018. Common practice is to train a model on a sample set and test the model on an out-of-sample set. The out-of-sample accuracy is the primary gauge for the forecasting power and generalizability of a model. The last 36 months of each period were selected as the out-of-sample testing set.

It is standard practice in econometrics to take the log-difference of economic variables to satisfy stationarity tests. While many researchers utilizing FTSMs disregard stationarity as unimportant (Duru and Yoshida, 2012), the log difference of all variables in this work was taken in accordance to works like Carroll et al. (1994), although formal stationarity tests were not conducted.

The time series were then smoothed with a 3-month moving average. Finally, the values were normalized between 0 and 1. A Granger Causality test concluded that all variables were significant at the 95% confidence level for predicting future vehicle sales. Partial autocorrelation tests of the variables were conducted to inform significance of lags 2, 3, and 4 as significant at the 95% confidence level. Lags 2, 3, and 4 are used as model parameters.

3.2. *Fuzzy Time Series Model Design*

The architecture of the model is a feed forward neural network with one input layer, one hidden dense layer, and one output layer. The dimensions of the input layer were dependent on the number of lags tested for each model. The optimal number of nodes and epochs for both the input and dense layer were determined by training and validating the model on each period, and selected based on the MSE of the testing sets. The network consisted of one input layer with 32 neurons, one dense layer with 4 neurons, and was trained on 1000 epochs.

Fuzzy intervals from 3 to 12 are tested in this work, describing highly-imprecise linguistic economic perception (e.g. "low", "average", "high") to highly-precise linguistic economic perception (e.g. "historically low", "significantly low", "moderately low", "slightly below average", "average", ..., etc.).

The model was trained on the fuzzy time series of the three economic variables as well as vehicle sales, which gave an output of the predicted fuzzy time series for all four variables. For the purposes of this analysis, only the predicted fuzzy time series of vehicle sales was studied. Specifics of the process are outlined below.

Step 1. Define fuzzy sets by partitioning each time series into n subintervals. For this step, the time series were bounded by an additional 5% of their respective range of values. The intervals were evenly spaced and non-overlapping. n = [3,12] were tested.

Step 2. Fuzzify the observations. Each quantitative value of a time series was mapped to the corresponding fuzzy set to establish the fuzzy time series.

Step 3. Establish the fuzzy relationships and train the model. To establish fuzzy relationships, a feed forward neural network was employed. The model was trained on the training set for each period.

Step 4. Testing the model. Each trained model was fed the respective period's fuzzified testing set and was set aside for defuzzification.

Step 5. Defuzzification of the model outputs. The output of the FTSMs was a non-integer number for each time step, representing a specific point of precision within a fuzzy interval. This differs from previous work, where the output of FTSMs was the integer corresponding to the midpoint of a fuzzy interval. Consequently, the defuzzification for a fuzzy-forecast $\mathrm{X_i}$ is as follows:

Defuzzification. Let N be the fuzzy interval corresponding to the rounded-down integer of $\mathrm{X_i}$. Let M be the fuzzy interval corresponding to the rounded-up integer of $\mathrm{X_i}$. Let IN and IM be the midpoint of the respective fuzzy-intervals N and M. Defuzzification is calculated as:

$$((X_{\mathrm{i}} - N) \cdot (IM - IN)) + IN \tag{3}$$

3.3. *LSTM Model Design*

The LSTM model consisted of one input LSTM layer, one dense layer, and one output layer. The dimensions of the input layer were dependent on the number of lags for each model. The output layer consisted of four nodes, one for each of the four variables. The output type was a continuous number, the predicted monthly percentage change of the variables. Only the output of vehicle sales was considered for this work. Similarly to the FTSM, the optimal number of nodes and epochs was for both the input and dense layer were determined by training and testing the model on each different period and selecting the optimal combination for the testing set. The optimal nodes for the LSTM and dense layer respectively were 32 and 8. The model was trained on 1000 epochs.

3.4. *Performance Metrics*

Mean squared error (MSE; Equation 4) and mean absolute percentage error (MAPE; Equation 5) were used to calculate error on the training and testing sets for both models. Both error metrics are standard in both the econometrics and machine learning literature.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

4. Results

Unsurprisingly, for the fits of the FTSM and LSTM models in the training period (see Tables 1 and 2), there is a general trend of improved fit as the number of model parameters increases. However, for the FTSMs, the results indicate that the best model fit is not necessarily the one with the highest level of precision (i.e. the greatest number of fuzzy intervals). Interval I(10) had the lowest MAPE of the FTSMs in Period 1, while I(7) and I(11) had the lowest MAPE in the training set for Periods 2 and 3, respectively. Granted, this performance in error is largely marginal.

The LSTM and VAR models produce more accurate results in the training sets than most FTSMs in Periods 2 and 3, with LSTM models outperforming all others significantly in Period 1. Interestingly, an order of VAR(2) produces some of the most accurate VAR models in the three training periods, while an order of VAR(3) produces the worst model in all three training periods. PACF tests show significance of all three lag orders tested. The difference in model accuracy of the training sets appears to largely be a function of capturing more of the sudden variance during periods of sharp changes in monthly vehicle sales (See Figures 3, 4, 5). The LSTM models capture this variance most accurately, while the FTSMs modestly follow the shape of time series during these periods of volatility, and the VAR models often approximating a straight line through the mean of the time series.

The FTS and LSTM models were evaluated twenty times in the testing sets, with the average of the error metrics recorded below. Since VAR estimates coefficients using OLS, there is no reason to run the VAR model testing sets more than once; the error metrics for the VAR trials are included as well. In the testing sets, the FTSMs broadly outperformed the LSTM models; the FTSMs outperformed the VAR models in Periods 1 and 2, but were bettered by the VAR(2) model in Period 3, albeit marginally (see Tables 3 4). The best FTSMs for Periods 1 and 2 are approximately twice as accurate as the best LSTM model fits, which is surprising given their similar architecture. In Period 3, all three modeling paradigms performed exceptionally well with single-digit MAPE error rates, the difference in error between them being largely marginal.

It is important to note the range of errors between the three periods, which is correlated directly to the respective variance of the time series in the different testing periods (see Figure 7 7 for more information). For example, the Period 2 testing set contained both the maximum and minimum value of the entire period set (both training and testing), which occur within just a few sequential months. The error rate of the Period 2 testing set is thus skewed by this massive volatility and absence in the training set, while, conversely, the Period 2 training period has the lowest error.

The most complex FTSMs do not appear to be optimal given that the results in Tables 3 and 4 indicate that simpler models are highly generalizable and accurate. For instance, in Period 1, I(7) produces the most accurate model in the testing set and performs strong across all lags. For Period 2, I(5) and I(9) produce accurate models across all lags, while again I(7) produces the best model measured by MAPE. For Period 3, the I(9) through I(12) all produce consistently accurate models, while I(6) produces the best model measured by MAPE.

Additional investigation into the best subset models was conducted. Table 5 is a direct comparison of the best model parameters, according to training MSE. One exception is for Period 3, where the FTSM with Interval 11 and Lag 2 was selected over the one with Interval 12 and Lag 2 due to the poor training fit of Interval 12. The best models were then trained and tested 100 times to quantify the variance of the out-of-sample test results. The FTSM models had significantly less variance in the model test results, as shown in Figures 9, 10, and 11).

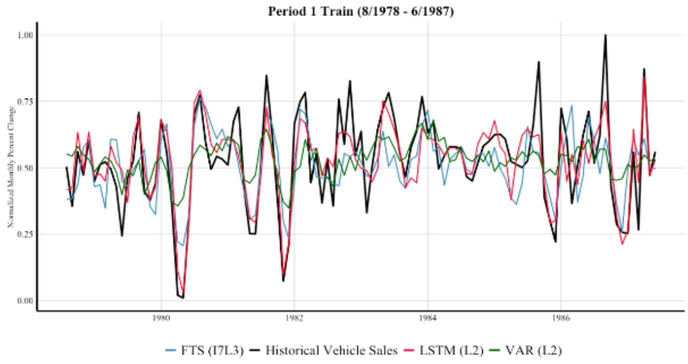

Figure 3: Period 1 Train Time Series

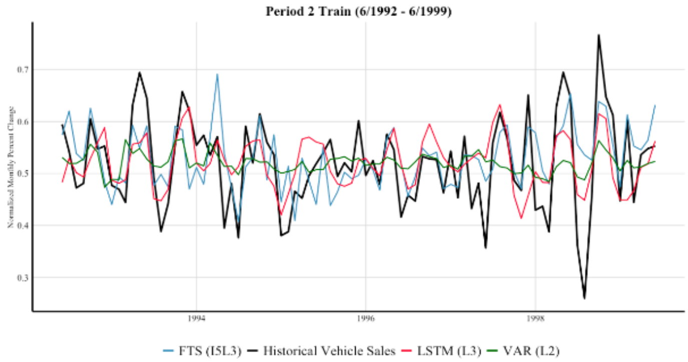

Figure 4: Period 2 Train Time Series

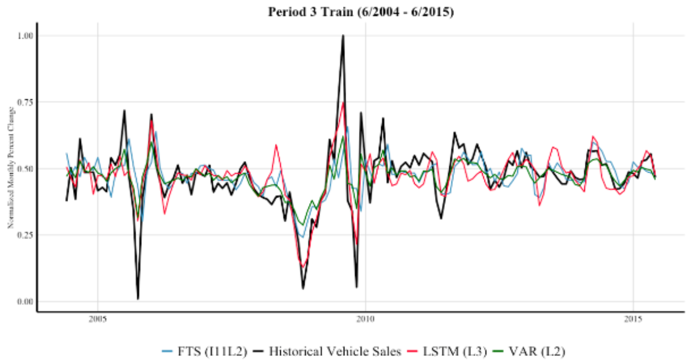

Figure 5: Period 3 Train Time Series

Table 1: Training MSE

	Period 1 Train (8/78 - 6/87)			Period 2 Train (6/92 - 6/99)			Period 3 Train (6/04 - 6/15)		
	Lag 2	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**
Interval 3	0.02370	0.02425	0.02339	0.00821	0.00867	0.00836	0.01341	0.01297	0.01106
Interval 4	0.02515	0.01885	0.01861	0.01009	0.00845	0.00885	0.01164	0.01151	0.01110
Interval 5	0.02526	0.01837	0.01749	0.00740	0.00739	0.00688	0.01146	0.01069	0.01078
Interval 6	0.02272	0.02102	0.01835	0.00923	0.00812	0.00734	0.01021	0.01059	0.00888
Interval 7	0.02421	0.01877	0.01694	0.00843	0.00777	0.00528	0.01002	0.01080	0.00963
Interval 8	0.02410	0.01840	0.01680	0.00867	0.00771	0.00619	0.01093	0.01023	0.01017
Interval 9	0.02264	0.02100	0.01994	0.00841	0.00802	0.00633	0.01086	0.01040	0.01020
Interval 10	0.02421	0.01778	0.01625	0.00804	0.00791	0.00661	0.00939	0.00922	0.00892
Interval 11	0.02528	0.02148	0.01858	0.00852	0.00848	0.00719	0.00983	0.01010	0.00973
Interval 12	0.02223	0.01820	0.01532	0.00837	0.00824	0.00610	0.01004	0.00984	0.00896
LSTM	0.01875	0.01178	0.00810	0.00701	0.00480	0.00429	0.00838	0.00491	0.00458
VAR	0.01936	0.04427	0.03519	0.00614	0.01423	0.00755	0.00790	0.01439	0.01293

Table 2: Training MAPE

	Period 1 Train (8/78 - 6/87)			Period 2 Train (6/92 - 6/99)			Period 3 Train (6/04 - 6/15)		
	Lag 2	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**
Interval 3	0.63417	0.57731	0.58571	0.14174	0.15022	0.15023	0.67620	0.63456	0.59505
Interval 4	0.77942	0.66432	0.66618	0.15453	0.13970	0.15012	0.61454	0.48417	0.45942
Interval 5	0.63586	0.51818	0.51689	0.13790	0.13436	0.13352	0.54959	0.40032	0.41441
Interval 6	0.58304	0.49050	0.50515	0.14597	0.13646	0.12943	0.51818	0.42748	0.41456
Interval 7	0.64176	0.54127	0.53987	0.14600	0.13718	0.11781	0.40637	0.49686	0.47034
Interval 8	0.67647	0.56826	0.57079	0.14016	0.13344	0.12167	0.53792	0.39233	0.39192
Interval 9	0.54317	0.49278	0.48444	0.14178	0.13921	0.12546	0.56325	0.44960	0.44759
Interval 10	0.58228	0.49010	0.45035	0.13896	0.14192	0.12877	0.38292	0.37365	0.39046
Interval 11	0.64872	0.60525	0.59464	0.14497	0.13817	0.13285	0.34993	0.33912	0.34381
Interval 12	0.56994	0.50015	0.46590	0.14174	0.14019	0.12532	0.52458	0.42041	0.41491
LSTM	0.44864	0.40152	0.29926	0.13281	0.10990	0.10466	0.46774	0.32723	0.33364
VAR	0.73307	1.04238	0.90426	0.12830	0.19070	0.13705	0.56782	0.76164	0.70258

Table 3: Testing MSE

	Period 1 Test (11/87 - 7/90)			Period 2 Test (11/99 - 7/02)			Period 3 Test (11/15 - 7/18)		
	Lag 2	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**
Interval 3	0.01331	0.01593	0.01822	0.02947	0.03494	0.03309	0.00521	0.00502	0.00464
Interval 4	0.01690	0.01532	0.01474	0.03745	0.04718	0.04703	0.00332	0.00215	0.00179
Interval 5	0.01360	0.01681	0.02366	0.02075	0.02229	0.02872	0.00399	0.00422	0.00528
Interval 6	0.01646	0.01615	0.01766	0.03346	0.03404	0.02911	0.00240	0.00190	0.00260
Interval 7	0.01421	0.01188	0.01304	0.02974	0.03379	0.02697	0.00233	0.00321	0.00413
Interval 8	0.01540	0.01697	0.01609	0.03301	0.03426	0.02908	0.00278	0.00265	0.00285
Interval 9	0.01745	0.01514	0.01552	0.02713	0.02439	0.02303	0.00206	0.00225	0.00210
Interval 10	0.01573	0.01794	0.01988	0.03013	0.03043	0.02919	0.00269	0.00277	0.00210
Interval 11	0.01578	0.01432	0.01862	0.03070	0.03278	0.03197	0.00220	0.00240	0.00211
Interval 12	0.01645	0.01723	0.01560	0.03297	0.03248	0.02775	0.00176	0.00200	0.00207
LSTM	0.02323	0.02542	0.04978	0.03374	0.03305	0.04753	0.00212	0.00208	0.00241
VAR	0.01249	0.02659	0.02334	0.01941	0.03926	0.02067	0.00085	0.00165	0.00170

Table 4: Testing MAPE

	Period 1 Test (11/87 - 7/90)			Period 2 Test (11/99 - 7/02)			Period 3 Test (11/15 - 7/18)		
	Lag 2	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**	**Lag 2**	**Lag 3**	**Lag 4**
Interval 3	0.20352	0.22721	0.23891	1.34452	1.85266	1.86666	0.15340	0.15029	0.14151
Interval 4	0.24105	0.24220	0.23625	1.67323	1.46380	1.43177	0.09862	0.08449	0.08135
Interval 5	0.20462	0.21845	0.24492	1.34524	1.04728	0.93655	0.12380	0.12938	0.14464
Interval 6	0.23515	0.22418	0.24062	1.47179	1.43226	1.59098	0.08890	0.07450	0.09133
Interval 7	0.21600	0.18327	0.19242	1.44139	1.46222	0.85533	0.08555	0.10820	0.12489
Interval 8	0.21502	0.22411	0.21126	1.53537	1.26548	1.24811	0.10086	0.10262	0.10458
Interval 9	0.24392	0.22444	0.22486	1.33403	1.30160	1.06942	0.08366	0.08848	0.08569
Interval 10	0.23651	0.25194	0.23974	1.55611	1.30446	1.23910	0.10061	0.09823	0.08495
Interval 11	0.23264	0.22131	0.23019	1.51870	1.37338	1.49411	0.08631	0.09432	0.08342
Interval 12	0.22674	0.22392	0.19982	1.54320	1.26848	1.02940	0.07490	0.08435	0.08335
LSTM	0.28074	0.29888	0.36989	1.68853	1.60820	2.14253	0.08067	0.08081	0.08313
VAR	0.20777	0.29246	0.29426	1.35408	2.07079	1.48217	0.05351	0.07512	0.07402

Table 5: Best Subsets

Period 1 Test (11/87 - 7/90)

Best Model	MSE	MAPE
FTSM (I7L3)	0.01188	0.18327
LSTM(L2)	0.02323	0.28074
VAR (L2)	0.01249	0.20777

Period 2 Test (11/99 - 7/02)

Best Model	MSE	MAPE
FTSM (I5L3)	0.02229	1.04728
LSTM (L3)	0.03374	1.60820
VAR (L2)	0.01941	1.35408

Period 3 Test (11/15 - 7/18)

Best Model	MSE	MAPE
FTSM (I11L2)	0.00220	0.08631
LSTM (L3)	0.00208	0.08081
VAR (L2)	0.00085	0.05351

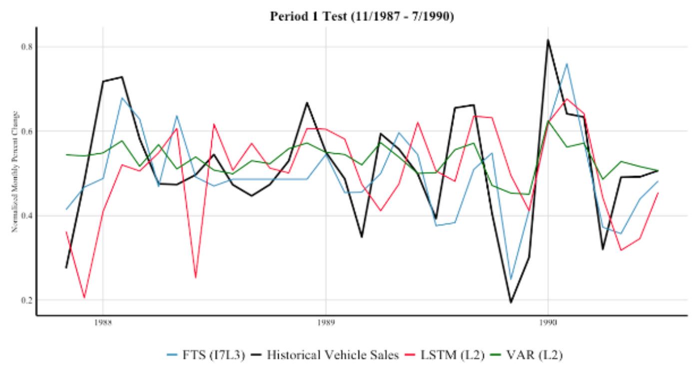

Figure 6: Period 1 Test Time Series

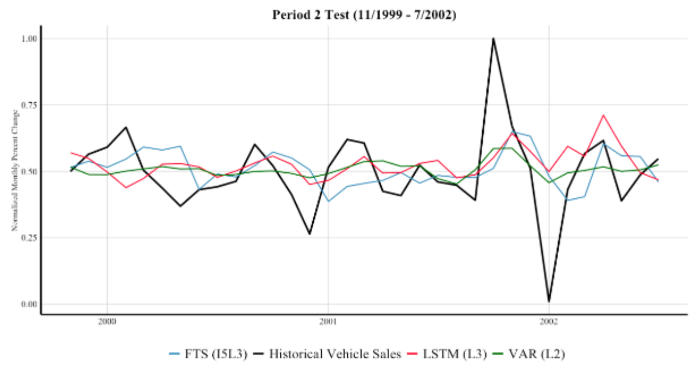

Figure 7: Period 2 Test Time Series

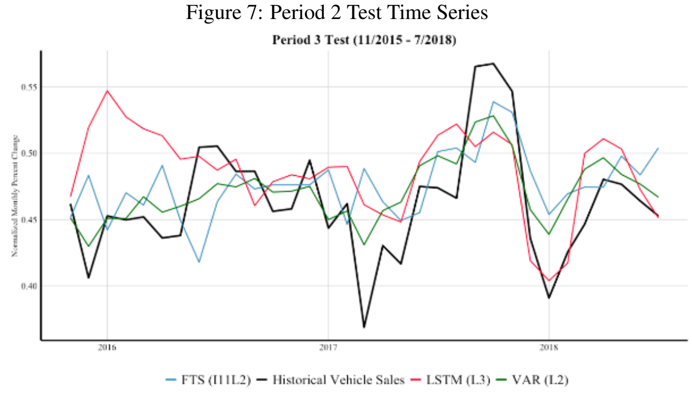

Figure 8: Period 3 Test Time Series

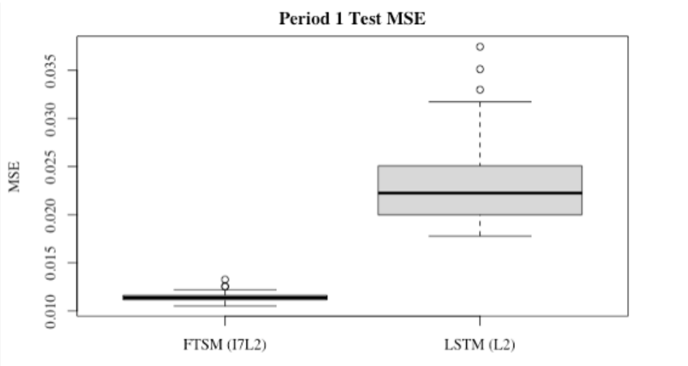

Figure 9: Period 1 Test MSE

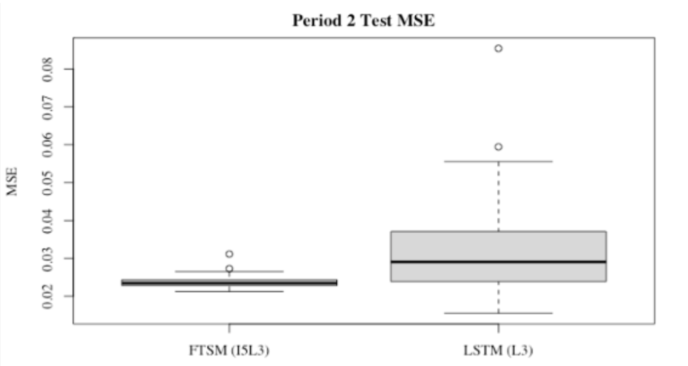

Figure 10: Period 2 Test MSE

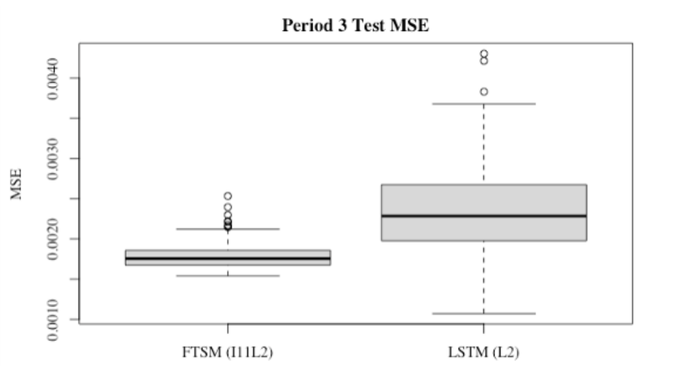

Figure 11: Period 3 Test MSE

5. Discussion and Conclusion

The FTSM paradigm produced accurate consumer spending models that performed comparatively or better than the LSTM and VAR counterparts in out-of-sample forecasting. This finding was validated across the consistency of three different test and training sets, which accounted for varying economic conditions and adjustments in consumer perception. However, the inconsistency of the results does not indicate strongly that, at least in this small sample size, consumers universally perceive economic sensitivity under a particular degree of precision. While the purpose of this paper was to introduce the application of fuzzy time series to a consumer rationality domain, a much larger study utilizing a range of independent and dependant variables is required to make a conclusion on consumers aggregate precision of economic perception.

The main conclusion of this work is that the dynamics inherent in fuzzy time series permit an accurate method of predicting consumer purchases, and should be considered as a time series modeling alternative to Long-Short-Term-Memory and Vector Autoregression. This paper shows that simulating consumers' imprecise perceptions of the economy can be leveraged to capture patterns in their future spending habits with a high degree of accuracy. These results suggest that the fuzzy time series approach could have significant future utility for forecasting aggregate consumer purchasing trends, an area that requires more research.

The forecasting power attributed to the fuzzy logic mechanics of the FTSM is additionally supported by relative comparison of the neural-network model architecture of the LSTM models. Both models were trained on identical node and epoch parameters; all else equal, the LSTM model has features of precision absent in the FTSM, while the FTSM intentionally muddies the precision of the model inputs. Formal statistical tests of comparison between the model accuracies were not conducted due to the lack of power in only three testing periods. VAR models, despite their relative simplicity, continue to be highly dependable economics time series models, and the results of this work support this.

Other important findings of this work concern the selection of the number of fuzzy intervals as a model parameter. Firstly, the results of the training and test fits show that the optimal test parameters can be informed by the accuracy of the parameters in the training sets. For example, in Period 2, I(7) was an outlier of accuracy in both the test and training sets; in Period 3, the same could be said about I(11). This should not be considered a hard-set rule, however, as the results are not so consistent as to inform the best testing model purely on training fit. Secondly, simplistic FTSMs, i.e. those with a fewer number of intervals, are highly generalizable and produce accurate out-of-sample fits. It is not clear why the optimal number of intervals in the FTSMs varied widely between the periods. This is an area that requires additional investigation.

Another potential area of future research would incorporate increased levels of imprecision common to traditional fuzzy logic models. Specifically, overlapping fuzzy intervals that could quantify membership between more than one interval. Overlapping intervals, or membership functions as they are referred to in traditional fuzzy logic literature, permit ambiguous quantification and description of fuzzy predicates. While this is not a standard approach in current FTS research, it fits well within the humanistic decision making framework this paper leverages, and is a potential area for increased model accuracy.

While it is difficult to pinpoint if the power of the FTSMs in this domain is a function of mirroring consumer rationality or purely a function of model mechanics, the results of this work show that FTSMs are a promising modeling alternative to forecasting consumer purchases, and should motivate additional research in this area.

References

Ala'raj, M., Abbod, M.F., Majdalawieh, M., 2021. Modelling customers credit card behaviour using bidirectional lstm neural networks. Journal of Big Data 8, 1–27.

Bao, W., Yue, J., Rao, Y., 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one 12, e0180944.

Barnes, M.L., Olivei, G.P., 2017. Consumer attitudes and their forecasting power for consumer spending. Journal of Money, Credit and Banking 49, 1031–1058.

Boydstun, A.E., Highton, B., Linn, S., 2018. Assessing the relationship between economic news coverage and mass economic attitudes. Political Research Quarterly 71, 989–1000.

Bram, J., Ludvigson, S.C., 1998. Does consumer confidence forecast household expenditure? a sentiment index horse race. Economic Policy Review 4.

Carroll, C.D., Fuhrer, J.C., Wilcox, D.W., 1994. Does consumer sentiment forecast household spending? if so, why? The American Economic Review 84, 1397–1408.

Curtin, R., 2007. Consumer sentiment surveys: worldwide review and assessment. Journal of business cycle measurement and analysis 2007, 7–42.

Curtin, R., 2019. Consumer expectations: a new paradigm. Business Economics 54, 199–210.

Dominitz, J., Manski, C.F., 2004. How should we measure consumer confidence? Journal of Economic Perspectives 18, 51–66.

Duru, O., Yoshida, S., 2012. Modeling principles in fuzzy time series forecasting, in: 2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), IEEE. pp. 1–7.

Egrioglu, E., Aladag, C.H., Yolcu, U., Uslu, V.R., Basaran, M.A., 2009. A new approach based on artificial neural networks for high order multivariate fuzzy time series. Expert Systems with Applications 36, 10589–10594.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.

Huang, Y.L., Hsieh, W.K., et al., 2019. A trend based forecasting model using fuzzy time series and pso algorithm, in: 2019 IEEE International Conference on Computation, Communication and Engineering (ICCCE), IEEE. pp. 21–24.

Index Calculations, University of Michigan, . Index calculations. `https://data.sca.isr.umich.edu/fetchdoc.php?docid=24770`.

Pickering, J., Harrison, J., Cohen, C., 1973. Identification and measurement of consumer confidence: Methodology and some preliminary results. Journal of the Royal Statistical Society: Series A (general) 136, 43–63.

Siami-Namini, S., Tavakoli, N., Namin, A.S., 2018. A comparison of arima and lstm in forecasting time series, in: 2018 17th IEEE international conference on machine learning and applications (ICMLA), IEEE. pp. 1394–1401.

Song, Q., Chissom, B.S., 1993. Forecasting enrollments with fuzzy time series—part i. Fuzzy sets and systems 54, 1–9.

Van Houdt, G., Mosquera, C., Nápoles, G., 2020. A review on the long short-term memory model. Artificial Intelligence Review 53, 5929–5955.

Zadeh, L.A., 1988. Fuzzy logic. Computer 21, 83–93.

Zhang, X., Wu, Q., Zhang, J., 2010. Crude oil price forecasting using fuzzy time series, in: 2010 Third International Symposium on Knowledge Acquisition and Modeling, IEEE. pp. 213–216.